

---

## CAPÍTULO 1. INTRODUCCIÓN

### 1.1 IMPORTANCIA DE LOS EMBALSES

En los últimos años la calidad del agua se está convirtiendo en una de las variables más importantes cuando se habla de gestionar recursos hídricos. La recién aprobada Directiva Marco del Agua de la UE (Directiva 2000/60/CE) establece por primera vez la necesidad de conocer el estado ecológico de las aguas continentales. Dicha directiva obliga también a que cada país miembro restablezca las aguas continentales a su estado natural para que mejore la calidad del agua (Armengol, 2000).

Los embalses o pantanos son sistemas de almacenamiento de aguas superficiales provenientes de ríos afluentes, lluvia o manantiales, para su posterior aprovechamiento (Markiegi et al., 1999). Son estructuras diseñadas y construidas para el beneficio de las poblaciones humanas. Entre estos beneficios están la generación de energía, recreación, la pesca comercial y deportiva, control de avenidas, suministro de agua y uso industrial. Cualquier efecto que dañe a estos usos tiene un impacto directo y notable (Tufford et al., 1999).

Hay más de 800.000 presas en el mundo actualmente, de las cuales 47.450 son clasificadas como grandes presas de acuerdo a la definición de la Comisión Internacional de Grandes Presas (ICOLD) la cual define a las grandes presas como las que poseen una altura de 15 metros o más, o presas entre 10 y 15 metros de altura que tengan una descarga por vertedor mayor de 2.000 metros cúbicos o un volumen embalsado mayor a un millón de metros cúbicos (Kabir, 2000).

Los embalses son lagos hechos por el hombre, los cuales en las regiones Mediterráneas no son solamente muy comunes sino que probablemente son los grandes cuerpos de agua dulce más importantes, desde un punto de vista económico (Rueda et al., 2006).

Actualmente hay más de 1.200 grandes presas en España que aportan una capacidad de unos 56.000 hm<sup>3</sup>. Es el quinto país del mundo con mayor número de ellas (CEDEX, 2006). España tiene la mayoría de las reservas de agua superficiales en embalses (Armengol, 2000).

Los embalses pueden tener algunos efectos adversos sobre el medio ambiente y el sistema económico y social (afectaciones agua abajo, inundación de tierras útiles,

---

calidad de las aguas, etc.) (Adler, 2005). Sin embargo, con una buena gestión de los recursos hídricos incluyendo la calidad del agua, estos problemas pueden ser evitados.

Las demandas en la calidad del agua de sueltas y extracciones de embalses para el suministro de agua y lagos son a menudo muy diversas y astringentes cuando usos adicionales y sus requerimientos relacionados deben cumplirse (Gelda et al., 2007).

En el presente trabajo se analizan algunas características de la calidad de las aguas en almacenamientos utilizando métodos multivariantes como el análisis de correlación canónica y el PLSR para estimar temperatura del agua en el embalse, concentraciones de oxígeno disuelto, pH y conductividad; y un método mecanicista, para determinar la temperatura del agua en el embalse. Con esto se puede conocer el comportamiento de los métodos multivariantes en comparación con el método mecanicista, establecer como varían estos parámetros, e investigar la posibilidad de efectuar inferencias y así contar con herramientas de gestión para evitar afectaciones en los usos tanto en el propio embalse, como río abajo o en el suministro.

## 1.2 CALIDAD DEL AGUA EN ALMACENAMIENTOS

En contraste con las aguas fluviales los lagos y embalses no fueron destacados en los años en que inició la modelación de la calidad del agua. Esto se debió, con la excepción de los grandes sistemas navegables como los Grandes Lagos, a que no habían estado históricamente en el foco principal del desarrollo urbano (Chapra, 1997).

Entre los primeros modelos de calidad del agua creados en la década de 1960, los modelos verticales unidimensionales de lagos y embalses han probado ser herramientas útiles en la gestión de la calidad del agua (Martin et al., 1999). Los modelos unidimensionales siguen siendo atractivos, apropiados y convincentes para describir la calidad del agua y la ecología de lagos y almacenamientos (Antonopoulos et al., 2003). Claramente un ajuste detallado de parámetros, crucial en un modelo físico-biológico acoplado, es extremadamente costoso en el caso tridimensional. Por esta razón y por el hecho de que los datos hidrodinámicos tridimensionales son aún muy escasos para calibrar y verificar estos modelos, se considera que la modelación unidimensional sigue siendo un planteamiento apropiado y convincente para la predicción de la calidad del agua (Joehnk et al, 2001).

---

En la década de 1970 se fue reconociendo que los lagos naturales y los hechos por el hombre eran igual de, si no más, importantes que los estuarios y ríos desde un punto de vista recreacional. Adicionalmente su empleo para suministro de agua, generación de energía y control de avenidas también contribuyeron a su significancia (Chapra, 1997).

Además de los usos mencionados anteriormente los embalses y su entorno presentan unos claros valores paisajísticos y como recurso recreativo (pesca, baño, excursión, navegación, etc.). Un embalse es un sistema dinámico que tiene elementos de vida, es decir, incluye flora, fauna y microorganismos vivos (Markiegi et al., 1999). Debido a todos estos factores se genera la importancia del estudio de la calidad del agua en estos sistemas.

Con el fin de aportar elementos para la protección de los usos y el sistema biótico del embalse se analizarán aquellas componentes de calidad del agua de las cuales se tiene información en perfil vertical, esto con el fin de realizar inferencias acerca de estas componentes. Las predicciones aparte de ser importantes *per se*, interesarán para revisar el cumplimiento de los máximos establecidos por cada Normatividad y además pueden servir de base al desarrollo de modelos de calidad del agua o ecológicos más sofisticados (por ejemplo para estudiar las condiciones del desarrollo de una especie en específico o la evolución de las características del agua a ser suministrada).

Las componentes de calidad del agua que se evaluarán en el presente trabajo son: temperatura, concentración de oxígeno, pH y conductividad.

### 1.2.1 Temperatura

La temperatura del agua es una de las características físicas de mayor importancia de los sistemas acuáticos. Afecta a numerosos parámetros de calidad del agua. La solubilidad de los gases disminuye y la solubilidad mineral se incrementa cuando la temperatura aumenta. Los coeficientes de crecimiento y respiración son dependientes de la temperatura y la mayoría de los organismos tienen distintos rangos de temperatura dentro de los cuales ellos se reproducen y compiten (Deas et al., 2000).

El oxígeno disuelto (OD), el ciclo de nutrientes, la productividad biológica y los desarrollos piscícolas son algunos de los parámetros de calidad del agua severamente controlados por los cambios de temperatura (Antonopoulos et al., 2003). La

---

temperatura tiene una fuerte influencia en los coeficientes de las reacciones químicas y biológicas (Chapra, 1997). Los coeficientes de estas reacciones se incrementan cuando la temperatura del agua aumenta (Deas et al., 2000).

El reconocimiento de distribución de la temperatura está en la base de la comprensión del desempeño y funcionamiento de ecosistemas de embalses (Han et al., 2000).

La temperatura de un cuerpo de agua es relevante desde una óptica de calidad del agua por tres razones: los derrames térmicos afectan al ecosistema hídrico, interviene en las reacciones químicas y biológicas y las variaciones en la temperatura perturban a la densidad del agua, afectando por lo tanto a su movimiento (Dolz et al., 1995).

Además, de los componentes de calidad, la temperatura del agua es el más fácil y a menudo el menos costoso de monitorizar en condiciones de campo (Deas et al., 2000).

La temperatura dentro del ecosistema puede cambiar el espectro de las especies que pueden vivir dentro del mismo. Su modificación y su cambio con el paso del tiempo (tanto por día como por año) afectan a seres vivos acuáticos alterando sus poblaciones debido a una modificación en su desarrollo, ciclo de reproducción, conducta, fuentes de alimentación, etc. (Dolz et al., 1995).

El periodo de apareamiento y crecimiento de las especies piscícolas son las etapas más sensibles para modificaciones de temperatura, pero hay temperaturas que, si bien, no son perjudiciales en ese mismo instante pueden favorecer el desarrollo de parásitos, a las especies competidoras, etc. (Dolz et al., 1995).

Un factor como la temperatura del agua también es relevante en la dinámica de la población de algas (Whigham et al., 2001).

El entorno ambiental determina el sistema térmico de un embalse y del agua almacenada, afectando a las características de la misma. El comportamiento térmico de un embalse depende de distintos factores: volumen y morfología (con especial atención a la profundidad media), acción eólica y características climáticas de la región, desplazamiento del agua en el almacenamiento (que está relacionado con varios factores, como son el tamaño del mismo, los caudales que ingresan y los caudales que egresan), régimen térmico del agua que ingresa, elevación de las tomas de donde se extrae el agua, etc. (Dolz et al., 1995).



---

El régimen de temperaturas de una corriente puede ser muy afectado por la construcción de un almacenamiento en su cauce. Río abajo del vertido del almacenamiento la temperatura del recurso hídrico experimentará un cambio que la llevará a recobrar el equilibrio térmico con el medio que lo rodea (Dolz et al., 1995).

En un embalse no se tiene, de ordinario, una temperatura fija para una profundidad. Además la enorme cantidad de agua dentro del almacenamiento, que casi en la totalidad de los casos se puede considerar como inmóvil, muestra una gran tendencia a conservar su temperatura presente (Dolz et al., 1995).

En la mayoría de los embalses se presenta un fenómeno llamado estratificación termal. Consiste en que durante el inicio de la primavera el agua a todas las profundidades está cercana a la temperatura de máxima densidad, por lo que hay muy poca resistencia termal al mezclado y sólo se requieren pequeñas cantidades de energía eólica para mezclar la columna de agua. Conforme avanza la primavera, las aguas superficiales de los almacenamientos son calentadas más rápidamente que el calor distribuido por mezclado. Como las aguas superficiales al aumentar su temperatura llegan a ser menos densas, la resistencia termal relativa al mezclado se incrementa marcadamente. Una diferencia de sólo unos pocos grados es entonces suficiente para evitar la circulación completa. Desde este momento la columna de agua está termalmente dividida en tres regiones, las cuales exceden la resistencia al mezclado entre ellas: el estrato más bajo, el hipolimnion; el periodo de estratificación de verano está caracterizado por un estrato superior más o menos uniforme, el epilimnion; el estrato entre el epilimnion y el hipolimnion es llamado el metalimnion y exhibe una marcada discontinuidad termal. Al terminar el verano la disminución de la temperatura del aire genera un ingreso negativo de calor al embalse y la pérdida de calor excede la entrada por energía solar. Las aguas superficiales se vuelven más frías y llegan a ser más frías que las subyacentes, por lo que descienden y ocurre una mezcla por combinación de corrientes de convección y circulación inducida por el viento. Una erosión progresiva del metalimnion desde arriba puede observarse y el epilimnion homogéneo incrementa su espesor. La circulación continúa con el enfriamiento gradual de la columna de agua. Cuando la temperatura del agua alcanza el punto de máxima densidad (4°C) la estratificación de agua fría que yace sobre agua a mayor temperatura es fácilmente rota por una pequeña cantidad de energía proveniente del viento y un mezclado completo vuelve a instaurarse (Wetzel, 2001).

La distribución de temperaturas y la configuración de la densidad del almacenamiento tenderán a modificar el comportamiento, en lo que a su dinámica se refiere, de los

---

caudales que ingresen a él, sus componentes intrínsecos y el nivel de interrupción de la continuidad de la temperatura del agua de las salidas del embalse sobre el arroyo (Dolz et al., 1995).

Asimismo la estratificación de la temperatura del agua fija en gran parte los procesos químicos y biológicos que ocurren y se manifiestan en el almacenamiento, lo que a su vez incide en las propiedades del agua del mismo embalse y de la corriente río abajo (Dolz et al., 1995).

### 1.2.2 Concentración de oxígeno

Si una sola variable fuera a ser medida en sistemas acuáticos para proporcionar la máxima información acerca del estado del sistema, esta variable sería el oxígeno disuelto (OD) (Environmental Laboratory, 1995).

La temperatura y el oxígeno son dos de los principales factores de calidad del agua en lagos y embalses (Antonopoulos et al., 2003).

El oxígeno controla muchas reacciones químicas a través de la oxidación y es una variable indicadora de la salud general de los sistemas acuáticos (Environmental Laboratory, 1995).

El oxígeno es un importante componente regulatorio; niveles muy bajos pueden generar mortalidad en masa de peces y otros organismos, movilización de nutrientes y metales y disminución en la degradación de materiales orgánicos tóxicos (Park et al., 2004).

El término de oxígeno posee fuentes y pérdidas las cuales incluyen los procesos de producción por la fotosíntesis de las plantas, las pérdidas por la respiración de los organismos y el consumo debido a la demanda bioquímica de oxígeno (DBO). El proceso de producción de oxígeno fotosintético es junto con la reaeración, de los procesos más importantes en la modelación del oxígeno disuelto (Antonopoulos et al., 2003). Los tributarios que arriban al propio cuerpo de agua también son una fuente de OD (Thomann et al., 1987).

Cuando la concentración de oxígeno disuelto en el agua es inferior a la de saturación se produce una transferencia de oxígeno desde la atmósfera hacia el seno del agua. Este proceso es llamado reaeración superficial (Martín et al., 1999).

---

La distribución resultante de OD en el almacenamiento afecta fuertemente la solubilidad de muchos constituyentes químicos inorgánicos. A menudo los planteamientos de control de la calidad del agua o gestión son formulados para mantener un ambiente aeróbico u óxico (o sea, que contenga oxígeno) (U.S. Army Corps of Engineers, 1987).

La solubilidad del oxígeno en el agua depende principalmente de la temperatura del agua, la presión parcial de oxígeno en la atmósfera y el contenido de sal (Manahan, 2000).

El impacto de las bajas concentraciones de OD o de las condiciones anaeróbicas se refleja en un ecosistema desbalanceado, mortalidad de peces, olores y otras molestias estéticas (Thomann et al., 1987).

La degradación de la materia orgánica presente en los sedimentos hace disminuir la concentración de oxígeno en el agua intersticial de los mismos, promoviendo la transferencia de oxígeno desde la columna de agua hacia los mismos (Martín et al., 1999). Esto ocasiona que bajo determinadas condiciones las concentraciones de oxígeno del agua cercana al fondo sean muy bajas.

En los almacenamientos la demanda de oxígeno disuelto puede ser dividida en dos por separado, pero altamente interactivas fracciones: demanda de oxígeno del sedimento (DOS) y demanda de la columna de agua (U.S. Army Corps of Engineers, 1987).

### 1.2.3 pH

Para medir la presencia del ion hidrógeno en el agua, se usa la notación pH (abreviatura en el idioma francés de *puissance d'Hydrogène*, el cual se puede traducir como “el potencial o la intensidad de hidrógeno”) (Chapra, 1997):

$$\text{pH} = -\log_{10} [\text{H}^+] \quad (1)$$

Debido al signo negativo y a la transformación logarítmica de  $[\text{H}^+]$ , un pH alto representa una baja actividad del ion hidrógeno mientras que un pH bajo representa una actividad alta del ion hidrógeno. En consecuencia la escala de pH refleja la intensidad de la condición de acidez o básica (alcalina) del agua. Un bajo pH connota acidez y un alto pH connota aguas alcalinas, con una condición neutral en un pH de

---

7,0. El rango típico de variación para aguas naturales es de 6,0 a 9,0 (Chapra, 1997).

La escala pH mide la concentración logarítmica de los iones hidrógeno ( $H^+$ ) e hidróxido ( $OH^-$ ), los cuales configuran el agua ( $H^+ + OH^- = H_2O$ ). Cuando ambos tipos de iones están en igual concentración el pH es 7,0 o neutral. Por debajo de 7,0 el agua es ácida (hay más iones hidrógeno que iones hidróxido). Cuando el pH está por arriba de 7,0 el agua es alcalina o básica (hay más iones hidróxido que iones hidrógeno). Debido a que la escala es logarítmica, una caída en el pH de 1,0 (una unidad) es equivalente a un incremento de 10 veces en la acidez. Entonces una muestra de agua con un pH de 5,0 es 10 veces más ácida que una con un pH de 6,0 y un pH de 4,0 es 100 veces más ácido que un pH de 6,0 (EPA, 2007).

La biota acuática es sensible a los extremos del pH. Existen por supuesto rangos de pH en los cuales viven los organismos. Frecuentemente estos rangos pueden ser representados mediante una curva razonablemente simétrica dentro de la cual a lo largo de los bordes un organismo puede vivir sin desarrollarse adecuadamente (Manahan, 2000).

El pH del agua ejerce una poderosa influencia en la distribución de especies. Sólo una minoría de los organismos pueden crecer por debajo de un pH de 4,5, como se evidencia en el descenso en el número de especies en aguas ácidas (Stiling, 2002).

Los nuevos modelos desarrollados para analizar los efectos de los controles alternativos sobre las entradas de acidez a los ambientes acuáticos sensibles requieren de la predicción del pH (Bowie et al., 1985).

Un cambio del pH puede tener efectos indeseables como aumentar la proporción de amonio no ionizado, o variar la concentración de metales disponibles en disolución (Martin et al., 1999).

Además, la hidrólisis de los químicos orgánicos puede ser sensible al pH. Adicionalmente la viabilidad de organismos y la bioacumulación y toxicidad de químicos orgánicos pueden ser dependientes del pH (Park et al., 2005).

El equilibrio del dióxido de carbono ( $CO_2$ ) – bicarbonato ( $HCO_3^-$ ) – carbonato ( $CO_3^{2-}$ ) es el mayor sistema de control en ambientes acuáticos. Este equilibrio afecta directamente al pH, el cual a su vez puede afectar a los constituyentes biológicos y químicos del sistema. Por ejemplo puede llegar a ser necesario simular el pH y la

---

alcalinidad con el propósito de calcular el amonio no ionizado (tóxico) o determinar las concentraciones disponibles de metales (Bowie et al., 1985).

El sedimento en el fondo de los almacenamientos atrapa y acumula determinadas sustancias químicas relativamente poco solubles cuya concentración aumenta con el tiempo pero que pueden redisolverse en otras condiciones (según el pH del agua por ejemplo); pertenecen a esta categoría los metales pesados (Markiegi et al., 1999).

#### 1.2.4 Conductividad

La conductividad es una medida de la capacidad del agua para transmitir una corriente eléctrica. La conductividad en el agua es afectada por la presencia de sólidos disueltos inorgánicos como los aniones cloruro, nitrato, sulfato y fosfato (iones que tienen una carga negativa); o los cationes sodio, magnesio, calcio, hierro y aluminio (iones que tienen una carga positiva). La conductividad es también afectada por la temperatura: cuanto más caliente el agua mayor conductividad (EPA, 2007).

La conductividad del agua depende de la concentración total de sustancias disueltas ionizadas en el agua y de la temperatura a la cual se haga la determinación (Romero Rojas, 1999).

La conductividad en corrientes y ríos (que pueden ser afluentes de un embalse) está primariamente afectada por la geología del área a través de la cual transitan sus aguas. Por ejemplo las corrientes que fluyen a través de áreas con lecho rocoso de granito tienden a mostrar baja conductividad debido a que el granito está compuesto en su mayoría de materiales inertes que no se ionizan (no se disuelven en presencia de compuestos iónicos) cuando son arrastrados hacia el agua. Por otra parte las corrientes que atraviesan áreas con suelos arcillosos tienden a tener alta conductividad debido a la presencia de materiales que se ionizan cuando son llevados por el agua (EPA, 2007).

El valor de la conductividad es muy usado en análisis de aguas para obtener un estimativo rápido del contenido de sólidos disueltos (Romero Rojas, 1999).

Las descargas a ríos pueden cambiar la conductividad dependiendo de su procedencia (sistemas de drenaje, vertidos, etc.) (EPA, 2007).

Se denomina Siemens, símbolo (S), a la unidad derivada del SI para la medida de la conductancia eléctrica. Esta unidad también se denomina mho, palabra inversa de

---

ohm u ohmio, porque la conductancia es la inversa de la resistencia (es.wikipedia.org, 2007).

La conductividad es medida en micromhos por centímetro ( $\mu\text{mhos/cm}$ ) o microsiemens por centímetro ( $\mu\text{S/cm}$ ). El agua destilada tiene una conductividad en el rango de 0,5 a 3  $\mu\text{mhos/cm}$ . La conductividad de los ríos en los Estados Unidos tiene un rango generalmente entre 50 y 1500  $\mu\text{mhos/cm}$ . Los estudios en aguas dulces continentales indican que las corrientes mantienen una buena diversidad de especies piscícolas cuando se encuentran en un rango entre 150 y 500  $\mu\text{mhos/cm}$ . La conductividad fuera de este rango podría indicar que el agua no es adecuada para ciertas especies de peces o macroinvertebrados (EPA, 2007).

La conductividad es útil como una medida general de calidad del agua. Cambios significativos en la conductividad podrían entonces ser un indicador de que una descarga o alguna otra fuente de contaminación ha entrado al cuerpo de agua (EPA, 2007). La medida de la conductividad constituye un parámetro básico de evaluación de la aptitud del agua para el riego (Romero Rojas, 1999).

### 1.3 EL FENÓMENO DE LA EUTROFIZACIÓN

La acción del hombre y su inacción son las causas principales de la rápida degeneración de lagos de agua dulce y embalses (Henderson-Sellers et al., 1987).

El término eutrofización derivado del término griego que significa “bien nutrido”, describe una condición de los lagos y embalses que involucra un exceso de crecimiento algal. La eutrofización es a menudo un fenómeno natural (Manahan, 2000).

La adición de nutrientes a las aguas naturales estimula el crecimiento de plantas. Esto en pequeñas cantidades, puede ser bueno. Por ejemplo debido a la gran cantidad de población humana en su cuenca de drenaje, el lago Michigan recibe más nutrientes que el Lago Superior. Consecuentemente el Lago Michigan presenta más crecimiento de plantas lo cual a la postre permite que albergue más variedad de peces. Sin embargo, cuando los cuerpos de agua son sobre-fertilizados el excesivo crecimiento de plantas resultante puede llegar a ser un problema serio de calidad del agua (Chapra, 1997).

Aunque el proceso de eutrofización es de carácter natural el tiempo asociado de

---

duración puede ser medido en el marco temporal geológico. Sin embargo, en poco menos de dos siglos el hombre ha incrementado rápidamente su uso de nutrientes, especialmente de fertilizantes agrícolas y detergentes, muchos de los cuales terminan dentro de los cuerpos de agua y por lo tanto aceleran el proceso de eutrofización (Henderson-Sellers et al., 1987).

Los embalses aparecen unidos a la civilización y llevan consigo la tendencia a la eutrofización y aun a la contaminación, que nacen de su proximidad a las actividades humanas relacionadas con el manejo del agua (Margalef, 1983).

La acumulación de nutrientes es un problema serio en lagos y en aguas de lento movimiento. Debe tenerse presente que en muchos proyectos de captaciones de ríos para abastecimiento, el agua cruda se almacena en grandes embalses poco profundos antes del tratamiento. Incluso con contenidos de nutrientes bajos puede haber un crecimiento prolífico de algas, lo que hace más difícil el tratamiento de la misma (Tebbutt, 1990).

El proceso de eutrofización describe el enriquecimiento de nutrientes del cuerpo de agua. De una condición inicial pobre en nutrientes (u oligotrófica) la adición de nutrientes al cuerpo de agua gradualmente aumenta las concentraciones (tanto en la columna de agua como en los sedimentos) hasta que una condición es probablemente alcanzada, donde el detrimento de la calidad del agua ocurre; en dicho estado el cuerpo de agua es denominado eutrófico (o sea, rico en nutrientes). La localización de cualquier cuerpo de agua a lo largo de esta continua escala de eutrofización, desde oligotrófico a eutrófico, es llamado su estado trófico. Originalmente sólo eran de uso común tres categorías discretas (oligotrófico, mesotrófico, eutrófico) (Henderson-Sellers et al., 1987). Actualmente se ha sumado a estas el término hiper-eutrófico (nutrido en exceso).

Según el destino que pretenda darse al agua embalsada su calidad deseable es diferente. En el caso de que se utilice para el abastecimiento de la población, las aguas deben ser oligotróficas (Markiegi et al., 1999).

Las consecuencias inmediatas del incremento de cargas de nutrientes en un lago o embalse son estimular el crecimiento de plantas (y subsecuentemente de animales) (Henderson –Sellers et al., 1987).

El crecimiento excesivo de plantas acuáticas tanto fijas como planctónicas, en altos

---

niveles, se considera que son una interferencia con los usos convenientes del agua (Thomann et al., 1987). Una causa común de menoscabo en el uso de un embalse es la eutrofización que puede impactar usos como la recreación y la conservación del hábitat o incrementar el costo de su uso como agua para beber (Tufford et al., 1999).

La eutrofización es un problema serio en muchos lagos y embalses. El enriquecimiento, natural o artificial, de los cuerpos de agua continentales usualmente causa eutrofización, con los típicos blooms algales, que generan el deterioro de la calidad del agua para el uso humano. Prediciendo el comportamiento de los cuerpos de agua enriquecidos se proporcionará una oportunidad de estar preparados durante los periodos críticos y combatir con los efectos adversos (Karul et al., 2000).

La predicción del comportamiento de los cuerpos de agua enriquecidos con nutrientes es difícil debido a los complejos procesos físicos, químicos y biológicos involucrados (Karul et al., 2000).

Debido a que otros parámetros son más fáciles de medir que las concentraciones de nutrientes, estos son a menudo utilizados como un sustituto para indicar los niveles de nutrientes actuales. Ejemplos de estos parámetros son la disminución de oxígeno disuelto, el incremento de clorofila-a o el incremento en la turbidez (Henderson-Sellers et al., 1987).

#### 1.4 OBJETIVOS

Los objetivos que busca alcanzar el presente trabajo son los que se enuncian a continuación:

- Utilizar algunos métodos multivariantes para estimar los valores de ciertas variables de calidad del agua en almacenamientos, haciendo uso de la información disponible (se usarán las variables de las que se dispone de datos). Con este proceder se busca tener una primera perspectiva acerca de si la información que se registra en el Embalse Amadorio es suficiente para pronosticar las variables de calidad del agua analizadas o poder determinar si para su estimación es necesario registrar información de otras variables que afecten al embalse (como las concentraciones en las entradas, etc.).
- Efectuar una comparación entre la aplicación de métodos multivariantes y un método mecanicista para obtener estimaciones de los valores de algunos parámetros de la calidad del agua en embalses y tener la posibilidad de



---

establecer su efectividad, facilidad de uso y sus respectivas ventajas e inconvenientes.

- Establecer por medio del análisis de factores las relaciones existentes entre variables meteorológicas e hidrológicas del embalse y entre estas y las concentraciones medidas en perfil vertical de algunos componentes de calidad del agua en el almacenamiento. Se buscará identificar factores (variables latentes) que expliquen el comportamiento y variación de los datos observados.
- Proponer una metodología para la predicción de algunas componentes de calidad del agua en base al Método de regresión en mínimos cuadrados parciales (PLSR). La predicción se hará haciendo uso de la matriz de variables meteorológicas e hidrológicas y el método citado.
- Establecer de ser necesario un planteamiento para el pre-procesado de los datos existentes (variables predictivas) y su clasificación, con el propósito de que la metodología propuesta en el punto previo alcance un mejor desempeño.

## 1.5 ESTRUCTURA DEL DOCUMENTO

En el capítulo 2 se presenta el desarrollo a través del tiempo de los métodos multivariantes y en especial del PLSR, así como algunos ejemplos de aplicación de los mismos en la caracterización de la calidad del agua en cuerpos de agua dulce.

El capítulo 3 expresa la evolución histórica de los modelos mecanicistas, sus primeros desarrollos, hasta llegar a las últimas aplicaciones de los modelos más actuales a lagos y presas.

Una descripción concisa del significado y pasos a seguir para la aplicación de algunos métodos multivariantes como son: análisis de componentes principales, análisis factorial, análisis cluster, análisis discriminante, el modelo lineal de regresión múltiple, el análisis de correlación canónica y la regresión parcial en mínimos cuadrados, se efectúa en el capítulo 4.

Los datos requeridos por el modelo mecanicista, así como la descripción detallada de los procesos que realiza y la secuencia en que los efectúa, así como también el esquema de solución aplicado en diferencias finitas y su metodología específica de solución son mostrados en el capítulo 5.

La información del sitio de estudio, comprendiendo ésta la ubicación geográfica, descripción de la zona y la cuenca donde se encuentra el Embalse Amadorio, datos

---

acerca del mismo; clima, geología y usos del suelo imperantes; información demográfica y sus posibles tendencias; así como vertidos hacia el embalse y normatividad aplicable; las demandas existentes y una breve descripción de la sonda multiparamétrica instalada en uno de sus paramentos, se puede consultar en el capítulo 6.

El capítulo 7 contiene la aplicación de diversos métodos multivariantes (comentados previamente en el capítulo 4) con varios propósitos: conocer la estructura de las variables predictivas y las variables respuesta y las relaciones existentes entre ellas; saber si pueden ser agrupadas las variables predictivas para mejorar la estimación de las variables respuesta mediante las técnicas de regresión y efectuar estimaciones de las variables de calidad del agua mediante diversas técnicas de regresión, así como evaluar los resultados obtenidos con las distintas regresiones mediante el coeficiente de determinación y el error de predicción para un conjunto de datos de validación.

La aplicación del modelo mecanicista se realizó en el capítulo 8. Primero se revisaron los datos y se efectuó la propuesta de una metodología específica para obtener la cobertura de nubes y la temperatura superficial del embalse. También se hizo la calibración y la validación del modelo.

En el capítulo 9 se efectúa el análisis de resultados y su discusión. Se analizan comportamientos relacionados de algunas de las variables, así como los resultados de las regresiones obtenidos con los métodos multivariantes y el método mecanicista, además se efectúa un comparativo de las estimaciones obtenidas mediante ambos enfoques. También se hacen propuestas acerca de cómo estimar las variables predictivas para alimentar los modelos (datos de entrada al modelo).

Las conclusiones generadas al desarrollar las distintas metodologías y las líneas futuras que se pueden seguir a partir del trabajo realizado se mencionan en el capítulo 10.

Para concluir las referencias empleadas más útiles se agrupan en el capítulo 11.

---

## CAPÍTULO 2. ESTADO DEL ARTE DE LOS MÉTODOS MULTIVARIANTES

### 2.1 DESARROLLO HISTÓRICO DE LOS MÉTODOS MULTIVARIANTES

Durante la Edad Media y hasta los principios del siglo XVII la Estadística era únicamente descriptiva. Más adelante surgieron dos escuelas: la descriptiva alemana, en la que Achenwall era su miembro más sobresaliente y considerado por sus compatriotas como el padre de la Estadística; siendo la otra la escuela de matemáticos políticos que desarrolló técnicas de previsión y principios acerca de la regularidad de algunos fenómenos sociales. En el año de 1801 se realizó en Francia el primer censo de población (Gil Pascual, 2003).

El matemático y astrónomo francés Pierre-Simon Laplace en su teoría analítica de las probabilidades hizo patentes las ventajas que podía proporcionar el cálculo de las probabilidades en el análisis de los hechos naturales de cierta complejidad (Gil Pascual, 2003).

El interés en el análisis de datos multivariantes ha crecido tremendamente desde el advenimiento de los ordenadores debido a la gran cantidad de cálculos involucrados. Los primeros pasos fueron dados con el artículo de Eugenio Beltrami, matemático italiano, acerca de la descomposición de valores singulares y con el trabajo referente a la regresión lineal de Sir Francis Galton (Umetrics, 2007).

Sir Francis Galton, científico y explorador británico, también fue pionero en el estudio de mezclas de distribuciones normales (Peña, 2002). Galton también abordó la regresión multivariante pero no pudo terminar este trabajo (Umetrics, 2007). Él desarrolló también la idea de correlación y enfocó su atención en la necesidad de un mayor conocimiento de las posibles formas de la distribución multivariante (Kotz, 2000). Francis Galton y Karl Pearson hicieron trabajos pioneros en las áreas de la correlación y el análisis de regresión (Johnson et al., 2004).

Karl Pearson, científico británico y colaborador de Galton, efectuó trabajos sobre los ajustes ortogonales que precedieron al análisis de componentes principales (Peña, 2002). Algunas veces es considerado como el iniciador del análisis de componentes principales con el artículo que presentó en 1901, además de haber realizado otras contribuciones importantes al análisis estadístico (Umetrics, 2007).

Charles Spearman, psicólogo inglés, propuso el primer modelo factorial basado en un

---

factor común, el factor  $g$  y un componente específico (Peña, 2002). Spearman y Pearson están entre los primeros que trabajaron en el análisis factorial (Johnson et al., 2004).

Ronald Aylmer Fisher, científico británico de los campos de la Genética y la Estadística, entre otros, fue el inventor del análisis discriminante, del método de máxima verosimilitud y del diseño estadístico de experimentos (Peña, 2002). La derivación que propuso de la distribución exacta del coeficiente de correlación muestral y cantidades relacionadas proporcionaron el ímpetu para la teoría de la distribución multivariante (Johnson et al., 2004). También fue el iniciador del análisis de la varianza (Gil Pascual, 2003).

P. C. Mahalanobis, estadístico hindú, tiene como su contribución más conocida la distancia de Mahalanobis, una medida de distancia estadística multivariante independiente de la escala (es.wikipedia.org, 2007).

Harold Hotelling, estadístico americano, fue el creador de los componentes principales, el análisis de correlaciones canónicas y nuevos métodos de inferencia multivariante. Es una de las figuras centrales en el desarrollo de la estadística multivariante y es el artífice principal de los métodos de reducción de la dimensión (Peña, 2002).

John Wishart, estadístico británico, primero fue ayudante de Pearson y luego de Fisher. En 1928 encontró la distribución que lleva su nombre (Peña, 2002). Esta distribución conjunta varianzas y covarianzas muestrales (Johnson et al., 2004). Hizo contribuciones destacadas al diseño de experimentos (Peña, 2002).

Samuel S. Wilks, estadístico estadounidense, construyó generalizaciones multivariantes para el análisis de la varianza y el coeficiente de correlación múltiple (Peña, 2002). También hizo otras contribuciones significativas al análisis, como procedimientos generales para probar ciertas hipótesis multivariantes (Johnson et al., 2004).

M. S. Bartlett y G. E. P. Box, estadísticos ingleses, contribuyeron a la gran teoría muestral asociada con ciertas pruebas estadísticas multivariantes (Johnson et al., 2004).

John Tukey, estadístico estadounidense, fue el creador de la transformada rápida de Fourier y de nuevos métodos para el diagnóstico de modelos y la estimación robusta (Peña, 2002).

---

Calyampudi R. Rao, estadístico hindú y discípulo de Mahalanobis, hizo contribuciones fundamentales a la inferencia estadística (como la cota de Cramer-Rao) y al análisis multivariante (análisis discriminante de Fisher para varios grupos, extensiones de correlaciones canónicas) (Peña, 2002).

Jean-Paul Benzecri, estadístico francés, fue el creador del análisis de correspondencias y la persona más destacada de la escuela francesa de análisis de datos (Peña, 2002).

La clasificación moderna (también denominada *taxonomía numérica*) de objetos o registros se inicia en 1957 con el imperativo de proponer criterios objetivos de clasificación (Sokal, Sneath, Michener) (Cuadras, 2007; es.wikipedia.org, 2008). Posteriormente Benzecri, Jardine, Sibson y Johnson relacionaron las clasificaciones jerárquicas con los espacios ultramétricos (Cuadras, 2007).

Para resolver el problema de la multicolinealidad muchos métodos estadísticos han sido propuestos. La regresión ridge (RR) fue sugerida por Hoerl y Kennard (1970) como un método para estabilizar las estimaciones de regresión en presencia de multicolinealidad, la cual asume que los coeficientes de regresión probablemente no son muy grandes (Cheng et al., 2006).

La regresión de componentes principales (PCR) introducida por Massy (1965), intenta reducir la dimensión y evitar la multicolinealidad por medio del uso de unos pocos componentes, las combinaciones lineales de variables explicativas (Cheng et al., 2006).

El advenimiento de los ordenadores en los años 60 origina el surgimiento final del análisis multivariante. La dificultad de los procesos de cálculo queda como algo secundario y comienza a usarse el análisis multivariante en ciencias como la Psicología, Educación, Biología, Medicina, Economía, Geología, etc. (Gil Pascual, 2003).

## 2.2 SURGIMIENTO DEL MÉTODO PLSR

En 1966 el estadístico y economista sueco Herman Wold presentó el algoritmo de Mínimos Cuadrados Parciales Iterativos no Lineal o NIPALS. El concepto no era enteramente nuevo porque hallazgos similares habían sido publicados ya en 1923 por Fischer et al. Pero la originalidad en el trabajo de H. Wold fue la interpretación de los mínimos cuadrados parciales y la capacidad de manejar datos faltantes. Durante los

---

años que siguieron, H. Wold evolucionó el algoritmo NIPALS hacia el método de regresión en Mínimos Cuadrados Parciales (PLSR) (Umetrics, 2007).

Los primeros artículos de Herman, aplicando el método PLSR, aparecieron alrededor de 1975 con aplicaciones a bloques múltiples de datos de las ciencias económicas y sociales (Wold, 2001).

La regresión PLS es una técnica que generalizada y combina características del análisis de componentes principales y de la regresión múltiple. Es particularmente útil cuando es necesario predecir un conjunto de variables dependientes usando un (muy) extenso conjunto de variables independientes (o sea, los predictores). Esta técnica se originó en las ciencias sociales (específicamente en la Economía) pero ha llegado a ser popular primero en quimiometría (química computacional) debido al hijo de Herman Wold, Svante, además de para la evaluación sensorial (Abdi, 2003a). Svante y Harald Martens modificaron el método originalmente propuesto por Herman Wold para su uso en la regresión, con el propósito de emplearlo en aplicaciones químicas, después de una aplicación inicial por parte de Kowalski et al. (Geladi et al., 1986)

La versión quimiométrica del PLSR con extracción de  $a$  factores, fue originalmente desarrollada como un algoritmo en dos bloques consistente en una secuencia simple de modelos parciales ajustados por mínimos cuadrados. Esto fue motivado intuitivamente por la experiencia práctica (debido a que “todas las variables pueden tener errores”) y consideraciones geométricas (una secuencia de proyecciones ortogonales). Este planteamiento orientado empíricamente y por el algoritmo fue satisfactorio para muchos químicos (Martens et al., 1989).

Pero para los estadísticos y algunos otros investigadores esta versión era difícil de comprender o de seguir. Una presentación ligeramente diferente y más compacta del método fue adicionada por Höskuldsson (1988) y Frank (1987), que resultó conceptualmente más simple para alguna gente. Esta fue continuada por más aportaciones de Manne (1987) y Helland (1988). Este trabajo mostró que los modelos del PLSR podrían ser calculados con un autovalor-ecuación de autovalores que hizo obsoleto al procedimiento iterativo (Geladi, 2002).

Además la regresión PLSR está convirtiéndose en una herramienta seleccionada en las Ciencias Sociales como técnica multivariada para datos experimentales y no experimentales (Abdi, 2007).

---

En la década de los noventas del siglo XX esta metodología fue extendiéndose para modelar y monitorear los procesos industriales, con éxito (Li et al., 2003).

De entre los muchos métodos estadísticos para modelos lineales con el problema de multicolinealidad, la regresión en mínimos cuadrados parciales (PLSR) ha llegado a ser en años recientes cada vez más popular y muy a menudo la mejor opción. Siendo comparativamente un método nuevo ha llegado a ser el método de regresión más popular en quimiometría (Cheng et al., 2006). Para algunos investigadores el PLSR es el método de calibración lineal más importante (Balabin et al., 2007). Esta técnica ha sido muy bien forjada con un gran potencial también en las Ciencias Naturales y en la Ingeniería (Wold, 2001).

El PLSR maneja el problema de la multicolinealidad por medio del cálculo de vectores latentes (comparable a los componentes del PCA) los cuales explican tanto los predictores o variables independientes, como los datos correspondientes a las medidas o variables dependientes (Abdi, 2003b).

El PLSR ha sido estudiado a gran detalle. Frank (1993) y Goutis (1996) verificaron propiedades de la estimación con PLSR. Ruscio (2000) estudió la relación entre el algoritmo PLSR para datos univariantes y la expresión polinomial de Cayley-Hamilton. Stone (1990) introdujo la regresión continua basada en OLS, PLSR y PCR. Goutis (1996) introdujo una modificación del PLSR usando penalización desigual. Wold (1992) y Durand (1997) extendieron el PLSR hacia la no linealidad usando funciones spline. Al año 2006 se tiene que el PLSR ha sido aplicado en muchos campos, especialmente con mucha amplitud en Química como “el uso de las Matemáticas y la Estadística en los datos químicos” (Cheng et al., 2006).

El PLSR ha sido también combinado con las redes neuronales como una nueva materia en el análisis no lineal (Cheng et al., 2006).

### 2.3 APLICACIÓN DE MÉTODOS MULTIVARIANTES EN DETERMINACIÓN DE CALIDAD DEL RECURSO HÍDRICO EN CUERPOS DE AGUA CONTINENTALES

A continuación se presentan algunas aplicaciones que se han efectuado de métodos multivariantes para estudios de calidad en aguas continentales.

Bartels et al. (1985) mostraron que el reconocimiento de patrones proporciona una forma de clasificar varios tipos de aguas superficiales en base a los datos fisicoquímicos. Basados en siete características relevantes se encontraron cinco

---

distintos conglomerados en una evaluación de cerca de 150.000 observaciones hechas durante un periodo de 8 años (1975-1982) en 110 ubicaciones en la provincia de Utrecht (Holanda). La discriminación entre estos conglomerados de calidad del agua requirió solamente siete características químicas y dos físicas. Para una supervisión ideal del sistema la relación entre las características físicas, químicas y biológicas debe conocerse; los métodos de reconocimiento de patrones ofrecen una posible solución a este problema.

La dependencia estacional de la composición de aniones inorgánicos del agua de la laguna de Venecia fue investigada por Marengo et al. (1995) a lo largo del periodo junio-noviembre de 1993. Desafortunadamente en este periodo el clima se comportó de forma atípica en lo que concierne tanto en lluvias como en temperatura y los datos obtenidos para las variables consideradas no permitieron correlacionarlas directamente con la temporada. Sólo el tratamiento de análisis de componentes principales permitió caracterizar las muestras como función de la temporada y el sitio de muestreo y permitió seleccionar los factores discriminantes significativos.

Momen et al. (1996) utilizaron el análisis cluster y el análisis de componentes para identificar patrones temporales y espaciales de química del agua en el Lago George, un lago en el noreste de Nueva York (Estados Unidos) durante el periodo 1981-1993. El lago posee dos almacenamientos principales que tienen similar área y volumen pero diferente estructura de la comunidad biológica, grupos de plancton, área de cuenca y desarrollo de actividades económicas dentro de sus límites. El análisis cluster basado en las componentes principales o en las variables originales, indicó que había varias diferencias en la química del agua entre los dos almacenamientos principales del lago. Se encontró que el análisis cluster fue una herramienta útil para detectar los cambios tanto bruscos (abruptos) y monótonos (graduales) en el tiempo y en el espacio.

Vega et al. (1998) realizaron una investigación sobre un total de 22 variables físico-químicas que fueron analizadas en muestras de agua tomadas cada tres meses durante dos años y medio (1990-1992) en tres estaciones de muestreo localizadas a lo largo de una sección de 25 km del Río Pisuerga, que pertenece a la cuenca del Río Duero (España). El análisis exploratorio de los datos experimentales fue llevado a cabo mediante gráficas de caja (box plots), ANOVA, análisis de componentes principales y reconocimiento no supervisado de patrones (análisis cluster) en un intento de discriminar fuentes de variación de calidad del agua. El PCA permitió la identificación de un número reducido de factores latentes con un significado hidro-químico: contenidos minerales, polución debida al hombre y temperatura del agua.



---

Además, las muestras de agua del río fueron clasificadas en grupos de similar calidad.

Modelos en mínimos cuadrados parciales (PLS) fueron utilizados por Einax et al. (1999) para examinar las relaciones entre las distribuciones de elementos en diferentes compartimentos en el Río Saale y el Río Elba (Alemania). Estas relaciones permitirían hacer predicciones de las concentraciones de elementos en un compartimento al conocerse las concentraciones en otros. La materia de estudio fueron las concentraciones de elementos medidos en el agua y los sedimentos del Río Saale, así como también en el agua y la materia suspendida en el Río Elba. Se dio especial énfasis a la comparación entre el PLS de dos vías y el de tres. La capacidad de predicción del modelo de tres vías no fue claramente superior al PLS de dos vías. Se encontró que no es posible predecir valores que difieran demasiado de los otros valores en el conjunto de datos. Además, la modelación se hizo problemática si la condición del río en su perfil longitudinal es muy variable o si la condición del río es muy distinta en el tiempo de modelación que es desarrollado y en el tiempo cuando las predicciones son efectuadas.

Petersen et al. (2001) aplicaron la técnica de análisis de componentes principales (PCA) a series temporales de concentraciones de nutrientes y parámetros relacionados con la calidad del agua en varias ubicaciones a lo largo del Río Elba (la parte que se encuentra en territorio alemán). Para efectuar el análisis consideraron el río dividido en dos partes (la ribereña y la de estuario), debido al azud Geesthacht. Sus objetivos eran: (1) identificar empíricamente los procesos principales de transporte de nutrientes y (2) proporcionar una descripción reducida del sistema, usando para ello pocos componentes del PCA significativos e interpretables y que puedan ser usados en modelos de diagnóstico simplificado.

Wunderlin et al. (2001) monitorearon 22 parámetros en diferentes estaciones en las partes alta, media y al comienzo de la cuenca baja del Río Suquía (Argentina) durante al menos dos años, incluyendo 232 muestras distintas. Obtuvieron una compleja matriz de datos la cual fue tratada usando las técnicas de reconocimiento de patrones del análisis cluster (CA), análisis factorial/componentes principales (FA/PCA) y el análisis discriminante (DA). El CA rindió buenos resultados como primer método exploratorio para evaluar las diferencias tanto espaciales como temporales. El FA/PCA necesitó 13 parámetros para señalar el 71% de los cambios tanto temporales como espaciales; en consecuencia la reducción de datos del FA/PCA en este caso no fue considerable como se esperaba. La técnica DA mostró los mejores resultados para la reducción de datos y el reconocimiento de patrones para los análisis temporal y

---

espacial. El DA permitió reducir grandemente la dimensionalidad de la matriz de datos inicial señalando unos pocos parámetros que indican los grandes cambios en la calidad del agua, así como también la variación de patrones asociados con las variaciones estacionales, el escurrimiento urbano y las fuentes de polución.

Un total de 207 muestras fueron tomadas y se midieron 19 variables físicas y químicas en el Río Mura en Eslovenia. Para manejar los resultados de todas las medidas Brodnjak-Vončina et al. (2002) emplearon distintos métodos: (i) los métodos estadísticos básicos para la determinación de los valores de la media y la mediana, las desviaciones estándar, los valores mínimo y máximo de las variables medidas y sus coeficientes de correlación mutuos, (ii) el análisis de componentes principales (PCA) y (iii) el método de agrupamiento basado en la red neuronal de Kohonen. De los resultados se concluyó que el método PCA no discriminó lo suficiente debido a que las variables estaban débilmente correlacionadas. Para la clasificación de este tipo de datos el método de redes neuronales artificiales fue más confiable.

Simeonov et al. (2002) elaboraron un estudio que trata de la modelación e interpretación de datos de monitoreo del agua fluvial de la cuenca del Río Saale y sus tributarios, el Ilm y el Unstrut (Alemania). Para un periodo de un año de observación entre septiembre de 1993 y agosto de 1994, fueron medidas veintisiete propiedades químicas y fisicoquímicas para estimar la calidad del agua. La aplicación del análisis cluster, análisis de componentes principales y modelación distribuida en puntuaciones absolutas de componentes principales revelaron información importante acerca del estado ecológico de la región de interés: (i) identificaron dos patrones separados de polución (corriente arriba y corriente abajo de los ríos); (ii) identificaron seis factores latentes responsables por la estructura de datos, con contenidos distintos, para los dos tipos de contaminación indicados y (iii) determinaron la contribución de cada factor latente (fuente de emisión) para la formación de la concentración total de carga química del agua del río.

Montes-Botella et al. (2003), valoraron los parámetros fisicoquímicos, de contaminación, la distribución de metales pesados y su variación estacional en el alto Río Odiel y en las minas El Lomero (España). Para efectuar esta investigación tomaron tres muestras de agua y analizaron entre julio de 1998 y noviembre de 1999. El análisis de componentes principales (PCA) mostró que los parámetros estaban relacionados con los productos de la oxidación de la pirita y sales que eran solubilizadas por la alta acidez generalizada en la oxidación de sulfuros, agrupados en la primera componente, que comprende un 40,88% de la varianza total.

---

Mendiguchía et al. (2004) midieron varios parámetros físico-químicos en 26 estaciones de muestreo localizadas a lo largo del Río Guadalquivir (España), en tres campañas diferentes de 2001 a 2002. Con los resultados construyeron una matriz de datos la cual fue analizada por medio del análisis factorial/análisis de componentes principales (FA/PCA) y el análisis cluster (CA). Este análisis permitió la identificación de cuatro zonas diferentes en el río con distinta calidad del agua. Los resultados obtenidos por PCA y CA se reforzaron mutuamente, confirmando la existencia de las zonas mencionadas.

Parinet et al. (2004) analizaron un sistema de diez lagos eutróficos localizados en la Costa Ivory (Francia), empleando el PCA con coeficientes de correlación lineal. Ellos sugirieron que las relaciones entre las variables analíticas eran capaces de generar mejores descriptores que las propias variables. Mostraron para este caso que es posible reducir de 18 a tan sólo cuatro el número de parámetros analíticos sin un deterioro notable en la representación de la calidad por medio del PCA, lo que hace posible simplificar la descripción del sistema.

La calidad del agua del Río Gomti (India), un tributario principal del Río Ganges, fue monitoreada en ocho sitios diferentes seleccionados en regiones de polución baja, moderada y alta. El monitoreo se hizo regularmente por un periodo de 5 años (1994-1998) para 24 parámetros. La compleja matriz de datos (17.790 observaciones) fue tratada por Singh et al. (2004) con distintas técnicas multivariantes, como el análisis cluster (CA), análisis de componentes principales/análisis factorial (FA/PCA) y análisis discriminante (DA). El CA mostró buenos resultados al representar tres diferentes grupos de similitudes entre los sitios de muestreo, reflejando los distintos parámetros de calidad del agua del sistema fluvial. El FA/PCA identificó seis factores los cuales fueron responsables de explicar el 71% de la varianza total del conjunto de datos. El DA mostró los mejores resultados para la reducción de datos y reconocimiento de patrones, tanto durante el análisis temporal como el espacial. El DA permitió la reducción en la dimensionalidad del extenso conjunto de datos, delineando unos pocos parámetros indicadores responsables de grandes variaciones en la calidad del agua.

De Carlo et al. (2004) investigaron la variabilidad en las concentraciones de elementos traza en arroyos de cuencas en Oahu, Hawai (Estados Unidos). El foco de atención está en el agua y la materia particulada suspendida recolectadas de la cuenca del Canal Ala Wai en Honolulu y también de la cuenca del Arroyo Kaneohe. El análisis de componentes principales de las concentraciones de elementos traza particulados

---

establece autovalores que contabilizan cerca del 80% de la varianza total y separan los elementos traza en tres factores.

Las relaciones entre clorofila a y 16 variables de calidad del agua (químicas, físicas y biológicas) en el Embalse Çamlýdere (Turquía) fueron estudiadas por medio del uso de puntuaciones de componentes principales (PCS) en un análisis de regresión lineal múltiple (MLR), con el propósito de predecir niveles de clorofila a por Çamdevýren et al. (2005). El PCA fue utilizado para simplificar la complejidad de las relaciones entre las variables de calidad del agua.

Una valoración multivariante fue adaptada por McNeil et al. (2005) para la clasificación de un conjunto de datos grande e irregular, de aproximadamente 34.000 muestras de agua superficial acumuladas por más de 30 años. Un método cluster de K-medias en dos etapas fue diseñado para analizar datos químicos en forma de porcentajes de iones principales; la primera etapa de agrupamiento produjo 347 grupos los cuales fueron entonces reagrupados para generar al final nueve clasificaciones de tipos de agua. Los hallazgos de este estudio confirman que la aplicación de estos métodos analíticos puede proporcionar una valoración útil de los controles sobre la composición del agua y soporte a la gestión a nivel regional.

Ulrich Callies (2005) analiza las observaciones de parámetros de calidad del agua, descarga y temperatura a lo largo del Río Elba. Comenta de inicio el estudio efectuado por Petersen et al. (2001), mencionado anteriormente. En su artículo aplica métodos multivariantes más avanzados, la modelación gráfica y el análisis de correlación canónica (CCA), al mismo conjunto de datos. Los resultados del CCA tienden a ser consistentes con las componentes dominantes obtenidas del PCA. A su vez los modelos gráficos permiten una representación más detallada de estructuras de interacción al ser capaces de condensar considerablemente la estructura de asociaciones entre datos.

Técnicas estadísticas multivariantes como el análisis cluster (CA), análisis factorial (FA), análisis de componentes principales (PCA) y análisis discriminante (DA), fueron aplicados a un conjunto de datos de calidad del agua en el Río Gomti (India) por Singh et al. (2005). Los datos se generaron durante tres años (1999-2001) monitoreando en ocho sitios distintos 34 parámetros (9792 observaciones). El DA mostró los mejores resultados para la reducción de datos y el reconocimiento de patrones tanto en el análisis temporal como espacial (necesitó sólo cinco parámetros para la asignación en el análisis temporal y diez parámetros en el espacial). El DA permitió delinear unos

---

pocos parámetros indicadores responsables de las grandes variaciones en la calidad del agua. El FA/PCA ayudó en la identificación de las fuentes responsables de las variaciones de la calidad del agua del río en tres regiones diferentes.

Se tomaron muestras mensualmente en el Río Reconquista (que es un importante tributario del Río de la Plata en Argentina) de cinco sitios del río; entre 18 y 20 parámetros fisicoquímicos fueron determinados en cada muestra. Esta campaña de monitoreo se llevó a cabo en 1995. Castañé et al. (2006) utilizaron el PCA en la ordenación de las muestras (por sitios, temporada y parámetros fisicoquímicos). Con el propósito de identificar tendencias estacionales en la concentración de contaminantes puntuaciones de los casos etiquetadas por temporada fueron expresadas en gráficas. Los resultados sugirieron que la contaminación antropogénica fue la mayor fuente para la predominancia de los parámetros de contaminación y no cambió significativamente a lo largo del tiempo.

Arhonditsis et al. (2006) presentan un estudio con dos casos de aplicación, el Lago Mendota y el Lago Washington (ambos localizados en Estados Unidos) donde la modelación con la ecuación estructural (el cual es un método que proporciona una técnica robusta para estudiar la interdependencia entre un conjunto de variables correlacionadas) es empleada para explorar las estructuras ecológicas (la dinámica del fitoplancton epilimnético en verano). Una metodología jerárquica Bayesiana también es introducida para suavizar las restricciones identificables clásicas y darles un tratamiento estocástico.

El principal objetivo del trabajo realizado por Kowalkowski et al. (2006) está enfocado a la clasificación de la calidad del agua del Río Brda (Polonia) y la evaluación de los datos de polución obtenidos por el monitoreo de mediciones durante el periodo 1994-2002. Las técnicas aplicadas fueron el análisis cluster, análisis de componentes principales, análisis discriminante y análisis factorial. Los resultados obtenidos permitieron determinar grupos naturales y grupos de localizaciones monitoreadas con polución característica similar e identificar las variables discriminantes importantes. También se encontró que algunas localizaciones estaban bajo fuerte influencia de contaminación municipal y algunas otras bajo la influencia de la agricultura dentro del periodo de tiempo observado.

Terrado et al. (2006) identificaron las fuentes de contaminación principales por metales pesados, compuestos orgánicos y otros parámetros físico-químicos en las aguas superficiales del Río Ebro (España) y describieron sus distribuciones temporal y

---

espacial. Para hacer esto usaron técnicas multivariantes y métodos geoestadísticos. Los datos históricos que fueron investigados estaban disponibles por parte de la Confederación Hidrográfica del Ebro (CHE) y cubren años diferentes, desde 1992 para agua, sedimentos y muestras de peces. Debido a la gran información en estas bases de datos y su complejidad la modelación multivariante (usando análisis de componentes principales, PCA) unida a los sistemas de información geográfica (GIS) se propuso para evaluar la calidad ambiental de la cuenca del río Ebro.

El objetivo del trabajo efectuado por Kowalik et al. (2006) fue presentar el uso de técnicas estadísticas establecidas, como el análisis cluster o el análisis de factores y el progreso efectuado en técnicas modernas básicas (el cluster con imágenes, la regresión en mínimos cuadrados parciales de estado múltiple, la proyección pursuit, o la teoría de la información). Estas técnicas fueron aplicadas al proyecto "Leitprojekt Elbe 2000" que consiste en investigaciones extensas desarrolladas acerca de la calidad del agua y los sedimentos de los tributarios germánicos del Río Elba. Las técnicas establecidas sirvieron para resumir los datos y reducir el número de variables necesarias para describir la calidad del agua y los sedimentos. Y los nuevos métodos permitieron la visualización de grandes conjuntos de datos, la extracción de información latente de los mismos y la detección de fuentes de polución o patrones entre diferentes compartimentos ambientales.

Un enfoque basado en el análisis de componentes principales (PCA) fue propuesto por Giussani et al. (2006) con el propósito de estudiar las características químicas (iones principales y secundarios, alcalinidad total, sílice disuelto) de las aguas superficiales y profundas del Lago Como (el cual representa una fuente importante de abastecimiento de agua potable y está ubicado en el Norte de Italia). El método permitió investigar las correlaciones entre los iones analizados de una manera muy obvia y sencilla, usando gráficas bidimensionales. Esta clase de modelo podría ser un mecanismo de validez general, no sólo para propósitos de investigación sino también para auxiliar a las agencias regionales en la supervisión rutinaria de la calidad del agua.

Técnicas estadísticas multivariantes como el análisis cluster (CA), el análisis de componentes principales (PCA), el análisis factorial (FA) y el análisis discriminante (DA), fueron aplicados por Shrestha et al. (2007) para la evaluación de las variaciones temporales y espaciales y la interpretación de un conjunto de datos de calidad del agua muy complejo de la cuenca del Río Fuji, generado durante 8 años (1995-2002), con el monitoreo de 12 parámetros en 13 sitios distintos (14976 observaciones). El análisis cluster jerárquico agrupó los 13 sitios de muestreo en tres conglomerados, que

---

fueron denominados relativamente poco contaminado (LP), contaminado intermedio (MP) y altamente contaminado (HP). El análisis discriminante dio los mejores resultados tanto en el análisis temporal como espacial y proporcionó una importante reducción de datos, delineando unos pocos parámetros indicadores responsables de las grandes variaciones en la calidad del agua.

Kannel et al. (2007) estudiaron la aplicación de técnicas seleccionadas a la cuenca del Río Bagmati en el Valle Kathmandu en Nepal: el análisis cluster, análisis de componentes principales, análisis de factores y el análisis discriminante, para clasificar los datos de calidad del agua en el río y evaluarlos. Los resultados permitieron determinar agrupamientos naturales de estaciones de monitoreo con características similares de polución e identificar las variables discriminantes principales que son importantes para la variación de la calidad del agua en la región, así como revelar las posibles fuentes de contaminación que están afectando a la calidad del agua del río. El análisis permitió agrupar 17 sitios de monitoreo dentro de 3 regiones, con 5 variables discriminantes principales (seleccionadas de 16): Conductividad eléctrica, OD, cloruros, NO<sub>2</sub>N (Nitrógeno como nitrito) y DBO. Los resultados revelaron que algunas localizaciones estaban bajo fuerte influencia de contaminación municipal y algunas otras bajo la influencia de minerales.

Felipe-Sotelo et al. (2007) estudiaron las variaciones locales y temporales de 21 parámetros físico-químicos que se midieron en 23 puntos de muestreo, en tres campañas efectuadas en los ríos Mero, Mendo y Mandeo, cerca de A Coruña (España). La evolución temporal de la calidad del agua fue valorada por medio del análisis de componentes principales de la matriz aumentada (MA-PCA) y el análisis de factores paralelos (PARAFAC). Más aún, el análisis clásico de componentes principales PCA (uno por campaña de muestreo) fue aplicado con propósitos exploratorios y de comparación. Los tres modelos aplicados llevaron a conclusiones similares, sin embargo el MA-PCA fue superior, debido a que el reordenamiento de las puntuaciones proporcionó una visión más directa y conveniente de las variaciones temporales y geográficas de las muestras, además de que podría ser más flexible y adaptable para estudios ambientales.

Un conjunto de datos de calidad del agua superficial de un río en la India, recopilados durante 10 años, fueron analizados por Singh et al. (2007) usando modelos de regresión en mínimos cuadrados parciales (PLSR). Estos fueron aplicados como una forma de valorar y comparar sus capacidades para pronosticar la demanda bioquímica de oxígeno (DBO) del agua del río en términos de sus errores cuadrados medios

---

relativos de validación cruzada. Tanto el modelo tri-PLS y el cuadri-PLS produjeron un error de validación relativamente bajo en comparación con el PLS abierto y capturaron una alta varianza en el modelo. Más aún, el análisis de las gráficas de puntuaciones y saturaciones de los modelos N-PLS pudieron proporcionar información temporal de la evolución de la calidad del agua del río.

Se obtuvieron conjuntos de datos del análisis cuantitativo de diecisiete pesticidas en muestras de agua de una red de canales de irrigación y drenaje en el delta del Río Ebro (España), los cuales fueron analizados mediante técnicas multivariantes y métodos geoestadísticos. Las muestras fueron tomadas en catorce localizaciones desde mayo hasta agosto de 2005. Terrado et al. (2007) por medio del análisis de componentes principales, investigaron la distribución espacial y temporal de los patrones de polución principales, causados por la aplicación de pesticidas en la región bajo estudio.



---

## **CAPÍTULO 3. ESTADO DEL ARTE DE LA MODELACIÓN CON MÉTODOS MECANICISTAS DE LA CALIDAD DEL AGUA EN EMBALSES**

### **3.1 NECESIDAD DE LOS MODELOS MECANICISTAS DE CALIDAD DEL AGUA PARA EMBALSES**

Los modelos matemáticos son herramientas valiosas para la gestión y evaluación de sistemas acuáticos (Deas et al., 2000).

Se puede definir un modelo matemático como una formulación idealizada que representa la respuesta de un sistema físico a un estímulo externo (Chapra, 1997). La principal ventaja de estos modelos estriba en el hecho de que son herramientas generales aplicables a diferentes condiciones locales (Deas et al., 2000).

Con mayor frecuencia se usan “modelos” para predecir el futuro. Estos típicamente incluyen un conjunto de relaciones que, ya sea a través de correlaciones o a través de funciones de causa y efecto, pretenden generar una mayor comprensión del proceso o procesos. A varios niveles, los modelos proporcionan representaciones de sistemas naturales complejos (Deas et al., 2000).

Los propósitos al efectuar una modelación son tener una base más racional para tomar decisiones dentro del control de la calidad del agua (Atanasova et al., 2006).

### **3.2 ESTADO DEL ARTE**

Entre los primeros intentos de caracterización matemática del comportamiento de un embalse están los de los limnólogos que intentaban explicar el ciclo termal anual y la formación de la termoclina. McEwen (1929) y Hutchinson (1957) estimaron los coeficientes de transferencia de calor usando perfiles de temperatura propuestos. Ertel (1954) diseñó un modelo difusional de la formación de la termoclina asumiendo un coeficiente constante de difusión, independientemente del tiempo o la profundidad. Sin embargo, estos esfuerzos no fueron exitosos en producir una representación creíble del ciclo anual completo de energía termal en lagos o presas (Orlob, 1981).

La necesidad de simular el ciclo anual de temperatura del agua superficial de almacenamientos creció con la construcción postergada, durante el periodo posterior al fin de la Segunda Guerra Mundial, de grandes presas en el occidente de los Estados Unidos (Orlob, 1981).

---

Durante el comienzo de los años 60's del siglo XX la preocupación de los grupos de conservación y biólogos por la supervivencia de la vida acuática en los entornos sensibles a la temperatura estimuló a las agencias de gobierno a patrocinar investigaciones para mejorar los métodos para la predicción del ciclo de temperatura anual en las presas en operación (Novotny, 2003).

El primer modelo matemático, una técnica manual, fue empleado en el Central Valley en los años 60's. Raphael (1962) aplicó esta técnica para el cálculo del balance de energía termal en almacenamientos propuestos, la cual empleó con éxito en la Presa Oroville sobre el Río California's Feather y a varios embalses en el Sistema del Río Columbia. El método permitía una estimación razonable de las temperaturas río abajo para estos proyectos pero falló en proporcionar una descripción de la distribución vertical de calor dentro del embalse (Deas et al., 2000).

La modelación matemática de almacenamientos de agua superficial se desarrolló en los Estados Unidos a mediados de los 60's cuando los esfuerzos para el control de la contaminación fueron dirigidos hacia la solución de problemas de enriquecimiento de nutrientes en los almacenamientos naturales y artificiales. La atención inicial fue para ejemplos notorios de avanzado deterioro de la calidad del agua como el Lago Erie, pero esta preocupación por la protección de otros lagos y embalses creció cuando comenzaron a mostrarse los primeros signos de eutrofización acelerada (Orlob, 1981).

Precisamente en esta época se inició un esfuerzo por parte de dos grupos independientes con el propósito de crear modelos para la predicción de cambios en la energía termal en los embalses profundos estratificados y corrientes de agua. Uno de los grupos era una firma privada de consultoría, Water Resources Engineers Inc. (WRE), que desarrolló los conceptos fundamentales para la predicción de la distribución de energía termal en corrientes y embalses. El otro grupo era del Parson's Hydraulic Lab en el Massachusetts Institute of Technology, que elaboró un modelo funcional para la simulación de almacenamientos profundos (Deas et al., 2000).

A comienzos de 1966 WRE y TVA (Tennessee Valley Authority) colaboraron en el desarrollo de un modelo, donde TVA contribuyó con las bases necesarias para cuantificar el intercambio de calor. Este esfuerzo combinado culminó en un reporte al CDFG (California Department of Fish and Game) en junio de 1967, el cual fue revisado un año más tarde, donde la Presa Fontana fue utilizada como caso de prueba (Orlob, 1981).

---

En ese mismo año Harleman y sus colaboradores desarrollaron en el MIT y con una colaboración estrecha con el TVA Laboratory experimentos de laboratorio. Este trabajo fue extendido a modelar embalses prototipo usando los datos del TVA de la Presa Fontana (Orlob, 1981).

Uno de los primeros estudios en el Central Valley fue elaborado por WRE en la cuenca del Río Feather a fines de los 60's por medio de un modelo de ordenador que fue usado para predecir la temperatura del agua para un almacenamiento –el proyecto Marysville (Orlob, 1981).

A comienzos de la década de 1970 el US Bureau of Reclamation (USBR) adoptó y fue aplicando activamente un modelo de simulación computacional de temperatura del agua en varios embalses sobre el cauce principal de la cuenca del Río Sacramento. Las primeras aplicaciones de modelos para ordenador se dirigían a las condiciones de temperatura del agua río abajo de los embalses para restaurar y/o mantener determinadas especies de peces (el mismo propósito continúa motivando la modelación de temperatura hasta el día de hoy) (Orlob, 1981).

En la misma época el USBR aplicó el modelo RTS (Reservoir Temperature Stratification) a los embalses Shasta y Folsom para simular a escala mensual las condiciones y respuesta termal (Deas et al., 2000).

El SNTMP, un modelo estacionario con fundamento físico, es un modelo de transporte de calor en una dimensión que predice las temperaturas media y máxima del agua como función de la distancia a la corriente y el flujo de calor ambiental. Fue desarrollado por Theurer et al. en 1984 (Orlob, 1981).

El Water Quality for River-Reservoir Systems (WQRRS, 1986) y el Water Quality Simulation Module HEC5-Q (HEC-5Q, 1987) se desarrollaron a partir del trabajo hecho por el WRE (1969), Chen y Orlob (1972), así como Beard y Willey (1972), además de otros modelos anteriores. Ambos modelos han sido ampliamente aplicados en el análisis de temperaturas en el Central Valley (Deas et al., 2000).

El USACE-HEC (United States Army Corps of Engineers-Hydrologic Engineering Center) aplicó el HEC-5Q al Sistema de presas del Sacramento Valley para ilustrar la aplicación de este modelo embalse-río al análisis de la calidad del agua (1987) (Deas et al., 2000).

---

En 1988 Smith aplicó el HEC-5Q a un conjunto similar de componentes embalse-río, pero no incluyó el Embalse de Oroville (Deas et al., 2000)

Christiansen y Orlob (1989) aplicaron los modelos del USBR a los embalses Shasta, Trinity, Whiskeytown y Folsom, así como los modelos de ríos asociados para los ríos Trinity, Sacramento y American, con el fin de valorar la predicción de temperatura y el desempeño del caudal. Estos modelos han sido los más ampliamente aplicados y usados continuamente para modelar temperatura en el Central Valley y posiblemente en los Estados Unidos (Orlob, 1981).

El Box Exchange Transport Temperature and Ecology of Reservoirs (BETTER) es un modelo bidimensional para temperatura y calidad del agua que se comenzó a emplear en 1990 (Deas et al., 2000).

La Presa New Bullards Bar en el Río Yuba, fue modelada con el CE-QUAL-R1 por Bookman-Edmonston (1991). El CE-QUAL-R1 es un modelo vertical unidimensional para embalses que fue una extensión del WQRRS (Deas et al., 2000).

Orlob et al. (1993) y Meyer y Orlob (1994) usaron el modelo WQRRS para investigar los efectos del cambio climático en la calidad del agua, incluida la temperatura (Orlob, 1981).

El RMA-6 (Resource Management Associates Inc.) es un modelo bidimensional uniformemente mezclado lateralmente, que fue aplicado a la Presa Keswick por Anderson (1994) (Deas et al., 2000).

Deas et al. (1997) aplicó los modelos desarrollados por Meyer y Orlob para simular la respuesta en la temperatura del agua para operaciones alternativas en la restauración de peces en el Río Sacramento, río abajo del embalse Keswick.

Deas (1998) aplicó el WQRRS en la Presa Trinity para examinar la extracción selectiva y cuestiones de almacenamiento para control de la temperatura del agua en el Río Trinity, río abajo del Embalse Lewiston (Deas et al., 2000).

Jensen et al. (1999) implementó el RMA-10 y el RMA-11 para caracterizar la compleja hidrodinámica y el régimen termal del almacenamiento Whiskeytown en tres dimensiones (Deas et al., 2000).

Periodo	Modelo	Aplicaciones	Investigador principal o Desarrollador del Modelo
1958-1963	BURT	Presas del Río Snake	Burt (Water Reserach Assoc.)
1961-1962	RAPHAEL	Embalses del Río Columbia	Raphael
1965-1968	WRE	Presas Fontana, Lago Washington, entre otros	Norton, Selna, Orlob, Roesner, et al. (WRE)
1965-1968	MIT	Modelos de laboratorio, Presas Fontana	Harleman, Dake, Huber, Ryan (MIT)
1968-1970	CORNELL	Lago Cayuga	Sundaram, et al. (Cornell Univ. y Aerospace Lab.)
1968-1973	LAKECO	Lago Washington, Lago Lincoln, Isla Tocks	Chen, Norton, King, Orlob, et al. (WRE)
1968-1969	WRE	Lago Roosevelt	Norton, Roesner, et al. (EPA y WRE)
1969-1975	2D-FEM	No conocidas	Liggett, Gallagher et al. (Cornell University)
1971-1973	MIT	Presas Fontana	Harleman, Markofsky (MIT-EPA)
1972-1975	HYDROSCIENCE	Lago Erie occidental, Lago Ontario	Di Toro, O'Connor, Toman, et al. (Great Lakes Basin Comm. Y EPA)
1972-1990	HEC-USBR	Presas Shasta, Whiskeytown, Oroville, Folsom, Trinity, Lewiston, Keswick, Ríos Sacramento, Feather, American	Rowell
1972-1976	SIMONS	Lago Ontario	Simons (Canada Centre for Inland Waters)
1973-1975	2D-FEM	Embalse Lower Granite	Norton, King et al. (WRE y Resource Management Associates, Inc.)
1974-1975	BACA-ARNETT	Lago Washington, Lago Mendota	Baca, Arnett (Batelle Memorial Inst.)
1974-1975	LK ONTARIO	Lago Ontario	Chen, Lorenzen, Smith (Tetra Tech-NOAA)
1979	SELECT	Presas Shasta, Folsom	Rowell
1986	HEC-5Q	Presas Shasta, Oroville, Folsom, New Melones, Tulloch	Smith, USACE-HEC,
1986	WQRRS	Presas Shasta, Trinity	Meyer, Deas
1990	FLOW-3D	Presas Shasta	Rowell
1990	BETTER	Embalses Whiskeytown, Lewiston	Rowell, JSA
1994	RMA	Presas Keswick	Anderson
1997	RMA	Presas Keswick	Deas
1999	CE-QUAL-W2	Embalse Shasta	Hanna et al.
1999	RMA	Embalse Whiskeytown	Jensen et al.

**TABLA 1** Breve cronología del desarrollo de aplicaciones de los modelos matemáticos en almacenamientos estratificados (Fuentes: Orlob, G. T. 1981. "Models for stratified impoundments". Y Deas, M. L., Lowney, C. L. 2000. "Water temperature modelling review")

---

Han et al. (2000) usaron el modelo 1D DYRESM para aplicarlo al embalse Sau, un embalse en un valle fluvial en el noreste de España.

Hornung, R. (2002) utilizó el modelo hidrodinámico 1-D DYRESM para modelar la estratificación en el Lago Constance, uno de los lagos más grandes de Europa.

Gal et al. (2003) usaron el modelo DYRESM para simular la estructura termal del Lago Kinneret ubicado en Israel, para un periodo amplio de 45 meses.

Bruce et al. (2006) cuantificaron el rol del zooplancton en los ciclos de nutrientes en el Lago Kinneret usando los modelos DYRESM y CAEDYM.

De la Fuente et al. (2008) presentaron una simulación eco-hidrodinámica de periodo amplio para un embalse dendrítico, el Rapel, localizado en Chile central. Para ello emplearon el modelo DYRESM-CAEDYM y el modelo AQUASEA.

---

## CAPÍTULO 4. PANORAMA DE LOS MÉTODOS MULTIVARIANTES

### 4.1 INTRODUCCIÓN

Se puede definir a los métodos multivariantes como el conjunto de métodos estadísticos para el tratamiento conjunto de variables con el propósito de analizar, describir e interpretar datos de múltiples variables (Gil Pascual, 2003). Los métodos estadísticos multivariantes pueden usarse para visualizar grandes cantidades de datos y extraer información latente (Kowalik et al., 2006).

Para efectuar una investigación con métodos multivariantes se supone que se han medido  $p$  variables numéricas en un conjunto de  $n$  elementos. Cada una de esas variables se denomina *variable escalar* (o *univariante*) y la agrupación de las  $p$  variables constituyen una *variable multivariante*. Las magnitudes de las  $p$  variables univariantes en los  $n$  elementos pueden representarse por medio de una matriz  $\mathbf{X}$ , de tamaño  $(n \times p)$  (Peña, 2002):

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_p] \quad (2)$$

Donde cada variable  $\mathbf{x}_i$  es un vector columna  $(n \times 1)$ , que representa a la variable escalar  $i$  medida en los  $n$  elementos del conjunto (Peña, 2002).

En los programas de monitoreo de la calidad del agua se generan datos multidimensionales complejos que requieren un tratamiento estadístico multivariante para el análisis e interpretación de la información subyacente (Singh et al., 2004).

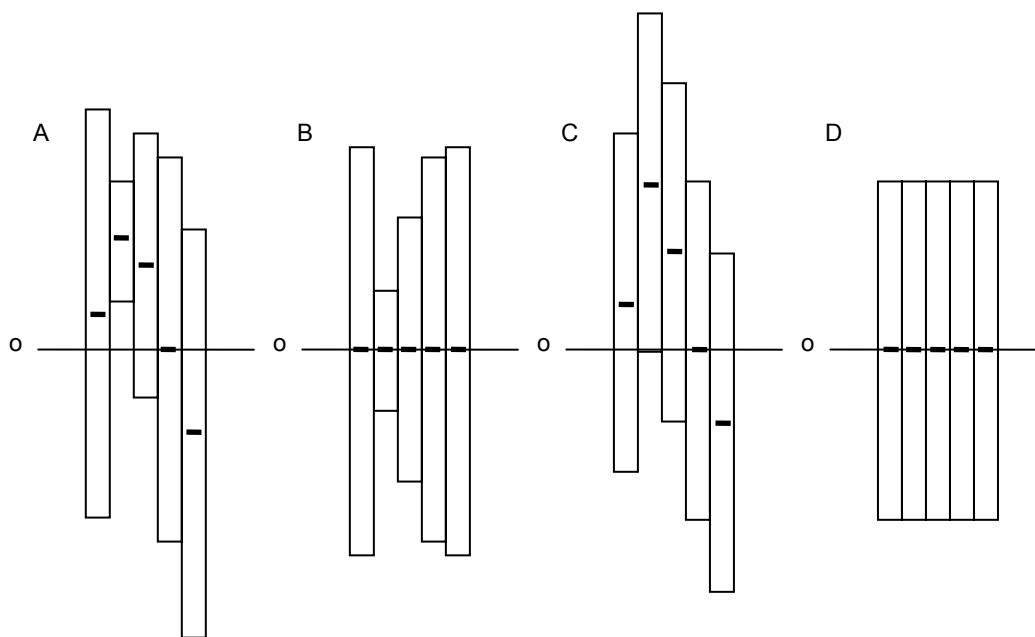
El mejor planteamiento para evitar la interpretación equívoca de los datos de monitoreo ambiental es la aplicación de métodos multivariantes (Quimiométricos o Chemometrics) (Simeonov et al., 2002). No requieren de conocimiento previo de los valores de las constantes o coeficientes involucrados en el fenómeno en estudio. Lo que se busca generalmente con estas técnicas es la simplificación (Blanco et al., 1995).

Los tipos de investigaciones en las que son utilizados los métodos multivariantes se podrían agrupar en los siguientes casos (Martínez Arias, 1999):

- 
1. Reducción de datos.
  2. Clasificación y agrupación.
  3. Análisis de las relaciones de dependencia.
  4. Construcción de modelos y pruebas (tests).

Los datos multivariantes se pueden organizar en arreglos (matrices) lo cual facilita su exposición y permite cálculos numéricos que pueden ser desarrollados de una forma ordenada y eficiente. La eficiencia conseguida se manifiesta en dos sentidos: (i) la descripción de los cálculos numéricos como operaciones matriciales y (ii) su implementación en ordenadores (Johnson et al., 1992).

Con el propósito de que los niveles de variación de las diferentes variables tengan algún ordenamiento de nivel equivalente, condición que es requerida en algunos métodos multivariantes, se efectúa el escalado (Bro et al., 2003).



**FIGURA 1** Pre-procesamiento de datos. Los datos para cada variable son representados por una barra de varianza y su centro. (A) Los datos sin tratar se observan así. (B) El resultado después de centrado en media solamente. (C) El resultado después de escalar la varianza solamente. (D) El resultado después de centrado en media y escalamiento de la varianza (Fuente: Geladi et al., 1986)

El escalado de la varianza es utilizado cuando las variables en el conjunto son medidas en unidades distintas ( $m$ ,  $m^3/s$ , etc.); el escalado es llevado a cabo dividiendo todos los valores de una cierta variable por la desviación estándar de dicha variable de manera que la varianza para cada variable es la unidad (Figura 1). Asimismo, se puede tomar la decisión de que ciertas variables son de menor importancia y por lo tanto no deberán influenciar mucho al modelo, por lo que se les puede dar un menor



---

peso. Sin embargo, se considera que lo más apropiado es que todas las variables tengan el mismo tipo de escalado (Geladi et al., 1986).

El escalado de las variables es usado por varias razones. Algunas de las más importantes son (Bro et al., 2003):

- Ajustar las diferencias de escala.
- Arreglar la heteroscedasticidad.

El autoescalado -o normalización, estandarización- consiste en centrar las variables en la media y escalar la desviación estándar a la unidad para que la varianza de cada variable sea idéntica inicialmente (Bro et al., 2003). Se mencionará expresamente cuando en alguno de los métodos multivariantes utilizados a lo largo de esta tesis sea necesario aplicar la estandarización.

Muchos de los métodos multivariantes tienen que cumplir una serie de suposiciones: normalidad, linealidad y homoscedasticidad (Martínez Arias, 1999).

#### 4.1.1 Suposición de distribución normal multivariante

La mayoría de las técnicas multivariantes están basadas en la suposición de que los datos (los vectores de observaciones  $\mathbf{x}_i$ ) fueron generados de una distribución normal multivariante, es decir, que cada variable y todas las combinaciones lineales de las variables están normalmente distribuidas. En algún grado la calidad de las inferencias realizadas con algunos métodos multivariantes depende de que tan cercana este la población en estudio a parecerse a una forma normal multivariante (Johnson et al., 2004).

Otros métodos se pueden aplicar sin contemplar esta suposición y están justificados por argumentos lógicos o de sentido común (Johnson et al., 2004).

Cuando se cumple el supuesto los residuos o la parte de la variación de las variables originales no explicada por el modelo siguen una distribución normal (Martínez Arias, 1999).

El supuesto de normalidad multivariante es requerido cuando la técnica se usa con un propósito inferencial y no así cuando su fin es solamente exploratorio (Martínez Arias, 1999).

---

Gráficos especiales llamados gráficos Q-Q, pueden ser empleados para valorar la suposición de normalidad. Estos gráficos pueden ser desarrollados para las distribuciones marginales de las observaciones muestrales en cada variable. Cuando los puntos yacen muy cercanos a lo largo de una línea recta la suposición de normalidad permanece. La normalidad es sospechosa si los puntos se desvían de una línea recta (Johnson et al., 1992).

Los gráficos Q-Q no son particularmente informativos a menos que el tamaño de muestra sea moderadamente grande; por ejemplo  $n \geq 20$  (Johnson et al., 1992).

Algunos paquetes estadísticos proporcionan resultados gráficos para examinar la normalidad univariante, además de pruebas estadísticas basadas en el contraste de Kolmogorov-Smirnov (con la corrección de Lilliefors). Los gráficos más utilizados son el histograma, el diagrama de tallo y hojas (Stem and leaf Plot), el diagrama de caja y bigotes (Box and Whisker Plot) y especialmente el gráfico Q-Q (Q-Q Plot), que es una linealización de la distribución normal (Martínez Arias, 1999).

#### 4.1.2 Suposición de homoscedasticidad

Este supuesto se aplica a los métodos de dependencia y supone, para el caso de una única variable, la igualdad de las varianzas en todos los grupos o en las diferentes combinaciones de valores de las variables independientes. Cuando hay varios grupos involucrados el supuesto se extiende a la “igualdad de las matrices de varianzas y covarianzas” de los distintos grupos en el análisis (Martínez Arias, 1999).

En este supuesto se requiere que las variables dependientes muestren varianzas iguales para todo el rango de valores de las variables predictoras, para el caso de que las variables dependientes sean cuantitativas. La prueba del supuesto suele realizarse por medio de un análisis de la distribución de los residuos (Martínez Arias, 1999).

#### 4.1.3 Suposición de linealidad

Esta suposición de linealidad establece que las relaciones entre todas las variables analizadas es de tipo lineal exclusivamente (Martínez Arias, 1999).

En algunos métodos la presencia de la no linealidad puede diagnosticarse a partir de algunos diagramas de los residuos. Sin embargo, la manera más común de obtener información acerca de la linealidad de las relaciones entre variables es revisando los diagramas de dispersión entre pares de variables. Cuando interesa inspeccionar

---

algunas variables la manera más rápida de proceder es utilizando en algunos paquetes estadísticos la opción matricial que genera diagramas de dispersión para grupos de variables (Martínez Arias, 1999).

Cuando las variables exhiban multicolinealidad los datos proporcionan información redundante por lo que no se requiere contar con todas en el mismo análisis (Martínez Arias, 1999). Por lo tanto las variables redundantes deben ser eliminadas del análisis.

#### 4.1.4 Transformaciones de datos

Las transformaciones de los datos proporcionan un medio de modificar las variables para enmendar el no cumplimiento de algunos de los supuestos antes mencionados (Martínez Arias, 1999).

Se ha demostrado teóricamente que las observaciones pueden frecuentemente ser consideradas más apegadas a una distribución normal tomando sus raíces cuadradas. Esto es (Johnson et al., 1992):

Escala original	Escala transformada
Variable de escala, $y$	$\sqrt{y}$

**TABLA 2** Transformación para aproximar los datos a una normal

Cuando las variables difieren moderadamente de la normalidad mostrando una ligera asimetría (skewness) positiva y no tienen curtosis, la transformación anterior, raíz cuadrada, generalmente proporciona buenos resultados. En el caso de asimetría negativa, una mejora similar puede conseguirse elevando al cuadrado los valores originales ( $x_i^2$ ). En las situaciones de asimetría positiva fuerte y curtosis elevada, produce buenos resultados la transformación logarítmica ( $\log x_i$ ). En los casos de asimetría extrema con distribuciones en forma de L, se emplea la transformación recíproca ( $1/x_i$ ) (Martínez Arias, 1999).

Una familia útil de transformaciones para este propósito es la familia de transformaciones potencia. Supóngase que  $x$  represente a una observación arbitraria. Las transformaciones potencia tienen como parámetro a la letra del alfabeto griego  $\lambda$ . Un valor dado para  $\lambda$  implica una transformación particular. Por ejemplo, considerando a  $x^\lambda$  con  $\lambda = -1$ , se tendría  $x^{-1} = 1/x$ , en este caso la selección de  $\lambda$  corresponde a la transformación recíproca. Cálculos de prueba y error con algunas de las transformaciones indicadas anteriormente podrían producir una mejora. La selección

---

final deberá siempre ser examinada por medio de un gráfico Q-Q u otras pruebas para ver si la suposición tentativa de normalidad es satisfactoria (Johnson et al., 1992).

Box y Cox consideraron una familia de transformaciones potencia ligeramente modificada (Johnson et al., 1992):

$$\mathbf{x}^{(\lambda)} = \begin{cases} \frac{\mathbf{x}^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln \mathbf{x} & \lambda = 0 \end{cases} \quad (3)$$

La cual es continua en  $\lambda$  para  $x > 0$ . Dadas las observaciones  $x_1, x_2, \dots, x_n$ , la solución de Box-Cox busca una potencia  $\lambda$  apropiada que maximice la expresión (Johnson et al., 1992):

$$\ell(\lambda) = -\frac{n}{2} \ln \left[ \frac{1}{n} \sum_{j=1}^n (\mathbf{x}_j^{(\lambda)} - \bar{\mathbf{x}}^{(\lambda)})^2 \right] + (\lambda - 1) \sum_{j=1}^n \ln \mathbf{x}_j \quad (4)$$

Y la definición de  $\bar{\mathbf{x}}^{(\lambda)}$  es:

$$\bar{\mathbf{x}}^{(\lambda)} = \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j^{(\lambda)} = \frac{1}{n} \sum_{j=1}^n \left( \frac{\mathbf{x}_j^\lambda - 1}{\lambda} \right) \quad (5)$$

Para encontrar el valor óptimo de  $\lambda$  se proponen varios valores y entonces se utiliza el que maximice la expresión (4) (Johnson et al., 1992).

La familia de transformaciones potencia de Box y Cox (1964) es ampliamente utilizada para obtener una transformación que normalice una variable respuesta de valores positivos (Edwards et al., 1995).

La transformación de Box-Cox logra estabilizar la varianza en una serie temporal y aproximar la distribución que presenta a una normal. Una variante de la expresión (3) para el caso de la transformación logarítmica es (Pérez, 2005):

$$\mathbf{x}^{(\lambda)} = g \ln(\mathbf{x} + \ell) \quad (6)$$

Donde  $g$  es la media geométrica de  $\mathbf{x} + \ell$ . El parámetro  $\ell$  se elige de manera que el resultado de la expresión  $\mathbf{x} + \ell$  sea siempre positiva (mayor que cero). Así  $\ell$  tomará el valor cero en caso de que todos los valores de la variable sean positivos y tomará un

---

valor igual en valor absoluto al valor negativo de mayor valor absoluto observado en la variable correspondiente (Pérez, 2005).

Para variables aleatorias multivariantes el procedimiento más simple es transformar cada variable aleatoria usando el procedimiento univariante adecuado. Aunque este planteamiento no garantiza la normalidad multivariante generalmente es suficiente bueno en la práctica (Jobson, 1992).

Se recomienda aplicar los siguientes pasos al efectuar análisis multivariantes:

1. Representar gráficamente las variables individualmente mediante un histograma o un diagrama de caja. Estas representaciones son muy útiles para detectar:
  - Asimetrías.
  - Heterogeneidad.
  - Datos atípicos o anormales (valores fuera de rango). El primer paso para identificar las observaciones sospechosas es detectar aquellas que lo sean claramente respecto a una variable. Para efectuar esto se puede emplear el histograma o los diagramas de caja (Peña, 2002). Los datos atípicos pueden de esta manera ser más fácilmente detectados.
2. Detectar la cantidad y distribución de valores perdidos y plantear soluciones al problema, según las circunstancias del caso. Por ejemplo, si la cantidad de valores perdidos no es muy grande, se podrían tomar valores medios como una primera aproximación del análisis.
3. Las variables que tengan muy pocos valores por encima del límite de detección se removerán antes de la aplicación del análisis multivariante (Terrado et al., 2006).
4. Construir los diagramas de dispersión de las variables por pares. Este procedimiento está ya incorporado en muchos paquetes estadísticos. Estos gráficos son importantes para apreciar si existen relaciones no lineales entre las variables (Peña, 2002).
5. Identificar y tratar las variables no normales (Martínez Arias, 1999):
  - Análisis de la asimetría y la curtosis y revisar los gráficos de probabilidad normal.
  - Transformar las variables si la aplicación del punto anterior así lo sugiere y

---

probar los resultados de las transformaciones realizadas.

6. Examinar mediante contrastes estadísticos la homogeneidad de las varianzas y la igualdad de las matrices de varianza covarianza. Cuando las variables independientes son cuantitativas pueden examinarse los gráficos de residuos (Martínez Arias, 1999).
7. Evaluar las variables en cuanto a la multicolinealidad y singularidad (Martínez Arias, 1999). Muchos paquetes estadísticos incorporan la posibilidad de valorar estos aspectos.
8. Generalmente las variables se estandarizan, de manera que tengan media cero y desviación típica unitaria (Peña, 2002). Este paso habitualmente genera una mejora significativa en los resultados de los análisis aplicados.

Para observar como se aplicaron algunos de estos pasos se pueden consultar las secciones 7.3 y 7.4.

#### 4.2 ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)

Se pueden emplear todas las variables para obtener un pequeño conjunto de nuevas variables que puedan usarse como una aproximación del conjunto original. Estas nuevas variables reciben el nombre de *componentes principales* (Jobson, 1992). El PCA se usa como un procedimiento matemático que transforma un conjunto de variables correlacionadas en un nuevo conjunto de variables no correlacionadas (las *componentes principales*) (Johnson, 1992). Esta técnica está siendo cada vez más popular en la modelación ecológica y ambiental (Marengo et al., 1995; Vega et al., 1998; Wunderlin et al., 2000; Petersen et al., 2001; Brodnjak-Vončina et al., 2002; Simeonov et al., 2002; Montes-Botella et al., 2003; Mendiguchía et al., 2004; Singh et al., 2004; De Carlo et al., 2004; Parinet et al., 2004; Çamdevýren et al., 2005; Singh et al., 2005; Giussani et al., 2006; Kowalkowski et al., 2006; Terrado et al., 2006; Castañé et al., 2006; Felipe-Sotelo, 2007; Terrado et al., 2007; Shrestha et al., 2007; Kannel et al., 2007).

Las componentes principales son transformaciones lineales de las variables originales y son mutuamente ortogonales (Jobson, 1992).

Aunque se requieren  $p$  componentes para reproducir la variabilidad total del sistema, frecuentemente mucha de esta variabilidad puede ser representada por un pequeño número  $k$  de componentes principales. Si esto ocurre habrá casi tanta información en

---

las  $k$  componentes como en las  $p$  variables originales (Johnson et al., 1992). Una de las razones de mayor peso para realizar un análisis de componentes principales, PCA, es para emplearlo como instrumento para filtrar los datos multivariantes. Se pueden crear nuevas variables llamadas puntuaciones (*scores*) de las componentes principales que se pueden emplear como entrada para el trazado de gráficas y situación de los casos (*observaciones* o *registros*) y, a menudo, una revisión de estas gráficas pondrán de manifiesto las anomalías en algunos datos (*outliers*) (Johnson, 2000).

Por lo tanto la detección de observaciones anormales (y su posible descarte) es importante (Martens et al., 1989). Una de las herramientas que puede auxiliar en este proceso es el PCA.

Sin embargo, en la calibración un dato anómalo puede algunas veces ser altamente informativo y no sólo un error: puede ser anormal debido sólo a un lapso de cierto tipo de variabilidad importante en las variables predictivas (Martens et al., 1989).

El proceso para obtener las componentes principales es el siguiente: sean  $x_{11}, x_{12}, \dots, x_{1n}$ , las  $n$  medidas de la primera variable. La media aritmética de estas medidas representada por  $\bar{x}_1$ , esta dada por (Johnson et al., 1992):

$$\bar{x}_1 = \frac{1}{n} \sum_{j=1}^n x_{1j} \quad (7)$$

Si las  $n$  medidas representan un subgrupo del conjunto completo de medidas que podrían haber sido observadas, entonces  $\bar{x}_1$  es conocida también como la media muestral de la primera variable (Johnson et al., 1992).

La media muestral puede calcularse de las  $n$  medidas de cada una de las  $p$  variables, por lo que en general habrá  $p$  medias muestrales (Johnson et al., 1992):

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij} \quad i = 1, 2, \dots, p \quad (8)$$

Una medida de dispersión es proporcionada por la varianza muestral definida para  $n$  medidas en la primera variable como (Wackerly et al., 2002):

$$S_i^2 = \frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 \quad (9)$$

En general para  $p$  variables se tiene (Johnson et al., 1992; Wackerly et al., 2002):

$$S_i^2 = \frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 \quad i = 1, 2, \dots, p \quad (10)$$

Una observación debe efectuarse respecto a las ecuaciones (9) y (10). Generalmente se define la varianza muestral con un divisor  $n - 1$ . Existen razones teóricas para hacer esto y es particularmente apropiado si el número de mediciones,  $n$ , es pequeño (Johnson et al., 1992).

Puede mostrarse que (Wackerly et al., 2002):

$$\sum_{j=1}^n (x_j - \bar{x})^2 = \sum_{j=1}^n x_j^2 - \frac{1}{n} \left( \sum_{j=1}^n x_j \right)^2 \quad (11)$$

Y después (Wackerly et al., 2002):

$$\sum_{j=1}^n (x_j - \bar{x})^2 = \sum_{j=1}^n x_j^2 - n\bar{x}^2 \quad (12)$$

Por lo tanto:

$$E \left[ \sum_{j=1}^n (x_j - \bar{x})^2 \right] = E \left( \sum_{j=1}^n x_j^2 \right) - nE(\bar{x}^2) \quad (13)$$

Y así:

$$E \left[ \sum_{j=1}^n (x_j - \bar{x})^2 \right] = \sum_{j=1}^n E(x_j^2) - nE(\bar{x}^2) \quad (14)$$

Nótese que  $E(x_j^2)$  es el mismo para  $i = 1, 2, \dots, n$ . Usando esta información y el hecho de que la varianza de una variable aleatoria está dada por  $V(\mathbf{x}) = E(\mathbf{x}^2) - \boldsymbol{\mu}^2$ , se concluye que  $E(\mathbf{x}_i^2) = V(\mathbf{x}_i) + (E(\mathbf{x}_i))^2 = \sigma^2 + \mu^2$ ,  $E(\bar{x}^2) = V(\bar{x}) + (E(\bar{x}))^2 = \sigma^2 / n + \mu^2$ , llegando a (Wackerly et al., 2002):



---


$$E \left[ \sum_{j=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})^2 \right] = \sum_{i=1}^n (\sigma^2 + \mu^2) - n \left( \frac{\sigma^2}{n} + \mu^2 \right) \quad (15)$$

Donde:

$\mu$  = Media de la población.  
 $\sigma^2$  = Varianza de la población.

Reduciendo:

$$E \left[ \sum_{j=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})^2 \right] = n(\sigma^2 + \mu^2) - \sigma^2 + n\mu^2 \quad (16)$$

Simplificando (Wackerly et al., 2002):

$$E \left[ \sum_{j=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})^2 \right] = n\sigma^2 - \sigma^2 \quad (17)$$

Y factorizando:

$$E \left[ \sum_{j=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})^2 \right] = (n-1)\sigma^2 \quad (18)$$

Despejando la varianza de la población (Wackerly et al., 2002):

$$\sigma^2 = \frac{1}{n-1} E \left[ \sum_{j=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})^2 \right] \quad (19)$$

Es necesario aclarar que la ecuación (19) constituye un estimador imparcial (no sesgado) de la varianza de la población (Wackerly et al., 2002).

Como la notación  $S^2$  es tradicionalmente empleada para indicar la varianza muestral, al considerarse arreglos (matrices) de cantidades las varianzas muestrales caen a lo largo de la diagonal principal. En este caso es conveniente usar dobles subíndices en las varianzas, con el fin de indicar sus posiciones en el arreglo. Por lo tanto se tendrá (Johnson et al., 1992):

$$S_i^2 = S_{ii} = \frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 \quad i = 1, 2, \dots, p \quad (20)$$

---

La covarianza muestral, una medida de asociación entre las variables  $i$  e  $k$  se calcula con la expresión siguiente (Johnson et al., 1992):

$$S_{ik} = \frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \bar{x}_i)(x_{kj} - \bar{x}_k) \quad i = 1, 2, \dots, p, \quad k = 1, 2, \dots, p \quad (21)$$

Se observa que la covarianza se reduce a la varianza muestral cuando  $i = k$ . Más aun,  $S_{ik} = S_{ki}$  para todas las  $i$ , y las  $k$  (Johnson et al., 1992).

El Coeficiente de correlación muestral (o el Coeficiente de correlación de productos de momentos de Pearson), es una medida de la asociación lineal entre dos variables que no depende de las unidades de medida de ellas. El Coeficiente de correlación muestral para las variables  $i$  e  $k$  se define como (Johnson et al., 1992):

$$r_{ik} = \frac{S_{ik}}{\sqrt{S_{ii}}\sqrt{S_{kk}}} = \frac{\sum_{j=1}^n (x_{ij} - \bar{x}_i)(x_{kj} - \bar{x}_k)}{\sqrt{\sum_{j=1}^n (x_{ij} - \bar{x}_i)^2} \sqrt{\sum_{j=1}^n (x_{kj} - \bar{x}_k)^2}} \quad (22)$$

Para  $i = 1, 2, \dots, p$ , y  $k = 1, 2, \dots, p$ . Nótese que  $r_{ik} = r_{ki}$  para todo  $i$  e  $k$  (Johnson et al., 1992).

El Coeficiente de correlación muestral es una versión estandarizada de la covarianza muestral, donde el producto de las raíces cuadradas de las varianzas muestrales proporciona la estandarización. Nótese que  $r_{ik}$  tiene el mismo valor si  $n$  o  $n-1$  es seleccionado como divisor común para  $S_{ii}$ ,  $S_{kk}$ , y  $S_{ik}$  (Johnson et al., 1992). Debido a lo anterior y a que no depende de las unidades de medida de las variables (como ya se había mencionado), se prefiere emplear el arreglo de coeficientes de correlación muestral en lugar del arreglo de covarianzas muestrales para los procedimientos subsecuentes relacionados con la obtención de las componentes principales.

Aunque los signos de la correlación muestral y la covarianza muestral son idénticos la correlación es generalmente más fácil de interpretar, debido a que su magnitud esta limitada (Johnson et al., 1992).

Las estadísticas descriptivas calculadas de  $n$  mediciones en  $p$  variables pueden organizarse en arreglos (matrices) (Johnson et al., 1992):

---


$$\text{Medias muestrales} \quad \bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_p \end{bmatrix} \quad (23)$$

$$\text{Varianzas y covarianzas muestrales} \quad \boldsymbol{\Sigma} = \mathbf{S}_{n-1} = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1p} \\ S_{21} & S_{22} & \cdots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p1} & S_{p2} & \cdots & S_{pp} \end{bmatrix} \quad (24)$$

$$\text{Correlaciones muestrales} \quad \mathbf{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix} \quad (25)$$

La matriz de varianzas y covarianzas muestrales es denotada por la expresión  $\mathbf{S}_{n-1}$ . El tamaño de todas las matrices es determinado por el número de variables  $p$  (Johnson et al., 1992).

Las componentes principales dependen solamente de la matriz de covarianza  $\boldsymbol{\Sigma}$  (o la matriz de correlación  $\mathbf{R}$ ) de  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ . Su desarrollo no requiere asumir una distribución normal multivariante (Johnson et al., 2004).

Para obtener las componentes principales dada la matriz de varianzas y covarianzas  $\mathbf{S}_{n-1}$  de orden  $p$ , los valores de los escalares  $\lambda$  y los vectores  $\mathbf{V}$  (de dimensiones  $p \times 1$ ), se usa la siguiente ecuación que satisfacen (Jobson, 1991):

$$\mathbf{S}_{n-1} \mathbf{V} = \lambda \mathbf{V} \quad (26)$$

Son llamados los autovalores ( $\lambda$ ) y los autovectores ( $\mathbf{V}$ ) de la matriz  $\mathbf{S}_{n-1}$ . Es común imponer la restricción adicional  $\mathbf{V}'\mathbf{V} = 1$  y así el autovector será normalizado para que tenga una longitud de 1 (Jobson, 1991).

Los autovalores proporcionan una medida de la varianza explicada por los autovectores correspondientes y son calculados para todas las componentes (folk.uio.no, 2007)

La ecuación (26) se puede reescribir como (Jobson, 1991):

$$(\mathbf{S}_{n-1} - \lambda \mathbf{I}) \mathbf{V} = \mathbf{0} \quad (27)$$

---

Con esto se obtiene un sistema de ecuaciones homogéneas. Una solución no trivial ( $\mathbf{V} \neq 0$ ) requiere que (Jobson, 1991):

$$|\mathbf{S}_{n-1} - \lambda \mathbf{I}| = 0 \quad (28)$$

Esta ecuación produce un polinomio en  $\lambda$  de grado  $p$  y es comúnmente llamado el polinomio característico. Las soluciones de la ecuación son las raíces del polinomio y reciben el nombre de *valores característicos*, aunque el término más comúnmente utilizado en Estadística es *autovalores* (Jobson, 1991).

Las direcciones de las componentes principales están marcadas por los autovectores de las matrices de covarianza (Flury, 2006).

Geoméricamente las componentes principales representan la selección de un nuevo sistema coordinado obtenido por la rotación del sistema original con  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$  como ejes coordinados. Los nuevos ejes representan las direcciones con máxima variabilidad y proporcionan una descripción más simple y más parsimoniosa de la estructura de covarianza (Johnson et al., 2004).

La primera componente principal es la combinación lineal de las variables con varianza máxima (Johnson et al., 2004).

La proporción de varianza total debida a la componente principal  $k$  es (Johnson et al., 1992):

$$\left( \begin{array}{l} \text{Proporción de la varianza total de} \\ \text{la población debida a la } k\text{-ésima} \\ \text{componente principal} \end{array} \right) = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_p} \quad k = 1, 2, \dots, p \quad (29)$$

Si la mayoría (por ejemplo 80% a 90%) de la varianza total para un valor grande de  $p$ , puede ser atribuida a la primera, segunda o tercera componentes, entonces estas componentes pueden “reemplazar” a las  $p$  variables originales sin mucha pérdida de información (Johnson et al., 1992).

Cuando las escalas de medida de las variables son muy distintas las variables con valores más grandes tendrán más peso en el análisis. Para evitar este problema conviene estandarizar las variables antes de calcular las componentes de manera que las magnitudes de los valores numéricos de las variables  $\mathbf{x}$  sean similares (Peña, 2002).

---

En la gran mayoría de los casos la obtención de las componentes principales se hace en las variables estandarizadas (también llamadas tipificadas) (Pérez, 2005).

Si solamente se consideran las primeras  $k$  componentes principales es de interés determinar que tanto de la variación en cada variable individual  $\mathbf{x}_i$  es explicado por esta aproximación. En el caso de la matriz de correlación, las varianzas de las  $\mathbf{x}_i$  son unitarias y por lo tanto la varianza explicada por las primeras  $k$  componentes es también la proporción de la varianza total. A esta proporción se le llama la comunalidad de la variable  $\mathbf{x}_i$  (Jobson, 1992).

La proporción de variación explicada por  $\lambda_p^R$  será (Peña, 2002):

$$\frac{\lambda_p^R}{p} \quad (30)$$

Concluyendo, cuando las variables  $\mathbf{x}$  originales están expresadas en distintas unidades conviene aplicar el análisis en la matriz de correlaciones  $\mathbf{R}$  o análisis normado (Peña, 2002). De aquí en adelante al hablar del análisis se hace referencia al análisis normado; al hablar de valores propios o autovalores se hace referencia a los obtenidos en dicho análisis.

Las variables iniciales  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ , se resumen en un menor número de variables  $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_k$  (componentes principales) que se pueden expresar así (Pérez, 2005):

$$\begin{aligned} \mathbf{c}_1 &= a_{11}\mathbf{x}_1 + a_{12}\mathbf{x}_2 + \dots + a_{1p}\mathbf{x}_p \\ \mathbf{c}_2 &= a_{21}\mathbf{x}_1 + a_{22}\mathbf{x}_2 + \dots + a_{2p}\mathbf{x}_p \\ &\vdots \\ \mathbf{c}_p &= a_{p1}\mathbf{x}_1 + a_{p2}\mathbf{x}_2 + \dots + a_{pp}\mathbf{x}_p \end{aligned} \quad (31)$$

En donde solamente se consideran  $k$  componentes principales que representan un porcentaje alto de la variabilidad de las variables originales, esto es  $k < p$  (Pérez, 2005).

Los coeficientes obtenidos del modelo de regresión son llamados los pesos de los componentes y son una expresión cuantitativa del rol de las componentes principales (Simeonov et al., 2002).

Para seleccionar el número de componentes se aplicará el criterio del autovalor con magnitud uno o criterio de Kaiser, el cual es muy común (Jobson, 1992; Shrestha et

al., 2007).

Una forma de medir la información que contiene una componente es por medio de su varianza. Debido a esto la primera componente es la que posee la mayor varianza (Pérez, 2005).

La primera componente principal, al igual que las demás, se representa por medio de una combinación lineal de las variables originales (Pérez, 2005):

$$c_{1i} = a_{11}x_{1i} + a_{12}x_{2i} + \dots + a_{1p}x_{pi} \quad i = 1, \dots, n \quad (32)$$

En la expresión (32) el subíndice  $i$  se refiere al número de observación considerado (Pérez, 2005).

Para las  $n$  observaciones de la muestra y todas las componentes principales se tiene la siguiente expresión matricial (Pérez, 2005):

$$\begin{bmatrix} c_{11} \\ c_{12} \\ \vdots \\ c_{1n} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{p1} \\ x_{12} & x_{22} & \dots & x_{p2} \\ \vdots & & & \vdots \\ x_{1n} & x_{2n} & \dots & x_{pn} \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1p} \end{bmatrix} \quad (33)$$

Una componente principal es una función lineal de todas las variables iniciales, que puede tener más correlación con algunas de ellas y menos con las restantes (Pérez, 2005).

Las ecuaciones que expresan las variables en función de las componentes principales son ventajosas para la interpretación de las componentes (Pérez, 2005):

$$\begin{aligned} \mathbf{x}_1 &= r_{11}\mathbf{c}_1 + r_{21}\mathbf{c}_2 + \dots + r_{k1}\mathbf{c}_k \\ \mathbf{x}_2 &= r_{12}\mathbf{c}_1 + r_{22}\mathbf{c}_2 + \dots + r_{k2}\mathbf{c}_k \\ &\vdots \\ \mathbf{x}_p &= r_{1p}\mathbf{c}_1 + r_{2p}\mathbf{c}_2 + \dots + r_{kp}\mathbf{c}_k \end{aligned} \quad (34)$$

Para  $\mathbf{x}_1$  la comunalidad se obtiene con  $h_1^2 = V(\mathbf{x}_1) = r_{11}^2 + r_{21}^2 + \dots + r_{k1}^2 \leq 1$ . La comunalidad proporciona una medida de la representación de una variable; así variables completamente representadas tienen comunalidad con valor uno (Pérez, 2005).

---

La interpretación de las componentes principales se facilita representando las proyecciones de las observaciones sobre un espacio bidimensional compuesto por las dos componentes principales más importantes (Peña, 2002). Así, el gráfico de las componentes principales puede revelar observaciones sospechosas (anómalas) y proporcionar una comprobación de la suposición de normalidad (Johnson et al., 1992).

La interpretación se favorece representando en el mismo plano de las dos componentes principales además de las observaciones a las variables originales. Esto puede hacerse utilizando como coordenadas su coeficiente de correlación con cada uno de los ejes (Peña, 2002).

El PCA constituye en algunos casos una antesala para la aplicación de otros métodos multivariantes, en los que se suple al grupo de variables iniciales por las componentes calculadas. Debido a esto en ocasiones se requiere conocer los valores que toman las componentes en cada una de las observaciones. Las puntuaciones  $z_{hi}$  se pueden calcular con la siguiente expresión (Pérez, 2005):

$$z_{hi} = a_{h1}x_{1i} + a_{h2}x_{2i} + \dots + a_{hp}x_{pi} \quad h = 1, \dots, p \quad i = 1, \dots, n \quad (35)$$

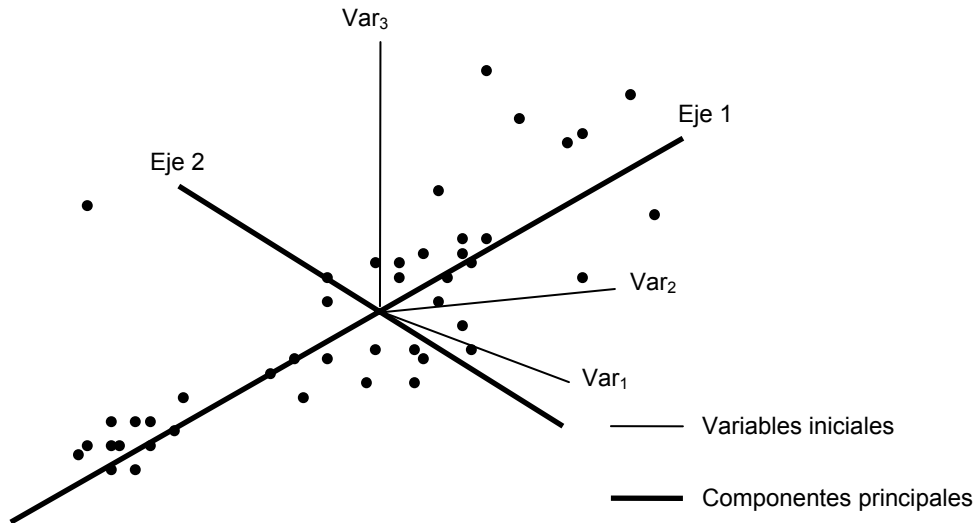
Donde  $i$  es el número de observación,  $h$  es el número de la componente principal seleccionada,  $p$  es el número de variables independientes,  $n$  es el número de observaciones,  $z_{hi}$  es el valor de la puntuación estandarizada de la observación  $i$  en la componente principal  $h$ ,  $x_{hi}$  es el valor estandarizado de la variable  $h$  en la observación  $i$  (Camdevýren et al., 2005).

El PCA por medio de los gráficos de puntuaciones permite extraer información útil en una forma fácil de ver, usando gráficos en lugar de interpretar una gran cantidad de datos (Giussani et al., 2006).

Los valores propios o autovalores proporcionan directamente las inercias proyectadas sobre cada una de las direcciones buscadas. Los vectores propios proporcionan las direcciones de máxima inercia y los ejes definidos por estos vectores unitarios son las componentes principales (Figura 2). Sobre estas direcciones se proyectan los individuos obteniendo las llamadas *puntuaciones en las componentes principales* (Aluja Banet et al., 1999).

Al rotar un grupo de componentes principales no se modifica la cantidad de inercia (variabilidad) total explicada y no cambian las comunalidades de las variables. La

rotación sólo modifica las saturaciones de las variables originales en las componentes (Pérez, 2005). Una explicación más amplia de los tipos de rotación se verá en la próxima sección de análisis factorial.



**FIGURA 2** Cambio de ejes y reducción de la dimensionalidad de los datos (Fuente: Aluja Banet et al., 1999)

### 4.3 ANÁLISIS FACTORIAL (FA)

#### 4.3.1 Descomposición factorial

La matriz aleatoria observable  $\mathbf{X}$  con  $p$  componentes tiene matriz de medias  $\boldsymbol{\mu}$  y matriz de varianzas-covarianzas  $\boldsymbol{\Sigma}$ . El modelo factorial postula que  $\mathbf{X}$  es linealmente dependiente de unas pocas variables aleatorias  $\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_m$ , llamadas factores comunes y  $p$  fuentes de variación adicionales  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ , llamados errores. El modelo del análisis factorial es (Johnson et al., 1992):

$$\begin{aligned}
 \mathbf{x}_1 - \mu_1 &= l_{11}\mathbf{f}_1 + l_{12}\mathbf{f}_2 + \dots + l_{1m}\mathbf{f}_m + \varepsilon_1 \\
 \mathbf{x}_2 - \mu_2 &= l_{21}\mathbf{f}_1 + l_{22}\mathbf{f}_2 + \dots + l_{2m}\mathbf{f}_m + \varepsilon_2 \\
 &\vdots \\
 \mathbf{x}_p - \mu_p &= l_{p1}\mathbf{f}_1 + l_{p2}\mathbf{f}_2 + \dots + l_{pm}\mathbf{f}_m + \varepsilon_p
 \end{aligned} \tag{36}$$

En notación matricial:

$$\underset{(p \times 1)}{\mathbf{X}} - \underset{(p \times 1)}{\boldsymbol{\mu}} = \underset{(p \times m)}{\mathbf{L}} \underset{(m \times 1)}{\mathbf{F}} + \underset{(p \times 1)}{\boldsymbol{\varepsilon}} \tag{37}$$

El coeficiente  $l_{ij}$  recibe el nombre de *carga* o *saturación* de la variable  $i$  en el factor  $j$ ,



así como la matriz  $\mathbf{L}$  es llamada la *matriz de cargas o saturaciones factoriales* (Johnson et al., 1992).  $\mathbf{F}$  es el *vector de factores comunes* linealmente independientes y  $\boldsymbol{\varepsilon}$  es el vector de errores o *factores únicos*. Los factores  $\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_m$  son comunes a todas las variables  $\mathbf{X}$  mientras que el error o factor único  $\varepsilon_i$  es único para la variable  $\mathbf{x}_i$  (Jobson, 1992).

La ecuación (37) implica que dada una muestra aleatoria simple de  $n$  elementos generada por el modelo factorial cada dato  $x_{ij}$  puede escribirse así (Peña, 2002)

$$x_{ij} = \mu_j + \lambda_{j1}f_{1i} + \dots + \lambda_{jm}f_{mi} + \varepsilon_{ij} \quad i = 1, \dots, n \quad j = 1, \dots, p \quad (38)$$

Los coeficientes  $\lambda_{j1}, \dots, \lambda_{jm}$ , dependen de la relación entre cada factor y la variable  $j$  (y que son los mismos para todos los elementos de la muestra). Los  $f_{1i}, \dots, f_{mi}$ , son los valores de los  $m$  factores en el elemento muestral  $i$  (Peña, 2002).

Por otra parte el modelo factorial hace algunas otras suposiciones (Johnson et al., 1992):

$$\text{Cov}(\mathbf{F}) = \mathbf{I}_{(m \times m)} \quad (39)$$

$$\text{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}_{(p \times p)} = \begin{bmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_p \end{bmatrix} \quad (40)$$

Además,  $\mathbf{F}$  y  $\boldsymbol{\varepsilon}$  son independientes (están incorrelacionados) por lo que se cumple (Johnson et al., 1992):

$$\text{Cov}(\boldsymbol{\varepsilon}, \mathbf{F}) = \mathbf{0}_{(p \times m)} \quad (41)$$

Las suposiciones (39), (40), (41) y la expresión (37), constituyen el modelo factorial ortogonal (Johnson et al., 1992).

El modelo factorial expresado en (37) es lineal en los factores comunes. La suposición de linealidad es inherente en la formulación del modelo factorial (Johnson et al., 1992).

La varianza de cada variable  $\mathbf{x}_i$  puede expresarse así (Jobson, 1992):

$$\sigma_i^2 = \sum_{j=1}^m \ell_{ij}^2 + \sigma_{\epsilon_i}^2 \quad (42)$$

Y por lo tanto la varianza esta dividida en dos partes (Jobson, 1992).

La porción de la varianza de la variable  $i$  que recibe la contribución de los  $m$  factores comunes es llamada la comunalidad  $i$ . La porción de  $\text{Var}(\mathbf{x}_i) = \sigma_{ii}$  debida a un factor específico es llamada comúnmente varianza específica. Llamando  $h_i^2$  a la comunalidad  $i$  se tiene (Johnson et al., 1992):

$$\underbrace{\sigma_{ii}}_{\text{Var}(\mathbf{x}_i)} = \underbrace{\ell_{i1}^2 + \ell_{i2}^2 + \dots + \ell_{im}^2}_{\text{Comunalidad}} + \underbrace{\psi_i}_{\text{Varianza específica}} \quad (43)$$

O también:

$$h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 + \dots + \ell_{im}^2 \quad (44)$$

La comunalidad  $i$  es la suma de los cuadrados de las saturaciones de la variable  $i$  en los  $m$  factores comunes (Johnson et al., 1992). La comunalidad es la parte de la variabilidad de las variables que es explicada sólo por los factores comunes (Cuadras, 2007).

En el modelo factorial la parte explicada de la varianza de una variable es debida a los factores y la parte no explicada se debe al ruido (Peña, 2002).

#### 4.3.2 Análisis factorial

Esta técnica es una rama del análisis multivariante que busca las relaciones internas de un conjunto de variables cuando estas relaciones pueden ser expresadas linealmente o aproximadamente así (Maxwell, 2006).

El análisis factorial, FA, tiene como propósito simplificar las numerosas relaciones que pudieran darse dentro del conjunto de variables medidas  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ . Para realizar esto busca *factores comunes*  $\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_k$ , que conecten a las variables originales. Para en verdad hacer más sencilla la estructura que pueda explicar el fenómeno en estudio se debe cumplir que  $k < p$ . Los factores deben representar a las variables originales de tal forma que se pierda un mínimo de información (Pérez, 2005).

Tradicionalmente dos modelos han estado disponibles para explorar las relaciones

---

internas de datos multivariantes. Uno es el modelo PCA y el otro es el FA. Con el interés de tener claridad es recomendable distinguir entre estos dos planteamientos. En el PCA un conjunto de  $p$  variables correlacionadas observadas es transformado en un número igual de nuevas variables (componentes) que tienen la propiedad de estar incorrelacionadas; en este modelo todas las  $p$  componentes son necesarias para reproducir con precisión los coeficientes de correlación de las variables  $\mathbf{X}$ . En contraste el objetivo del FA es explicar estas correlaciones en términos de un mucho menor número  $m$  de variables hipotéticas o factores (Maxwell, 2006).

Como en el análisis de componentes principales el propósito esencial del FA es describir la variación entre muchas variables pero al contrario que en el PCA, el modelo subyacente en el análisis factorial especifica un pequeño número de factores comunes. Todas las covarianzas o correlaciones son explicadas por los factores comunes. Una porción de la varianza no explicada por los factores comunes es asignada a los términos de error residual, los cuales son llamados *factores únicos*. Se supone que los factores únicos están mutuamente incorrelacionados (Jobson, 1992).

Los factores buscados deben cumplir el *principio de interpretabilidad* (sencillos de interpretar) y el principio de *parsimonia* (que se necesite la menor cantidad de ellos) (Pérez, 2005).

El principio de parsimonia aplicado al FA indica que de entre varias soluciones admisibles se opta por la más simple de ellas. El modelo factorial será entonces el que conlleve un número mínimo  $m$  de factores comunes (Cuadras, 2007).

Un objetivo fundamental del FA es determinar si las variables muestran patrones de relación entre sí, de manera que las variables se puedan dividir en subconjuntos y de forma que las variables en un subconjunto estén fuertemente correlacionadas con cada una de las otras y que las variables en subconjuntos distintos posean bajas correlaciones entre sí. Debido a esto el FA es usado frecuentemente para analizar la estructura de correlación de las variables en un conjunto de datos (Johnson, 2000).

El FA tiene relación con las componentes principales pero hay algunas diferencias. Primero las componentes principales se construyen para explicar las varianzas, en cambio los factores se elaboran para explicar las covarianzas o correlaciones entre las variables. Segundo, las componentes principales son una herramienta de descripción, en tanto que el FA esta fundamentado en que los datos se generaron de un modelo estadístico establecido (Peña, 2002).

---

Los vínculos entre las variables originales  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ , son expresados por su matriz de correlaciones y el determinante de esta matriz debe tener un valor pequeño para que sea aplicado con éxito el FA (Pérez, 2005).

El FA calcula e interpreta los factores a partir de la matriz de correlaciones entre las variables  $\mathbf{R}$  (Cuadras, 2007). Se considerarán los dos métodos más populares con el propósito de extraer los factores, el de componentes principales y el método de máxima verosimilitud. La solución de ambos métodos puede ser rotada con el propósito de simplificar la interpretación de los factores (Johnson et al., 1992).

La solución con componentes principales del modelo factorial recibe este nombre del hecho de que las saturaciones factoriales son los coeficientes escalados de las primeras componentes principales muestrales (Johnson et al., 1992).

El análisis factorial de la matriz de correlación muestral por componentes principales se obtiene comenzando con  $\mathbf{R}$  en lugar de  $\mathbf{S}$  (Johnson et al., 1992).

La selección del número de factores comunes puede basarse en los autovalores estimados, de la misma manera que con las componentes principales. Idealmente las contribuciones de los primeros factores de las varianzas muestrales deben ser grandes. Se tiene (Johnson et al., 1992):

$$\left( \begin{array}{c} \text{Proporción de la} \\ \text{varianza muestral} \\ \text{total debida al factor } j \end{array} \right) = \begin{cases} \frac{\hat{\lambda}_j}{s_{11} + s_{22} + \dots + s_{pp}} & \text{En el análisis factorial de } \mathbf{S} \\ \frac{\hat{\lambda}_j}{p} & \text{En el análisis factorial de } \mathbf{R} \end{cases} \quad (45)$$

Con base en la expresión anterior un criterio frecuentemente utilizado en los paquetes estadísticos es establecer  $m$  igual al número de autovalores de  $\mathbf{R}$  mayores en valor que uno ( $>1$ ) si la matriz de correlación muestral esta factorizada (Johnson et al., 1992). Este es el criterio más comúnmente utilizado. La varianza de cada una de las variables  $\mathbf{X}$  tiene valor uno y por lo tanto el criterio del autovalor uno indica que un factor será retenido si explica al menos tanto como una variable sola (Jobson, 1992).

El paso más importante en la estimación del modelo de análisis factorial es la estimación de  $m$  el número de factores. Si  $m$  es demasiado grande algunos de los residuos o factores únicos (factores de error) podrían estar mezclados con los factores comunes; y si  $m$  es demasiado pequeño factores comunes importantes podrían ser

omitidos (Jobson, 1992).

En el método de la máxima verosimilitud se considera que si los factores comunes  $\mathbf{F}$  y los factores específicos  $\boldsymbol{\varepsilon}$  se pueden suponer como distribuidos normalmente, entonces la estimación de máxima verosimilitud de las saturaciones factoriales y las varianzas específicas puede obtenerse. Cuando  $\mathbf{f}_j$  y  $\boldsymbol{\varepsilon}_j$  son conjuntamente normales, las observaciones  $\mathbf{x}_j - \boldsymbol{\mu} = \mathbf{L}\mathbf{f}_j + \boldsymbol{\varepsilon}_j$  son entonces normales y la probabilidad es (Johnson et al., 1992):

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= (2\pi)^{\frac{np}{2}} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \mathbf{e}^{-\left(\frac{1}{2}\right)\text{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' + n(\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})'\right)\right]} \\ &= (2\pi)^{\frac{-(n-1)p}{2}} |\boldsymbol{\Sigma}|^{-\frac{(n-1)}{2}} \mathbf{e}^{-\left(\frac{1}{2}\right)\text{tr}\left[\boldsymbol{\Sigma}^{-1}\left(\sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})'\right)\right]} \\ &\quad \times (2\pi)^{\frac{-p}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \mathbf{e}^{-\left(\frac{n}{2}\right)(\bar{\mathbf{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu})} \end{aligned} \quad (46)$$

La cual depende de  $\mathbf{L}$  y  $\boldsymbol{\Psi}$  a través de  $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' + \boldsymbol{\Psi}$ . Este modelo aún no está bien definido debido a la multiplicidad de elecciones de  $\mathbf{L}$  que son posibles por transformaciones ortogonales. Es deseable hacer a  $\mathbf{L}$  bien definida mediante la imposición de una condición de unicidad conveniente (Johnson et al., 1992):

$$\mathbf{L}'\boldsymbol{\Psi}^{-1}\mathbf{L} = \mathbf{I} \quad (47)$$

La máxima verosimilitud estima que  $\hat{\mathbf{L}}$  y  $\hat{\boldsymbol{\Psi}}$  deben obtenerse por maximización numérica. Por fortuna programas de ordenador eficientes existen actualmente para permitir estimar la máxima verosimilitud de una forma rápida (Johnson et al., 1992).

En el FA pueden efectuarse contrastes que se aplican antes de la extracción como la prueba de esfericidad de Bartlett y la medida de adecuación muestral de Kaiser, Meyer y Olkin (KMO) (Pérez, 2005).

La prueba de esfericidad de Bartlett sirve para conocer si las variables están correlacionadas entre sí (Pérez, 2005). Con esta prueba se confirma que las variables no son ortogonales pero están correlacionadas y por lo tanto se puede explicar la variabilidad de los datos con un menor número de variables (Vega et al., 1998). Esta prueba indica si la matriz de correlación es una matriz identidad lo cual indicaría que las variables no están relacionadas. El nivel de significación debe ser menor que 0,05 para que existan relaciones significativas entre variables (Parinet et al., 2004).

---

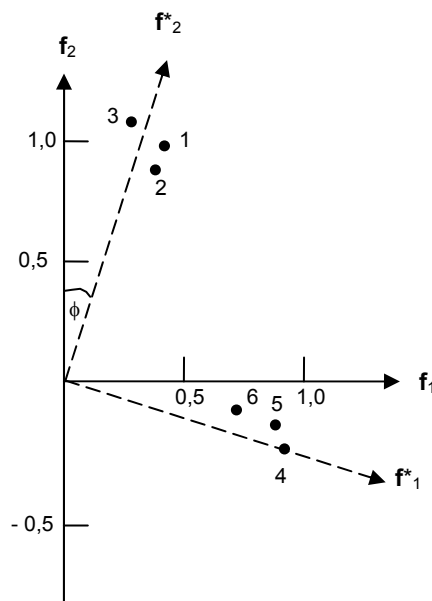
La medida de Kaiser, Meyer y Olkin se basa en los coeficientes de correlación observados entre cada par de variables y en sus coeficientes de correlación parcial. Valores del KMO inferiores a 0,5 no son aceptables, considerándose que para ese caso no es adecuado realizar el análisis factorial. Los valores superiores a 0,5 son aceptables, lo que indica que es pertinente efectuar un FA con los datos. En tanto los valores de la medida KMO estén más cercanos a 1 mejor es la adecuación de la información a un modelo factorial (Pérez, 2005).

#### 4.3.3 Rotación de factores

En el análisis factorial se busca que los factores sean fácilmente entendibles en términos del fenómeno en estudio. Desafortunadamente en pocos casos es fácil encontrar una interpretación sencilla a los factores obtenidos en un primer momento. La rotación de factores partiendo del primer análisis busca obtener factores que tengan una interpretación sencilla (Figura 3) (Pérez, 2005).

Los factores **F** no son únicos. Por medio de la introducción de una matriz de transformación ortogonal **T** tal que  $\mathbf{T}'\mathbf{T} = \mathbf{T}\mathbf{T}' = \mathbf{I}$ , nuevos factores  $\mathbf{G} = \mathbf{F}\mathbf{T}$  pueden ser definidos de manera que también satisfagan el modelo de análisis factorial (Jobson, 1991).

La cantidad total de varianza explicada por el conjunto de factores transformados (rotados) permanece constante (Browne, 2006).



**FIGURA 3** Rotación de factores. Los puntos representan variables (Fuente: Johnson et al., 1992)

---

Idealmente se quisiera tener un patrón de cargas en las que cada variable saturara fuertemente en un factor simple y tenga saturaciones de pequeñas a moderadas en los restantes factores (Johnson et al., 1992). Se prefiere que cada factor englobe el comportamiento de un grupo de variables originales con las que se le pueda relacionar (Tusell, 2007).

Algunos investigadores clasifican las saturaciones factoriales como “fuertes” (valores absolutos mayores de  $> 0,75$ ), “moderadas” (entre  $0,50$  y  $0,75$ ) y “débiles” (entre  $0,30$  y  $0,50$ ) (Singh et al., 2004).

Haciendo referencia a los procedimientos ortogonales el más empleado es el Varimax, propuesto por Kaiser y trata de minimizar el número de variables que hay con pesos o saturaciones elevadas en cada factor (Visauta et al., 2003).

En el método Varimax la matriz de pesos factoriales es rotada de manera que las cargas estimadas muestren un mayor contraste entre columnas. El objetivo es obtener un relativamente gran número de elementos próximos a ser nulos y un número limitado de grandes elementos, además de muy pocos elementos de tamaño intermedio. Como resultado, la conclusión de un análisis puede ser simplificada y algunas veces más significativa. Este método se deriva del Varimax de Kaiser (Neudecker, 2006).

El Varimax usa como criterio para la simplificación de la estructura la expresión (Neudecker, 2006):

$$\sum_{j=1}^m \left[ \left( \sum_{i=1}^p \ell_{ij}^4 \right) - \frac{1}{p} \left( \sum_{i=1}^p \ell_{ij}^2 \right)^2 \right] \quad (48)$$

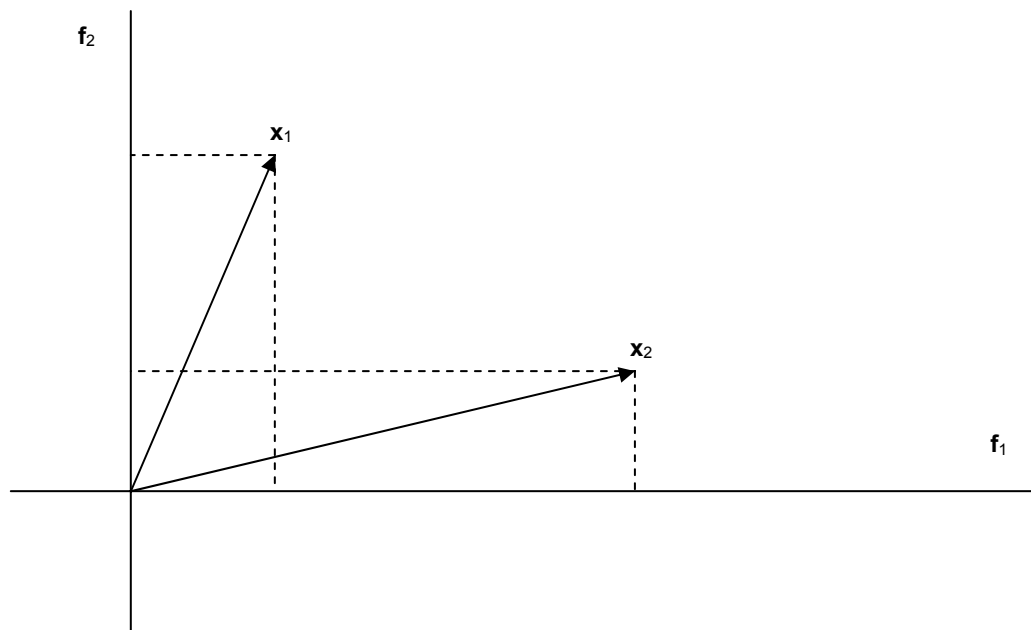
Donde las  $\ell_{ij}$  son las cargas factoriales estimadas. Esta expresión es proporcional a la varianza intragrupal de las estimaciones cuadráticas habiendo  $m$  grupos (factores) (Neudecker, 2006).

#### 4.3.4 Puntuaciones factoriales

Una vez efectuado el Análisis Factorial, construyendo los factores, es importante obtener las *puntuaciones factoriales (scores)* de las observaciones para saber cuánto puntúan en cada factor (Calvo, 1993). Es una forma de situar a cada observación con respecto a cada uno de los factores.

---

Las saturaciones, pesos o cargas factoriales, se pueden representar de manera gráfica mediante las proyecciones ortogonales de cada una de las variables en cada factor (Figura 4) (Pérez, 2005).



**FIGURA 4** Proyecciones ortogonales de las variables originales  $x_1$  y  $x_2$  sobre los factores  $f_1$  y  $f_2$  (Fuente: Pérez, 2005)

Los beneficios de analizar las puntuaciones factoriales de las observaciones son variados: identificar las observaciones extremas (anormales), localizar la ubicación de ciertos grupos de la muestra, en qué factor satura una observación y en cuáles no, etc. (Calvo, 1993).

Aunque la normalidad multivariante es a menudo supuesta para las variables en el análisis factorial es muy difícil justificar esta suposición para grandes números de variables. Las gráficas de puntuaciones factoriales deben de examinarse antes de utilizar éstas puntuaciones en otros análisis. Estas gráficas pueden revelar valores anómalos (Johnson et al., 1992).

#### 4.3.5 Aplicación del análisis factorial

En el análisis factorial no existe una estrategia simple a ser aplicada. Una opción razonable que se propone es (Johnson et al., 1992):

1. Para iniciar verificar la estructura de los datos que recogen las variables. Una forma de realizar esto es por medio de un gráfico de dispersión matricial de las variables. Se trata de hacer patente que el análisis factorial se basa en la



---

matriz de coeficientes de correlación  $r$  de Pearson entre esas variables. Pero la correlación  $r$  de Pearson plantea varios supuestos (Calvo, 1993):

- 1.1 Supone la existencia de una relación lineal entre las variables.
  - 1.2 La incidencia de la forma de la distribución que tienen las variables. Se acepta de inicio que todas las variables tienen una distribución normal.
  - 1.3 El número de variables debe seleccionarse para que los resultados ofrezcan una estructura factorial adecuada, fácil de entender y con contenido teórico relevante.
  - 1.4 La forma y características de los datos debido a que la correlación  $r$  de Pearson es distinta en la mayoría de los casos si se calcula con los datos sin tratar o con esos mismos datos transformados.
2. Desarrollar un análisis factorial con el método de componentes principales (Johnson et al., 1992).
  3. Buscar las observaciones sospechosas por medio del trazado de las gráficas de las puntuaciones factoriales (Johnson et al., 1992). Los puntos que se encuentren alejados del resto deberán ser investigados para saber si son efectivamente observaciones atípicas.
  4. Aplicar una rotación Varimax (Johnson et al., 1992).
  5. Desarrollar un análisis factorial por el método de la máxima verosimilitud incluyendo una rotación con el método Varimax (Johnson et al., 1992).
  6. Comparar las soluciones de los análisis factoriales (Johnson et al., 1992; Calvo, 1993). Uno de los aspectos más importantes a comparar es el valor de los autovalores y la varianza explicada por cada uno de ellos, así como la varianza acumulada.
  7. Repetir los cinco pasos anteriores para otros números de factores comunes  $m$  (Johnson et al., 1992).
  8. Para grandes conjuntos de datos dividirlos por la mitad y aplicar un análisis factorial en cada parte. Comparar las dos soluciones entre sí y con la conseguida con el conjunto completo de los datos, para comprobar la estabilidad de la solución (Johnson et al., 1992).

Más detalles de cómo desarrollar estos pasos se verán en la sección 7.6.

#### 4.4 ANÁLISIS CLUSTER (CA)

El objetivo del análisis cluster es dividir la información constituida por observaciones multivariantes en subgrupos llamados conglomerados (clusters), de forma que las observaciones que se encuentren en determinado conglomerado tengan

---

características semejantes entre sí con respecto a las variables medidas en tanto que aquellas observaciones que pertenezcan a otros conglomerados distintos no sean semejantes a las de otros grupos (Johnson, 2000).

El CA es una técnica no supervisada de reconocimiento de patrones que descubre la estructura intrínseca o comportamiento subyacente del conjunto de datos sin hacer suposiciones de inicio (*a priori*) acerca de los datos, con el propósito de clasificar los objetos del sistema dentro de categorías o conglomerados en base a su cercanía o similitud (Figura 5) (Singh et al., 2004).

En el caso de que las variables utilizadas para la formación de los conglomerados estén expresadas en escalas numéricas muy dispares es necesario para obtener buenos resultados en la clasificación estandarizar las variables, para evitar problemas de clasificación por estar las variables expresadas en unidades distintas o poseer diferentes órdenes de magnitud (tanto de los valores numéricos como de la varianza de los datos analizados). La estandarización de los datos asegura que cada variable tenga la misma influencia en el análisis. Asimismo, es perjudicial para el CA que existan variables altamente correlacionadas por lo que es necesario revisar previamente la multicolinealidad (Pérez, 2005).

El análisis cluster estudia tres tipos de problemas (Peña, 2002):

- a) Partición de los datos.
- b) Construcción de jerarquías.
- c) Clasificación de variables.

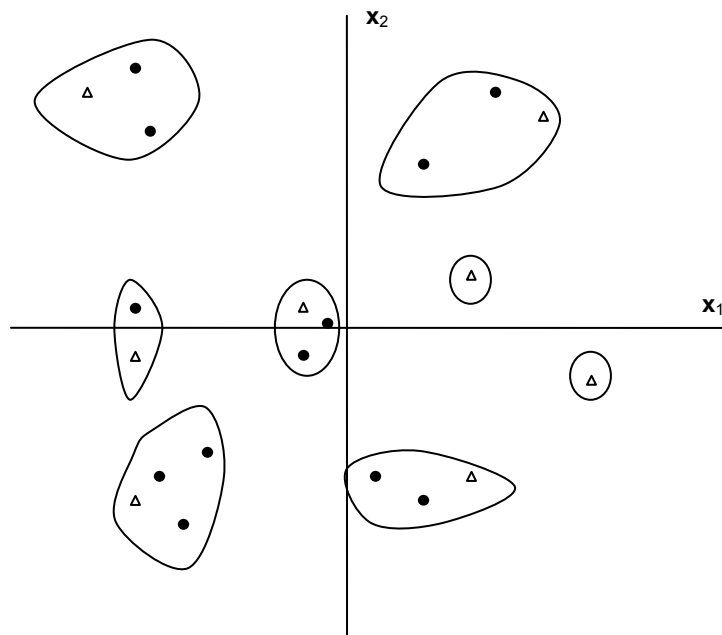
La cantidad de observaciones en cada conglomerado debe alcanzar un cierto valor mínimo ya que si existen valores anormales serían agrupados en conjuntos de una o muy pocas observaciones (Pérez, 2005).

Cuanto más grande es un conglomerado o subgrupo es más heterogéneo (Cuadras, 2007). Por lo que uno de los factores que afecta a la variabilidad entre los elementos de un conglomerado es su tamaño (número de elementos que contiene).

Los procedimientos en el CA pueden clasificarse en cinco tipos principales: jerárquico, separativo, tipo Q, densidad y de agrupación (Jobson, 1992):

1. En el enfoque jerárquico la técnica procede secuencialmente de manera tal que en cada paso sólo una observación o grupo de observaciones cambia su

- 
- pertenencia a un grupo y los grupos en cada paso son anidados con respecto a los conjuntos previos.
2. El método separativo inicia con un número dado de conglomerados, por ejemplo  $g$  como objetivo y entonces separa las observaciones para obtener el número requerido de conglomerados.
  3. Los métodos tipo Q incluyen una variedad de técnicas que son similares al análisis de factores.
  4. El modo de densidad o procedimiento de modo de búsqueda supone que los objetos estarán distribuidos en el espacio, de tal forma que existen varias áreas de densidad con regiones internas entre ellas que están muy dispersas.
  5. El método de agrupación (clumping) permite que los conglomerados se traslapen.



**FIGURA 5** Ilustración de la idea principal detrás del análisis cluster representado para dos variables (Martens et al., 1989)

Para efectuar un análisis de este tipo en primer lugar se debe poder medir la semejanza o desemejanza entre dos observaciones o registros y, a continuación, la semejanza o desemejanza entre los conglomerados de observaciones (Johnson, 2000). El agrupamiento es hecho en base a medidas de similaridad o distancias (disimilaridades) (Johnson et al., 1992).

La medida de similaridad más común es el coeficiente de correlación de Pearson. Debido a que un coeficiente de correlación puede tomar los valores del rango  $[-1, 1]$ ,

---

se acostumbra tomar el valor absoluto del valor del coeficiente (Jobson, 1992).

Una medida sencilla y muy comúnmente utilizada de desemejanza o disimilaridad en la distancia Euclidiana estándar, también llamada distancia métrica, que es la distancia entre dos observaciones en el espacio muestral  $p$ -dimensional. Su expresión es (Johnson, 2000):

$$d_{rs} = [(\mathbf{x}_r - \mathbf{x}_s)'(\mathbf{x}_r - \mathbf{x}_s)]^{1/2} \quad (49)$$

Donde:

$d_{rs}$  = Distancia Euclidiana estándar.

$\mathbf{x}_r$  = Primera observación considerada.

$\mathbf{x}_s$  = Segunda observación considerada.

La distancia Euclidiana cuadrática es la medida de distancia más empleada en estudios relacionados con calidad del agua en cuerpos de agua continentales (Vega et al., 1998; Mendiguchía et al., 2004; Singh et al., 2005; Kowalik et al., 2006).

Existen dos tipos básicos de búsqueda de conglomerados: los de naturaleza *jerárquica* o *no jerárquica* (Johnson, 2000).

En el agrupamiento jerárquico los resultados son mostrados como un dendograma; estos métodos son los más fáciles de aplicar a pequeños grupos de datos. Los métodos no jerárquicos (o partitivos) se aplican principalmente a grandes conjuntos de datos (McNeil et al., 2005).

#### 4.4.1 Métodos de agrupación no jerárquica

Una clasificación o agrupación *no jerárquica* de  $n$  objetos (registros) que constituyen una matriz de datos cuantitativos  $\mathbf{X}$ , consiste en obtener  $k$  grupos homogéneos y excluyentes entre sí (conglomerados). La información de partida es una tabla o matriz de dimensiones  $n \times p$  (Tusell, 2007).

En los métodos de agrupación *no jerárquica*, *partitivos* o de *optimización*, la forma de buscar agrupamientos es seleccionar de inicio un conjunto de puntos simientes de aquéllos y, después, construir esos conglomerados en torno a cada una de las simientes. Esto se efectúa al asignar cada punto del conjunto de datos a su simiente más cercana empleando las medidas de desemejanza para medir las distancias entre

---

cada unos de los puntos y esas simientes de conglomerados. A continuación se pueden dividir los conglomerados demasiado grandes y los que estén cercanos a otros se pueden combinar (Johnson, 2000).

En los métodos no jerárquicos que tienen por objeto realizar una sola partición de las observaciones en  $k$  grupos, previamente se debe fijar este número (Visauta et al., 2003).

Dentro de las diversas técnicas de agrupación no jerárquica la más utilizada es el *algoritmo de K-medias* (Pérez, 2005). La técnica de “K-Medias” de McQueen (1967) está basada en la distancia al centroide más próximo. Los centros de los diversos conglomerados se pueden conocer o no de inicio, lo que genera dos alternativas en el procedimiento que son (Visauta et al., 2003):

- Clasificación cuando los centros son conocidos.
- Clasificación cuando los centros son desconocidos.

La técnica de K-medias mide la proximidad entre grupos usando la distancia Euclidiana entre centroides de los grupos (Jobson, 1992).

El algoritmo de K-medias cuenta con las siguientes etapas (Peña, 2002):

1. Partir las observaciones en  $k$  conglomerados iniciales (Johnson et al., 1992).
2. En lugar de comenzar con la partición de todas las observaciones en  $k$  grupos preliminares en el paso 1 es posible optar por (Johnson et al., 1992) seleccionar  $k$  puntos como centros de los grupos iniciales.
3. Calcular las distancias Euclidianas de cada observación a los centros de los  $k$  grupos y asignar cada observación al conglomerado cuyo centro esté más cercano (Peña, 2002).
4. Definir un criterio de optimalidad y comprobar si reasignando alguna de las observaciones se mejora el desempeño (Peña, 2002).
5. Si no es posible mejorar el criterio de optimalidad, terminar (Peña, 2002).

El criterio empleado de homogeneidad o de optimalidad en el procedimiento de K-medias es minimizar la suma de los cuadrados dentro de los grupos para todas las variables en el estudio. Este criterio se expresa así (Peña, 2002):

$$\text{Criterio} = \sum_{k=1}^k \sum_{j=1}^p \sum_{i=1}^{n_g} (x_{ijk} - \bar{x}_{jk})^2 \quad (50)$$

Donde  $x_{ijk}$  es el valor de la observación  $j$  en el elemento  $i$  del grupo  $k$  y  $\bar{x}_{jk}$  la media de esta variable en el conglomerado (Peña, 2002).

Al emplear el método de K-medias hay que establecer el número de conglomerados. Se han propuesto diversos métodos para seleccionar el número de grupos. Un planteamiento que es muy empleado es efectuar un test F de reducción de variabilidad comparando la suma de cuadrados dentro de los grupos con  $k$  grupos, con al de  $k+1$  grupos y obteniendo la reducción relativa de variabilidad al aumentar un conglomerado más. La expresión es (Peña, 2002):

$$F = \frac{\text{Criterio}(k) - \text{Criterio}(k + 1)}{\text{Criterio}(k + 1)/(n - k - 1)} \quad (51)$$

Y comparar la disminución de variabilidad al agregar un conglomerado con la varianza media. Una regla práctica que proporciona resultados razonables propuesta por Hartigan en 1975 y empleada en algunos paquetes estadísticos, es utilizar un conglomerado más si el cociente anterior es mayor que 10 (Peña, 2002).

#### 4.4.2 Métodos de agrupación jerárquica

Una clasificación jerárquica es una sucesión de conglomerados, de manera que cada uno se obtiene agrupando conjuntos constituidos previamente (Cuadras, 2007).

En el agrupamiento jerárquico los conglomerados son formados secuencialmente, comenzando con el par de objetos más similares y formando conglomerados superiores paso a paso. Usualmente la distancia Euclidiana proporciona la similaridad entre las dos muestras o registros y es una “distancia” que puede representarse por la diferencia entre los valores analíticos de ambas muestras (Singh et al., 2004). Los métodos de agrupación jerárquica no resultan aconsejables para procesar grandes cantidades de observaciones (Pérez, 2005).

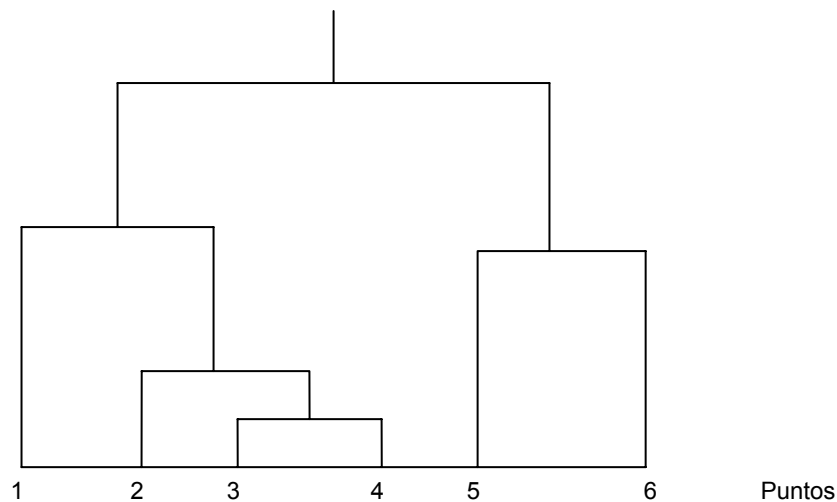
En los métodos de análisis de agrupación jerárquica para selección de conglomerados los datos observados se concentran en agrupamientos, en una sucesión anidada de conglomerados. Las técnicas más óptimas de agrupación jerárquica reciben el nombre de métodos de agrupación de un enlace. Uno de estos métodos es el del vecino más

---

cercano en el que se aplican cuatro pasos: (i) se inicia con  $N$  agrupamientos donde cada uno de ellos contiene solamente una observación (punto), (ii) se enlazan los dos puntos más cercanos según la medida seleccionada de la distancia, (iii) se determina la desemejanza entre este nuevo conglomerado y cualquier otro punto como la distancia mínima entre los dos puntos del agrupamiento y este único punto y (iv) se continua combinando los conglomerados que sean los más cercanos entre sí, de manera que en cada paso la cantidad de conglomerados se reduzca en uno (Johnson, 2000).

Si las variables son medidas en escala continua la medida de distancia más utilizada es la distancia Euclidiana entre las variables (Peña, 2002).

Una forma de decidir en que instante detener el proceso de agrupación es elaborar un *diagrama de árbol jerárquico* o *dendograma*. Este gráfico contiene ramas que unen puntos (observaciones) y muestra el orden en que se asignan los puntos a los conglomerados (Figura 6) (Johnson, 2000).



**FIGURA 6** Diagrama de árbol jerárquico o dendograma (Johnson, 2000).

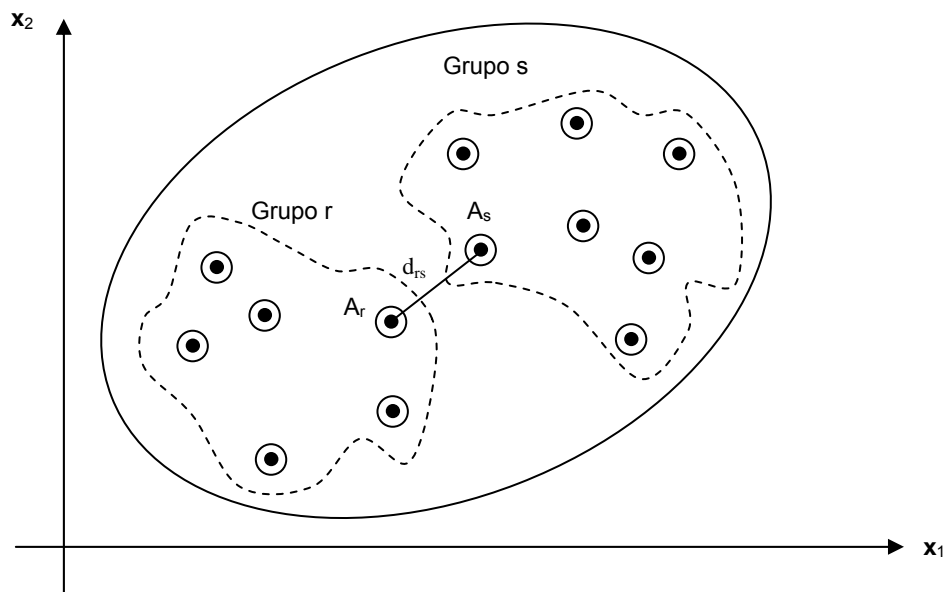
El dendograma provee un resumen visual del proceso de agrupamiento presentando un gráfico de los grupos y su proximidad (Wunderlin et al., 2000).

Otros métodos de agrupación jerárquica son: (i) el método del vecino más lejano, (ii) el método del centroide, (iii) el método del promedio y (iv) el método de la varianza mínima de Ward y Wishart (conocido como el método de Ward), en el cual la distancia entre dos conglomerados es el cuadrado de la distancia entre las medias de esos

---

conglomerados, dividida entre la suma de los recíprocos de la cantidad de observaciones que se encuentre en el interior de cada uno de éstos (Johnson, 2000).

El método del vecino más cercano (Figura 7) o método de unión simple tiende a maximizar la conexión de una pareja de conglomerados y va en el camino de generar una cantidad menor de grupos que el método del vecino más lejano (Johnson, 2000).



**FIGURA 7** Midiendo la proximidad entre grupos. Método del vecino más cercano usando la distancia Euclidiana (Fuente: Jobson, 1992)

En los métodos de agrupación jerárquica las fuentes de error y variación no están formalmente consideradas. Esto quiere decir que el método de agrupación será sensible a los valores anómalos (Johnson et al., 1992).

#### 4.4.3 Análisis cluster de variables

El análisis cluster de variables es una técnica exploratoria que puede auxiliar en la reducción de la dimensión. El propósito es obtener una matriz de distancias entre las variables como en el caso de las observaciones y entonces emplear un procedimiento jerárquico de agrupamiento (Peña, 2002).

Se puede proponer una forma de medir la distancia entre dos variables  $x_A$  y  $x_B$ , por medio de la distancia Euclidiana, considerando su valor para una observación como dos puntos (Peña, 2002):



---

$$d_{AB} = \sum_{i=1}^n (\mathbf{x}_A - \mathbf{x}_B)^2 \quad (52)$$

Con el fin de evitar que la distancia dependa de las unidades de medida las variables deben estandarizarse. Haciendo esto, la expresión (52) quedaría así (Peña, 2002):

$$d_{AB} = 2n(1 - r_{AB}) \quad (53)$$

Siendo  $r_{AB}$  el coeficiente de correlación entre las variables  $A$  y  $B$  (Peña, 2002).

#### 4.5 ANÁLISIS DISCRIMINANTE (DA)

Se usa principalmente para clasificar individuos (*registros*) o unidades experimentales en dos o más poblaciones definidas previamente (Johnson, 2000).

El objetivo del DA es describir las características que diferencian a las observaciones de varias poblaciones conocidas (Johnson et al., 1992). Y con esta información se responde a la pregunta: ¿a qué población pertenece una observación de la cual se tienen mediciones en algunas variables? (Calvo, 1993).

El DA es utilizado para determinar las variables que discriminan entre dos o más grupos que ocurren naturalmente. Opera sobre los datos originales y construye una función discriminante para cada grupo (Singh et al, 2004).

Este procedimiento también se conoce como *clasificación supervisada*, lo que significa que se conoce un conjunto de observaciones clasificadas correctamente que sirven de patrón para clasificar las observaciones posteriores (Peña, 2002).

En el DA se asigna una observación a un grupo definido previamente (que hace la función de variable dependiente) conforme a los valores de algunas variables medidas en dicha observación (variables independientes) (Visauta et al., 2003).

Aun más, el uso del DA está basado en varios supuestos fundamentales como son la normalidad multivariante, la homogeneidad de las matrices de varianza-covarianza (supuesto de homoscedasticidad), linealidad y la no presencia de multicolinealidad (Pérez, 2005). En los casos prácticos la técnica es muy robusta y no es necesario obligatoriamente que se cumplan todos los supuestos (Gil Pascual, 2003).

La distancia de Mahalanobis de una observación  $i$  al grupo  $k$  es, expresado en forma

---

matricial (Calvo, 1993):

$$\mathbf{D}^2 = (\mathbf{x}_i - \bar{\mathbf{x}}_k)' \mathbf{S}_n^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_k) \quad (54)$$

La cual debe ser mínima. Los coeficientes de las variables en las funciones discriminantes entre los probables pares de grupos  $j$  y  $k$ , será (Calvo, 1993):

$$\mathbf{A}_{jk} = \mathbf{S}_n^{-1} (\bar{\mathbf{x}}_j - \bar{\mathbf{x}}_k) \quad (55)$$

Y para obtener el término independiente (Calvo, 1993):

$$\mathbf{c}_{jk} = \frac{1}{2} (\bar{\mathbf{x}}_j - \bar{\mathbf{x}}_k)' \mathbf{S}_n^{-1} (\bar{\mathbf{x}}_j + \bar{\mathbf{x}}_k) \quad (56)$$

Así se puede determinar a que grupo pertenece una nueva observación  $i$ . La técnica conformada por las expresiones (55) y (56) no es recomendable cuando el número de grupos es mayor de 3 debido a que se complica demasiado (Calvo, 1993).

Fisher propuso un método discriminante para varias poblaciones cuyo propósito primario es separarlas. También puede ser empleado para clasificarlas. No es necesario suponer que las  $g$  poblaciones son normales multivariantes (Johnson et al., 1992).

Las funciones lineales Discriminantes de Fisher para cada grupo  $k$ , se obtienen con (Calvo, 1993):

$$\mathbf{A}_k = \mathbf{S}_n^{-1} \bar{\mathbf{X}}_{kv} \quad (57)$$

Y para el término independiente (Calvo, 1993):

$$\mathbf{c}_k = \frac{1}{2} \bar{\mathbf{X}}_{kv}' \mathbf{S}_n^{-1} \bar{\mathbf{X}}_{kv} \quad (58)$$

Así se puede saber la capacidad de discriminación de cada una de las  $m$  variables de cada subconjunto y también la asignación de una nueva observación a uno de los  $q$  grupos. Este planteamiento es de utilidad cuando el número de grupos es mayor de 3 (Calvo, 1993).

La expresión lineal discriminante tiene la forma siguiente (Visauta et al., 2003):

---

$$D = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p \quad (59)$$

Donde:

$x_i$  = Son las variables independientes.

$b_0$  = Es una constante.

$b_p$  = Coeficientes estimados de las observaciones originales de manera que los valores de la función difieran el máximo posible entre los grupos.

Para algunos estudios el DA fue aplicado a los datos experimentales con y sin estandarización, la habilidad discriminante fue la misma pero las  $b_i$  fueron diferentes para ambos casos (Wunderlin et al., 2000).

En algunas aplicaciones existen datos disponibles con un gran número de variables. En estos casos obviamente sería deseable seleccionar un relativamente pequeño subconjunto de variables que contuvieran casi tanta información como el conjunto original. Este es el objetivo del análisis discriminante por pasos (*stepwise*); varios paquetes estadísticos comerciales tienen la capacidad de efectuar este tipo de análisis (Johnson et al., 1992).

Uno de los métodos para elegir las variables significativas frecuentemente empleado, el llamado *método stepwise* (método paso a paso) puede elegir el proceso hacia delante (*forward stepwise*) o hacia atrás (*backward stepwise*). Para el método de elección paso a paso hacia delante la primera variable elegida es la que maximiza la separación entre conjuntos; en este modo las variables son incluidas paso a paso comenzando con las más significativas, hasta que cambios no significativos son obtenidos. Para el método de elección paso a paso hacia atrás se incluyen todas las variables de inicio y van siendo removidas una por una en cada paso, eliminando las que generan una menor disminución en la discriminación entre conjuntos hasta que son obtenidos cambios no significativos (Pérez, 2005).

La razón de error aparente puede calcularse fácilmente de la *matriz de confusión* la cual muestra la pertenencia actual a la población contra la pertenencia predicha (Johnson et al., 1992).

		Pertenenca predicha		
		Pob. 1	Pob. 2	
Pertenenca actual	Pob. 1	$n_{1C}$	$n_{1M} = n_1 - n_{1C}$	$n_1$
	Pob. 2	$n_{2M} = n_2 - n_{2C}$	$n_{2C}$	$n_2$

**TABLA 3** Matriz de confusión con 2 poblaciones (Fuente: Johnson et al., 1992)

Donde:

- $n_{1C}$  = Número de observaciones de la población 1 clasificadas correctamente.
- $n_{1M}$  = Número de observaciones de la población 1 incorrectamente clasificadas como de la población 2.
- $n_{2C}$  = Número de observaciones de la población 2 clasificadas correctamente.
- $n_{2M}$  = Número de observaciones de la población 2 incorrectamente clasificadas como de la población 1.
- $n_1$  = Número de observaciones de la población 1.
- $n_2$  = Número de observaciones de la población 2.

La razón de error aparente es (Johnson et al., 1992):

$$APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2} \quad (60)$$

La matriz de confusión o tabla de clasificación (también llamada tabla de asignación) es usada para valorar el desempeño del DA (Shrestha et al., 2007).

Es adecuado resaltar que el problema de la clasificación es completamente estadístico y no se puede esperar de ordinario una discriminación efectiva al 100% (Tusell, 2007).

Los datos utilizados para evaluar la función discriminante no deben ser los mismos datos que los que fueron empleados para estimar dicha función (Jobson, 1992).

Una técnica más intensiva en el uso del ordenador es el uso del Procedimiento jackknife, el cual remueve sólo una observación y usa las  $(n - 1)$  observaciones restantes para determinar la función discriminante. La función discriminante estimada es entonces utilizada para clasificar la observación que fue omitida. Este proceso se repite  $(n - 1)$  veces de manera que cada observación es dejada fuera una vez. La razón de error de clasificación incorrecta obtenida en este proceso está cercana a la estimación insesgada de la razón de error esperada de las muestras con  $n_1$  y  $n_2$  observaciones de los conjuntos 1 y 2, respectivamente (Jobson, 1992).

Para datos continuos las transformaciones para normalizar los datos no producen en

---

general una mejoría de los porcentajes del error de clasificación y, en algunos casos, los porcentajes de error son mayores que en los datos sin transformar (Jobson, 1992).

Una vez efectuado el análisis cluster un análisis discriminante basado en las  $p$  variables originales puede utilizarse para caracterizar las diferencias entre los grupos de conglomerados (Jobson, 1992). Además, la clasificación obtenida a través del análisis cluster puede ser confirmada por medio del DA (Kannel et al., 2007).

#### 4.6 MODELO LINEAL DE REGRESIÓN MÚLTIPLE (MLR)

La regresión múltiple tiene como objetivo analizar un modelo que explique el comportamiento de una variable  $y$  (variable dependiente), empleando la información proporcionada por un conjunto de variables explicativas  $x_1, x_2, \dots, x_k$  (variables independientes) (Pérez, 2005).

El modelo lineal tiene la forma (Pérez, 2005):

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k + u \quad (61)$$

Los *coeficientes o parámetros de regresión*  $b_1, b_2, \dots, b_k$  expresan el valor de la influencia que las variables independientes poseen sobre la variable dependiente  $y$ . El coeficiente  $b_0$  recibe el nombre de término independiente o constante del modelo. El error del modelo se denomina  $u$  (Pérez, 2005).

Se cuenta con un conjunto de  $T$  observaciones efectuadas sobre todas las variables. El modelo se puede expresar de la forma (Pérez, 2005; Pérez, 2007):

$$y_t = b_0 + b_1x_{1t} + b_2x_{2t} + \dots + b_kx_{kt} + e_t \quad t = 1, 2, 3, \dots, T \quad (62)$$

El modelo lineal se formula bajo las siguientes hipótesis (Pérez, 2005):

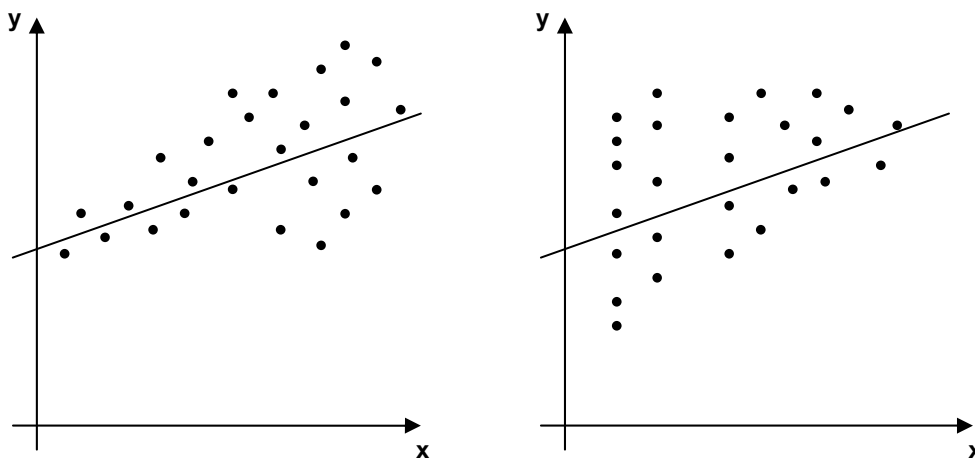
- Las variables  $x_1, x_2, \dots, x_k$  son deterministas (sus valores son conocidos) y sus valores provienen de una muestra registrada.
- Linealidad del modelo (Carbonell et al., 1983).
- El término de error  $e$  es una variable aleatoria con esperanza nula (media cero) y matriz de covarianzas constante y diagonal (ver la Figura 8, donde no se cumple esta condición).
- La variable  $y$  es aleatoria pues depende de  $e$ .
- Se supone que no hay errores de especificación, esto quiere decir que

---

todas las variables  $\mathbf{x}_i$  relevantes para la explicación de la variable  $\mathbf{y}$  están incluidas en el modelo.

- Las variables  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$  son linealmente independientes.
- En ocasiones se toma en cuenta la hipótesis de normalidad de los residuos consistente en que el conjunto de las  $e_t$  sean normales.

Así, si el modelo ajustado es correcto los residuos (errores,  $e_t$ ) deberán exhibir un comportamiento que tienda a confirmar las suposiciones anteriores o al menos no deberán exhibir una negación de estas suposiciones (Miket, 2007).



**FIGURA 8** Dos ejemplos de heteroescadasticidad (Fuente: Carbonell et al., 1983)

Los residuos  $e_i$  representan la cantidad de información que el modelo no ha sido capaz de explicar (Miket, 2007). Cuando se quiere estimar el modelo lineal de regresión múltiple se utiliza el planteamiento de mínimos cuadrados. El concepto de mínimos cuadrados establece que la función que mejor se ajusta a los datos es la que minimiza la varianza del error  $\mathbf{e}$ , lo que se expresa así (Pérez, 2005):

$$\min \sum_{t=1}^T e_t^2 = \sum_{t=1}^T (y_t - (b_0 + b_1 x_{1t} + b_2 x_{2t} + \dots + b_k x_{kt}))^2 \quad (63)$$

Se deriva respecto a  $b_0, b_1, \dots, b_k$  y se iguala con cero (Pérez, 2005):

$$\begin{aligned}
\sum_{t=1}^T y_t &= nb_0 + b_1 \sum_{t=1}^T x_{1t} + \dots + b_k \sum_{t=1}^T x_{kt} \\
\sum_{t=1}^T y_t x_{1t} &= b_0 \sum_{t=1}^T x_{1t} + b_1 \sum_{t=1}^T x_{1t}^2 + \dots + b_k \sum_{t=1}^T x_{1t} x_{kt} \\
&\vdots \\
\sum_{t=1}^T y_t x_{kt} &= b_0 \sum_{t=1}^T x_{kt} + b_1 \sum_{t=1}^T x_{kt} x_{1t} + \dots + b_k \sum_{t=1}^T x_{kt}^2
\end{aligned} \tag{64}$$

Las ecuaciones expresadas en (64) forman un sistema denominado de ecuaciones normales que puede resolverse para  $b_0, b_1, \dots, b_k$  (Pérez, 2005).

El modelo expresado en (62) se puede expresar matricialmente como (Pérez, 2005):

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_t \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{1T} & x_{2T} & \dots & x_{kT} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_t \end{bmatrix} \tag{65}$$

El cual se puede expresar de forma breve como:  $\mathbf{Y} = \mathbf{Xb} + \mathbf{u}$ . Al vector  $\mathbf{Y}$  se le puede llamar vector de observaciones,  $\mathbf{b}$  se denomina vector de parámetros,  $\mathbf{X}$  es la matriz de diseño y  $\mathbf{u}$  se denomina vector de errores o de desviaciones aleatorias (Cuadras, 2007).

Un primer objetivo del desarrollo del modelo es obtener estimaciones (valores numéricos) de los coeficientes  $b_0, b_1, b_2, \dots, b_k$  con base en la información de la muestra. Al contar con el vector de estimaciones  $\hat{\mathbf{b}}$  de los coeficientes se puede escribir (Pérez, 2005):

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{b}} = \hat{b}_0 + \hat{b}_1 \mathbf{x}_1 + \hat{b}_2 \mathbf{x}_2 + \dots + \hat{b}_k \mathbf{x}_k \tag{66}$$

O lo que es lo mismo (Pérez, 2005):

$$\hat{y}_t = \hat{b}_0 + \hat{b}_1 x_{1t} + \hat{b}_2 x_{2t} + \dots + \hat{b}_k x_{kt} \quad t = 1, 2, 3, \dots, T \tag{67}$$

Los residuos son las diferencias que existen entre los valores verdaderos y los valores estimados de la variable  $Y_t$ , o sea:  $\hat{u}_t = y_t - \hat{y}_t$  (Pérez, 2005).

Respecto a la distribución de los residuos comportamientos anómalos en su distribución podrán poner en duda el modelo. Por tanto es necesario conocer acerca

---

de su comportamiento normal, para ellos se supone que todas las suposiciones iniciales se cumplen, aunque en realidad el requerimiento de normalidad sólo es necesario para la distribución de los residuos (Carbonell et al., 1983).

Las estimaciones de los parámetros es posible calcularlas por el método de mínimos cuadrados lo que significa minimizar la suma de los cuadrados de los residuos (la suma residual SR), lo que se expresa así (Pérez, 2005):

$$SR = \sum_{t=1}^T \hat{u}_t^2 = \sum_{t=1}^T (y_t - \hat{y}_t)^2 \quad (68)$$

La expresión de las estimaciones de los parámetros es (Pérez, 2005):

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \quad (69)$$

Para estudiar el ajuste del modelo se requieren las siguientes definiciones (Pérez, 2005):

La *suma total* es la varianza muestral de la variable dependiente (o endógena) y es una medida del valor de las fluctuaciones de dicha variable alrededor de su media (Pérez, 2005):

$$ST = \sum_{i=1}^T (y_t - \bar{y})^2 \quad (70)$$

La *suma explicada* es la magnitud del cambio de la variable estimada  $\hat{y}_t$  alrededor de la media de  $y$ . Debido a esto la suma explicada es una medida de la variación explicada (Pérez, 2005):

$$SE = \sum_{i=1}^T (\hat{y}_t - \bar{y})^2 \quad (71)$$

La *suma residual* revela la magnitud del error del modelo al intentar representar el desempeño de la variable  $y_t$  (Pérez, 2005; Pérez, 2007):

$$SR = \sum_{i=1}^T (y_t - \hat{y}_t)^2 \quad (72)$$

Y la relación entre las tres cantidades anteriores es (Pérez, 2005; Pérez, 2007):



---

$$ST = SE + SR \quad (73)$$

Los tres términos anteriores reciben el nombre de suma de cuadrados (Pérez, 2005; Pérez, 2007).

El *coeficiente de determinación* ( $R^2$ ) es un valor que describe el ajuste global del modelo y se obtiene como el cociente de la variabilidad explicada (la suma explicada) y la variabilidad total (la suma total) (Pérez, 2005):

$$R^2 = \frac{SE}{ST} = \frac{ST - SR}{ST} = 1 - \frac{SR}{ST} \quad (74)$$

El coeficiente de determinación expresa (en tanto por ciento) la variación de la variable dependiente que es explicada por el modelo de regresión (Sánchez Carrión, 1999).

Un modelo tendrá mejor ajuste en cuanto mayor sea  $R^2$  (Pérez, 2005; Draper et al., 1998). El *Coefficiente de correlación múltiple*  $R$  es la raíz cuadrada del coeficiente de determinación (Pérez, 2005).

Se debe apuntar que, con criterio general, no constituye una buena estrategia introducir muchas variables independientes en el modelo. A pesar de que sea posible aumentar el coeficiente de determinación también puede aumentar el valor del error estándar, por lo que cabe pensar en algún sistema de búsqueda de “la mejor ecuación de regresión” que permita buscar dentro de todas las posibilidades aquella que de acuerdo a unos criterios preestablecidos pueda adecuarse mejor a los datos (Visauta Vinacua, 2002).

El coeficiente de determinación depende de las variables introducidas en el modelo. Este inconveniente se arregla sustituyéndolo por el coeficiente de determinación corregido. El coeficiente de determinación corregido por los grados de libertad es (Pérez, 2005; Pérez, 2007):

$$\bar{R}^2 = 1 - (1 - R^2) \frac{T - 1}{T - k - 1} \quad (75)$$

En la expresión anterior  $T$  es el número de observaciones en el conjunto y  $k$  es el número de variables independientes.  $\bar{R}^2$  es un valor adecuado que mide el éxito de la regresión que será mejor en tanto este valor se aproxime a uno (Pérez, 2005; Pérez, 2007).

---

El objetivo en este tipo de análisis es encontrar la ecuación con el menor número de variables que continúe explicando un porcentaje de varianza en la variable dependiente que sea comparable al porcentaje explicado con todas las variables en la ecuación (Miket, 2007).

El coeficiente de determinación  $R^2$  nunca decrecerá al añadir nuevas variables al modelo, no así el coeficiente de determinación corregido  $\bar{R}^2$  que no necesariamente se incrementa con la adición de nuevas variables al modelo y que por este motivo es preferible al momento de estimar el ajuste del mismo a los datos. Por otra parte, un incremento del  $R^2$  no necesariamente conlleva una disminución del error estándar de las estimaciones ya que cambia el número de grados de libertad de la regresión (Visauta Vinacua, 2002).

Se puede considerar a  $\bar{R}^2$  como una buena medida de la calidad de la regresión. El modelo será tanto mejor cuanto mayor sea el coeficiente de determinación corregido  $\bar{R}^2$  (Pérez López et al., 2007).

Para lograr el objetivo de encontrar la ecuación con el menor número de variables existen varios procedimientos: la *eliminación hacia atrás*, la *selección hacia delante* y la *regresión paso a paso* (Miket, 2007).

En el procedimiento de *eliminación hacia atrás* se inicia determinando la ecuación de regresión conteniendo todas las variables ( $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p$ ). Entonces las variables son probadas una por una y la menos significativa es removida del modelo en cada paso (Miket, 2007). Las variables son eliminadas del modelo de acuerdo con los criterios de salida. Se toma en cuenta sólo un criterio de los dos siguientes (Visauta Vinacua, 2002):

- La probabilidad asociada al estadístico F (con un valor de 0,10 por defecto), conocida como la probabilidad de F para remover.
- El valor del propio estadístico, conocido como F para remover y que tiene asignado un valor por defecto de 2,71.

El procedimiento se termina cuando todas las variables que permanecen en la ecuación proporcionan una contribución significativa a la predicción de la variable dependiente  $\mathbf{y}$ . El algoritmo procede así (Miket, 2007):

1. Ajustar una ecuación de regresión conteniendo todas las variables.

---

2. Una prueba F parcial (F para remover) es calculada por cada una de las variables independientes que sigan en la ecuación.

- El estadístico parcial F (F para remover) =  $[RSS_2 - RSS_1]/MSE_1$ .

Donde:

$RSS_1$  = Suma residual de cuadrados con todas las variables que están presentes en la ecuación.

$RSS_2$  = Suma residual de cuadrados con una de las variables removida.

$MSE_1$  = La media cuadrática para el error con todas las variables que están presentes en la ecuación.

3. El valor parcial de F más bajo (F para remover) es comparado con  $F_{\alpha}$  para algún  $\alpha$  pre-especificado. Si  $F_{\text{más bajo}} \leq F_{\alpha}$ , entonces se remueve esa variable y se regresa al paso 2. Si  $F_{\text{más bajo}} > F_{\alpha}$ , entonces se acepta la ecuación tal como está.

El procedimiento de *selección hacia delante* comienza sin variables en la ecuación. Entonces las variables son probadas una cada vez y la más significativa es agregada al modelo en cada paso (Miket, 2007), esto es, las variables son entradas en el modelo una a una de acuerdo con unos criterios de entrada. El criterio de entrada puede seleccionarse de entre estos dos (Visauta Vinacua, 2002):

- La probabilidad asociada al estadístico F (opción por defecto y con un valor de 0,05), conocida como probabilidad F para entrar.
- El valor del propio estadístico F, conocido como F para entrar y que tiene un valor por defecto de 3,84, que es el que corresponde a una probabilidad de 0,05 (Visauta Vinacua, 2002).

El proceso termina cuando todas las variables que no están en la ecuación no tienen efecto significativo en la variable dependiente  $y$ . El algoritmo es desarrollado así (Miket, 2007):

1. Sin variables en la ecuación, se calcula la prueba F parcial (F para entrar) para cada una de las variables independientes que no están en la ecuación.

- El estadístico F parcial (F para entrar) =  $[RSS_2 - RSS_1]/MSE_1$ .

Donde:

$RSS_1$  = Suma residual de cuadrados con todas las variables que

---

están presentes en la ecuación y la variable bajo consideración.

$RSS_2$  = Suma residual de cuadrados con todas las variables que están presentes en la ecuación.

$MSE_1$  = La media cuadrática para el error con las variables que están presentes en la ecuación y la variable bajo consideración.

2. El valor parcial de F más alto (F para entrar) es comparado con  $F_\alpha$  para algún  $\alpha$  pre-especificado. Si  $F_{\text{más alto}} > F_\alpha$ , entonces se añade la variable y se regresa al paso 1. Si  $F_{\text{más alto}} \leq F_\alpha$ , entonces se acepta la ecuación tal como está.

Para el procedimiento de *regresión paso a paso* (Stepwise Regression) en cada paso el modelo de ecuación puede verse incrementado con una nueva variable independiente, puede perder una variable seleccionada anteriormente o puede quedarse tal como se encuentra. Las variables son examinadas en cada paso para entrar o salir del modelo de acuerdo con los criterios especificados. Es una combinación de los dos métodos anteriores y es un procedimiento muy utilizado. Al seleccionar la primera variable de entre los posibles regresores  $x_j$  (variables independientes), se escoge aquel  $x_{(1)}$  que tiene una mayor correlación con la variable dependiente. Si la ecuación que posee la variable elegida en primer lugar no resulta significativa, a un nivel predeterminado, se termina el proceso de selección y la mejor predicción que puede efectuarse de la variable dependiente es su valor medio. En caso contrario, se continúa al paso siguiente. Para seleccionar la segunda variable de entre las restantes variables independientes se escoge aquella que tenga el coeficiente de correlación parcial más alto. El objetivo es encontrar la variable que maximice el porcentaje de variación explicada por el modelo que ya tenía incluido el primer regresor. Esto es equivalente a encontrar la variable con el valor de la F más elevado –valores que se consiguen con la hipótesis de que el nuevo coeficiente es igual a cero- y que superen un valor mínimo establecido (Sánchez Carrión, 1999).

Dentro de la regresión paso a paso la ecuación de regresión se determina sin contener variables en el modelo. Las variables son entonces probadas una a la vez, usando el coeficiente de correlación parcial como una medida de la importancia en la predicción de la variable dependiente  $y$ . En cada paso la variable con el coeficiente de correlación parcial más altamente significativo (F para entrar) es introducida al modelo. Una vez hecho esto el estadístico parcial F (F para remover) es calculado para todas las variables dentro del modelo para probar si alguna de las variables previamente introducida puede ahora ser borrada. Esta prueba del “predictor menos útil actualmente en la ecuación” es llevada a cabo en cada etapa del procedimiento paso a

---

paso. Un predictor que fue incluido como el mejor candidato en una etapa temprana, puede ser superfluo en una etapa posterior debido a las relaciones entre él y otras variables actualmente en la regresión (Draper et al., 1998). Este proceso sigue hasta que no hay más variables que puedan ser agregadas o borradas del modelo. El coeficiente de correlación parcial para una variable dada es la correlación entre esa variable y la respuesta, cuando las variables independientes presentes en la ecuación son mantenidas como fijas. Es también la correlación entre la variable dada y los residuales calculados del ajuste de una ecuación con las variables independientes presentes en la ecuación (Miket, 2007).

El programa SPSS utiliza por defecto como valor de F para entrar 3,84. Junto a este valor de la F, el programa también usa el criterio del nivel de significación de la F. Este criterio recibe el nombre de probabilidad de F para entrar. Por defecto, la probabilidad para entrar toma un valor de 0,05 (Sánchez Carrión, 1999).

SPSS tiene dos criterios para remover variables que fueron previamente seleccionadas. El primero, llamado F para remover es el mínimo valor de la F que ha de tener una variable para continuar en la nueva ecuación, formada por las variables antiguas más la nueva. Por defecto F para remover es igual a 2,71. El segundo criterio tiene que ver con la probabilidad asociada a la F, la probabilidad de F para remover. Si la probabilidad asociada a la F es mayor que 0,10 (por defecto), se elimina la variable. El hecho de que las variables que han sido anteriormente seleccionadas puedan ser removidas de la regresión, permite tratar a todas las variables con equidad, no importando en qué paso hayan sido seleccionadas (Sánchez Carrión, 1999).

Efectuar la comprobación del modelo por medio de sus residuales es parte de la validación (Geladi, 2002).

El gráfico de los residuos contra cada una de las variables independientes permite detectar a una variable con mayor responsabilidad por la heteroscedasticidad, observando aquella en la cual se presenta menos aleatoriedad. También es útil el gráfico de valores observados contra valores predichos cuyos puntos deben de ajustarse lo más posible a una diagonal del primer cuadrante (Pérez, 2005).

En muchos de los casos para solucionar el problema de la heteroscedasticidad es conveniente aplicar logaritmos a los valores. Además, pueden removerse del análisis las variables que más responsabilidad tienen en el problema de la heteroscedasticidad (Pérez, 2005).

---

Los residuales deben exhibir un patrón de independencia. Existen varias pruebas estadísticas que pueden detectar la autocorrelación entre los residuos. Las más comunes son (Miket, 2007):

- El contraste estadístico de Durbin Watson.
- La función de autocorrelación.
- La prueba de rachas.

Un contraste muy empleado para evaluar la presencia de autocorrelación es el contraste de Durbin-Watson para el cual la expresión (Pérez, 2005):

$$D = \frac{\sum_{t=2}^T (\hat{e}_t - \hat{e}_{t-1})^2}{\sum_{t=1}^T e_t^2} \quad (76)$$

Permite aplicar la regla (no tan estricta) de que si el valor de  $D$  es cero existe autocorrelación positiva perfecta, si  $D$  está próximo a 2 no hay autocorrelación y si  $D$  está próximo a 4 hay autocorrelación negativa perfecta. Existen tablas donde  $D$  se encuentra tabulado y según el rango que le corresponda a su valor, se acepta o rechaza la hipótesis de autocorrelación (Pérez, 2005).

La función de autocorrelación para un retardo  $k$  se define como (Miket, 2007):

$$r_k = \frac{\frac{1}{T-k} \sum_{i=1}^{T-k} (e_i - \bar{e})(e_{i+k} - \bar{e})}{\frac{1}{T} \sum_{i=1}^T (e_i - \bar{e})^2} = \frac{\frac{1}{T-k} \sum_{i=1}^{T-k} e_i e_{i+k}}{\frac{1}{T} \sum_{i=1}^T e_i^2} \quad (77)$$

Si los residuos son independientes entonces  $r_k$  deberá estar cercano a cero para todos los valores de  $k$  (Miket, 2007).

La prueba de rachas, que se basa en el orden en el que se obtienen las observaciones muestrales, es una técnica útil para probar la hipótesis nula  $H_0$  de que las observaciones en realidad se extraen al azar. También se puede utilizar para detectar desviaciones en la aleatoriedad de una secuencia de mediciones cuantitativas en el tiempo, ocasionadas por tendencias o periodicidades (Walpole et al., 1999). Es importante destacar que la realización de un contraste de aleatoriedad sobre los datos muestrales, exige que éstos estén dispuestos en el mismo orden en que fueron seleccionados los elementos que forman la muestra. Es decir, no se permite ningún

tipo de manipulación sobre los datos, ni tampoco su clasificación en tablas de frecuencias (Miró i Martínez et al., 2006).

Esta prueba usa el hecho de que los residuos oscilarán alrededor de cero en una proporción “normal” si las partidas o inicios aleatorios son independientes. Este procedimiento contrasta si es aleatorio el orden de aparición de dos valores de una variable (Pérez López, 2008). En esta prueba se analiza la secuencia temporal del signo de los residuos (+ + + – – + + – – – + + +) y se cuenta el número de rachas (o sea, el número de periodos en que los residuos conservan el mismo signo) (Miket, 2007). También se puede llamar racha a una sucesión de valores por encima o debajo de la mediana. La longitud de la racha es el número de observaciones consecutivas con esta propiedad (Peña, 2005b). Este número de rachas será bajo si los residuos están correlacionados positivamente y alto si están correlacionados negativamente (Miket, 2007). Se rechazará la hipótesis de independencia cuando el número de rachas sea significativamente pequeño o grande (Peña, 2005b).

Para resolver la presencia de multicolinealidad se puede (Pérez, 2005):

- Ampliar el número de observaciones.
- Transformar las variables.
- Remover del análisis algunas variables.
- Sustituir las variables independientes por sus primeras componentes principales (puntuaciones).

Cuando los datos no presenten linealidad se pueden hacer transformaciones para linealizar el modelo y efectuar todo el proceso con la estructura linealizada. Algunas transformaciones que pueden ser de ayuda son (Etxeberria, 1999):

Función	Transformación	Forma lineal
$y = \alpha x^b$	$y' = \log y \quad x' = \log x$	$y' = \log \alpha + \beta x'$
$y = \alpha e^{\beta x}$	$y' = \log y$	$y' = \alpha + \beta \log x$
$y = \frac{x}{(\alpha x - \beta)}$	$y' = \frac{1}{y} \quad x' = \frac{1}{x}$	$y' = \alpha - \beta x'$
$y = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}}$	$y' = \log \left( \frac{y}{1 - y} \right)$	$y' = \alpha + \beta x$

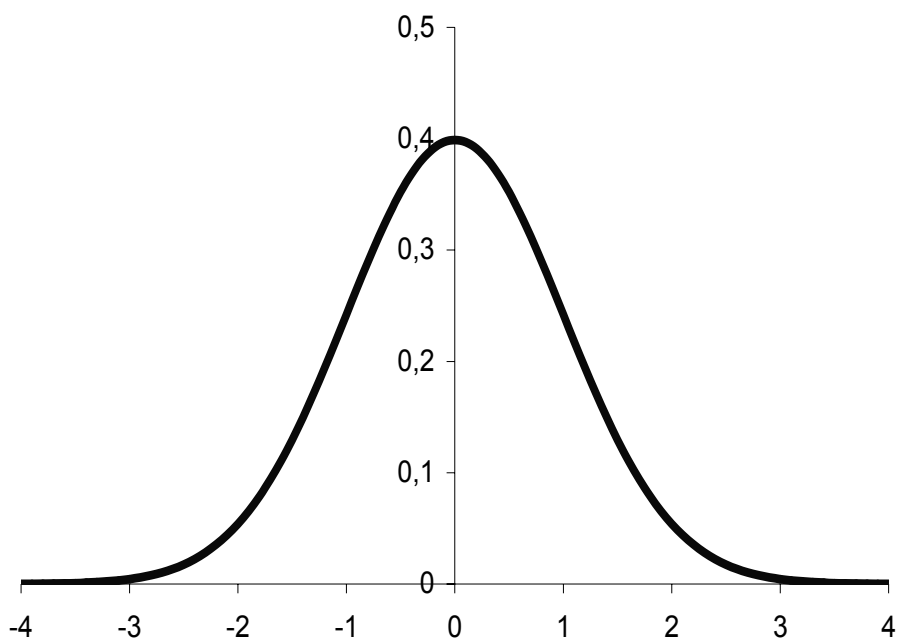
**TABLA 4** Transformaciones para linealizar el modelo (Fuente: Etxeberria, 1999)

---

Respecto a la normalidad, si se detectan desviaciones con respecto a la misma se debe revisar su significatividad mediante el uso de alguna prueba no paramétrica que permita revisar el ajuste de los datos a la curva normal (Etxeberria, 1999). Si el modelo propuesto es correcto sus residuos deberán (aproximadamente) parecerse a las observaciones de una distribución normal con media cero (Figura 9) (Miket, 2007).

Las pruebas estadísticas comunes para probar la normalidad son (Miket, 2007):

1. La prueba de Kolmogorov-Smirnov.
2. La prueba del Chi-cuadrado.



**FIGURA 9** Función densidad de probabilidad normal

La prueba de Kolmogorov-Smirnov utiliza la función de distribución acumulativa empírica como una herramienta para probar la efectividad del ajuste a la distribución. Si  $F_n(x)$  es la proporción de las observaciones en la muestra que son menores o iguales a  $x$  y  $F_0(x)$  denota la función de distribución acumulativa hipotética de la población (población normal), la prueba estadística de Kolmogorov-Smirnov es (Miket, 2007):

$$D_n = \sup_x |F_n(x) - F_0(x)| = \text{La máxima distancia entre } F_n(x) \text{ y } F_0(x) \quad (78)$$

Si los residuos no se ajustan a una distribución normal la observación  $D_n$  será grande (Miket, 2007).



---

La prueba de bondad de ajuste Chi-cuadrada usa el histograma como herramienta para probar la bondad del ajuste de una distribución. Si  $f_i$  expresa la frecuencia observada en cada uno de los intervalos de clase del histograma y  $E_i$  representa el número esperado de observaciones en cada intervalo de clase asumiendo la distribución supuesta, la distribución hipotética es rechazada si el estadístico  $\chi^2$  es grande (mayor que el valor crítico de una distribución Chi-cuadrada con  $m - 1$  grados de libertad, donde  $m$  es el número de intervalos de clase usados para construir el histograma).  $\chi^2$  se define así (Miket, 2007):

$$\chi^2 = \sum_{i=1}^m \frac{(f_i - E_i)^2}{E_i} \quad (79)$$

Si los residuos son plurimodales esto puede ser ocasionado por la existencia de distintas subpoblaciones en la población bajo análisis (Etxeberria, 1999). Una forma de resolver este problema podría ser dividir a la población en grupos por medio de una técnica de clasificación y entonces realizar el análisis a cada grupo por separado.

#### 4.7 ANÁLISIS DE CORRELACIÓN CANÓNICA (CCA)

El análisis de correlación canónica, CCA, es una generalización de la correlación múltiple en aspectos de regresión. Se requiere que las variables se dividan en dos grupos. La asignación de las variables en estos grupos debe depender de la naturaleza de las variables (por ejemplo en uno de los grupos podrían estar las variables fáciles de medir y en el otro las que sean difíciles de obtener). Un aspecto fundamental que se espera dilucidar con el CCA es si se pueden utilizar las variables de uno de los grupos para pronosticar las variables del otro grupo (Johnson, 2000).

El propósito de esta técnica es caracterizar las relaciones estadísticas independientes que existen entre dos (y posiblemente más) conjuntos de variables aleatorias. Las *correlaciones canónicas* asociadas miden la extensión (lineal) de la relación entre los conjuntos (Kettenring, 2006).

El análisis de correlación canónica permite una partición explícita de un conjunto de datos en variables explicativas y respuesta con el propósito de identificar pares de vectores (patrones) en estos espacios parciales de datos (Callies, 2005).

En el CCA el objetivo es buscar una combinación lineal de un grupo de variables independientes medidas en escala numérica y una combinación lineal de un segundo

conjunto de variables dependientes también medidas en escala numérica, correlacionadas al máximo (que tengan la correlación más grande) (Pérez, 2005).

La idea es de principio determinar un par de combinaciones lineales que posean la mayor correlación. Después se obtiene el par de combinaciones lineales que tenga la mayor correlación de entre los pares incorrelacionados con el par seleccionado inicialmente. Y así el método continúa. Los pares de combinaciones lineales reciben el nombre de *variables canónicas* y sus correlaciones son llamadas *correlaciones canónicas* (Johnson et al., 1992).

Si se dispone de un conjunto de datos de  $n$  observaciones y  $k$  variables pueden subdividirse en dos grupos: el primero con  $p_1$  variables y el segundo con  $p_2$  variables (Peña, 2002). Siendo  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_{p_1})$  e  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_{p_2})$  variables estadísticas centradas, que constituyen en conjunto la variable  $p$ -dimensional  $(\mathbf{x}_1, \dots, \mathbf{x}_{p_1}, \mathbf{y}_1, \dots, \mathbf{y}_{p_2})$  con el valor  $p = p_1 + p_2$ , y  $p_2 < p_1$ , con matriz de covarianzas cuadrada de orden  $p$  particionada así (Pérez, 2005):

$$\mathbf{S} = \left[ \begin{array}{ccc|ccc} \mathbf{S}_{11} & \cdots & \mathbf{S}_{1p_1} & \mathbf{S}_{1,p_1+1} & \cdots & \mathbf{S}_{1p} \\ \mathbf{S}_{21} & \cdots & \mathbf{S}_{2p_1} & \mathbf{S}_{2,p_1+1} & \cdots & \mathbf{S}_{2p} \\ \vdots & & \vdots & \vdots & & \vdots \\ \mathbf{S}_{p_11} & \cdots & \mathbf{S}_{p_1p_1} & \mathbf{S}_{p_1,p_1+1} & \cdots & \mathbf{S}_{p_1,p} \\ \hline \mathbf{S}_{p_1+1,1} & \cdots & \mathbf{S}_{p_1+1,p_1} & \mathbf{S}_{p_1+1,p_1+1} & \cdots & \mathbf{S}_{p_1+1,p} \\ \vdots & & \vdots & \vdots & & \vdots \\ \mathbf{S}_{p,1} & \cdots & \mathbf{S}_{p,p_1} & \mathbf{S}_{p,p_1+1} & \cdots & \mathbf{S}_{p,p} \end{array} \right] = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix} \quad (80)$$

En la expresión anterior  $\mathbf{S}_{11}$  y  $\mathbf{S}_{22}$  son las matrices de covarianzas del conjunto de las  $\mathbf{x}_i$  y del conjunto de las  $\mathbf{y}_i$ , respectivamente.  $\mathbf{S}_{12}$  es la matriz de covarianzas de las  $p_1$  variables del primer grupo con las  $p_2$  variables del segundo grupo. Además, se tiene que  $\mathbf{S}_{12} = \mathbf{S}_{21}'$  (Calvo, 1993).

Si en lugar de utilizar matrices de covarianzas se emplea la matriz de correlaciones se tiene (Calvo, 1993):

$$\mathbf{R} = \begin{matrix} & \begin{matrix} p_1 & p_2 \end{matrix} \\ \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{bmatrix} & \begin{matrix} p_1 \\ p_2 \end{matrix} \end{matrix} \quad (81)$$

En el CCA se determinan  $r$  ( $r < p_1$ ) pares sucesivos de nuevas variables ( $\mathbf{u}_i, \mathbf{v}_i$ ) llamadas *variables canónicas* que sean combinación lineal de las variables originales, de forma que cada par constituya la mejor explicación de cada conjunto respecto al otro, que no haya sido proporcionada por los pares anteriores. Estas combinaciones lineales tienen la forma (Pérez, 2005):

$$\begin{aligned} \mathbf{u}_i &= \alpha_{i1}\mathbf{x}_1 + \alpha_{i2}\mathbf{x}_2 + \dots + \alpha_{i,p1}\mathbf{x}_{p1} \\ \mathbf{v}_i &= \gamma_{i1}\mathbf{y}_1 + \gamma_{i2}\mathbf{y}_2 + \dots + \gamma_{i,p2}\mathbf{y}_{p2} \end{aligned} \quad (82)$$

Lo que se está haciendo es reemplazar las  $p_1 + p_2$  variables por dos nuevas variables (las variables canónicas)  $\mathbf{u}_i$  y  $\mathbf{v}_i$  (Calvo, 1993).

El primer par de variables canónicas es un par de combinaciones lineales  $\mathbf{u}_1$  y  $\mathbf{v}_1$ , que tienen varianzas unitarias, en las cuales la correlación entre los grupos se maximiza (Johnson et al., 1992).

El segundo par de variables canónicas es un par de combinaciones lineales  $\mathbf{u}_2, \mathbf{v}_2$ , que tienen varianzas unitarias y las cuales maximizan la correlación de entre todas las alternativas que están incorrelacionadas con el primer par de variables canónicas (Johnson et al., 1992). Y así sucesivamente para las restantes variables canónicas.

Suponiendo que no hay singularidades dentro de ambos conjuntos de variables el número de variables canónicas que puede ser definido en esta forma es igual al número de variables en el más pequeño de los dos conjuntos (Kettenring, 2006).

En total aplicando el CCA a los dos conjuntos de variables,  $\mathbf{X}$  con  $p_1$  variables,  $\mathbf{Y}$  con  $p_2$  variables se pueden obtener  $r$  combinaciones lineales (número de pares de variables canónicas), donde  $r = \min(p_1, p_2)$  que se pueden ordenar por su importancia (Peña, 2002). Consecuentemente hay  $r$  coeficientes de correlación canónica (Tusell, 2007):

$$\begin{array}{lll}
\mathbf{u}_1 = \mathbf{X}\mathbf{a}_1, & \mathbf{v}_1 = \mathbf{Y}\mathbf{b}_1, & r_1 = \text{cor}(\mathbf{u}_1, \mathbf{v}_1) \\
\mathbf{u}_2 = \mathbf{X}\mathbf{a}_2, & \mathbf{v}_2 = \mathbf{Y}\mathbf{b}_2, & r_2 = \text{cor}(\mathbf{u}_2, \mathbf{v}_2) \\
\vdots & \vdots & \vdots \\
\mathbf{u}_r = \mathbf{X}\mathbf{a}_r, & \mathbf{v}_r = \mathbf{Y}\mathbf{b}_r, & r_r = \text{cor}(\mathbf{u}_r, \mathbf{v}_r)
\end{array} \tag{83}$$

Cuando existen correlaciones canónicas altas (se obtienen varios coeficientes de correlación canónica) las conclusiones son más útiles debido a que se conoce que medidas de diferentes variables muestran relación entre sí y además se puede saber la influencia relativa que poseen cada una de las variables  $\mathbf{x}_i$  y cada una de las  $\mathbf{y}_i$  (Calvo, 1993).

Las  $r$  variables canónicas obtenidas en el CCA tienen algunas propiedades que se mencionan a continuación (Peña, 2002):

1. Las variables canónicas tienen por coeficientes a los vectores propios relacionados al valor propio de las matrices  $\mathbf{S}_{ii}^{-1/2}\mathbf{S}_{ij}\mathbf{S}_{jj}^{-1}\mathbf{S}_{ji}$ , para  $i = 1, 2$  y con la condición  $i \neq j$ .
2. Las variables canónicas constituyen representaciones de los dos grupos de variables y se obtienen en pares atendiendo al requerimiento de máxima correlación.
3. Se acostumbra tomar los signos de las variables canónicas de manera que las correlaciones entre las variables canónicas  $\mathbf{u}_i$  y  $\mathbf{v}_i$  sean de signo positivo.
4. El valor de las correlaciones canónicas  $\lambda_i^2$  es el coeficiente de correlación elevado al cuadrado, existente entre las dos variables canónicas que correspondan.
5. Las correlaciones canónicas son invariantes a las transformaciones lineales de las variables.
6. La primera correlación canónica  $\lambda_1^2$  tiene un valor mayor o igual que el mayor coeficiente de correlación simple elevado al cuadrado, entre variables individuales de cada conjunto  $\mathbf{X}$  e  $\mathbf{Y}$ .
7. Si se elige estandarizar las variables y se desarrolla el análisis con las matrices de correlación, las correlaciones canónicas no varían.

Una manera de aplicar el CCA es iniciar por el primer cálculo de las correlaciones canónicas. Una vez terminado se pueden excluir para análisis posteriores tanto las variables del conjunto  $\mathbf{X}$  como del conjunto  $\mathbf{Y}$  que tengan una menor preponderancia

---

explicativa. De esta manera se efectúa un segundo cálculo y se revisan los resultados generados que serán distintos de los obtenidos con el primer cálculo. Se deben seguir efectuando aplicaciones del CCA y comparando resultados entre sí hasta que se consigan conclusiones importantes tanto de la mejor correlación canónica observada, así como de aquellos grupos de variables que aparecen en dicha correlación. También es importante detectar las variables que fueron eliminadas por su débil aportación explicativa y señalar las posibles causas de que ello ocurriera (Calvo, 1993).

Es interesante mencionar que además de su propio campo de acción el CCA abarca como casos particulares, los métodos de regresión y, por extensión, las técnicas del análisis discriminante. Para el caso del análisis discriminante si se tienen  $G-1$  variables dependientes y estableciendo (Peña, 2002):

$$y_i = \begin{cases} 1 & \text{si la observación pertenece al grupo 1, } i = 1, \dots, G-1 \\ 0 & \text{en otro caso} \end{cases} \quad (84)$$

Estas  $G-1$  variables del tipo binario (pues sólo tiene cada una de ellas dos valores posibles) se pueden ordenar en una matriz  $\mathbf{Y}$  con dimensiones  $n \times (G-1)$ . Además se tiene la matriz  $\mathbf{X}$  con dimensiones  $n \times p$ , con  $p$  variables independientes. La correlación canónica existente entre las matrices  $\mathbf{X}$  e  $\mathbf{Y}$  lleva a la misma solución que la que proporciona el análisis discriminante (Peña, 2002).

#### 4.8 PLSR

La regresión en mínimos cuadrados parciales, PLSR, es un término libre para una familia de filosofías y técnicas relacionadas con los métodos de modelación multivariantes derivados de los conceptos básicos de Herman Wold de ajuste iterativo de modelos bi-lineales en varios bloques de variables (Martens et al., 1989).

La principal aplicación del modelado con PLS en la calibración multivariante es usar el modelo de regresión final para propósitos predictivos (De Jong, 1993).

El PLSR es una técnica de calibración multivariante que permite encontrar una relación entre un conjunto de datos predictores (independientes)  $\mathbf{X}$  y un conjunto de respuestas (dependientes)  $\mathbf{Y}$ . Este método es capaz de proporcionar predicciones estables aún cuando  $\mathbf{X}$  contenga variables altamente correlacionadas (Singh et al., 2007).

---

Este método fue una propuesta efectuada con el propósito de resolver el problema de multicolinealidad en la regresión o la calibración. Comparado con otros métodos de regresión para datos colineales, una ventaja con el PLSR es que “la información en la matriz  $\mathbf{Y}$  es utilizada” (Helland, 2006).

El PLSR es una extensión del modelo de regresión lineal múltiple (MLR). Busca un modelo lineal que describa unas variables a pronosticar en términos de otras variables observables (predictivas). Además realiza esta función sin imponer las restricciones de otros métodos ([www.statsoft.com](http://www.statsoft.com)).

El método PLSR es una buena alternativa a los métodos clásicos de MLR y regresión en componentes principales (PCR) porque es más robusto. Robusto significa que los parámetros del modelo no cambian mucho cuando nuevas muestras de calibración son tomadas de la población total (Geladi et al., 1986).

Dentro del ámbito de los modelos, otra forma de expresar el concepto de robustez es el siguiente: una propiedad deseable de un buen estimador para un parámetro  $\vartheta$  en el modelo  $f(x)$  es continuar siendo razonablemente bueno como estimador de  $\vartheta$  si el modelo experimenta una pequeña modificación. Cuando se da esta propiedad, se dice que el estimador es robusto para  $\vartheta$  (Peña, 2005b).

Cualquier análisis de datos efectuado con PLSR está basado en la suposición de *homogeneidad*. Esto significa que el proceso o sistema investigado debe de estar en un estado similar a través de toda la investigación y el mecanismo de influencia de  $\mathbf{X}$  sobre  $\mathbf{Y}$  debe ser el mismo (Wold et al., 2001b).

El método busca componentes de  $\mathbf{X}$  que sean también relevantes para  $\mathbf{Y}$ . Así, el PLSR busca un conjunto de componentes (los *vectores latentes*) que desarrollen una descomposición simultánea de  $\mathbf{X}$  e  $\mathbf{Y}$  con la limitación de que estos expliquen, tanto como sea posible, la covarianza entre  $\mathbf{X}$  e  $\mathbf{Y}$  (Abdi, 2007).

Los modelos de variables latentes como el PLSR, asumen la existencia de un conjunto de variables latentes subyacentes que conducen los cambios en el proceso o sistema investigado (Wold et al., 2001).

Las limitaciones del MLR no afectan al método PLSR. Este último aprovecha que la correlación entre las variables originales es elevada para sustituirlas por unas pocas variables que no están correlacionadas entre sí. Estas variables que sustituyen dentro

---

del método a las originales reciben el nombre de variables latentes y se obtienen por medio de la combinación lineal de todas las variables de inicio, como lo indica la siguiente expresión (Ferré, 2006):

$$\mathbf{t}_i = w_1\mathbf{x}_1 + w_2\mathbf{x}_2 + \dots + w_p\mathbf{x}_p \quad (85)$$

Donde  $w_1, w_2, \dots, w_p$  son los *pesos* o *cargas*. Y  $p$  es el número de variables originales. El valor que una observación adquiere en esta nueva variable,  $\mathbf{t}_i$ , recibe el nombre de *puntuación (score)*. En general son necesarias dos o más puntuaciones (dos o más variables latentes) para describir las observaciones de un fenómeno cualquiera (Ferré, 2006).

El método PLSR extrae de las variables predictivas ( $\mathbf{X}$ ) las puntuaciones  $\mathbf{t}$ , y de las variables respuesta ( $\mathbf{Y}$ ) las puntuaciones  $\mathbf{u}$ , las cuales maximizan su covarianza, esto es  $\max \{ \text{cov}(\mathbf{t}, \mathbf{u}) \}$  cada vez que se extrae una componente y relaciona cada par de saturaciones en una relación lineal. Debido a esto, el método es nombrado como “parcial”, pues se aplica a la parte restante de la información que está sin explicar después de extraer cada una de las componentes (Li et al., 2003).

Además el PLSR ha demostrado ser una poderosa técnica de regresión para problemas donde los datos tienen ruido (Baffi et al., 1999).

El uso de este método tiene varias ventajas. La primera es que con él se puede obtener una solución más robusta que con otros métodos (la solución no se ve muy afectada por el ruido en las observaciones al contrario de lo que sucede en otras técnicas). La segunda, no es necesario calcular matrices inversas. Tercera, la precisión puede ser mejorada por medio del ajuste del número de variables latentes (factores) (Li et al., 2003).

Varias aplicaciones han mostrado que este método resuelve los problemas de predicción multivariante para datos altamente colineales o linealmente dependientes, con una capacidad de predicción satisfactoria. El modelo resultante a menudo posee buenas propiedades de interpretación también debido a su parsimonia dimensional (De Jong, 1993).

El principal propósito del PLSR es construir un modelo lineal  $\mathbf{Y} = \mathbf{XB} + \mathbf{E}$ , donde  $\mathbf{Y}$  es una matriz respuesta de  $n$  casos por  $m$  variables,  $\mathbf{X}$  es una matriz de variables predictoras (o independientes) de  $n$  casos por  $p$  variables,  $\mathbf{B}$  es una matriz de

---

coeficientes de regresión con dimensiones  $p \times m$ , por último  $\mathbf{E}$  es un término de ruido para el modelo el cual posee las mismas dimensiones que  $\mathbf{Y}$ . Generalmente, las variables en  $\mathbf{X}$  e  $\mathbf{Y}$  están centradas, por medio de la sustracción de sus medias y están escaladas, lo cual se efectúa dividiéndolas entre sus desviaciones estándar (www.statsoft.com).

La modelación bi-lineal es un enfoque poderoso y flexible de la calibración multivariante. Puede producir predictores informativos y confiables  $\hat{\mathbf{Y}} = f(\mathbf{X})$ , por medio de la proyección de muchas variables (Martens et al., 1989):

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_K) \quad (86)$$

En unas pocas variables (Martens et al., 1989):

$$\mathbf{T} = (\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_A) \quad (87)$$

O sea (Martens et al., 1989):

$$\mathbf{T} = \mathbf{XW} \quad (88)$$

Y usando estas variables “comprimidas”  $\mathbf{T}$ , como regresores para  $\mathbf{Y}$  (Martens et al., 1989).

Por esta razón, las estructuras comunes en las variables  $\mathbf{X}$  son comprimidas dentro de modelos estabilizados, dejando fuera mucho del ruido como residuales. El problema importante es, sin embargo, como definir los factores más relevantes,  $\mathbf{T}$ , desde el punto de vista de la predicción (Martens, et al., 1989).

El número máximo de factores o variables latentes que puede ser calculado es el resultado de  $\min(N, K)$ , donde  $N$  es el número de observaciones y  $K$  es el número de variables independientes  $\mathbf{X}$  (Li et al., 2002).

El PLSR divide las variables independientes y las variables dependientes medidas en el producto de dos matrices más pequeñas. Esto involucra un paso de compresión de datos donde los datos medidos son comprimidos a un pequeño número de intensidades llamadas puntuaciones (*scores*), en un nuevo sistema coordinado. Los nuevos ejes son llamados *factores* y son combinaciones lineales de las variables originales. Un factor representa una variación sistemática encontrada en el conjunto de



---

datos. Los coeficientes de regresión de cada variable original en cada factor son llamados las *saturaciones* (Blanco, et al., 1995).

El PLSR produce una matriz de pesos o cargas **W** que refleja la estructura de covarianza entre las variables predictoras y las variables respuesta. La matriz de pesos **W** tiene dimensiones  $p \times c$ . Las columnas de **W** son los vectores de pesos para las columnas de **X** que al multiplicarse entre sí producen la matriz de puntuaciones factoriales **T** de dimensiones  $n \times c$ . Estos pesos o cargas son calculados de manera que cada uno de ellos maximice la covarianza entre las respuestas y las correspondientes puntuaciones factoriales. El procedimiento de mínimos cuadrados ordinario (OLS) para la regresión de **Y** sobre **T** es entonces desarrollado para producir **Q**, que son las cargas para **Y** (o los pesos para **Y**), de manera que  $Y = TQ + E$ . Una vez que **Q** es calculado se tiene la expresión  $Y = XB + E$ , donde  $B = WQ$  y la predicción con el modelo está completa (www.statsoft.com).

Una matriz adicional que es necesaria para una descripción completa del procedimiento PLSR es la matriz de saturaciones factoriales **P** de dimensiones  $p \times c$  la cual proporciona un modelo factorial  $X = TP + F$ , donde **F** es la parte sin explicar de las puntuaciones de **X** (www.statsoft.com).

El algoritmo estándar para calcular las componentes del PLSR (o sea, los factores) es el algoritmo NIPALS (*nonlinear iterative partial least squares* o algoritmo en mínimos cuadrados parciales iterativo no lineal). Existen muchas variantes del algoritmo NIPALS (www.statsoft.com).

El Algoritmo NIPALS estándar es el siguiente (Baffi et al., 1999):

0. Centrar en media y escalar **X** e **Y**.
1. Establecer las puntuaciones de las variables respuesta, **u**, iguales a una columna de **Y**.
2. Calcular los pesos o cargas de las variables entrada, **w**, por medio de la regresión de **X** sobre **u**:

$$\mathbf{w}^T = \frac{\mathbf{u}^T \cdot \mathbf{X}}{\mathbf{u}^T \cdot \mathbf{u}} \quad (89)$$

3. Normalizar **w** a longitud unitaria:

---

---

$$\mathbf{w} = \frac{\mathbf{w}}{\|\mathbf{w}\|} \quad (90)$$

4. Calcular las puntuaciones de las entradas,  $\mathbf{t}$ :

$$\mathbf{t} = \frac{\mathbf{X} \cdot \mathbf{w}}{\mathbf{w}^T \cdot \mathbf{w}} \quad (91)$$

5. Calcular las cargas de salida,  $\mathbf{q}$ , por medio de la regresión de  $\mathbf{Y}$  sobre  $\mathbf{t}$ :

$$\mathbf{q}^T = \frac{\mathbf{t}^T \cdot \mathbf{Y}}{\mathbf{t}^T \cdot \mathbf{t}} \quad (92)$$

6. Normalizar  $\mathbf{q}$  a longitud unitaria:

$$\mathbf{q} = \frac{\mathbf{q}}{\|\mathbf{q}\|} \quad (93)$$

7. Calcular las nuevas puntuaciones de la salida,  $\mathbf{u}$ :

$$\mathbf{u} = \frac{\mathbf{Y} \cdot \mathbf{q}}{\mathbf{q}^T \cdot \mathbf{q}} \quad (94)$$

8. Comprobar la convergencia de  $\mathbf{u}$ ; si el caso es que sí se está dando la convergencia, ir al paso 9, si no es así ir al paso 2.

9. Calcular las saturaciones de las entradas,  $\mathbf{p}$ , por medio de la regresión de  $\mathbf{X}$  sobre  $\mathbf{t}$ :

$$\mathbf{p}^T = \frac{\mathbf{t}^T \cdot \mathbf{X}}{\mathbf{t}^T \cdot \mathbf{t}} \quad (95)$$

10. Calcular el coeficiente interno de regresión del modelo  $b$ :

$$\mathbf{b} = \frac{\mathbf{t}^T \cdot \mathbf{u}}{\mathbf{t}^T \cdot \mathbf{t}} \quad (96)$$

11. Obtener la matriz residual de las entradas:

$$\mathbf{E} = \mathbf{X} - \mathbf{t} \cdot \mathbf{p}^T \quad (97)$$

12. Obtener la matriz residual de las salidas:

---


$$\mathbf{F} = \mathbf{Y} - \mathbf{b} \cdot \mathbf{t} \cdot \mathbf{q}^T \quad (98)$$

13. Si dimensiones adicionales son necesarias en el PLSR reemplazar  $\mathbf{X}$  e  $\mathbf{Y}$  por  $\mathbf{E}$  y  $\mathbf{F}$ , respectivamente y repetir los pasos del 1 al 13.

En algunas investigaciones el algoritmo anterior NIPALS es sintetizado así (Li et al., 2003):

- Desde  $i = 1$  hasta  $a$  (el número de componentes elegido).

$\mathbf{w}_i$  = Es el vector singular dominante restante de  $\mathbf{X}^T \mathbf{Y}$ .

- Calcular las puntuaciones y las saturaciones.

$$\begin{aligned} \mathbf{t}_i &= \mathbf{X} \mathbf{w}_i / \|\mathbf{X} \mathbf{w}_i\|, & \mathbf{p}_i &= \mathbf{X}^T \mathbf{t}_i \\ \mathbf{q}_i &= \mathbf{Y}^T \mathbf{t}_i / \|\mathbf{Y}^T \mathbf{t}_i\|, & \mathbf{u}_i &= \mathbf{Y} \mathbf{q}_i \end{aligned} \quad (99)$$

- Se obtiene el coeficiente interno de regresión.

$$\mathbf{b}_i = \mathbf{u}_i^T \mathbf{t}_i \quad (100)$$

- Se deflacionan las matrices de datos.

$$\mathbf{X} = \mathbf{X} - \mathbf{t}_i \mathbf{p}_i^T, \quad \mathbf{Y} = \mathbf{Y} - \mathbf{b}_i \mathbf{t}_i \mathbf{q}_i^T \quad (101)$$

- Se regresa al paso 1 para la próxima componente.

En situaciones prácticas el algoritmo NIPALS usualmente converge (Geladi et al., 1986).

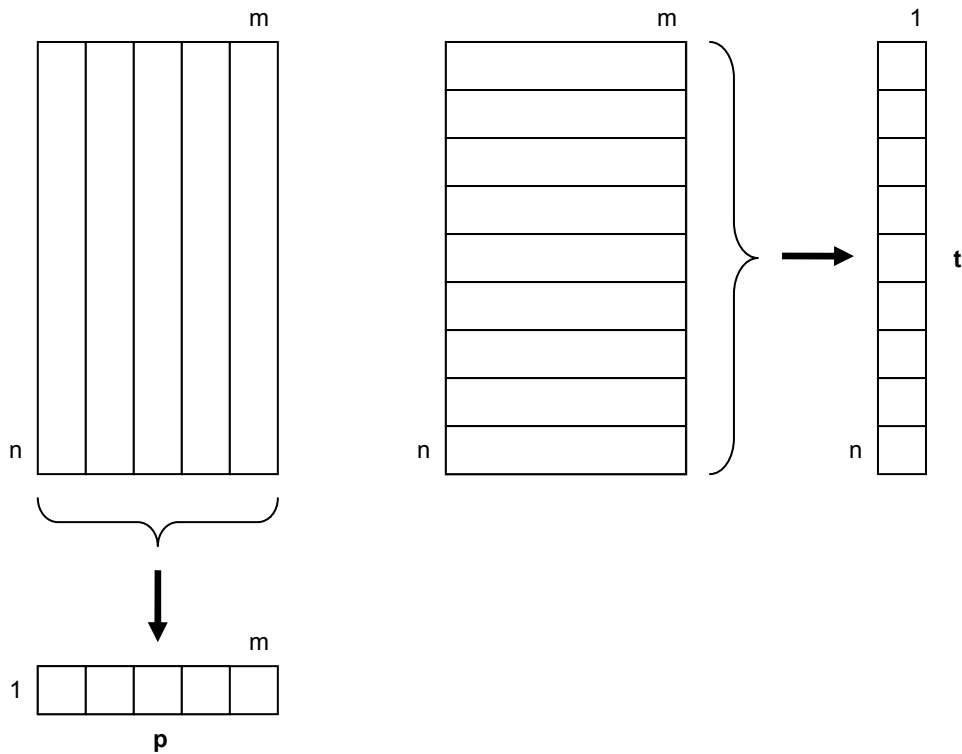
En el PLSR  $\mathbf{X}$  (Figura 10) e  $\mathbf{Y}$  son descompuestos en la suma de series de matrices de rango 1 (Li et al., 2003):

$$\begin{aligned} \mathbf{X} &= \mathbf{t}_1 \mathbf{p}_1^T + \dots + \mathbf{t}_a \mathbf{p}_a^T + \mathbf{E} = \mathbf{T} \mathbf{P}^T + \mathbf{E} \\ \mathbf{Y} &= \mathbf{u}_1 \mathbf{q}_1^T + \dots + \mathbf{u}_a \mathbf{q}_a^T + \mathbf{F} = \mathbf{U} \mathbf{Q}^T + \mathbf{F} \end{aligned} \quad (102)$$

Donde  $\mathbf{t}_i$  y  $\mathbf{u}_i$  son llamados los vectores de puntuaciones;  $\mathbf{T}$  y  $\mathbf{U}$  son las matrices de puntuaciones con las columnas  $\mathbf{t}_i$  y  $\mathbf{u}_i$ , respectivamente;  $\mathbf{p}_i$  y  $\mathbf{q}_i$  son los vectores de saturaciones;  $\mathbf{P}$  y  $\mathbf{Q}$  son las matrices de saturaciones con columnas  $\mathbf{p}_i$  y  $\mathbf{q}_i$ ;  $\mathbf{E}$  y  $\mathbf{F}$  son los residuales de  $\mathbf{X}$  e  $\mathbf{Y}$ . Y la relación entre  $\mathbf{X}$  e  $\mathbf{Y}$  es descrita por la relación interna entre  $\mathbf{t}_i$  y  $\mathbf{u}_i$ , donde  $\mathbf{u}_i = b_i \mathbf{t}_i$  y  $b_i$  es llamado el coeficiente de regresión parcial. En la

ecuación (102) el sufijo  $a$  representa el número de factores comunes utilizado (Li et al., 2003).

La fuerza de este método es la capacidad de eliminar colinealidades y ruido dentro de la matriz de predictores  $\mathbf{X}$ . Esto se logra mediante la extracción de matrices parciales  $\mathbf{X}_p$  de  $\mathbf{X}$ , donde  $\mathbf{X}_p$  contiene – en el caso óptimo – sólo la información que es relevante para la descripción de la matriz de respuestas  $\mathbf{Y}$ . Cada matriz parcial  $\mathbf{X}_p$  es representada por un conjunto de vectores ortogonales  $\mathbf{t}$  y  $\mathbf{w}$ . El vector  $\mathbf{t}$  contiene las puntuaciones de los objetos en  $\mathbf{X}_p$  sobre la variable latente  $\mathbf{W}$ . Así, dos vectores ortogonales  $\mathbf{u}$  y  $\mathbf{q}$  son calculados para reproducir  $\mathbf{Y}_p$  como una matriz parcial de  $\mathbf{Y}$ . Además, las variables latentes son calculadas bajo la restricción de máxima covarianza entre  $\mathbf{t}$  y  $\mathbf{u}$  con el propósito de optimizar la capacidad predictiva del modelo de regresión (Einax et al., 1999).



**FIGURA 10** Puntuaciones ( $\mathbf{t}$ ) y saturaciones ( $\mathbf{p}$ ) son obtenidas proyectando  $\mathbf{X}$  dentro de los vectores

Una forma de ver al PLSR es como un conformador de “nuevas variables” (variables latentes),  $\mathbf{t}_a$ , como las combinaciones lineales de las  $\mathbf{X}$  y luego usa las nuevas  $\mathbf{t}_a$  como variables predictoras de  $\mathbf{Y}$  (Wold et al., 2001b).

Finalmente el modelo de regresión puede expresarse como (Li et al., 2003):

$$\begin{aligned} \mathbf{X} &= \mathbf{TP}^T + \mathbf{E} \\ \mathbf{Y} &= \mathbf{TBQ}^T + \mathbf{F} = \mathbf{UQ}^T + \mathbf{F} \end{aligned} \quad (103)$$

Donde  $\mathbf{B}$  es una matriz diagonal con los elementos  $b_i$  (Li et al., 2003):

$$\mathbf{B} = \begin{bmatrix} b_1 & & & 0 \\ & b_2 & & \\ & & \ddots & \\ 0 & & & b_a \end{bmatrix} \quad (104)$$

En la cual  $a$  es el número de factores o regresiones parciales seleccionado (Li et al., 2003).

Si hablamos de estimaciones de las variables respuesta (dependientes) se puede utilizar la siguiente expresión (Abdi, 2007):

$$\hat{\mathbf{Y}} = \mathbf{XB}_{\text{PLS}} \quad (105)$$

Donde  $\mathbf{B}_{\text{PLS}}$  es la matriz  $\mathbf{B}$  obtenida durante el desarrollo del PLSR (Abdi, 2007).

Un método alternativo de estimación de las componentes del PLSR es el algoritmo SIMPLS el cual se describe a continuación ([www.statsoft.com](http://www.statsoft.com)):

Para cada  $h = 1, \dots, c$ , donde ([www.statsoft.com](http://www.statsoft.com)):

$$\begin{aligned} \mathbf{A}_0 &= \mathbf{X}'\mathbf{Y} \\ \mathbf{M}_0 &= \mathbf{X}'\mathbf{X} \\ \mathbf{C}_0 &= \mathbf{I} \end{aligned} \quad (106)$$

Y dado  $c$ , se inicia con ([www.statsoft.com](http://www.statsoft.com)):

1. Calcular  $\mathbf{q}_h$ , el autovector dominante de  $\mathbf{A}_h'\mathbf{A}_h$ .
2.  $\mathbf{w}_h = \mathbf{A}_h\mathbf{q}_h$ ,  $\mathbf{c}_h = \mathbf{w}_h'\mathbf{M}_h\mathbf{w}_h$ ,  $\mathbf{w}_h = \mathbf{w}_h/\sqrt{\mathbf{c}_h}$  y se almacena  $\mathbf{w}_h$  dentro de  $\mathbf{W}$  como una columna.
3.  $\mathbf{p}_h = \mathbf{M}_h\mathbf{w}_h$  y se almacena  $\mathbf{p}_h$  dentro de la matriz  $\mathbf{P}$  como una columna.
4.  $\mathbf{q}_h = \mathbf{A}_h'\mathbf{w}_h$  y se almacena  $\mathbf{q}_h$  dentro de la matriz  $\mathbf{Q}$  como una columna.
5.  $\mathbf{v}_h = \mathbf{C}_h\mathbf{p}_h$  y  $\mathbf{v}_h = \mathbf{v}_h/\|\mathbf{v}_h\|$ .
6.  $\mathbf{c}_{h+1} = \mathbf{c}_h - \mathbf{v}_h\mathbf{v}_h'$  y  $\mathbf{M}_{h+1} = \mathbf{M}_h - \mathbf{p}_h\mathbf{p}_h'$ .

---

7.  $\mathbf{A}_{h+1} = \mathbf{c}_h \mathbf{A}_h$

La matriz  $\mathbf{T}$  en el algoritmo SIMPLS es calculada como  $\mathbf{T} = \mathbf{XW}$  y la matriz  $\mathbf{B}$  para la regresión de  $\mathbf{Y}$  sobre  $\mathbf{X}$  es obtenida con  $\mathbf{B} = \mathbf{WQ}'$  (www.statsoft.com).

El SIMPLS constituye un algoritmo muy eficiente para un amplio rango de estructuras de datos (De Jong, 1993).

Otro método para desarrollar el PLSR es el Algoritmo Kernel que fue propuesto por Lindgren et al. (1993). Presupone que las columnas de  $\mathbf{X}$  e  $\mathbf{Y}$  están centradas en media y escaladas apropiadamente antes de efectuar la estimación del modelo PLSR. Los pasos de este algoritmo son (Dayal et al., 1997):

1. Calcular las matrices de covarianza  $\mathbf{X}^T \mathbf{X}$  y  $\mathbf{X}^T \mathbf{Y}$ . La matriz kernel  $\mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X}$  puede calcularse mediante la multiplicación de  $\mathbf{X}^T \mathbf{Y}$  por  $(\mathbf{X}^T \mathbf{Y})^T$ .
2. El vector de pesos o cargas  $\mathbf{w}_a$  es calculado como el autovector correspondiente al autovalor más grande de  $(\mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X})_a$  usando tanto el método de la potencia u otras técnicas (SVD, etc.):

$$\mathbf{w}_a \propto (\mathbf{X}^T \mathbf{Y} \mathbf{Y}^T \mathbf{X})_a \mathbf{w}_a \quad (107)$$

3. Los vectores de saturaciones  $\mathbf{p}_a$  y  $\mathbf{q}_a$  son calculados como:

$$\mathbf{p}_a^T = \frac{\mathbf{w}_a^T (\mathbf{X}^T \mathbf{X})_a}{\mathbf{w}_a^T (\mathbf{X}^T \mathbf{X})_a \mathbf{w}_a} \quad (108)$$

$$\mathbf{q}_a^T = \frac{\mathbf{w}_a^T (\mathbf{X}^T \mathbf{Y})_a}{\mathbf{w}_a^T (\mathbf{X}^T \mathbf{X})_a \mathbf{w}_a} \quad (109)$$

4. Después del cálculo de cada vector latente las matrices de covarianza  $\mathbf{X}^T \mathbf{X}$  y  $\mathbf{X}^T \mathbf{Y}$  pueden ser actualizadas así:

$$(\mathbf{X}^T \mathbf{X})_{a+1} = (\mathbf{I} - \mathbf{w}_a \mathbf{p}_a^T)^T (\mathbf{X}^T \mathbf{X})_a (\mathbf{I} - \mathbf{w}_a \mathbf{p}_a^T) \quad (110)$$

$$(\mathbf{X}^T \mathbf{Y})_{a+1} = (\mathbf{I} - \mathbf{w}_a \mathbf{p}_a^T)^T (\mathbf{X}^T \mathbf{Y})_a \quad (111)$$

Nótese que sólo la matriz  $\mathbf{X}$  esta siendo deflacionada en las ecuaciones anteriores (Dayal et al., 1997).

---

En la mayoría de los algoritmos del PLSR tanto la matriz  $\mathbf{X}$  como la matriz  $\mathbf{Y}$  son deflacionadas después del cálculo de cada vector latente (Dayal et al., 1997).

Algunas propiedades básicas de los parámetros del PLSR son (Li et al., 2003):

- a) Las puntuaciones,  $\mathbf{t}$ , son ortonormales o sea  $\mathbf{t}_i^T \mathbf{t}_j = \mathbf{0}$  para  $i \neq j$ , y  $\|\mathbf{t}_i\| = 1$ .
- b) Los pesos o cargas,  $\mathbf{w}$ , son ortogonales o sea  $\mathbf{w}_i^T \mathbf{w}_j = \mathbf{0}$ .

Por otra parte, se debe tener cuidado cuando se analicen grandes conjuntos de datos por medio del método PLSR para que el resultado sea el mejor posible. A continuación se proporcionan algunas recomendaciones:

- Una forma inicial recomendable de proceder cuando se tienen grandes conjuntos de datos es dividir las observaciones en categorías y entonces desarrollar análisis por separado para cada categoría (Kettaneh, 2005).
- Con respecto a la velocidad de procesamiento el algoritmo NIPALS o sus equivalentes continúan siendo los más rápidos (Kettaneh, 2005).
- Una ventaja con los grandes grupos de datos es que procedimientos que consumen mucho tiempo de ordenador, como la técnica de validación cruzada, no son necesarios (Kettaneh, 2005).
- Todos los grandes conjuntos de datos poseen datos anómalos los cuales es necesario identificar y eliminar del conjunto de calibración antes de efectuar el modelado (aplicar el procedimiento). Uno de los planteamientos para hacer esto es ordenar por cada variable y remover o modificar un pequeño porcentaje de los valores extremos de la variable. Típicamente este porcentaje está entre el 1 y el 5%. Por ejemplo, una forma de proceder sería, después de ordenar, reemplazar todos los valores mayores que el percentil 99 o menores que el percentil 1 por los valores de estos límites. A este planteamiento se le ha dado el nombre de “*recortado*”. El recortado remueve la gran mayoría de los datos anómalos (Srivastava et al., 2001; Lix et al., 2004; Kettaneh et al., 2005; Chang, 2007) y se recomienda que este proceso sea aplicado de manera rutinaria a cualquier conjunto grande de datos (Kettaneh, 2005).
- Muchos grandes conjuntos de datos muestran comportamientos no lineales substanciales debido a que muchas variables varían sobre un rango muy grande. Algunas de estas faltas de linealidad podrían desaparecer después de transformaciones adecuadas de datos (logarítmica, etc.). Sin embargo, la única forma de tratar con no linealidades muy fuertes es dividir los datos en partes

---

pequeñas –conglomerados- y entonces elaborar modelos por separado (Kettaneh, 2005).

Esta técnica está basada en la suposición de que la relación subyacente no lineal puede ser aproximada localmente por un modelo lineal y por lo tanto sobre estas regiones pequeñas (dominios de linealidad local) (Baffi et al., 1999).

Estas cadenas de modelos lineales o modelos lineales por intervalos, son llamados modelos jerárquicos en la dirección de la observación (Kettaneh, 2005).

La experiencia disponible en el uso del método PLSR con varias técnicas analíticas ha demostrado que permite algunos efectos no lineales, entregando buenas predicciones al respecto (Blanco et al., 1995)

Otro planteamiento para manejar la no linealidad de los datos es buscar subconjuntos lineales en los grandes conjuntos de datos (Figura 11) (Geladi, 2002). Un planteamiento simple para extender la matriz de entrada es incluir combinaciones no lineales de las variables originales  $X$  (como son logaritmos, valores cuadráticos, productos cruzados, etc.) y entonces desarrollar una regresión PLS lineal sobre la matriz extendida de entradas (variables independientes) y la matriz de salida (variables dependientes) (Li et al., 2005).

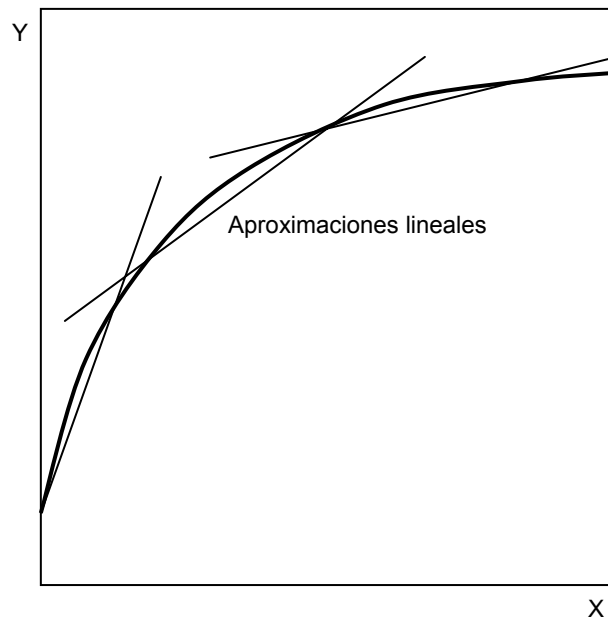
La utilización de factores en el PLSR tiene otro beneficio, además de reducir el número de variables: reduce el ruido de los datos. Los factores están ordenados por la variabilidad que explican: cada nuevo factor que se adiciona al modelo describe menos variabilidad sistemática (que es la parte de la variabilidad que se utiliza para predecir) y más variabilidad aleatoria (el ruido que incrementa la incertidumbre en el pronóstico). Debido a esto se deben tomar los factores necesarios para modelar la mayoría de la variabilidad sistemática, pero no incluir demasiados para impedir introducir ruido en exceso (Ferré, 2006).

La cantidad de observaciones necesarias para estimar un modelo PLSR dependerá de las predicciones que se quieran efectuar con él. Las observaciones utilizadas en la calibración deben mostrar todas las variaciones que después se podrían encontrar en las observaciones del intervalo de tiempo en el que se quiere efectuar el pronóstico. Esto permite al algoritmo poder identificar las variaciones que se pueden presentar



---

(Ferré, 2006).



**FIGURA 11** Aproximaciones lineales locales (Fuente: Geladi, 2002)

#### 4.8.1 El PLSR para una sola variable $y$ (Algoritmo PLS1)

El método de la regresión PLS aplicado en su forma básica a una variable individual  $y$  no es iterativo. Sea  $A_{\max}$  el número máximo de factores PLSR que serán calculados. Primero las variables de entrada  $\mathbf{X}$  son mejoradas por medio de un pretratamiento de escalado para tener niveles comparables de ruido y asegurarse de que ninguna variable encubrirá (ocultará) a las otras (Martens et al., 1989).

Al PLS1 en algunas publicaciones se le conoce como modelo PLSR de respuesta simple y al PLS2 como modelo PLSR multi-respuesta (Ergon, 2006).

Al utilizar el PLS1 para diferentes variables respuesta,  $y_i$ , se tendrán diferentes matrices de cargas  $\mathbf{w}_i$  y los diferentes estimadores  $\hat{\mathbf{b}}_i$  estarán localizados en sub-espacios diferentes. Al utilizar el PLS2 para todas las respuestas se asegura contar con estimadores individuales en el mismo sub-espacio, teniéndose a menudo predicciones degradadas (Ergon, 2006).

#### 4.8.2 El PLSR para varias variables $\mathbf{Y}$ (Algoritmo PLS2)

Algunas ocasiones varias variables  $y_1, y_2, \dots, y_J$  deben ser predichas del mismo conjunto de datos de entrada  $\mathbf{X}$ . El objetivo es utilizar un procedimiento de predicción

---

que no requiera  $J$  distintas repeticiones de casi el mismo modelado. Para no llegar a esto se utiliza la llamada regresión PLS2. Si se tuviera el caso  $J = 1$  se tendría el mismo proceso que con el algoritmo PLS1 (Martens et al., 1989).

Si existen varias variables  $y_i$  es recomendable usar un solo análisis PLS2 en lugar de  $j$  distintas ejecuciones del PLS1, en una primera etapa. El análisis PLS2 puede también ser útil durante la calibración si se sabe que las variables  $y_i$  están fuertemente correlacionadas entre sí. En la modelación PLS2 el efecto del ruido aleatorio en los datos de calibración  $Y$  puede ser reducido y podría proporcionar mejores predicciones de las variables  $Y$  que la calibración PLS1, para estas variables consideradas de manera individual (Martens et al., 1989).

Cuando se da el caso de que los problemas de no linealidad son un agente de más importancia que el ruido aleatorio podría ser ventajoso usar modelación PLS1 separada para cada variable de  $Y$  para obtener los resultados finales de calibración, la detección de datos anómalos y la predicción de los valores (Martens et al., 1989).

El PLS2 tiene la capacidad de analizar y modelar varias variables juntas,  $Y$ , lo cual tiene la ventaja de proporcionar un panorama completo y simple, en lugar de modelos por separado para cada variable  $y_i$ . Por tanto, cuando las variables  $y_i$  están correlacionadas deben ser analizadas en conjunto. Si las  $y_i$  en verdad miden aspectos enteramente distintos y por lo tanto son claramente independientes, un modelo PLS2 tenderá a mostrar muchas componentes y será difícil de interpretar. En ese caso procede efectuar el modelado con el PLS1 para cada variable y esto proporcionará un conjunto de modelos más simples, con pocas dimensiones, los cuales son más fáciles de interpretar (Wold et al., 2001).

Por lo tanto, para efectuar la elección correcta entre PLS1 y PLS2 se debe comenzar con un PCA solamente de la matriz  $Y$ . Este mostrará el número de componentes resultantes  $A_{PCA}$ . Si este valor es pequeño comparado con el número de variables  $y_i$  ( $M$ ), las  $y_i$  se encuentran correlacionadas y un modelo PLS2 para todas las  $y_i$  tendrá mayor garantía de éxito. Sin embargo, si el número de componentes significativas es grande, lo cual se observa en los gráficos de saturaciones del PCA, se deberán desarrollar modelos PLS1 para cada uno de los grupos configurados por las componentes obtenidas del PCA (Wold et al., 2001b).

---

#### 4.8.3 Error de predicción

Un paso muy importante cuando se ajustan modelos que serán utilizados para la predicción de observaciones futuras es verificar los resultados, o sea, aplicar los resultados actuales a un nuevo conjunto de observaciones que no hayan sido usadas para calcular estos resultados (estimación de los parámetros) ([www.statsoft.com](http://www.statsoft.com)).

Un análisis (o modelado) consiste de dos pasos: (i) se hace un intento por proponer un modelo (de la forma  $Y = f(X)$ ) entre dos grupos de variables, las independientes ( $X$ ) y las dependientes ( $Y$ ) para el comportamiento del sistema investigado. El conjunto de datos utilizado en este paso se llama un conjunto de *calibración* o *entrenamiento*. Los parámetros del modelo son llamados coeficientes de regresión o sensibilidades. (ii) Se obtienen las variables independientes de otras muestras o grupos de datos. Estas son usadas junto con las sensibilidades para predecir los valores de las variables dependientes. Este es el paso de predicción o prueba (Geladi et al., 1986).

A un conjunto de valores predichos de las variables respuesta,  $Y$ , se le representa por  $\hat{Y}$  (De Jong, 1993; Hadjiiski et al., 1999; Abdi, 2007).

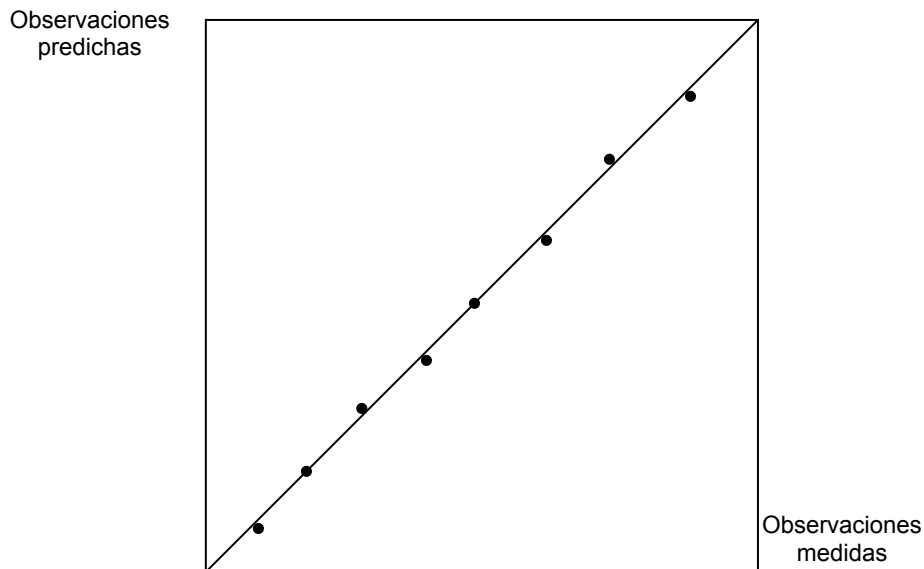
El objetivo más importante de la calibración del modelo es predecir correctamente. La situación ideal es cuando los valores predichos coinciden con las observaciones medidas. Entonces, todos los puntos en el gráfico predichos contra observados caen sobre una diagonal y sólo una pequeña cantidad de ruido genera desviaciones de esta línea (Figura 12) (Geladi, 2002).

Para llegar a un modelo correcto y altamente predictivo sólo las variables relacionadas con la propiedad bajo estudio deben ser incluidas. El modelo correcto más pequeño posible es aquel con el mínimo error de predicción para el conjunto de observaciones (Baumann et al., 2002).

Para reducir el error de predicción existen varias alternativas: (i) usar más datos porque se ha visto que el error de predicción (o estimación) va disminuyendo para grandes conjuntos de calibración, (ii) utilizar mejores datos, porque una forma de mejorar la capacidad de predicción es emplear mejores datos, (iii) simplificando la población para la calibración, la calibración multivariante está basada en la aproximación de datos con una población limitada de observaciones y no se puede esperar que el modelo calibrado ajuste igual de bien a todos los tipos de observaciones; en algunos casos podría darse una mejora en el ajuste y predicción de

---

resultados si la población para calibración es dividida en varias subpoblaciones pequeñas y (iv) usando un modelo mejorado se pueden hacer mejoras potenciales al modelo por medio del uso de modificaciones o pretratamiento de los datos y comprobar las suposiciones fundamentales del método (Martens, et al., 1989).



**FIGURA 12** Comportamiento lineal con poco ruido, homogéneo, sin sesgo, de un modelo utilizado para obtener las observaciones estimadas (Fuente: Geladi, 2002)

#### 4.8.4 Validación cruzada (Cross validation)

Un punto muy importante al hablar del tema de validación es que si usamos una muestra (separada) para efectuar la validación se espera (como hipótesis para que todo el proceso sea congruente y en verdad el modelo funcione adecuadamente para realizar con él predicciones útiles) que las mismas relaciones entre las variables que había al construir el modelo se mantengan aproximadamente iguales, cumpliéndose esto por lo menos para las muestras de validación y calibración (Sundberg, 2006).

En otras palabras, aparecen errores cuando los valores de  $Y$  seleccionados para la calibración no representan a la población para la cual se desea efectuar la predicción. Por lo tanto un objetivo de la selección de la muestra de calibración es que proporcione una buena extensión (esto es, que los valores que incluya muestren la mayor variedad de situaciones que se puedan presentar en el sistema o todas las condiciones que se puedan presentar en el periodo para el que se desea realizar la validación) (Sundberg, 2006).

Se debe de garantizar que la muestra seleccionada para efectuar la calibración

---

muestre una variación similar a la de la población natural para la cual se quiere realizar el pronóstico, en particular si la muestra es pequeña (Sundberg, 2006).

Con el propósito de construir un modelo con la aptitud de “ajustarse bien” y que efectúe predicciones “adecuadas” es necesario tener una prueba para la significancia predictiva de cada variable latente. La validación cruzada, CV, es una forma práctica y confiable de comprobar la significancia predictiva (Li, et al., 2003).

Se ha mostrado que la CV hace más eficiente el uso de los datos de entrenamiento disponibles para calibración en comparación con un conjunto de datos fijo. Los errores obtenidos por validación cruzada son siempre menores (Baumann et al., 2002).

La validación interna o validación cruzada, (CV), es un método de re-muestreo donde múltiples grupos son dejados fuera del cálculo de las sumas de cuadrados de las diferencias observadas menos estimadas. Debido a que se trata de un método de re-muestreo, la CV puede ser llevada a cabo de numerosas maneras (Geladi, 2002).

La CV es el mejor método de validación interna. Así como el enfoque de validación externa busca validar la calibración del modelo en un conjunto independiente de datos de prueba. Pero al contrario que la validación externa, la CV no desecha datos para probar solamente (Martens, et al., 1989); esto quiere decir que en la validación externa se usa solamente una parte de los datos, los correspondientes a la etapa de calibración para construir el modelo (que es probado con los datos reservados para validación), por lo que la validación externa no emplea todos los datos disponibles para elaborar el modelo. Por su parte la validación cruzada, CV, al considerar la información dividida en grupos, deja uno de estos fuera en cada cálculo, por lo que al final los utiliza todos para determinar el modelo.

La CV completa consiste en que cada observación es dejada fuera del proceso de regresión una vez y el modelo es calculado sin ella (Figura 13). Entonces el modelo desarrollado se usa para predecir esa observación (Davies, 2001).

Para validar el modelo se necesita obtener una matriz  $Y$  con valores conocidos. Para este propósito se puede usar la misma matriz  $Y$  usada en la calibración (validación interna) o si estuviera disponible, una matriz  $Y$  de prueba (Figura 13) (Einax et al., 1999).

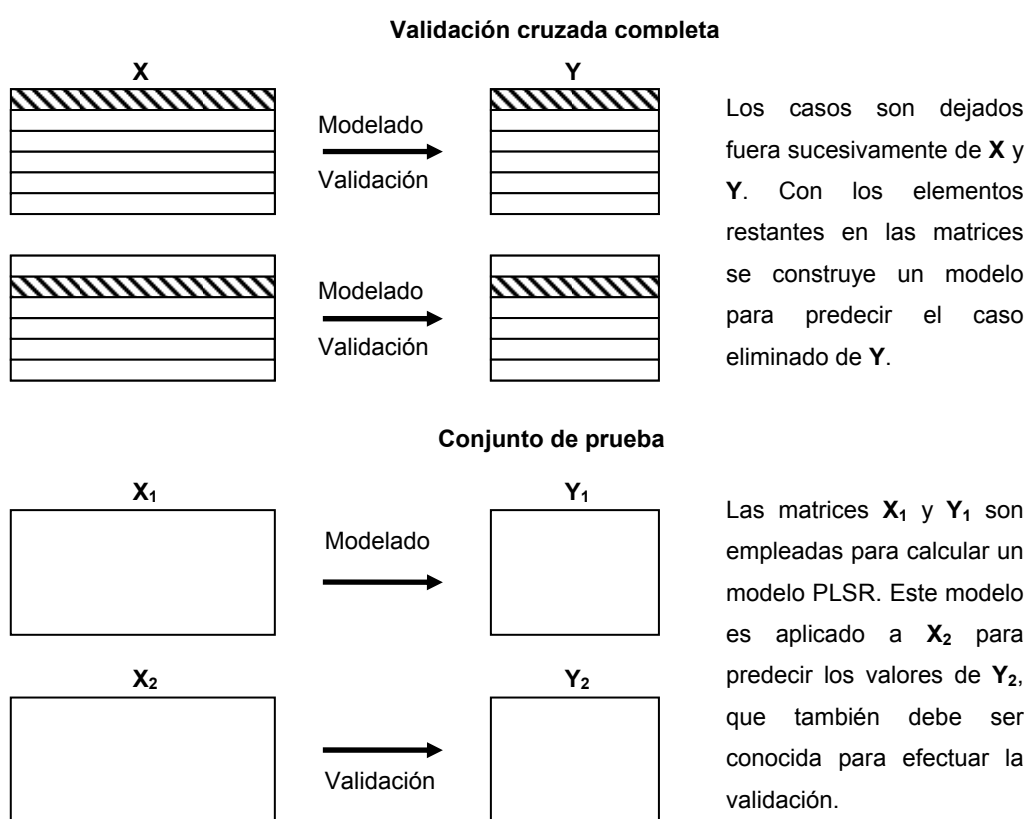
Una mejor indicación de la aplicabilidad de los modelos en los usos prácticos es obtenida cuando pueden ser probados con un conjunto de prueba (Einax et al., 1999).

---

Si no se sabe cuántos vectores latentes serían suficientes para un modelo PLSR se puede emplear la técnica CV para seleccionar el número óptimo (Dayal et al., 1997).

El número adecuado de factores es particular de cada fenómeno a ser calibrado y depende de la propiedad a ser pronosticada. Este número se obtiene, en principio, probando modelos con distinto número de factores (Ferré, 2006).

Cuando en el método PLSR son seleccionadas más componentes que las necesarias (número óptimo) el procedimiento se ve sometido a un “sobre ajuste”, con lo que el error de predicción se incrementa (Engelen et al., 2003).

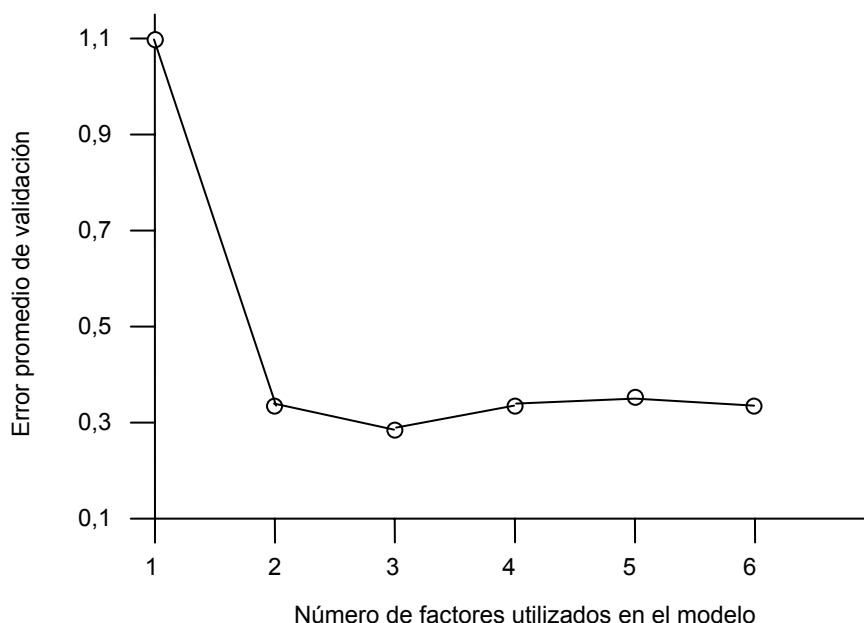


**FIGURA 13** Validación de modelos PLSR (Fuente: Einax et al., 1999)

Si en el método PLSR son seleccionadas menos componentes que las óptimas se dice que el modelo está “subajustado” y el error de validación será mayor que en el modelo con la cantidad adecuada de factores (Geladi, 2002).

La selección del número de componentes principales (vectores latentes) no es fácil y se basa a menudo en el procedimiento por estimación del error de predicción de la CV,

el cual es usualmente calculado por la variante del procedimiento de CV llamada “dejando uno fuera” (LOO). El procedimiento LOO usa un algoritmo iterativo para determinar el error de las  $m$  predicciones de la matriz  $\mathbf{Y}$  ( $\hat{\mathbf{Y}}$ ), desarrolla la calibración con  $(m - 1)$  muestras y usa la restante para cuantificación. El modelo es usualmente construido en base al número de PCs (componentes principales) que minimizan el error de predicción (Figura 14) (Blanco et al., 1995).



**FIGURA 14** Error promedio de validación para modelos con distinto número de factores (Fuente: Ferré, 2006)

Una forma de cuantificar la magnitud del fallo del modelo PLSR es por medio del *Error cuadrático medio de predicción* (MSEC), el cual se obtiene con la siguiente expresión (Blanco, et al., 1995):

$$\text{MSEC} = \frac{1}{n} \cdot \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^n (\hat{y}_{ij} - y_{ij})^2 \quad (112)$$

Donde  $y_{ij}$  es la concentración de la variable dependiente  $j$  en la muestra (observación)  $i$ ,  $\hat{y}_{ij}$  es la concentración estimada,  $n$  es el número de variables dependientes en la muestra y  $m$  el número de observaciones en el conjunto de calibración (Blanco et al., 1995).

Con el propósito de caracterizar la capacidad de predicción del modelo propuesto se puede calcular la *raíz del error cuadrático medio* (RMSECV) de validación cruzada

---

(Balabin et al., 2007):

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^N (\mathbf{y}_i - \hat{\mathbf{y}}_i)^2}{N}} \quad (113)$$

Donde  $\mathbf{y}_i$  y  $\hat{\mathbf{y}}_i$  se refieren al valor actual y al valor predicho para la observación  $i$ , respectivamente;  $N$  es el número de observaciones en el conjunto de validación. Esta medida de la capacidad del modelo para pronosticar valores está planteada para variables dependientes tomadas individualmente no para todo el conjunto de variables dependientes (Balabin et al., 2007).

Se sugiere que para seleccionar  $k$  (el número de componentes o factores del modelo) se tome en cuenta el valor mínimo que alcance el RMSECV (Engelen et al., 2003).

Con el propósito de seleccionar un modelo PLSR óptimo y parsimonioso (o sea, que represente la variabilidad en los datos sin causar sobre ajuste) también se cuenta con el Criterio de información de Akaike (AIC) (Viscarra Rossel, 2007):

$$\text{AIC} = N \log(\text{RMSECV}) + 2m \quad (114)$$

Donde  $N$  es el número de observaciones y  $m$  es el número de parámetros del modelo (Viscarra Rossel, 2007).

En la CV completa se repite la calibración  $i$  veces, tratando cada vez la  $i$ -ésima parte del conjunto de calibración total de observaciones predichas. Al final de la calibración las observaciones han sido tratadas como observaciones predichas y el RMSECV estimado (raíz del error cuadrado medio de validación cruzada) puede ser calculado (Martens et al., 1989).

Para el caso del PLSR, Geladi y Kowalski (1986) y Wold et al. (2001) diseñaron la siguiente propuesta (Li et al., 2003):

1. Dividir los datos ( $\mathbf{X}$  e  $\mathbf{Y}$ ) en  $G$  grupos.
2. Remover el grupo  $i$  de los  $G$  grupos anteriores y desarrollar el PLSR con  $k$  variables latentes en los  $G - 1$  grupos restantes. El resultado es un modelo  $i$  paralelo.
3. Introducir las variables independientes,  $\mathbf{x}_i$ , del grupo removido  $i$  dentro del modelo paralelo  $i$ . La suma de los errores cuadráticos de las diferencias entre



---

la respuesta,  $y_i$  y su estimación proveniente del modelo paralelo,  $\hat{y}_i$ , que recibe el nombre de  $PRESS_i$ , es calculada.

4. Se itera desde el paso 2 al 3, hasta que los  $G$  modelos paralelos son construidos, así como todos los  $PRESS_i$  son calculados.
5. Se suman todas las  $PRESS_i$  asociadas con  $k$  variables latentes y el resultado es llamado  $PRESS_k$ .
6. Se va al paso 2 para el caso de la siguiente variable latente.

Una medida muy útil para probar el ajuste de una regresión es el valor  $R^2$ , el *coeficiente de determinación*. La definición de  $R^2$  es (Li et al., 2003):

$$R^2 = 1 - \frac{RSS}{SS} = 1 - \frac{\sum(\mathbf{y} - \hat{\mathbf{y}})^2}{\sum(\mathbf{y} - \bar{\mathbf{y}})^2} \quad (115)$$

En la opinión de Hadjiiski et al., 1999, el coeficiente de determinación es una alternativa para seleccionar un modelo que reproduzca el sistema.

Para el caso de predicción una medida similar,  $Q^2$ , recibe el nombre de  $R^2$  de validación cruzada y generalmente indica la calidad de la predicción efectuada con el modelo elaborado. La definición de  $Q^2$  es (Li et al., 2003):

$$Q^2 = 1 - \frac{PRESS}{SS} = 1 - \frac{\sum(\mathbf{y} - \hat{\mathbf{y}})^2}{\sum(\mathbf{y} - \bar{\mathbf{y}})^2} \quad (116)$$

Cuando la validación cruzada es aplicada en modo secuencial, la CV es desarrollada para un componente y después el siguiente. Si  $k$  componentes son seleccionados para el modelo, el valor de validación cruzada,  $Q_{k,i}^2$ , es calculado para la variable respuesta  $i$  ( $y_i$ ) después de cada componente. La suma de las  $Q_{k,i}^2$ , llamada  $Q_k^2$ , se obtiene con la expresión (Li et al., 2003):

$$Q_k^2 = \sum Q_{k,i}^2 \quad (117)$$

Un valor más grande de  $Q_k^2$  quiere decir que la habilidad de predicción del modelo es mayor. Si el valor  $Q_{k+1}^2$  es mayor que  $Q_k^2$ , la componente adicional deberá ser incluida en el modelo. El número de variables latentes que proporciona el máximo  $Q_k^2$  deberá ser empleado en la modelación con PLSR (Li et al., 2003).

---

La capacidad de predicción de un modelo puede explicarse por el hecho de que las saturaciones del primer componente principal o factor, en las muestras de calibración y prueba, sean muy similares para ambos conjuntos. La mayor parte de la variación en el conjunto de prueba puede ser explicada por el primer componente principal (primer factor común) (Rodrigues et al., 2006).

#### 4.8.5 Metodología de aplicación

A continuación se proporciona un resumen de cómo desarrollar e interpretar un modelo PLSR (Wold et al., 2001b):

1. Tener una buena comprensión del estado del problema.
2. Conseguir datos adecuados tanto de  $Y$  (respuestas) como de  $X$  (predictores).
3. La primera parte de información acerca del modelo es el número de componentes significativas,  $A$ , que constituye la complejidad del modelo y por tanto del sistema.
4. La bondad del ajuste del modelo está dada por  $R^2$  y  $Q^2$ . La  $R^2$  proporciona un límite superior de qué tan bien explica el modelo los datos y predice las nuevas observaciones y el  $Q^2$  aporta un límite inferior para los mismos aspectos.
5. Las gráficas de puntuaciones ( $u$ ,  $t$ ) para las primeras dos o tres dimensiones del modelo mostrarán la presencia de datos anómalos, curvatura o grupos en los datos.
6. a. Si se observan problemas, por ejemplo valores pequeños de  $R^2$  y  $Q^2$ , o datos anómalos, grupos o curvatura en los gráficos de puntuaciones, se debe intentar remediar el problema. Después de posiblemente tener que transformar los datos, modificar el modelo, dividir los datos en grupos, eliminar datos anómalos o cualquier otra acción, se comienza en el paso 1.  
  
b. Si no se observan problemas y el modelo es interpretable, se puede intentar “podar” (recortar) ligeramente el modelo mediante la eliminación de términos no importantes, como por ejemplo coeficientes de regresión pequeños.

Para tener una impresión de qué tan bien predice un modelo elaborar la gráfica de valores  $Y$  observados en la muestra de calibración o en la muestra (separada) de validación contra los valores predichos correspondientes  $\hat{Y}$  (Sundberg, 2006).

Cuando es posible hacerlo, obtener un conjunto adicional (independiente) de observaciones para usar en la validación externa y comprobar así el modelo.

---

Desafortunadamente, esto no es a menudo factible y el planteamiento más común para la selección del modelo base es el método de validación cruzada dejando una observación fuera (cuando el número de datos no es muy grande) (Denham, 2000).

Aparte de las medidas ya mencionadas si tenemos un conjunto de prueba con  $j$  observaciones y la medida  $Y_i$  es conocida para este grupo, se puede utilizar para calcular el valor la *raíz del error cuadrático medio de predicción* (RMSEP) (Hadjiiski et al., 1999):

$$\text{RMSEP} = \left[ \frac{1}{J} \sum_{i=1}^J (\mathbf{y}_i - \hat{\mathbf{y}}_i)^2 \right]^{1/2} \quad (118)$$

Entre más cercanos estén  $\mathbf{y}_i$  y la estimación  $\hat{\mathbf{y}}_i$ , el valor de RMSEP se hace más pequeño y esto significa que el modelo elegido predice bien para el conjunto de prueba (Hadjiiski et al., 1999).

---

---

## CAPÍTULO 5. DESCRIPCIÓN DEL MODELO MECANICISTA

### 5.1 INTRODUCCIÓN

Los modelos matemáticos para aplicación en embalses pueden clasificarse de manera amplia como fundamentados físicamente, empíricos o combinados. Los modelos fundamentados físicamente utilizan las ecuaciones que gobiernan el transporte de calor, de los constituyentes de calidad y de flujo del agua, para simular temperatura y concentraciones en el agua, sobre la base de la geometría del sistema (por ejemplo la tabla calados-volúmenes-áreas), caudales y condiciones climáticas (Hydrologic Engineering Center, 1986; Martín et al., 1999; Martin et al., 1999; Deas et al., 2000).

Teóricamente un modelo con fundamento físico es aplicable a un amplio rango de sistemas y es capaz de efectuar simulaciones bajo una variedad de circunstancias que podrían no estar presentes actualmente en el sistema (Martin et al., 1999). Uno de los modelos que se desarrolló en este trabajo es de este tipo.

El modelo planteado en este trabajo fue desarrollado a través de la revisión de la literatura disponible y programado especialmente en el lenguaje Visual Basic 2005 (integrante de la suite Visual Studio 2005); posee dos componentes fundamentales: una componente hidrodinámica (flujo de los caudales) y otra de temperatura del agua (flujo de calor y transporte).

La mayoría de los proyectos de modelación en almacenamientos pueden ser adecuadamente estimados con representaciones unidimensionales a lo largo del eje principal de variación. Muchos de los grandes embalses están estratificados y tienen apreciables gradientes verticales de temperatura y otros constituyentes, pero son uniformes lateralmente. Para almacenamientos el eje principal de variación es vertical (Martin et al., 1999; Hornung, 2002).

Un embalse puede ser modelado con varios grados de sofisticación. Los modelos unidimensionales son típicamente empleados en almacenamientos estratificados con grandes tiempos de residencia. La dinámica de muchos depósitos y lagos vista en escalas de tiempo mayores que aquellas de eventos extremos como tormentas y avenidas es bien descrita usando esta aproximación (Martin et al., 1999).

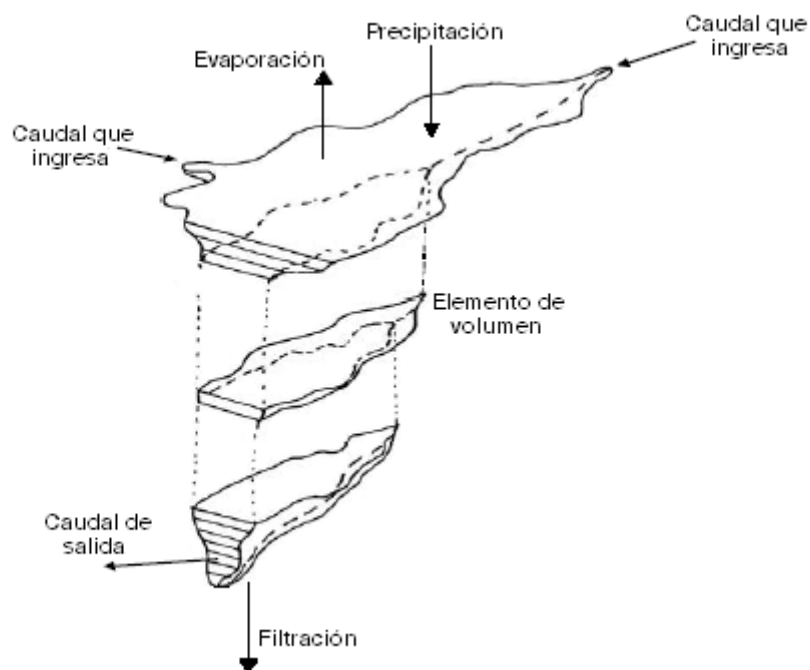
En la mayoría de los modelos unidimensionales el sistema es caracterizado como una serie de capas horizontales bien mezcladas (Figura 15). Cada “rebanada” horizontal

---

(elemento de volumen) se caracteriza por un área, espesor y volumen (Environmental Laboratory, 1995; Deas et al., 2000).

La energía térmica y la masa pasan a través de cada capa por advección y difusión. La evaporación, el calor sensible y los flujos de radiación de onda larga ocurren solamente en la capa superficial (Deas et al., 2000). La radiación solar penetra a través de las capas, decayendo exponencialmente con la profundidad (Ley de Beer-Lambert) (Hydrologic Engineering Center, 1986). Los caudales que ingresan son agregados a las capas a la profundidad adecuada donde la densidad del caudal que entra (como función de la temperatura) coincide con la densidad del agua en el embalse (Deas et al., 2000). También ingresa volumen de agua debido a la precipitación directa sobre el espejo de agua (Environmental Laboratory, 1995).

Los flujos de salida pueden ocurrir en cualquier capa, incluyendo la superficial donde se pierde masa debido a la evaporación, o la capa del fondo donde se pierde agua debido a la infiltración (Hydrologic Engineering Center, 1986; Deas et al., 2000).



**FIGURA 15** Esquema de modelo unidimensional de embalse (Fuente: Deas, M. L., Lowney, C. L. 2000. "Water temperature modelling review")

Dentro de cada capa la temperatura se considera isotermal (completamente mezclada) y cada una de las concentraciones de constituyentes de calidad del agua se considera homogénea (uniforme). Los caudales que ingresan o que salen acontecen como fuentes o pérdidas en cada capa y los que ingresan son mezclados instantáneamente

---

dentro de la capa. El transporte interno ocurre solamente en la dirección vertical, a través de los procesos de advección y difusión efectiva. El término de difusión efectiva engloba la difusión molecular y turbulenta, así como también el mezclado por procesos convectivos (Martin et al., 1999; Deas et al., 2000).

La advección se origina debido al movimiento del flujo y no cambia la identidad de la sustancia que está siendo transportada (Chapra, 1997). La advección vertical es gobernada tanto por la localización de los caudales que ingresan, como por los caudales que egresan del embalse (Hydrologic Engineering Center, 1986; Martin et al., 1999). Entonces el cálculo de las zonas de distribución de las entradas y toma de las sueltas es muy importante dentro del modelo.

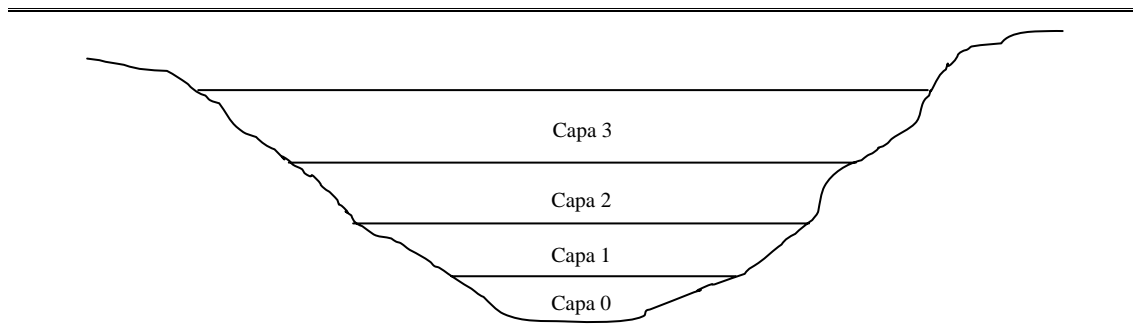
La advección vertical es uno de los dos mecanismos de transporte utilizados en el modelo para transportar constituyentes de calidad del agua entre elementos. La advección vertical se define como el flujo entre elementos, el cual es el resultado de la continuidad del flujo en todos los elementos (Environmental Laboratory, 1995).

La difusión se refiere al movimiento de masa debido al movimiento aleatorio del agua. En una escala microscópica la difusión molecular se origina debido al movimiento Browniano aleatorio de las moléculas de agua. Una clase similar de movimiento aleatorio ocurre a gran escala debido a vórtices y es llamado difusión turbulenta. Ambas tienen la tendencia a minimizar gradientes (esto es, diferencias de concentración) debido al movimiento de masa desde regiones de alta concentración hacia las de baja (Chapra, 1997).

La difusión horizontal es generalmente mucho mayor que la difusión vertical. Este es el motivo por el cual se considera en la práctica que las capas horizontales están uniformemente mezcladas (Martin et al., 1999).

Las capas son numeradas iniciando en el fondo del embalse y continuando hacia arriba (Figura 16).

Debido a que cada capa es considerada como uniforme en sus atributos térmicos, biológicos y químicos, los parámetros de la capa son asignados en el punto medio de esta cuando los resultados son presentados.



**FIGURA 16** Orden de numeración de las capas, iniciando desde el fondo del embalse

Suposiciones del modelo y limitaciones.- Todos los modelos son representaciones simplificadas de un prototipo real. Los mayores “supuestos” y limitaciones del modelo son:

- Supuesto unidimensional.- Las limitaciones en este aspecto: Las variaciones longitudinales y laterales en los constituyentes de la calidad del agua no pueden ser predichas, todas las cantidades que entran y los constituyentes son instantáneamente dispersados en las capas horizontales, las predicciones del modelo son probablemente más representativas de las condiciones de la calidad del agua cerca de la presa o en la parte donde tiene mayor profundidad el embalse (Martin et al., 1999).
- El modelo no cuenta con un algoritmo para el caso en que se forme una capa de hielo en la superficie del almacenamiento.
- Muchos de los programas de monitoreo obtienen datos mensualmente. Para obtener valores a usarse en los intervalos computacionales establecidos (diarios) como estos son menores que el lapso de muestreo, se requiere proponer valores diarios entre los datos de muestreo (es posible hacer esto mediante interpolación).

Una característica importante del modelo es el espesor variable de cada capa. Cada una de ellas puede expandirse o contraerse a lo largo del eje vertical para adaptarse al agua que fluye hacia dentro o fuera de dicha capa. Un exceso de entradas sobre las salidas origina un almacenamiento extra de agua en una capa y se produce su expansión. Un exceso de salidas sobre las entradas genera una contracción.

La ventaja de una formulación de capas variables es significativa. La dispersión numérica es reducida porque los flujos verticales entre capas no son requeridos (Environmental Laboratory, 1995).



---

Las capas que llegan a ser demasiado grandes o demasiado pequeñas pueden eventualmente llevar a problemas de resolución o a errores numéricos. Esta situación es controlada redimensionando capas que han excedido los límites de espesor especificados por el usuario. Cuando las capas son divididas o unidas sus atributos geométricos son recalculados y todas las concentraciones de energía y materiales son redefinidas. Los tamaños de las capas son repetidamente revisados para su división o unión hasta que todas las dimensiones están dentro de los límites especificados (Environmental Laboratory, 1995).

Existen dos formas para describir el movimiento de un fluido. En el primer método, la representación Lagrangiana, la variación temporal de las variables relevantes (velocidad, densidad, etc.) es descrita por un “elemento de fluido en movimiento”. En el segundo método, la representación Euleriana, las ecuaciones para las variables son formuladas en “puntos fijos” (como en un Sistema de Coordenadas Cartesianas) (Lerman et al., 1995). El presente modelo corresponde, evidentemente, a la representación Lagrangiana.

## 5.2 DATOS NECESARIOS

La información requerida para la implementación del modelo mecanicista en un embalse es la siguiente:

- Temperaturas iniciales. Dos lecturas a distintas profundidades como mínimo.
- Para las ocho tomas y el vertedor, datos acumulados diarios de caudal de salida para todo el periodo de simulación.
- Características de los caudales de salida. Para las tomas se requiere la altura sobre el fondo del embalse y el ángulo del puerto de extracción. Para el vertedor se necesita la altura de la cresta (medida desde el fondo) y la longitud libre del vertedor.
- Se requiere proponer, de inicio, las siguientes constantes generales:  $K_e$ , Coeficiente de extinción lumínica; Factor por viento;  $\beta$ , Proporción de la radiación absorbida en la superficie de la radiación entrante neta;  $K_{eSS}$ , Coeficiente de extinción lumínica por sólidos suspendidos;  $a$  y  $c$ , Coeficientes para calcular la difusión turbulenta vertical.
- Datos a escala diaria de: volumen en el embalse, velocidad del viento, radiación, evaporación, precipitación, filtración; para los flujos entrando por el río y (hasta) cuatro entradas: caudal de entrada y temperatura

---

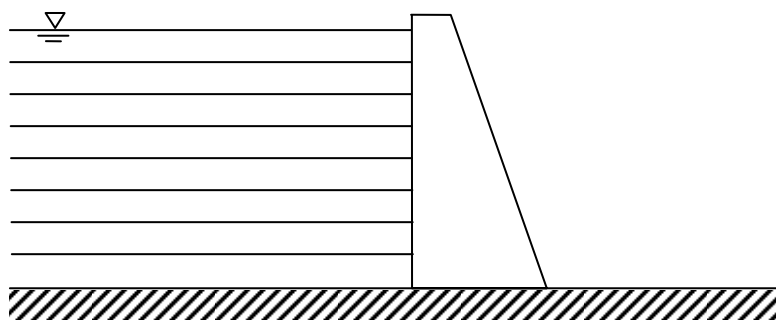
correspondiente.

- La tabla calado-volumen-superficie del almacenamiento.
- Las características de las capas: espesor mínimo y espesor máximo permitidos.
- Datos geográficos de la presa: latitud, longitud y altitud.
- Datos expresados como media del día correspondiente de: temperatura del aire sobre el embalse, humedad relativa, temperatura del agua en la superficie del embalse y cobertura de nubes ( $C_L$ ).

### 5.3 ORDEN DE LOS PROCESOS A REALIZAR EN LAS CAPAS DEL EMBALSE PARA CADA DÍA DEL PERIODO DE SIMULACIÓN

El flujo propuesto para los procesos a llevar a cabo en el modelo es:

- 1º Efectuar la Mezcla entre capas debida a Flotación. A este proceso también se le llama Comprobación de la Estabilidad de la Estructura de las capas y se efectúa comparando la densidad de las capas.
- 2º Unir capas que tengan un espesor por debajo del Espesor Mínimo y re-numerar las capas.
- 3º Dividir por un número entero de divisiones las capas que tengan un espesor superior al espesor máximo permitido y re-numerar las capas. Esto se hace para disminuir la dispersión numérica.
- 4º Calcular la Difusión Turbulenta en las capas del embalse, ya modificadas por los procesos anteriores.
- 5º Ubicar, en la capa correspondiente, cada uno de los caudales que ingresa al embalse.
- 6º Determinar los caudales a extraer de cada una de las capas debido a la extracción por tomas y desagüe del vertedor.



**FIGURA 17** Capas con las que se representa un embalse

- 
- 7° Calcular los flujos de calor en las capas del almacenamiento.
  - 8° Efectuar las modificaciones de la capa de fondo y en la superficial debido a precipitación, evaporación e infiltración.
  - 9° Obtener la temperatura en cada una de las capas para el siguiente día (Figura 17).

#### 5.4 ESTABILIDAD DE LA ESTRUCTURA DE CAPAS

La densidad termal se obtiene solamente en función de la temperatura del agua. La determinación de la densidad se puede hacer para cada una de las capas del embalse o para los flujos de entrada, en cada intervalo de tiempo.

No se involucra en el cálculo de la densidad las concentraciones de sólidos. Sin embargo, se supone de inicio que tanto en el embalse como en los caudales de entrada, las concentraciones de sólidos, tanto suspendidos como disueltos, son reducidas y no tienen una influencia significativa en la determinación de la densidad del agua (Hydrologic Engineering Center, 1986). Si se tuviera el caso de altas concentraciones de sólidos, ya sea en el embalse o en los flujos de entrada, la densidad de estos dependería en gran medida, aparte de la temperatura, de dichos valores de concentración.

Para calcular la densidad termal se tiene la siguiente expresión (U. S, Army Corps of Engineers, 1987):

$$\rho_T = 1000 - \frac{(T - 3,98)^2 (T + 283)}{503,57 (T + 67,26)} \quad (119)$$

Donde:

T = Temperatura del agua.

La obtención de la densidad de las capas es importante pues los siguientes procesos dependen de su determinación:

- Ubicación vertical, en la capa que corresponde, de los caudales que entran al embalse.
- Procesos de mezclado de capas, impulsados por procesos térmicos (inestabilidad por flotación y empuje hidrostático de capas).
- Determinación de características generales de las capas, al unirse o

---

dividirse.

Una vez que se cuenta con la densidad de cada una de las capas para un día de simulación la estabilidad de la estructura de capas es comprobada por comparación de la densidad de las capas (Centre for Water Research, 2007). Para que la estructura de capas sea estable las aguas de mayor densidad deben estar en el fondo del embalse y tienen sobre ellas aguas menos densas.

Esta comprobación es hecha comenzando con la capa superficial y comparando la densidad con la capa directamente debajo de ella. Si la densidad de la capa sobre yacente (de las dos capas, la que está arriba) es mayor que la densidad de la capa por debajo de ella, las capas son combinadas, las concentraciones de las dos capas son fusionadas de acuerdo a las ecuaciones que gobiernan la conservación de constituyentes y la nueva densidad se obtiene con la nueva temperatura. La densidad de esta capa es entonces comparada con la siguiente capa debajo de ella y el proceso se repite hasta que la capa con la cota más baja (la que está en el fondo del embalse) es alcanzada. Este proceso asegura que el perfil de la densidad sea siempre estable (Centre for Water Research, 2007).

Cuando se combinan dos capas, las leyes de conservación de energía térmica (temperatura) y masa (concentraciones de constituyentes de calidad del agua) pueden generalizarse así (Centre for Water Research, 2007):

$$C_i^* = \frac{C_i \cdot \Delta M_i + C_{i+1} \cdot \Delta M_{i+1}}{\Delta M_i + \Delta M_{i+1}} \quad (120)$$

Donde:

$C_i^*$  = Concentración de la nueva capa.

$C_i$  = Concentración en la capa  $i$ .

$C_{i+1}$  = Concentración en la capa  $i + 1$ .

$\Delta M_i$  = Masa de la capa  $i$ .

$\Delta M_{i+1}$  = Masa de la capa  $i + 1$ .

Los subíndices se refieren a los números de capa.

## 5.5 UNIÓN DE CAPAS CON ESPESOR POR DEBAJO DEL MÍNIMO

Las capas que llegan a ser demasiado grandes o demasiado pequeñas, pueden eventualmente conducir a problemas de resolución o a errores numéricos. Esta situación es reducida por medio del redimensionado de capas que han excedido los

---

límites de espesor especificados por el usuario (Hydrologic Engineering Center, 1986; Environmental Laboratory, 1995).

Son establecidos límites por el usuario para los espesores y volúmenes de cada capa. Los límites superior e inferior son señalados para asegurar que una resolución adecuada es alcanzada y que un número excesivo de capas no será usado (Centre for Water Research, 2007).

Cuando las capas son divididas o unidas sus atributos geométricos son recalculados y todas las concentraciones y temperaturas son elaboradas nuevamente. Los tamaños de las capas son continuamente revisados para su división o unión hasta que todos los tamaños están dentro de los límites especificados (Centre for Water Research, 2007).

El espesor mínimo debe de cumplir con la regla  $\text{Espesor}_{\max} > 2 \cdot \text{Espesor}_{\min}$ , de otra forma el algoritmo de unión y el de la división podrían no funcionar adecuadamente (Hornung, 2002).

Para las capas con espesor menor que el mínimo permisible son unidas con la capa vecina de menor espesor. Las capas se reenumeran y las características geométricas son determinadas (Centre for Water Research, 2007).

Los criterios que se propusieron en el modelo son:

- El límite inferior de espesor de una capa será de 0,6 m para el modelo propuesto. Este valor se plantea así debido a que se tomó una formulación para el flujo de radiación solar en las capas del embalse, donde la profundidad del agua se toma con el parámetro  $Z^*$ , comenzando 0,6 m por debajo de la superficie (Environmental Laboratory, 1995).
- Para todas las comprobaciones con el propósito de verificar que las capas estén dentro de los límites especificados sólo se atenderá a los espesores y no a los volúmenes para tener sólo un criterio de comparación, lo que hará más fáciles los cálculos. Después de poner en funcionamiento un algoritmo, como al unir o separar capas, todas las propiedades, tanto geométricas como de constituyentes de calidad del agua sí tomarán en cuenta el volumen de cada una de las capas involucradas.
- Se programará el mecanismo de unión de capas para las capas que tengan un espesor inferior a 0,6 m. Si el espesor de la capa del fondo es

---

inferior a 0,6 m se unirá con la capa que está por encima de ella. Si la capa superficial tiene un espesor inferior a 0,6 m se unirá a la capa que ésta por debajo de ella. Para todas las otras capas, la capa de espesor inferior a 0.6 m se unirá con la capa vecina, de las dos existentes, que tenga el espesor menor (Centre for Water Research, 2007).

## 5.6 DIVISIÓN DE CAPAS CON ESPESOR MAYOR AL ESPESOR MÁXIMO ESPECIFICADO

El valor del  $PLT_{max}$  (espesor máximo permitido de capa) controla fuertemente la resolución espacial del modelo. Esto se debe a que las capas se unirán y crecerán en espesor hasta que alcancen el límite superior de espesor de capa, serán divididas, se unirán otra vez, especialmente durante el mezclado en la superficie del agua. Un  $PLT_{max}$  de 2 a 3 m ha proporcionado la generación de los mejores resultados en modelos de múltiples capas en almacenamientos (Hornung, 2002).

Las capas con espesor mayor al espesor máximo admisible son divididas en las porciones requeridas, creando nuevas capas con propiedades idénticas pero con espesores reducidos (espesores menores que el máximo permitido) (Hornung, 2002). Las capas son reenumeradas y las propiedades geométricas determinadas.

Los criterios que se aplicaron para efectuar la división de capas fueron:

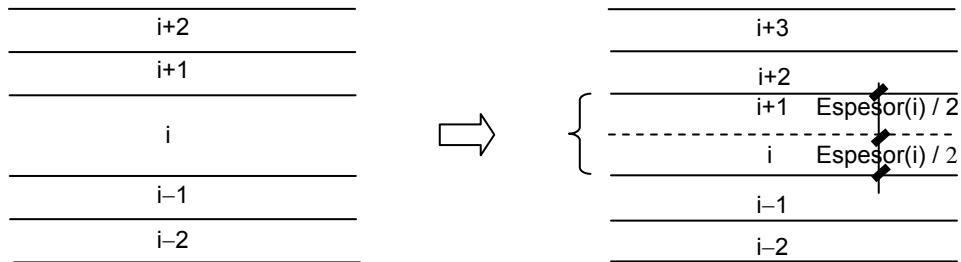
- El límite superior de espesor de una capa será de 2,5 m haciendo un promedio del rango antes mencionado. Este límite cumple a su vez el requerimiento de que el Límite Superior  $> 2 \cdot$  (Límite Inferior) para que el algoritmo de la división de capas funcione adecuadamente. (Hornung, 2002)
- Para todas las comprobaciones con el propósito de verificar que las capas estén dentro de los límites especificados sólo se atenderá a los espesores, para tener sólo un criterio de comparación. Después de poner en funcionamiento un algoritmo, como al unir o separar capas, todas las propiedades tanto geométricas como de constituyentes de calidad del agua sí tomarán en cuenta el volumen de cada una de las capas involucradas.

El algoritmo propuesto para efectuar la división de capas es:

- Se comienza la revisión del espesor las capas para que no rebasen el

límite superior permisible. Se iniciará con la capa del fondo. Al dividir una capa se re- numeran las capas por arriba de ella (Figura 18).

- El espesor de la capa que se vaya a dividir se fraccionará por medio de un número entero que dependerá de cuántas veces rebase el espesor el límite máximo establecido.



**FIGURA 18** División de la capa  $i$  en dos capas y re- numeración de las capas por encima de la capa  $i$

- Al dividir una capa las capas resultantes tendrán los mismos valores para los constituyentes de calidad del agua que la capa original, por lo que no es necesario efectuar cálculos de mezclado (Hornung, 2002).
- Al final del procedimiento teniendo ya los nuevos espesores, se obtendrá para cada capa su volumen correspondiente y el área de su superficie superior (frontera superior).

## 5.7 CÁLCULO DE LA DIFUSIÓN TURBULENTA

La difusión turbulenta es un proceso de transporte que reduce los gradientes y es calculada usando un Coeficiente de Difusión Turbulenta que es dependiente de la velocidad del viento, magnitudes de los caudales de entradas y salidas, así como de la estratificación de las densidades (Martin et al., 1999).

La clave para simular la estratificación se relaciona en como el coeficiente de difusión vertical varía con la profundidad. Una variedad de procedimientos han sido propuestos. Uno de los primeros fue el de Munk y Anderson (1948). Este modelo asume que el coeficiente de difusión es una función del número de Richardson (Chapra, 1997):

$$E(z) = \frac{E_0}{(1 + a R_i)^{3/2}} \quad (121)$$

Donde (Chapra, 1997):

---

$E_0$  = El coeficiente de difusión para estabilidad neutral.  
 $R_i$  = Número de Richardson.  
 $a$  = Una constante (será una constante general a calibrar).

Una variedad de formulaciones han sido propuestas para relacionar  $E_0$  con la profundidad y la velocidad del viento. Una formulación simple es (Chapra, 1997):

$$E_0 = c \omega_0 \quad (122)$$

Donde (Chapra, 1997):

$c$  = Una constante empírica (será una constante general a calibrar).  
 $\omega_0$  = Velocidad cortante del viento.

$$\omega_0 = \sqrt{\frac{\tau_s}{\rho_w}} \quad (123)$$

Donde (Chapra, 1997):

$\rho_w$  = Densidad del agua.  
 $\tau_s$  = Esfuerzo cortante en la interfase aire-agua.

$$\tau_s = \rho_{\text{aire}} \cdot c_d \cdot U_w^2 \quad (124)$$

Donde (Chapra, 1997):

$\rho_{\text{aire}}$  = Densidad del aire.  
 $c_d$  = Coeficiente de arrastre.

$$c_d = 0,00052 U_w^{0,44} \quad (125)$$

Donde (Chapra, 1997):

$U_w$  = Velocidad del viento.

La densidad del aire para una temperatura dada se calcula con la expresión (Llorca et al., 2006):

$$\rho_{\text{aire}} = \frac{P \cdot P_m}{R T} \quad (126)$$

Donde (Llorca et al., 2006):

$\rho_{\text{aire}}$  = Densidad del aire, en gramo/litro =  $\text{kg/m}^3$ .  
 $P$  = Presión atmosférica, en atmósferas.



$P_m$  = Peso molecular del aire = 28,96 g/mol.  
 $R$  = Constante de los gases = 0,082 (Atm·litro)/(K·mol).  
 $T$  = Temperatura termodinámica ambiente, en K.

La presión atmosférica se puede obtener con la fórmula (Llorca et al., 2006):

$$P = 1013,25 \cdot \left(1 - 22,5569 \times 10^{-6} \cdot z\right)^{5,25611} \quad \text{hPa} \quad (127)$$

Donde (Llorca et al., 2006):

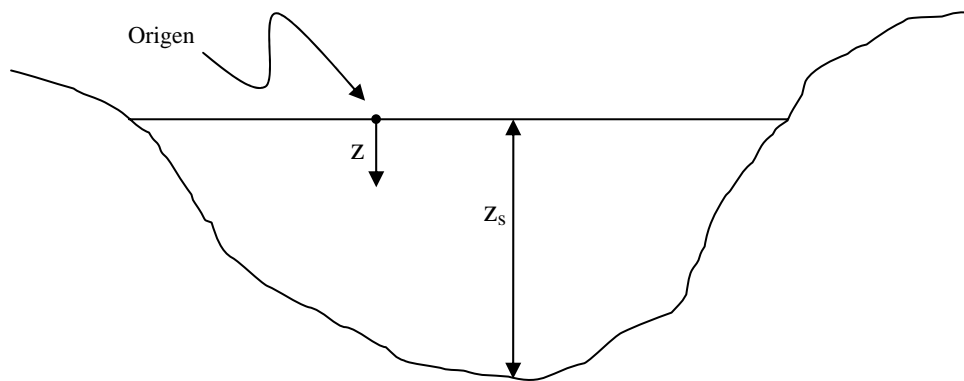
$z$  = Altitud local, en m.

Una formulación para poder calcular el número de Richardson es la siguiente (Chapra, 1997):

$$R_i = \frac{-\left(\frac{g}{\rho}\right)\left(\frac{\partial \rho}{\partial x}\right)}{\frac{\omega_0}{(z_s - z)^2}} \quad (128)$$

Donde (Chapra, 1997):

$z_s$  = Calado total del lago o almacenamiento (Figura 19).  
 $z$  = Profundidad medida desde la superficie del agua (Figura 19).



**FIGURA 19** Representación de algunas variables para obtener el número de Richardson

Y una estimación propuesta de la derivada parcial de la densidad con respecto a la profundidad (el gradiente vertical de la densidad) es:

$$\left(\frac{\partial \rho}{\partial x}\right)_i = \frac{\rho_i - \rho_{i-1}}{\frac{\text{Espesor}_i}{2} + \frac{\text{Espesor}_{i+1}}{2}} \quad (129)$$

Donde:

$\left(\frac{\partial \rho}{\partial x}\right)_i$  = Tasa de cambio de la densidad con respecto a la profundidad

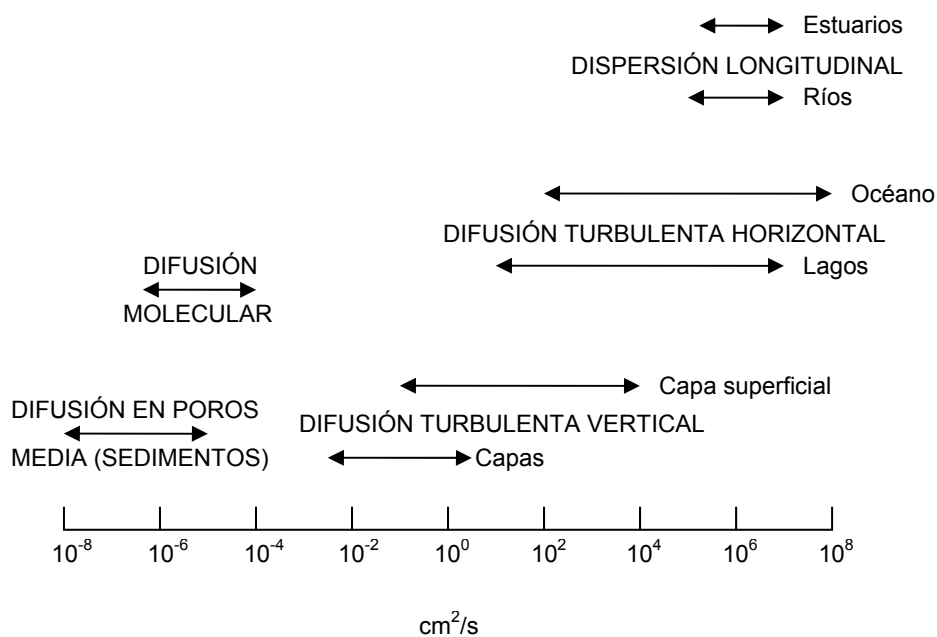
(también llamado gradiente de la densidad).

$\rho_i$  = Densidad de la capa i.

Esesor<sub>i</sub> = Espesor de la capa i.

Si la densidad aumenta su valor en la dirección vertical hacia abajo el gradiente es de signo positivo.

Para tener una idea del orden de magnitud de los coeficientes de difusión turbulenta se presenta la Figura (20) (Chapra, 1997):



**FIGURA 20** Rangos típicos del Coeficiente de difusión en aguas naturales y sedimentos (Fuente: Chapra, S. C. 1997. "Surface water-quality modeling")

Los coeficientes de difusión turbulenta son varios órdenes de magnitud mayores que los de escala molecular. Nótese también que la difusión horizontal es generalmente mucho mayor que la difusión vertical (Chapra, 1997).

El rango típico de la Difusión Vertical Turbulenta (definida como propiedad) es el siguiente (Thomann et al., 1987; Wüest et al., 2003):

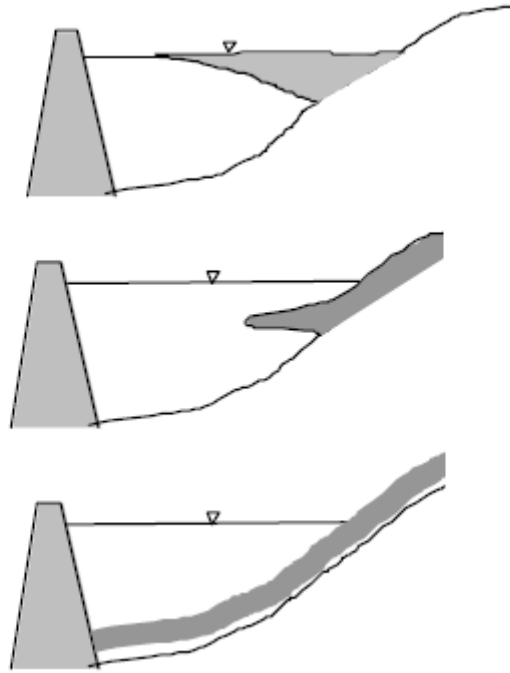
$$10^{-2} \text{ a } 10^2 \text{ cm}^2/\text{s}$$

La Difusión Vertical Turbulenta es la suma de la difusión molecular y la difusión vorticosa turbulenta (Wüest et al., 2003).

---

## 5.8 UBICACIÓN DE CAUDALES QUE INGRESAN AL EMBALSE

Dinámica de las entradas.- La entrada del río, de una conducción o cualquier otro elemento que aporte flujo al embalse es modelada por la inserción de este volumen dentro del número de capas existentes, a la profundidad apropiada (Deas et al., 2000). Si el volumen de la capa llega a ser excesivo nuevas capas son formadas en el siguiente paso de tiempo.



**FIGURA 21** Esquema de posibles ubicaciones, a distintos niveles, de un flujo que ingresa a un embalse (Fuente: Deas, M. L., Lowney, C. L. 2000. "Water temperature modelling review")

La comparación inicial que se efectúa para ubicar correctamente el flujo que ingresa es la siguiente:

- Si Temperatura del caudal de entrada  $\geq$  Temperatura de la superficie del embalse, entonces el volumen total es añadido a la capa superior. Como el agua que ingresa es menos densa que la del embalse flotará sobre la superficie del embalse, como un sobre-flujo (Figura 21). Los sobre flujos son comunes en primavera. Los vientos superficiales a menudo provocan que los sobre-flujos sean rápidamente disipados.
- Si Temperatura del caudal de entrada  $<$  Temperatura de la superficie del embalse, ocurre flujo sumergido y son necesarios cálculos para ubicar el flujo de entrada. Al ser el agua que ingresa más densa que el agua de superficie la flotación negativa causa que el flujo que ingresa se movilice

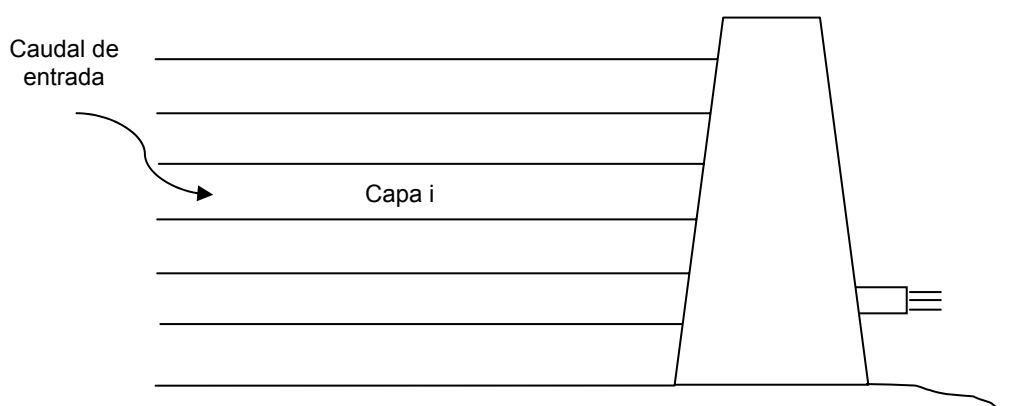
---

por debajo de la superficie del embalse convirtiéndose en un sub-flujo (Figura 21). Los sub-flujos son comunes en el otoño. Las aguas fluviales someras se enfrían más rápido que las aguas embalsadas y las tormentas otoñales causan influjos de alta turbidez. Si los sub-flujos son más densos que las aguas del fondo podrían permanecer en el fondo del almacenamiento (Figura 21). Sin embargo, la mayoría de los sub-flujos son menos densos que la capa del fondo. En ese caso el sub-flujo tiende a separarse del fondo y se llama ínter flujo.

El centroide del flujo que entra al depósito se considera que estará ubicado en la capa cuya densidad tenga una magnitud que esté más cercana a la densidad del flujo de entrada (Figura 22).

Si un volumen que ha ingresado alcanza su nivel de flotación neutral es instantáneamente insertado en la capa que está ubicada en ese nivel. Así es como se obtiene el volumen diario (caudal diario), que ingresa a cada una de las capas.

El algoritmo para asignar un caudal de entrada a una capa específica del embalse es el siguiente:



**FIGURA 22** Ubicación de un caudal de entrada en la capa *i*

1. Verificar que exista el caudal de entrada.
2. Calcular densidad del agua en el caudal de entrada. Se utiliza la fórmula (119).
3. Determinar en que capa ingresa ese flujo. Para esto se utiliza la expresión:

---

---

$$\text{Comprobación} = \left| \frac{\rho_{\text{entrada}} - \rho_{\text{capa}}}{\rho_{\text{capa}}} \right| \cdot 100 \quad (130)$$

Se aplica esta fórmula en todas las capas en las que se encuentre dividido el embalse y en la capa donde el resultado sea menor es donde se asume que ingresará el caudal, debido a que se presentará la mayor similitud de las densidades del caudal que ingresa y la capa del embalse correspondiente.

4. Al final del procedimiento se modifica el volumen de las capas para el día siguiente de simulación.

## 5.9 EXTRACCIÓN DE CAUDALES POR TOMAS Y VERTEDOR

### Procesos de mezclado por los flujos de salida

Los flujos de salida de un embalse contribuyen al mezclado a través de corrientes que ellos inducen. La magnitud y localización de estas corrientes controlan no sólo la calidad de las sueltas, sino también la calidad dentro del lago o almacenamiento (Martin et al., 1999). Dado el control de los caudales de salida para la calidad del agua en el embalse varios aspectos de la gestión se presentan:

- ¿Qué pasa con la calidad del agua de las sueltas si el diseño o la operación de las estructuras de toma es modificada?
- ¿Qué pasa con la calidad del agua en el almacenamiento si el diseño o la operación de las estructuras de toma es cambiada?
- ¿Pueden las estructuras de toma ser operadas para cumplir con las normas de calidad del agua para el suministro solicitado?
- ¿Pueden las estructuras de suelta selectiva (de nivel especificado) mejorar la calidad del agua en el embalse?
- Finalmente, ¿dentro del almacenamiento, de dónde viene el agua?

El modelo planteado permite la simulación (no así por el momento la optimización) de la operación de las tomas existentes, para simular la variable temperatura. A continuación se expresan las características de los métodos desarrollados en el modelo para lograr esto.

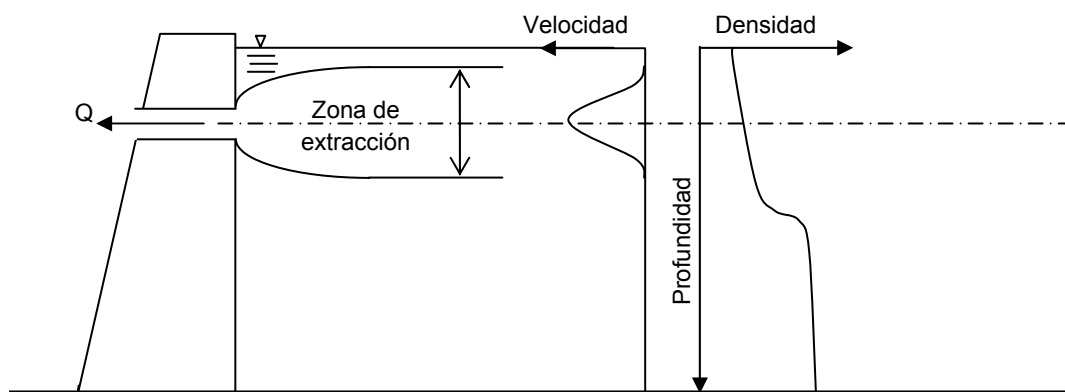
### Características de los procesos de mezcla por flujos de salida

Cuando el agua es sustraída del embalse una zona de extracción se desarrolla. La

---

extensión de esta zona depende del valor de la suelta, la localización de la extracción, la batimetría del almacenamiento y el ambiente de estratificación del lago. La estratificación, la cual limita la extensión vertical de las capas de extracción, tiene una influencia particularmente fuerte. Para una cantidad de caudal de suelta y localización la zona de extracción disminuye a la vez que el gradiente de la densidad aumenta. Dependiendo del grado de estratificación, la magnitud de la suelta, su localización y otros factores relacionados al diseño de la presa y la batimetría del almacenamiento, la zona de extracción puede ser bastante estrecha o extendida al fondo del embalse o a la superficie del agua (Martin et al., 1999).

Un patrón típico de extracción es mostrado en la Figura (23) (Martin et al., 1999).



**FIGURA 23** Patrón de extracción (Fuente: Martin, J. L., McCutcheon, S. C. 1999. "Hydrodynamics and transport for water quality modelling")

Dentro de la zona de extracción la distribución de velocidades puede variar de un punto de máxima velocidad a velocidad cero en el límite de esta zona. El punto de máximas velocidades no está necesariamente centrado en el puerto de extracción sino que depende de la forma del perfil de la densidad. Identificando el límite de la zona de extracción y la distribución de velocidades dentro de esta zona se hace relativamente sencillo estimar la calidad de las sueltas del almacenamiento de los perfiles verticales de calidad del agua (Martin et al., 1999). Las velocidades de las extracciones pueden también ser utilizadas directamente en modelos para predecir los efectos de estas en la calidad del agua del embalse.

#### Análisis del proceso de extracción

Numerosos métodos para predecir la extensión de las zonas de extracción y las

distribuciones de velocidad resultantes han sido propuestos. Por ejemplo Fischer et al. (1979) describieron métodos para calcular patrones de extracción similares a los utilizados en el análisis de inter-flujos. El modelo BETTER aplicado a un número de embalses de la Tennessee Valley Authority calcula el espesor de la zona de extracción ( $\Delta h$ ) arriba y debajo de la elevación de la toma con (Hydrologic Engineering Center, 1986):

$$\Delta h = c_w \sqrt{Q_{out}} \quad (131)$$

Donde (Hydrologic Engineering Center, 1986):

$Q_{out}$  = Es el valor del caudal de suelta.  
 $c_w$  = Coeficiente de espesor.

Se asume una distribución de flujo triangular o Gaussiana para repartir flujos dentro de la zona de extracción (Bender et al. 1990).

Bohan and Grace (1973) desarrollaron ecuaciones para calcular los límites teóricos de la extracción los cuales fueron modificados por Smith et al. (1985) para incluir el ángulo de la extracción (Martin et al., 1999):

$$\frac{Q}{Z^3 N} = \frac{\theta}{\pi} \quad (132)$$

Donde (Martin et al., 1999):

$Q$  = Caudal.  
 $Z$  = Distancia desde el centro de línea del puerto a los límites superior e inferior de extracción ( $b$  o  $D'$  en la Figura 24).  
 $\theta$  = Ángulo del puerto de extracción (en radianes).  
 $N$  = Flotación.

$$N = \sqrt{\frac{\Delta \rho g}{\rho Z}} \quad (133)$$

Donde (Martin et al., 1999):

$\Delta \rho$  = Diferencia de densidad entre los límites superior o inferior de la extracción y la densidad al centro de línea del puerto de extracción.  
 $\rho$  = Densidad (en  $\text{kg/m}^3$ ) al centro de línea del puerto (Figura 24).

La convención es que  $\Delta \rho$  es positivo para flujos estratificados estables, donde (Martin et al., 1999):

---

---

$$\Delta\rho = \rho(\text{límite superior}) - \rho(\text{puerto desagüe}) \quad (134)$$

$$\Delta\rho = \rho(\text{puerto desagüe}) - \rho(\text{límite inferior}) \quad (135)$$

Para resolver la ecuación (132) deben conocerse la elevación de la superficie del agua, la elevación del fondo del almacenamiento, la elevación del puerto de extracción y el perfil de la densidad. Sin embargo, debido a que la ecuación (132) contiene términos como la distancia desde el centro de línea del puerto ( $Z$ ) y la densidad (la cual es función de  $Z$ ), es necesario resolver esta ecuación iterativamente. Un procedimiento de solución típico donde las zonas de extracción superior o inferior pueden formarse libremente dentro del embalse (no hay interferencia en la superficie o en el fondo) es (Martin et al., 1999):

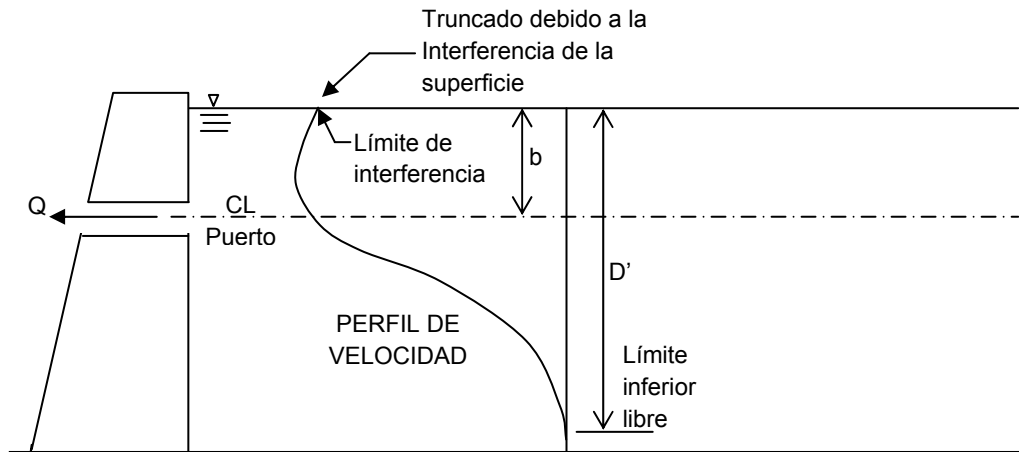
1. Re-arreglar la ecuación (132) así:

$$Q - \frac{Z^3 N\theta}{\pi} = 0 \quad (136)$$

2. Comprobar si existe interferencia primero empleando un valor de  $Z$  igual a la distancia desde el centro de línea del puerto a la superficie. Estimar la densidad en el centro de línea del puerto de extracción y en la superficie del agua y sustituir estos valores en la ecuación (136). Si la solución no es cero y es positiva, la interferencia de la superficie existe. Similarmente, sustituyendo la distancia desde el centro de línea del puerto al fondo junto con la densidad en el fondo del embalse y determinar si la interferencia del fondo existe.
3. Si ambas evaluaciones del paso 2 son negativas, la zona de extracción se forma libremente en el embalse. El límite de la superficie de extracción (la zona por arriba del puerto) puede determinarse usando estimaciones iterativas de valores para  $Z$  y la densidad a la altura  $Z$  encima del centro de línea hasta que la ecuación (136) se aproxime a cero, dentro de la tolerancia. El límite inferior de la extracción (por debajo del centro de línea del puerto) puede ser determinado de manera similar.
4. Si la interferencia de la superficie o del fondo existen un límite teórico de extracción puede determinarse. Este límite es deducido usando valores de  $Z$  calculados usando elevaciones por encima de la superficie del agua (para interferencia de la superficie) o debajo del fondo del



almacenamiento (para interferencia del fondo). Sin embargo, esta solución requiere un estimado de la densidad para regiones fuera de los límites del embalse. Davis et al. (1987) estimó esas densidades por interpolación lineal usando la densidad al centro de línea del puerto y la densidad en la superficie o el fondo del almacenamiento.



**FIGURA 24** Definición de las características de la extracción (Fuente: Martin, J. L., McCutcheon, S. C. 1999. "Hydrodynamics and transport for water quality modelling")

Para el caso donde una frontera experimente interferencia y la otra no el límite de extracción que se forma libremente no puede ser estimado con precisión usando la ecuación (136). Smith et al. (1985) propusieron una extensión de la ecuación (136) para estimar el límite libre de formación de la capa. La solución continúa de manera similar a como se describió anteriormente usando la formulación descrita por Smith (1985) (Martin et al., 1999):

Sólo para el límite libre (Figura 24) (Martin et al., 1999):

$$\frac{Q}{D^3 N} = \frac{0.125 \phi \theta}{X^3 \pi} \quad (137)$$

Donde (Martin et al., 1999):

$$\phi = \frac{1}{2} \left[ 1 + \frac{1}{\pi} \sin \left( \frac{\frac{b}{D'} \pi}{1 - \frac{b}{D'}} \right) + \frac{\frac{b}{D'}}{1 - \frac{b}{D'}} \right] \quad (138)$$

Y (Martin et al., 1999):

$$X = \frac{1}{2} \left[ 1 + \frac{\frac{b}{D'}}{1 - \frac{b}{D'}} \right] \quad (139)$$

b = Distancia desde el centro de línea del puerto a la frontera de interferencia.  
D' = Distancia entre el límite de extracción libre y la frontera de interferencia (Figura 24).

El valor de N es (Martin et al., 1999):

$$N = \sqrt{\frac{\Delta\rho}{\rho} \frac{g}{D'}} \quad (140)$$

Donde (Martin et al., 1999):

$\Delta\rho$  = Diferencia de densidad entre el valor de la superficie (para interferencia de superficie) y el límite inferior libre o la densidad en el fondo (para interferencia desde el fondo) y el límite libre superior. La convención es que:

$$\Delta\rho = \rho(\text{capa superior}) - \rho(\text{límite libre}) \quad (141)$$

O también (Martin et al., 1999):

$$\Delta\rho = \rho(\text{límite libre superior}) - \rho(\text{capa del fondo}) \quad (142)$$

Una vez que los límites de la extracción son establecidos, la distribución de velocidades puede ser estimada. Para la estimación del perfil de velocidad dentro de los límites de extracción el embalse es dividido en capas y la densidad determinada al centro de línea de cada capa. El cálculo de la distribución vertical de velocidades se basa en una estimación de la localización de la velocidad máxima la cual se puede obtener con (Bohan and Grace 1973):

$$Y_L = H \left[ \text{sen} \left( 1.57 \frac{Z_L}{H} \right) \right]^2 \quad (143)$$

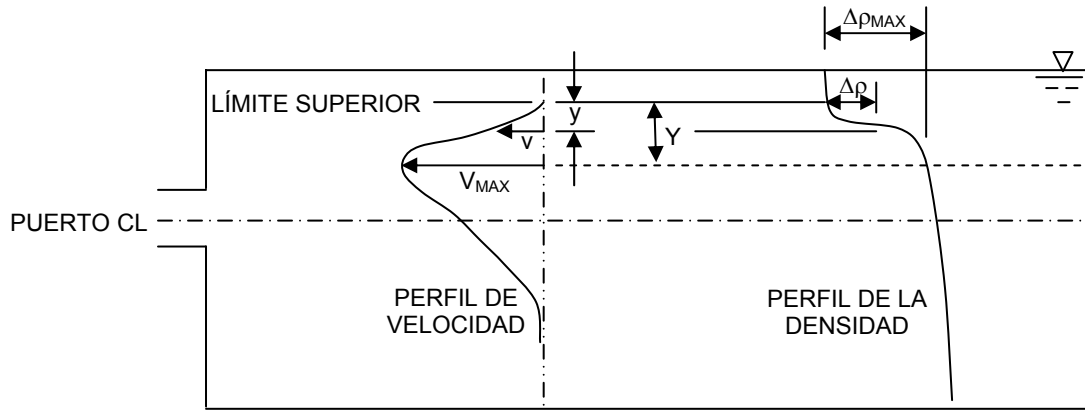
Donde (Martin et al., 1993):

$Y_L$  = Es la distancia desde el límite inferior a la elevación de velocidad máxima (Figura 25).

H = Es la distancia vertical entre los límites de extracción superior e inferior.

$Z_L$  = Es la distancia vertical entre el centro de línea de la toma de salida y el límite inferior de la extracción.

Donde la interferencia existe los límites de extracción teóricos son empleados, los cuales pueden ser por encima de la superficie del agua (para interferencia de la superficie) o por debajo del fondo del embalse (para interferencia del fondo) (Martin et al., 1999).



**FIGURA 25** Definición esquemática de variables para calcular los perfiles de velocidad (Fuente: Martin, J. L., McCutcheon, S. C. 1999. "Hydrodynamics and transport for water quality modelling")

Una vez que la localización de la velocidad máxima es determinada, la velocidad en cada capa I puede ser obtenida con (Martin et al., 1999):

$$V_N(I) = 1 - \left( \frac{y(I)}{Y} \frac{\Delta\rho(I)}{\Delta\rho_{MAX}} \right)^2 \quad (144)$$

La expresión (144) es para el caso de zonas de extracción experimentando interferencia (Martin et al., 1999).

Y para el caso donde no exista interferencia (Bohan and Grace 1973):

$$V_N(I) = \left( 1 - \frac{y(I)}{Y} \frac{\Delta\rho(I)}{\Delta\rho_{MAX}} \right)^2 \quad (145)$$

Donde (Martin et al., 1999):

$V_N(I)$  = Velocidad normalizada ( $V(I)/V_{MAX}$ ).

$y(I)$  = Distancia vertical desde la elevación de velocidad máxima al centro de línea de la capa I.

$Y$  = Distancia vertical desde la elevación de velocidad máxima a el límite de extracción superior o inferior (se determina por si el centro de línea de la capa I está por encima o por debajo del punto de velocidad máxima).

$\Delta\rho(I)$  = Diferencia de la densidad entre la elevación de velocidad máxima y el centro de línea de la capa I.

---

$\Delta\rho_{MAX}$  = Diferencia de la densidad entre el punto de velocidad máxima y el límite de extracción superior o inferior (como corresponda).

Nótese que si la interferencia existe serán pronosticadas velocidades para localizaciones tanto por arriba de la superficie del agua o por debajo del fondo del almacenamiento. Estas distribuciones de velocidades son entonces truncadas en los límites del embalse para producir la distribución final de velocidades (Martin et al., 1999).

Una vez que la distribución de velocidades es conocida, el valor del caudal en cada capa I puede ser estimado con (Hydrologic Engineering Center, 1986):

$$Q(I) = \frac{V_N(I)}{\sum_{I=1}^N V_N(I)} Q_T \quad (146)$$

Donde (Hydrologic Engineering Center, 1986):

$Q(I)$  = Caudal que se extrae de la capa I.  
 $Q_T$  = Magnitud total de la suelta.

La calidad de la extracción puede ser determinada de un simple promedio ponderado de caudal:

$$C_R = \frac{Q(I)C(I)}{\sum_{I=1}^N Q(I)} \quad (147)$$

Donde:

$C_R$  = Concentración del constituyente de calidad del agua en la suelta.  
 $C(I)$  = Concentración en cada capa.

### Caudales de salida por el vertedor

El análisis del flujo en vertedor se basa en los resultados de Bohan y Grace (1973), Harleman y Elder (1965) y experimentos conducidos por el WES más recientemente. Para la extracción sobre un vertedor el límite superior de extracción se asume que será en la superficie del agua y el límite inferior de extracción es calculado de (Environmental Laboratory, 1995):

$$\frac{Q}{L} - \left[ C - \frac{DH_w}{Z_0 + H_w} \right] \sqrt{\frac{\Delta\rho_0}{\rho_w} g (Z_0 + H_w)^3} = 0 \quad (148)$$

---

Donde (Environmental Laboratory, 1995):

C = Coeficiente adimensional,  $C = 0.54$  para  $Z_0 \geq H_w$   
 $C = 0.78$  para  $Z_0 < H_w$

Q = Descarga sobre el vertedor,  $m^3/hr$ .

D = Coeficiente adimensional,  $D = 0.0$  para  $Z_0 \geq H_w$   
 $D = 0.70$  para  $Z_0 < H_w$

L = Longitud libre del vertedor, m.

$H_w$  = Distancia vertical desde la superficie del agua a la cresta del vertedor, m.

$Z_0$  = Distancia vertical desde la cresta del vertedor al límite inferior de la extracción, m.

$\Delta\rho_0$  = Diferencia de la densidad entre la elevación de la cresta del vertedor y el límite inferior de extracción,  $kg/m^3$ .

$\rho_w$  = Densidad a la elevación de la cresta del vertedor,  $kg/m^3$ .

Las variables anteriores son mostradas en la Figura (26). La ecuación (148) requiere una solución iterativa para  $Z_0$  (Environmental Laboratory, 1995).

Las descargas de las capas por flujo sobre el vertedor son determinadas de acuerdo al tipo de vertedor. Para un vertedor sumergido el perfil de velocidad por abajo del punto de velocidad máxima está dado por (Environmental Laboratory, 1995):

$$u = u_m \left[ 1 - \left( \frac{y \cdot \Delta\rho}{Y \cdot \Delta\rho_m} \right) \right]^3 \quad (149)$$

Donde (Environmental Laboratory, 1995):

u = Velocidad en el punto considerado.

$u_m$  = Velocidad máxima.

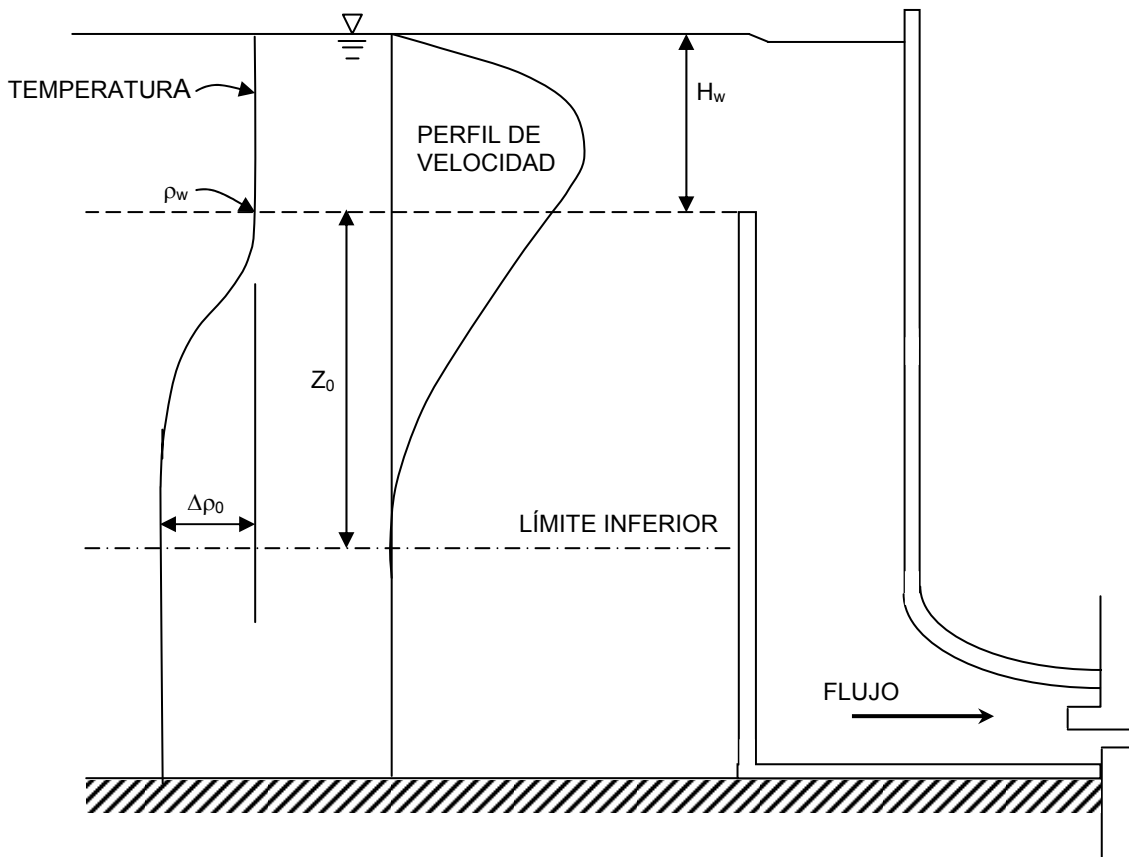
Y arriba del punto de máxima velocidad con (Environmental Laboratory, 1995):

$$u = u_m \left[ 1 - \left( \frac{y \cdot \Delta\rho}{Y \cdot \Delta\rho_m} \right)^2 \right] \quad (150)$$

Para un vertedor libre la velocidad máxima ocurre en la superficie. El perfil de velocidad por debajo de la superficie está dado por (Environmental Laboratory, 1995):

$$u = u_m \left[ 1 - \left( \frac{y \cdot \Delta\rho}{Y \cdot \Delta\rho_m} \right)^P \right] \quad (151)$$

Donde el exponente  $P$  es definido empíricamente en términos del Coeficiente de descarga para vertedor libre como 4.35 menos el valor de 1.04 veces el Coeficiente de entrada (Environmental Laboratory, 1995).



**FIGURA 26** Extracción por vertedor (Fuente: Environmental Laboratory, 1995. "CE-QUAL-R1: A numerical one-dimensional model of reservoir water quality; user's manual")

## 5.10 FLUJOS DE CALOR EN LAS CAPAS DEL ALMACENAMIENTO

La distribución de energía en un embalse tiene influencia significativa en las operaciones de almacenamiento y calidad del agua. La temperatura del agua es un factor dominante en la determinación de la densidad, que a su vez, controla la ubicación del agua que entra, la carga de constituyentes asociada y la distribución del flujo de salida (Environmental Laboratory, 1995). La temperatura del agua tiene una fuerte influencia en los valores de las reacciones químicas y biológicas (Chapra, 1997). Por estas razones, la estructura de temperatura en las capas es tan importante.

Todos los flujos de calor actúan solamente en la superficie del agua y en consecuencia afectan solamente a la capa superficial, con excepción de la radiación solar de onda corta, que no puede ser considerada exclusivamente así pues es absorbida exponencialmente con la profundidad (Environmental Laboratory, 1995).

Dentro de los procesos más dominantes en el intercambio de energía térmica se tiene la advección y el intercambio de calor a través de la interfase aire-agua (Hydrologic

---

Engineering Center, 1986; Martin et al., 1999).

El régimen termal de almacenamientos templados es el resultado, primeramente, de la interacción de dos procesos: (1) transferencia de energía térmica y movimiento a través de la superficie del embalse y (2) la fuerza de gravedad actuando sobre las diferencias de la densidad en el interior del almacenamiento (Martin et al., 1999).

Para poder calcular la temperatura en las capas de un almacenamiento un paso previo es obtener los flujos de energía térmica en las capas. Este proceso se inicia con la radiación de onda corta en superficie (Hydrologic Engineering Center, 1986; Martin et al., 1999).

Radiación solar neta de onda corta.- La radiación solar puede ser medida de una forma relativamente no costosa y es reportada por algunas estaciones climatológicas. Sólo si la radiación solar medida no está disponible deberá ser calculada (Hydrologic Engineering Center, 1986).

Del Instituto Valenciano de Investigaciones Agrarias (I.V.I.A) se obtuvieron los datos de radiación solar global (medidos en la estación de Villajoyosa, en  $W/m^2$ ) (Paredes et al., 2006b). Para los cálculos a efectuar estas mediciones se requieren en unidades de  $cal/(día \cdot m^2)$  (Hydrologic Engineering Center, 1986; Martin et al., 1999). Con el propósito de efectuar la conversión se emplea el siguiente factor (Serway et al., 2005):

$$1 \frac{W}{m^2} \times \frac{1 \frac{J}{s}}{1 W} \times \frac{1 cal}{4,186 J} \times \frac{86400 s}{1 día} = 20640 \frac{cal}{m^2 \cdot día}$$

Energía térmica en una capa  $i$  (excepto la capa superficial) y el Coeficiente de extinción lumínica.- La radiación solar es absorbida en la superficie del agua y selectivamente con la profundidad, dependiendo de la longitud de onda de la luz, propiedades del agua y la materia suspendida en la misma. Esta absorción es usualmente supuesta como de comportamiento exponencial con la profundidad (Ley de Beer-Lambert), pero efectos de superficie generan discrepancias menores en el primer metro superficial del embalse. Para incorporar estas discrepancias en el modelo una cierta fracción,  $\beta$ , de la radiación solar se asume que será absorbida en una zona de la superficie de 0,6 m de espesor (Figura 27). La fracción remanente,  $(1-\beta) \cdot$  (Radiación solar neta), es absorbida exponencialmente en la profundidad con un coeficiente de extinción lumínica,  $\eta$ . Esto es (Environmental Laboratory, 1995):

$$\Phi(Z) = (1 - \beta) Q_{ns} e^{-\eta Z^*} \quad (152)$$

Donde (Environmental Laboratory, 1995):

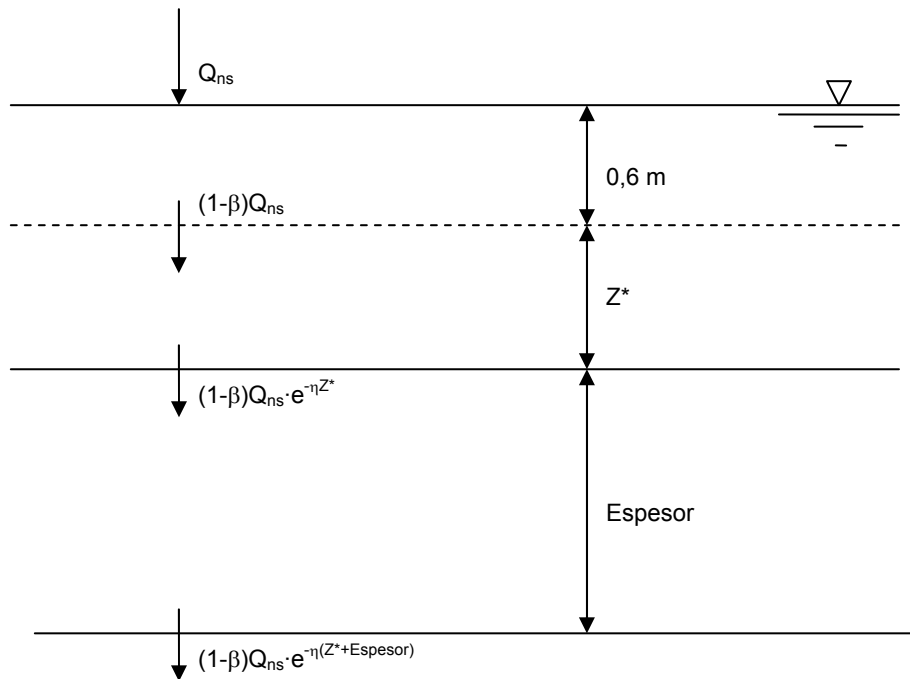
$\Phi(Z)$  = Flujo de radiación solar a la profundidad  $Z$ .

$Z^*$  = Profundidad del agua, comenzando 0,6 m por debajo de la superficie (esto es,  $Z^* = Z - 0,6$  m).

$Q_{ns}$  = Radiación solar.

Para el caso general de una capa,  $i$ , que tiene un espesor y está limitada por las superficies horizontales  $A(i + 1)$  y  $A(i)$  y localizada 0,6 m +  $Z$  debajo de la superficie, la energía absorbida es (Environmental Laboratory, 1995):

$$\text{Energía térmica}(i) = \left[ (1 - \beta) Q_{ns} e^{-\eta Z^*} \cdot \text{Area}(i + 1) \right] - \left[ (1 - \beta) Q_{ns} e^{-\eta(Z^* + \text{Espesor})} \cdot \text{Area}(i) \right] \quad (153)$$



**FIGURA 27** Flujos de radiación solar con la profundidad (Fuente: Environmental Laboratory. 1995. "CE-QUAL-R1: A numerical one-dimensional model of reservoir water quality; user's manual")

El coeficiente de extinción,  $\eta$ , es dependiente de la calidad y tipo de materia suspendida en el agua. Este coeficiente de extinción es entonces dividido en dos partes (Environmental Laboratory, 1995):

$$\eta = \eta_w + \sum k_j C_j \quad (154)$$



---

Donde (Environmental Laboratory, 1995):

- $\eta$  = Coeficiente de extinción lumínica, 1/m.
- $\eta_w$  = Coeficiente de extinción para agua clara, 1/m.
- $k_j$  = Coeficiente de extinción para el constituyente  $C_j$ , (1/m)/(g/m<sup>3</sup>).
- $C_j$  = Constituyente de calidad, (g/m<sup>3</sup>).

Las constantes generales que se requerirán, según lo antes mencionado, para efectuar la distribución de energía termal en las capas del embalse son:

- La fracción de la radiación absorbida en los primeros 0,6 m.
- Coeficiente de extinción lumínica para agua clara.

Flujo de energía térmica en la capa superficial del embalse.- El intercambio de calor en la superficie es modelado como una combinación de cinco procesos (Hydrologic Engineering Center, 1986; Chapra, 1997; Martin et al., 1999):

$$J_n = I(0) + J_{an} - J_{br} - J_c - J_e \quad (155)$$

Donde (Hydrologic Engineering Center, 1986; Chapra, 1997; Martin et al., 1999):

- $J_n$  = Intercambio de calor en la superficie del agua. Si tiene signo positivo (+), la capa de agua en la superficie del embalse recibe energía en forma de calor; si el signo es negativo (-), la pierde.
- $I(0)$  = Radiación solar neta de onda corta en la superficie del agua.
- $J_{an}$  = Radiación atmosférica neta de onda larga.
- $J_{br}$  = Radiación de onda larga que sale del agua.
- $J_c$  = Conducción.
- $J_e$  = Evaporación.

Todos los flujos son expresados en cal/(día·m<sup>2</sup>) (Hydrologic Engineering Center, 1986; Martin et al., 1999).

1.- Radiación solar neta de onda corta.- Se obtiene con la siguiente expresión, que es la radiación solar total que ingresa en la superficie del embalse menos la radiación que pasa a las capas por debajo de la superficial (Environmental Laboratory, 1995; Martin et al., 1999):

$$I(0) = Q_{ns} \cdot Area_n - (1 - \beta) \cdot Q_{ns} \cdot Area_{n-1} \cdot e^{-\eta \cdot Z^*} \quad (156)$$

Donde (Environmental Laboratory, 1995; Martin et al., 1999):

$Area_n$  = Área de la superficie del embalse.

---

$Area_{n-1}$  = Área de la frontera inferior de la capa superficial del embalse.

2.- Radiación atmosférica de onda larga.- El flujo de radiación de onda larga que desciende de la atmósfera es uno de los mayores términos en el balance de calor en superficie. Este flujo puede calcularse usando la Ley de Stefan-Boltzmann (Martin et al., 1999):

$$J_{an} = \sigma \cdot (T_{aire} + 273)^4 \cdot \epsilon_{sky} (1 - R_L) \cdot 10000 \cdot Area_n \quad (157)$$

Donde (Martin et al., 1999):

$J_{an}$  = Radiación atmosférica de onda larga.

$\sigma$  = La constante de Stefan-Boltzmann, donde:

$$\sigma = 11,702 \times 10^{-8} \frac{\text{cal}}{\text{cm}^2 \cdot \text{d} \cdot \text{K}^4}$$

$T_{aire}$  = Temperatura del aire (°C).

$\epsilon_{sky}$  = Emisividad efectiva de la atmósfera (adimensional).

$R_L$  = Coeficiente de reflexión de onda larga (adimensional).  $R_L = 0,03$ .

Y se tiene que (Martin et al., 1999):

$$\epsilon_{sky} = \epsilon_{clear} (1 + 0,17 C_L^2) \quad (158)$$

Donde (Martin et al., 1999):

$C_L$  = Fracción del cielo cubierta de nubes.

El valor de  $\epsilon_{clear}$  se calcula con la expresión de Brunt (1932):

$$\epsilon_{clear} = 0,6 + 0,031 \sqrt{e_{aire}} \quad (159)$$

Donde (Martin et al., 1999):

$e_{aire}$  = Presión de vapor del aire (en mmHg).

La presión de vapor del aire se puede calcular con (Martin et al., 1999; Hornung, 2002):

$$e_{aire}(T_{aire}) = \frac{h_r}{100} \cdot \exp \left[ 2,3026 \cdot \left( \frac{7,5 \cdot T_{aire}}{T_{aire} + 237,3} + 0,7858 \right) \right] \quad (160)$$

Donde (Martin et al., 1999; Hornung, 2002):

---

$e_{\text{aire}}$  = Presión de vapor del aire, en mbar.  
 $h_r$  = Humedad relativa (en %).

Para convertir de mbar a mmHg, se tiene que  $1 \text{ mbar} = 0,75006 \text{ mmHg}$ .  $e_{\text{aire}}$  se deberá de multiplicar por este factor antes de introducirla a la ecuación (159) (Serway et al., 2005).

3.- Radiación de onda larga proveniente del agua.- La radiación que emite la superficie del agua es representada por la Ley de Stefan-Boltzmann (Martin et al., 1999):

$$J_{br} = \varepsilon\sigma(T + 273)^4 \cdot 10000 \cdot \text{Area}_n \quad (161)$$

Donde (Martin et al., 1999):

$J_{br}$  = Radiación de onda larga proveniente del agua.  
 $\varepsilon$  = Emisividad del agua.  $\varepsilon = 0,97$ .  
 $T$  = Temperatura del agua (en °C).

4.- Conducción y convección.- Conducción es el proceso de transferencia de calor de molécula a molécula cuando materiales a distintas temperaturas son puestos en contacto. Convección es el proceso de transferencia de calor que ocurre debido al movimiento de masa en los fluidos (Serway et al., 2005). Ambos pueden ocurrir en la interfase aire-agua y es posible describirlos con la expresión (Martin et al., 1999):

$$J_c = c_1 f(U_w)(T - T_{\text{aire}}) \cdot 10000 \cdot \text{Area}_n \quad (162)$$

Donde (Martin et al., 1999):

$J_c$  = Flujo de calor debido a conducción y convección.  
 $c_1$  = Coeficiente de Bowen.  $c_1 = 0,47 \text{ mmHg}/^\circ\text{C}$   
 $f(U_w)$  = Define la dependencia de la transferencia de la velocidad del viento sobre la superficie del agua.  
 $U_w$  = Velocidad del viento medida a una distancia fija sobre la superficie del agua.  
 $T$  = Temperatura del agua en la superficie del embalse (en °C).

La opción que se usará para calcular  $f(U_w)$  es, de Brady, Graves y Geyer (1969) (Chapra, 1997; Martin et al., 1999):

$$f(U_w) = 19,0 + 0,95U_w^2 \quad (163)$$

En estas expresiones el viento irá multiplicado por un factor llamado Factor por viento (es una constante general) (Hornung, 2002).

---

5.- Evaporación y condensación.- El calor perdido por evaporación puede ser representado por la Ley de Dalton (Martin et al., 1999):

$$J_e = f(U_w)(e_s - e_{aire}) \cdot 10000 \cdot Area_n \quad (164)$$

Donde (Martin et al., 1999):

$J_e$  = Calor perdido por evaporación.  
 $e_s$  = Presión de saturación de vapor en la superficie del agua (en mmHg).  
 $e_{aire}$  = Presión de vapor del aire (en mmHg).

La presión de saturación de vapor es calculada con (Chapra, 1997; Martin et al., 1999; Llorca et al., 2006):

$$e_s = 4,596e^{\frac{17,27T}{237,3+T}} \quad (165)$$

Donde (Chapra, 1997; Martin et al., 1999; Llorca et al., 2006):

$T$  = Temperatura del agua (en °C).

Transferencia de calor entre el sedimento y el agua.- No se incluirá en los cálculos. La información que se requiere del sedimento para incluirlo en los cálculos de flujo de energía térmica es: la temperatura de los sedimentos del fondo, la densidad, el calor específico de los sedimentos y el espesor efectivo de la capa de sedimento (Chapra et al., 2003).

### 5.11 PRECIPITACIÓN, EVAPORACIÓN E INFILTRACIÓN

En el flujo de calor que ocurre dentro del embalse que se considera dividido en capas, intervienen, entre otros, tres factores (Martin et al., 1999; Hornung, 2002):

- Evaporación.
- Precipitación.
- Filtración.

Evaporación.- La evaporación se resta en volumen de la capa superior del embalse. La ecuación genérica de pérdida de calor por evaporación es (Martin et al., 1999; Serway et al., 2005):

$$H_L = \rho \cdot L_w \cdot E \quad (166)$$

---

Donde (Martin et al., 1999; Serway et al., 2005):

$H_L$  = Pérdida de calor, en  $W/m^2$ .

$\rho$  = Densidad del agua que está siendo evaporada, en  $kg/m^3$ .

$L_w$  = Calor latente del agua, en  $J/kg$ .

$E$  = Tasa de evaporación, en  $m/s$ .

Con solamente una proporción pequeña de moléculas (en comparación al volumen total del embalse) localizada cerca de la superficie, moviéndose en la dirección correcta para escapar del líquido y con una velocidad superior a la velocidad media en un cierto instante, la tasa de evaporación es limitada. Además, como las moléculas de mayor energía escapan y las que quedan tienen menor energía cinética media, la temperatura del líquido se reduce. Este fenómeno también es llamado de enfriamiento evaporativo (Martin et al., 1999).

Al considerar la energía térmica perdida por evaporación en la capa superficial del embalse se está tomando en cuenta la evaporación y, en consecuencia, el volumen evaporado no tienen influencia ya en la ecuación de energía diferencial de energía termal. Para efectos de caudales de salida el volumen evaporado se sustrae de la capa superficial del almacenamiento.

Precipitación.- El volumen de agua que ingresa al embalse por precipitación directa sobre el espejo de agua se considera que entra a la capa superior. La temperatura del agua de lluvia se considera igual a la temperatura del aire cercano a la superficie del embalse.

Filtración.- Se asume que la filtración ocurre por completo en la capa 0 (capa que se ubica en el fondo del embalse). El agua que se infiltra deja el embalse conteniendo la misma temperatura y propiedades que la capa del fondo.

## 5.12 PROCESO PARA EL CÁLCULO DE TEMPERATURA EN CADA UNA DE LAS CAPAS

Debido a que la temperatura es una concentración de energía termal, la ecuación general de conservación de masa es aplicable. Para cualquier capa  $i$  la ecuación general de energía termal es (Environmental Laboratory, 1995):

$$\rho_i C_p \frac{\partial}{\partial t} (T_i V_i) = \frac{\partial}{\partial Z} \left( \rho_i C_p D_i A_i \frac{\partial T}{\partial Z} \right) \Delta Z_i + \frac{\partial}{\partial Z} (\phi_i A_i) \Delta Z_i + \rho_i C_p (Q_{in} T_{in} + Q_p T_p - T_i Q_{out}) \quad (167)$$

---

Donde (Environmental Laboratory, 1995):

- $\rho_i$  = Densidad del agua,  $\text{kg/m}^3$ .
- $C_p$  = Calor específico del agua,  $\text{cal}/(^{\circ}\text{C}\cdot\text{kg})$ .
- $T_i$  = Temperatura,  $^{\circ}\text{C}$ .
- $V_i$  = Volumen de la capa,  $\text{m}^3$ .
- $D_i$  = Difusividad,  $\text{m}^2/\text{d}$ .
- $A_i$  = Área,  $\text{m}^2$ .
- $\Delta Z_i$  = Espesor,  $\text{m}$ .
- $\phi_i$  = Flujo de radiación solar,  $\text{cal}/\text{m}^2/\text{d}$ .
- $Q_{in}$  = Volumen que ingresa,  $\text{m}^3/\text{d}$ .
- $T_{in}$  = Temperatura del volumen que ingresa,  $^{\circ}\text{C}$ .
- $Q_p$  = Volumen bombeado de regreso,  $\text{m}^3/\text{d}$ .
- $T_p$  = Temperatura del volumen bombeado de regreso,  $^{\circ}\text{C}$ .
- $Q_{out}$  = Volumen que sale,  $\text{m}^3/\text{d}$ .

Los términos de la ecuación (167) representan varios procesos que ocurren dentro del almacenamiento en relación con los flujos de energía térmica en las capas, como son los siguientes (Environmental Laboratory, 1995):

$$\left[ \begin{array}{c} \text{Razón de} \\ \text{cambio} \\ \text{cal / d} \end{array} \right] = \left[ \begin{array}{c} \text{Difusión} \\ \text{cal / d} \end{array} \right] + \left[ \begin{array}{c} \text{Absorción interna} \\ \text{de radiación solar} \\ \text{cal / d} \end{array} \right] + \left[ \begin{array}{c} \text{Advección} \\ \text{cal / d} \end{array} \right] \quad (168)$$

La advección incluye términos para salidas, tributarios y bombes de regreso, que ingresan hacia el almacenamiento. El último paso es calcular el nuevo perfil vertical de temperatura. Para la capa superficial del embalse el flujo neto de calor a través de la interfase aire-agua debe ser agregado en el lado derecho de la ecuación (167) (Environmental Laboratory, 1995).

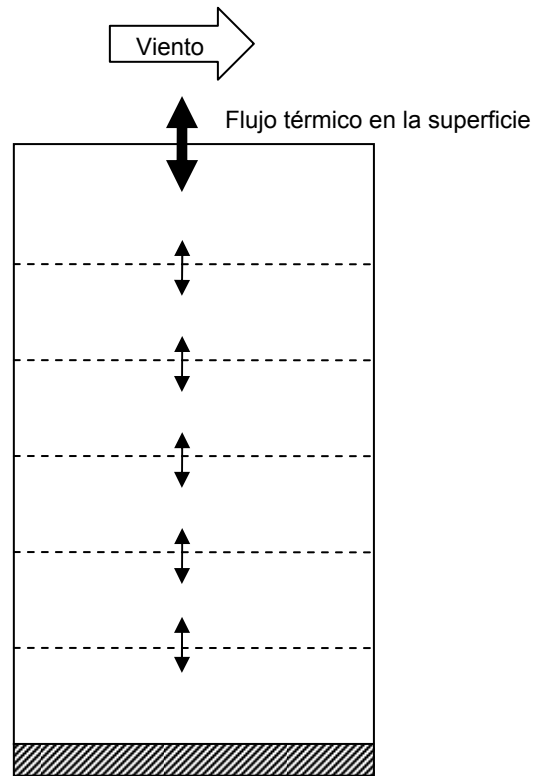
Dos métodos han sido desarrollados para modelar la distribución vertical de temperaturas. El primero está basado en un planteamiento de difusión turbulenta (Figura 28). Las soluciones numéricas son desarrolladas dividiendo la dimensión vertical en segmentos. El calor es suministrado en la superficie y entonces es distribuido a través de la profundidad por difusión (Chapra, 1997).

Un segundo método, llamado planteamiento de capa mezclada, utiliza un balance de energía mecánica para predecir el espesor del epilimnion. Este espesor de la capa superficial es entonces modelado como un segmento bien mezclado. El hipolimnion es modelado como una serie de capas usando el planteamiento de difusión turbulenta (Chapra, 1997).

Debido a su simplicidad se recurre más el planteamiento de difusión turbulenta

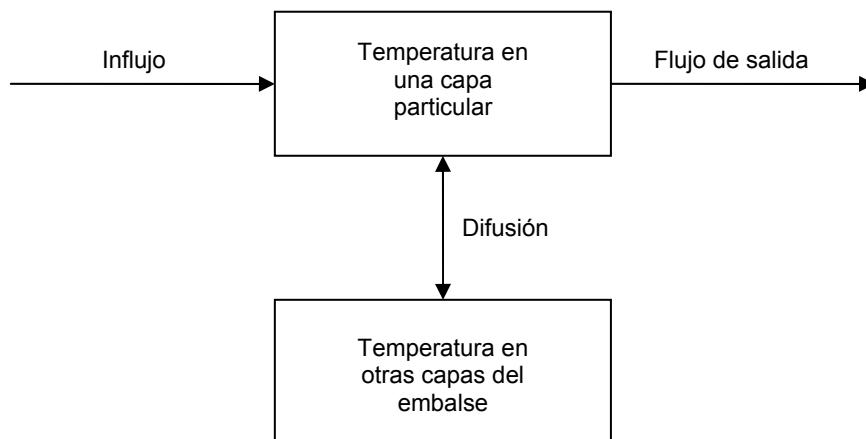
---

(Chapra, 1997), que es el que se utilizó en este trabajo (Figura 28).



**FIGURA 28** Difusión turbulenta (Fuente: Chapra, S. C. 1997. "Surface water-quality modeling")

Un ejemplo de diagrama de bloques para cálculo de temperatura en una capa es el de la Figura (29):



**FIGURA 29** Factores que intervienen en la obtención de la temperatura de una capa

El desarrollo en diferencias finitas de la expresión de conservación de la energía térmica para la obtención de la temperatura de una capa  $i$  es (Figura 30):

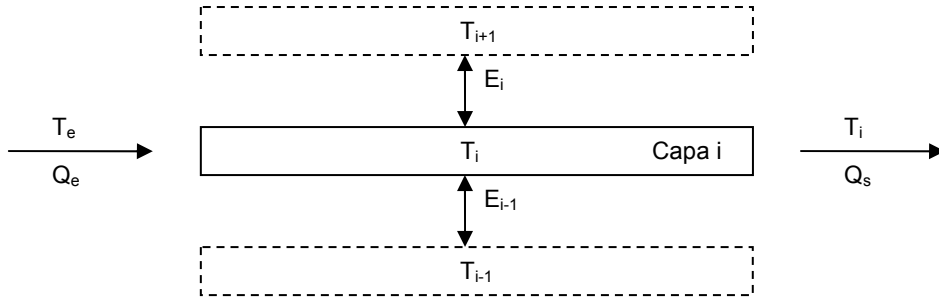


FIGURA 30 Flujos de energía térmica en una capa *i*

La nomenclatura que se utiliza es (Figura 31):

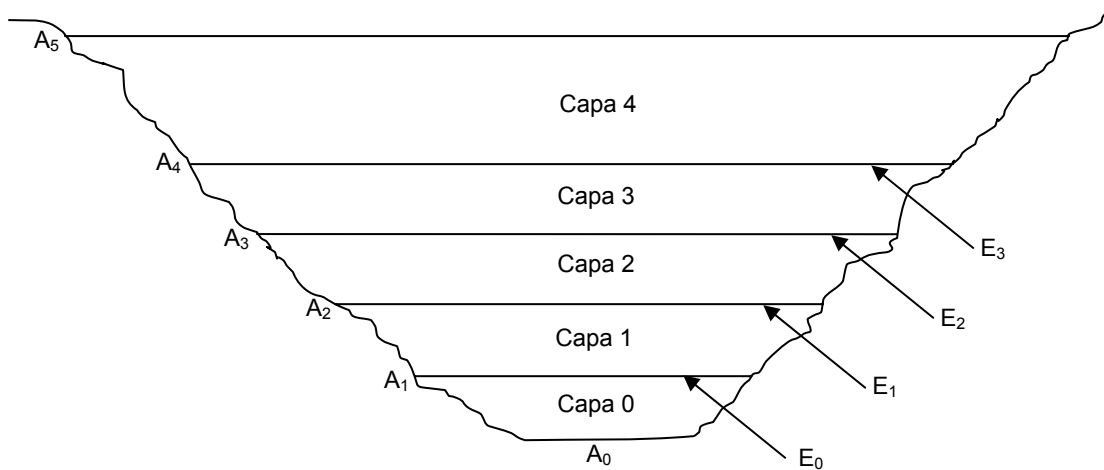
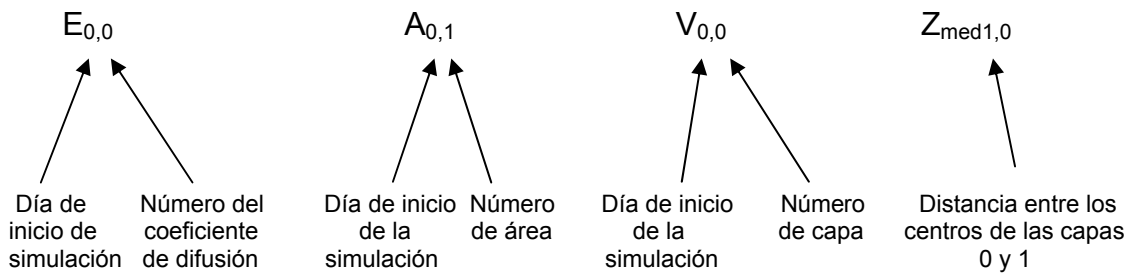


FIGURA 31 Algunas de las variables empleadas en las ecuaciones en diferencias finitas

En el caso de temperatura se tiene la expresión (Environmental Laboratory, 1995):

$$\frac{d(T_i V_i)}{dt} = T_i \cdot \frac{d(V_i)}{dt} + V_i \cdot \frac{d(T_i)}{dt} \quad (169)$$



Y desarrollando para una capa  $i$  cualquiera (Environmental Laboratory, 1995):

$$\begin{aligned} \frac{d(T_i V_i)}{dt} = T_i \cdot \frac{d(V_i)}{dt} + V_i \cdot \frac{d(T_i)}{dt} = \\ Q_e \cdot T_e - Q_s \cdot T_i + E_{i,i} (T_{i+1} - T_i) \cdot \frac{A_{i,i}}{Z_{med_{i+1,i}}} \\ - E_{i,i-1} (T_i - T_{i-1}) \cdot \frac{A_{i,i-1}}{Z_{med_{i,i-1}}} + \text{Energía (+/-) por radiación} \end{aligned} \quad (170)$$

Para obtener el cambio en la temperatura de capa se despeja de la ecuación (170):

$$\begin{aligned} V_i \cdot \frac{d(T_i)}{dt} = Q_e \cdot T_e - Q_s \cdot T_i + E_{i,i} (T_{i+1} - T_i) \cdot \frac{A_{i,i}}{Z_{med_{i+1,i}}} \\ - E_{i,i-1} (T_i - T_{i-1}) \cdot \frac{A_{i,i-1}}{Z_{med_{i,i-1}}} + \text{Energía (+/-) por radiación} - T_i \cdot \frac{d(V_i)}{dt} \end{aligned} \quad (171)$$

Y se puede expresar el último término de la ecuación (171) así:

$$\frac{d(V_i)}{dt} = \frac{V_{i+1} - V_i}{\Delta t} \quad (172)$$

Por lo tanto, la ecuación (171) quedaría así:

$$\begin{aligned} V_i \cdot \frac{d(T_i)}{dt} = Q_e \cdot T_e - Q_s \cdot T_i + E_{i,i} (T_{i+1} - T_i) \cdot \frac{A_{i,i}}{Z_{med_{i+1,i}}} \\ - E_{i,i-1} (T_i - T_{i-1}) \cdot \frac{A_{i,i-1}}{Z_{med_{i,i-1}}} + \text{Energía (+/-) por radiación} - T_i \cdot \frac{V_{i+1} - V_i}{\Delta t} \end{aligned} \quad (173)$$

La energía ganada o perdida por radiación se puede calcular así (Martin et al., 1999):

$$H = \rho V_i c (T_2 - T_1) \quad (174)$$

Donde (Martin et al., 1999):

$H$  = Energía ganada o perdida por radiación para la capa completa, en cal/día.

$\rho$  = Densidad de la capa, en  $\text{kg/m}^3$ .

$V_i$  = Volumen de la capa  $i$ , en  $\text{m}^3$ .

$c$  = Calor específico del agua =  $1 \frac{\text{cal}}{\text{g}^\circ\text{C}}$ .

$T_2$  = Temperatura final de la capa, en  $^\circ\text{C}$ .

$T_1$  = Temperatura inicial de la capa, en  $^\circ\text{C}$ .

$m_i$  = Masa de la capa  $i$ , en kg. Se tiene  $m = \rho V_i$ .

Para expresar en las mismas dimensiones de la ecuación (173) la ecuación (174):

$$\frac{H}{\rho c} = V_i (T_2 - T_1) \quad (175)$$

$$\text{Energía (+/-) por radiación} = \frac{H}{\rho c} \quad (176)$$

La expresión (176) será la que se va a sustituir en la (173). Sólo falta expresarla en las mismas unidades. Para esto se obtiene el factor de conversión correspondiente:

$$\frac{H}{\rho c} = \frac{H \left\{ \frac{\text{cal}}{\text{d}} \right\}}{\rho \left\{ \frac{\text{kg}}{\text{m}^3} \right\} \left\{ 1 \frac{\text{cal}}{\text{g}^\circ\text{C}} \right\}} \times \frac{1}{\left\{ \frac{1000 \text{ g}}{\text{kg}} \right\}} = \frac{H}{1000 \rho} \left\{ \frac{\text{m}^3 \cdot ^\circ\text{C}}{\text{día}} \right\} \quad (177)$$

La expresión (177) es la que se usará dentro de la fórmula (173) para obtener la Energía (+/-) por radiación. Por lo que para una capa  $i$  cualquiera se tiene, haciendo la sustitución (Figura 32):

$$\begin{aligned} V_i \cdot \frac{d}{dt}(T_i) = & Q_e \cdot T_e - Q_s \cdot T_i + E_{i,i} (T_{i+1} - T_i) \cdot \frac{A_{i,i}}{Z_{\text{med}}}_{i+1,i} \\ & - E_{i,i-1} (T_i - T_{i-1}) \cdot \frac{A_{i,i-1}}{Z_{\text{med}}}_{i,i-1} + \frac{H_i}{1000 \rho_i} - T_i \cdot \frac{V_{i+1,i} - V_{i,i}}{\Delta t} \end{aligned} \quad (178)$$

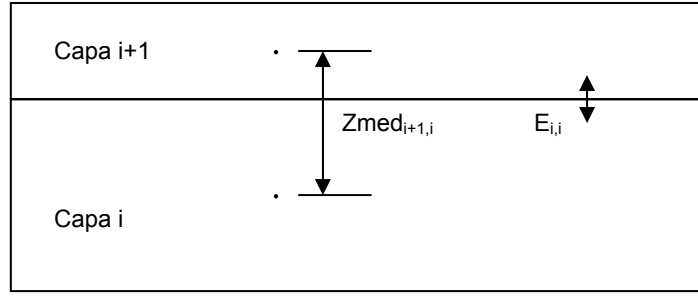
Para la capa superior ( $n-1$ ) la ecuación es:

$$\begin{aligned} V_{n-1} \cdot \frac{d}{dt}(T_{n-1}) = & Q_e \cdot T_e - Q_s \cdot T_{n-1} - E_{i,n-2} (T_{n-1} - T_{n-2}) \cdot \frac{A_{i,n-2}}{Z_{\text{med}}}_{n-1,n-2} \\ & + \frac{H_{n-1}}{1000 \rho_{n-1}} - T_{n-1} \cdot \frac{V_{i+1,n-1} - V_{i,n-1}}{\Delta t} \end{aligned} \quad (179)$$

Para la capa del fondo (capa 0) la ecuación es:

$$\begin{aligned} V_0 \cdot \frac{d}{dt}(T_0) = & Q_e \cdot T_e - Q_s \cdot T_0 + E_{i,0} (T_1 - T_0) \cdot \frac{A_{i,1}}{Z_{\text{med}}}_{1,0} \\ & + \frac{H_0}{1000 \rho_0} - T_0 \cdot \frac{V_{i+1,0} - V_{i,0}}{\Delta t} \end{aligned} \quad (180)$$

El valor de un constituyente para una capa se considera ubicado en el centro, respecto al eje vertical de la capa.



**FIGURA 32** Significado físico de la variable Zmed

Al agrupar las ecuaciones (178), (179) y (180), con respecto a las temperaturas, se llega a las ecuaciones siguientes:

Capa del fondo (Capa 0):

$$V_0 \cdot \dot{T}_0 = \left[ -Q_s - E_{i,0} \cdot \frac{A_{i,1}}{Z_{med1,0}} + \frac{V_{i,0} - V_{i+1,0}}{\Delta t} \right] T_0 + \left[ E_{i,0} \cdot \frac{A_{i,1}}{Z_{med1,0}} \right] T_1 + \left[ Q_e \cdot T_e + \frac{H_0}{1000\rho_0} \right] \quad (181)$$

El subíndice  $i$  se refiere a un día específico del periodo para el cual se está efectuando la simulación.

Para una capa intermedia  $j$ :

$$V_j \cdot \dot{T}_j = \left[ +E_{i,j-1} \cdot \frac{A_{i,j}}{Z_{medj,j-1}} \right] T_{j-1} + \left[ -Q_s - E_{i,j} \cdot \frac{A_{i,j+1}}{Z_{medj+1,j}} - E_{i,j-1} \cdot \frac{A_{i,j}}{Z_{medj,j-1}} + \frac{V_{i,j} - V_{i+1,j}}{\Delta t} \right] T_j + \left[ E_{i,j} \cdot \frac{A_{i,j+1}}{Z_{medj+1,j}} \right] T_{j+1} + \left[ Q_e \cdot T_e + \frac{H_j}{1000\rho_j} \right] \quad (182)$$

Capa superior ( $n-1$ ):

$$V_{n-1} \cdot \dot{T}_{n-1} = \left[ +E_{i,n-2} \cdot \frac{A_{i,n-1}}{Z_{medn-1,n-2}} \right] T_{n-2} + \left[ -Q_s - E_{i,n-2} \cdot \frac{A_{i,n-1}}{Z_{medn-1,n-2}} + \frac{V_{i,n-1} - V_{i+1,n-1}}{\Delta t} \right] T_{n-1} + \left[ Q_e \cdot T_e + \frac{H_{n-1}}{1000\rho_{n-1}} \right] \quad (183)$$

La ecuación (182) expresa el hecho de que cada capa es influenciada sólo por sus vecinas inmediatas (arriba y debajo). La expresión (181) indica que la capa del fondo sólo es influenciada por la capa arriba de ella y la (183) que la capa superior sólo se ve afectada por la capa que yace inmediatamente por debajo de ella.

Por definición (Environmental Laboratory, 1995):

$$T^* = T(t + \Delta t) \quad (184)$$

Donde (Environmental Laboratory, 1995):

$T^*$  = Temperatura en el próximo paso de tiempo (Temperatura en  $t + \Delta t$ ).

Agrupando las ecuaciones que se producen al aplicar (181), (182) y (183), se genera el siguiente sistema de ecuaciones diferenciales (se ejemplifica para cinco capas horizontales):

$$\begin{cases} V_0 \dot{T}_0 = A_{00} T_0 + A_{01} T_1 & 0 & 0 & 0 & +P_0 & \dots(0) \\ V_1 \dot{T}_1 = A_{10} T_0 + A_{11} T_1 + A_{12} T_2 & 0 & 0 & 0 & +P_1 & \dots(1) \\ V_2 \dot{T}_2 = 0 + A_{21} T_1 + A_{22} T_2 + A_{23} T_3 & 0 & 0 & 0 & +P_2 & \dots(2) \\ V_3 \dot{T}_3 = 0 & 0 & +A_{32} T_2 + A_{33} T_3 + A_{34} T_4 & +P_3 & \dots(3) \\ V_4 \dot{T}_4 = 0 & 0 & 0 & +A_{43} T_3 + A_{44} T_4 & +P_4 & \dots(4) \end{cases} \quad (185)$$

El término  $P_j$  es un factor que incluye el influjo (caudal de entrada) y otras transferencias hacia la capa  $j$  que son independientes de las temperaturas en las capas.

La estrategia numérica para resolver el sistema de ecuaciones (185) involucra una formulación de ecuaciones en diferencias homogéneas, en las cuales las temperaturas para el próximo punto en el tiempo son expresadas directamente en términos de las temperaturas presentes y sus valores de cambio e implícitamente en términos de los valores de cambio en el próximo punto (paso de tiempo) (Hydrologic Engineering Center, 1986). La naturaleza implícita de esta estrategia es manifestada en la aproximación siguiente (Hydrologic Engineering Center, 1986; Environmental Laboratory, 1995):

$$\begin{aligned}
T_0^* &= T_0 + \frac{\Delta t}{2} \dot{T}_0 + \frac{\Delta t}{2} \dot{T}_0^* \\
T_1^* &= T_1 + \frac{\Delta t}{2} \dot{T}_1 + \frac{\Delta t}{2} \dot{T}_1^* \\
T_2^* &= T_2 + \frac{\Delta t}{2} \dot{T}_2 + \frac{\Delta t}{2} \dot{T}_2^* \\
T_3^* &= T_3 + \frac{\Delta t}{2} \dot{T}_3 + \frac{\Delta t}{2} \dot{T}_3^* \\
T_4^* &= T_4 + \frac{\Delta t}{2} \dot{T}_4 + \frac{\Delta t}{2} \dot{T}_4^*
\end{aligned} \tag{186}$$

Si las expresiones de (186) son sustituidas en el sistema (185), el planteamiento de solución es (Environmental Laboratory, 1995):

$$\begin{bmatrix}
V_0 - A_{00} \cdot \frac{\Delta t}{2} & -A_{01} \cdot \frac{\Delta t}{2} & 0 & 0 & 0 \\
-A_{10} \cdot \frac{\Delta t}{2} & V_1 - A_{11} \cdot \frac{\Delta t}{2} & -A_{12} \cdot \frac{\Delta t}{2} & 0 & 0 \\
0 & -A_{21} \cdot \frac{\Delta t}{2} & V_2 - A_{22} \cdot \frac{\Delta t}{2} & -A_{23} \cdot \frac{\Delta t}{2} & 0 \\
0 & 0 & -A_{32} \cdot \frac{\Delta t}{2} & V_3 - A_{33} \cdot \frac{\Delta t}{2} & -A_{34} \cdot \frac{\Delta t}{2} \\
0 & 0 & 0 & -A_{43} \cdot \frac{\Delta t}{2} & V_4 - A_{44} \cdot \frac{\Delta t}{2}
\end{bmatrix} \cdot \begin{bmatrix} \dot{T}_0^* \\ \dot{T}_1^* \\ \dot{T}_2^* \\ \dot{T}_3^* \\ \dot{T}_4^* \end{bmatrix} = \tag{187}$$

$$= \begin{bmatrix} A_{00} & A_{01} & A_{02} & A_{03} & A_{04} \\ A_{10} & A_{11} & A_{12} & A_{13} & A_{14} \\ A_{20} & A_{21} & A_{22} & A_{23} & A_{24} \\ A_{30} & A_{31} & A_{32} & A_{33} & A_{34} \\ A_{40} & A_{41} & A_{42} & A_{43} & A_{44} \end{bmatrix} \cdot \begin{bmatrix} T_0 + \frac{\Delta t}{2} \cdot \dot{T}_0 \\ T_1 + \frac{\Delta t}{2} \cdot \dot{T}_1 \\ T_2 + \frac{\Delta t}{2} \cdot \dot{T}_2 \\ T_3 + \frac{\Delta t}{2} \cdot \dot{T}_3 \\ T_4 + \frac{\Delta t}{2} \cdot \dot{T}_4 \end{bmatrix} + \begin{bmatrix} P_0 \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{bmatrix}$$

Donde (Environmental Laboratory, 1995):

$V_j$  = Volumen de la capa j.

$\Delta t$  = Intervalo de tiempo (un día).

$\dot{T}_j^* = \dot{T}_j^{i+1}$  = Derivada de la temperatura respecto al tiempo en el siguiente paso de tiempo.

$T_j$  = Temperatura de la capa j en el paso de tiempo presente.

$\dot{T}_j$  = Derivada de la temperatura respecto al tiempo en el paso de tiempo presente.

$P_j$  = Términos independientes de la temperatura de las capas en la expresión (185).

Para proyectar desde el punto presente en el tiempo hacia el próximo todos los términos en el lado derecho de (187) son conocidos debido a que el resultado es una matriz de constantes (Environmental Laboratory, 1995).

---

Cuando se expresan en forma matricial las ecuaciones contienen un coeficiente matricial de forma tridiagonal. Las ecuaciones son integradas cada intervalo computacional mediante un método numérico implícito que consiste en eliminación Gaussiana implementada para minimizar el crecimiento del error y los requerimientos de almacenamiento (Algoritmo de Thomas) (Hydrologic Engineering Center, 1986).

Por lo tanto el sistema (187) se puede expresar así ([http://www.cfd-online.com/Wiki/Tridiagonal\\_matrix\\_algorithm\\_-\\_TDMA\\_\(Thomas\\_algorithm\)](http://www.cfd-online.com/Wiki/Tridiagonal_matrix_algorithm_-_TDMA_(Thomas_algorithm))):

$$\begin{bmatrix} b_0 & c_0 & 0 & 0 & 0 & d_0 \\ a_1 & b_1 & c_1 & 0 & 0 & d_1 \\ 0 & a_2 & b_2 & c_2 & 0 & d_2 \\ 0 & 0 & a_3 & b_3 & c_3 & d_3 \\ 0 & 0 & 0 & a_4 & b_4 & d_4 \end{bmatrix} \quad (188)$$

Y el algoritmo de Thomas se aplica así ([http://www.cfd-online.com/Wiki/Tridiagonal\\_matrix\\_algorithm\\_-\\_TDMA\\_\(Thomas\\_algorithm\)](http://www.cfd-online.com/Wiki/Tridiagonal_matrix_algorithm_-_TDMA_(Thomas_algorithm))):

Desde  $i = 1$  hasta  $n-1$  (cuatro en este ejemplo):

$$b_i = b_i - a_i \cdot \frac{c_{i-1}}{b_{i-1}} \quad (189)$$

$$d_i = d_i - a_i \cdot \frac{d_{i-1}}{b_{i-1}} \quad (190)$$

Para la derivada de la temperatura en el siguiente paso de tiempo  $n - 1$  (para este ejemplo, el correspondiente a 4) ([http://www.cfd-online.com/Wiki/Tridiagonal\\_matrix\\_algorithm\\_-\\_TDMA\\_\(Thomas\\_algorithm\)](http://www.cfd-online.com/Wiki/Tridiagonal_matrix_algorithm_-_TDMA_(Thomas_algorithm))):

$$\dot{T}_4^* = \dot{T}_{n-1}^* = \frac{d_{n-1}}{b_{n-1}} \quad (191)$$

Las otras derivadas de la temperatura en el siguiente paso de tiempo se obtienen así desde  $i = n - 2$  hasta  $i = 0$  ([http://www.cfd-online.com/Wiki/Tridiagonal\\_matrix\\_algorithm\\_-\\_TDMA\\_\(Thomas\\_algorithm\)](http://www.cfd-online.com/Wiki/Tridiagonal_matrix_algorithm_-_TDMA_(Thomas_algorithm))):

$$\dot{T}_i^* = \frac{d_i - c_i \cdot \dot{T}_{i+1}^*}{b_i} \quad (192)$$

---

Finalmente la temperatura en el siguiente paso de tiempo para cada una de las capas se obtiene con las expresiones (186) (Environmental Laboratory, 1995).



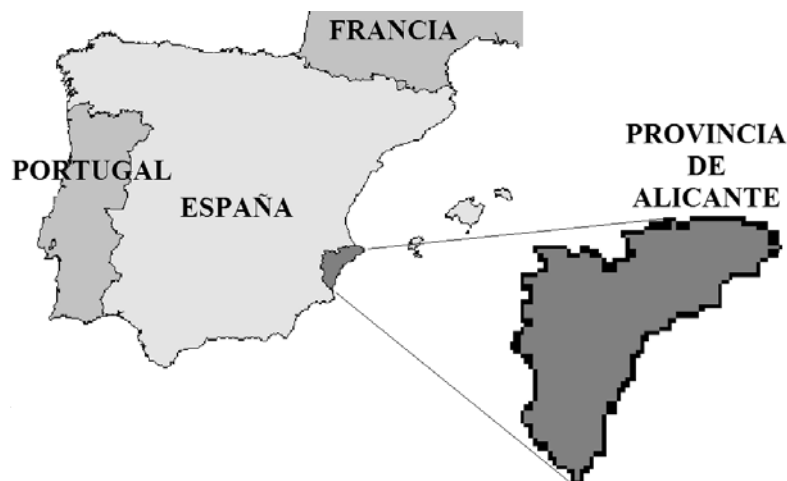


---

## CAPÍTULO 6. DESCRIPCIÓN DEL SITIO DE ESTUDIO

### 6.1 UBICACIÓN GEOGRÁFICA

El Embalse Amadorio está ubicado en la Provincia de Alicante (Figura 33) y está dentro del dominio de la Confederación Hidrográfica del Júcar (CHJ). Pertenece al sistema de explotación de la Marina Baja (Figura 34) ([www.chj.es](http://www.chj.es)).



**FIGURA 33** Ubicación geográfica de la Provincia de Alicante

El Sistema de Explotación Marina Baja se sitúa en la provincia de Alicante, entre los 1.100 m.s.n.m. y el mar Mediterráneo; comprende las cuencas propias de los ríos Algar y Amadorio y las subcuencas litorales comprendidas entre el río Algar y el límite sur del término municipal de Villajoyosa ([www.chj.es](http://www.chj.es)).

Las coordenadas del Embalse Amadorio, en el Sistema de Coordenadas Universal Transversal de Mercator (UTM 30) son (Paredes Arquiola et al., 2006):

X: 738 390 m

Y: 4 272 435 m



**FIGURA 34** Ubicación del Sistema de Explotación Marina Baja (Fuente: <http://www.chj.es/>)

## 6.2 ANTECEDENTES

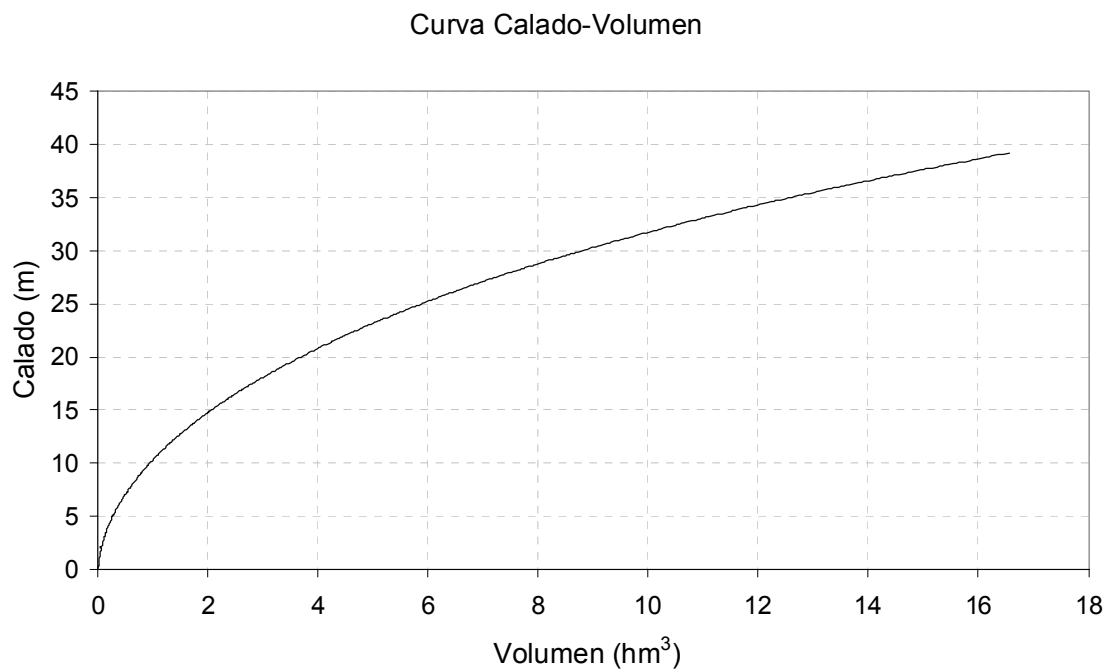
Los datos correspondientes a ubicación de las fuentes de información, geografía, hidrología, clima, características de la cuenca y el embalse (Figura 35), vertidos y demandas, se obtuvieron del “Tomo III. Estudio de la calidad ambiental de las aguas en el embalse de Amadorio. Propuesta de actuaciones”, el cual se incluye en las referencias (Capítulo 11).

La información de geología se obtuvo de [www.geoiberica.com](http://www.geoiberica.com). La información geográfica de la dirección electrónica [www.dip-alicante.es](http://www.dip-alicante.es).

También de la CHJ se obtuvo la información de hidrología, calidad del agua y morfología del embalse.

Son de especial mención los datos de la Sonda multiparamétrica, en el Embalse Amadorio, registrados por la CHJ. Los parámetros que se recolectan a diferentes profundidades son: temperatura, pH, oxígeno disuelto y conductividad.

De la Estación Villajoyosa se obtuvieron los datos meteorológicos.



**FIGURA 35** Curva calado-volumen del Embalse Amadorio

### 6.3 INFORMACIÓN BÁSICA DE LA ZONA

El sistema de explotación Marina Baja comprende las cuencas de los ríos Algar y Amadorio y las subcuencas litorales comprendidas entre el río Algar y el límite sur del término municipal de Villajoyosa, ocupando una superficie de 583 km<sup>2</sup> ([www.chj.es](http://www.chj.es)).

Cuenta con sólo 49 hm<sup>3</sup> de recursos hídricos. En cuanto a los usos consuntivos, alcanzan el 62,64 hm<sup>3</sup> al año de los que 30,49 representan la demanda urbana, 31,15 la agrícola y 1,00 hm<sup>3</sup> la industrial ([www.chj.es](http://www.chj.es)).

La Marina Baja está considerada como un área de alto riesgo en lo referente a las sequías. Esto produce un impacto socioeconómico grave ya que se trata de una de las zonas con mayor extensión de regadío y es un polo turístico ([www.chj.es](http://www.chj.es)).

El período de sequía que hubo en los años 80-85 afectó gravemente a esta zona y hubo restricciones de agua para abastecimiento ([www.chj.es](http://www.chj.es)).

Con el fin de solucionar esta situación se han realizado diversas actuaciones tendientes a extender la oferta de abastecimiento mediante infraestructuras de captación y regulación, aprovechamiento de sobrantes. La batería de actuaciones puesta en juego puede decirse que ha llegado al techo teórico de aprovechamiento de los recursos, que alcanzan un buen nivel de servicio para la situación actual en un año

---

medio, existiendo problemas en caso de sequía ([www.chj.es](http://www.chj.es)).

El déficit previsto para el año horizonte es de unos 6 hm<sup>3</sup>, no existiendo más fuente alternativa que el aumento en la utilización de aguas depuradas generadas por el propio proceso de urbanización en toda la provincia de Alicante, por ejemplo, para los riegos de Aspe y Hondón de las Nieves ([/www.chj.es](http://www.chj.es)).

#### 6.4 DESCRIPCIÓN DE LA CUENCA

El Río Amadorio de 28,4 km de longitud y 205,0 km<sup>2</sup> de cuenca, nace en las sierras de Retamar de la unión de varios barrancos, siendo los más importantes Rogull, Monferri y Escuders ([www.chj.es](http://www.chj.es)).

El Amadorio sigue al sureste, entre las sierras de la Grana y del Aguilar donde recibe por su margen izquierda las aguas procedentes de los barrancos de la Surca y de los Bortolons ([www.chj.es](http://www.chj.es)).

Seguidamente llega a la ciudad de Relleu, de la que recibe sus aguas procedentes de la estación depuradora de aguas residuales (EDAR) ([www.chj.es](http://www.chj.es)).

Antes de alcanzar el embalse de Amadorio el Río Amadorio pasa por la Presa de Relleu, que actualmente no se encuentra en servicio. Aguas abajo de Relleu el río recoge por su margen izquierda las aguas del barranco de la Pedrera y por la derecha del barranco de la Cueva. Finalmente se une al río Sella en el Embalse de Amadorio, en el término municipal de Villajoyosa ([www.chj.es](http://www.chj.es)).

El Río Sella, principal afluente del Amadorio, nace al pie de la Sierra de Retamar en la unión del barranco de El Seguro y el de Tangarina. En esta zona destaca el pico de Aitana con más de 1500 m de altitud ([www.chj.es](http://www.chj.es)).

El río llega a la población de Sella, donde recibe las aguas de la EDAR de esta población. Aguas abajo afluye por la margen izquierda el barranco del Arch y más abajo, también por la izquierda, los barrancos de Charquer y el arroyo Salado. Posteriormente pasa por la población de Orcheta y finalmente confluye con el Río Amadorio y llega al embalse ([www.chj.es](http://www.chj.es)).

A partir del embalse el caudal del río es mínimo. Aparece en este tramo la depuradora de aguas residuales de Villajoyosa (Figura 36) ([www.chj.es](http://www.chj.es)).

Desemboca en el mar Mediterráneo atravesando la ciudad de Villajoyosa. Está previsto el acondicionamiento del cauce del río (www.chj.es).



FIGURA 36 Cuenca del Río Amadorio (Fuente: <http://www.chj.es/>)

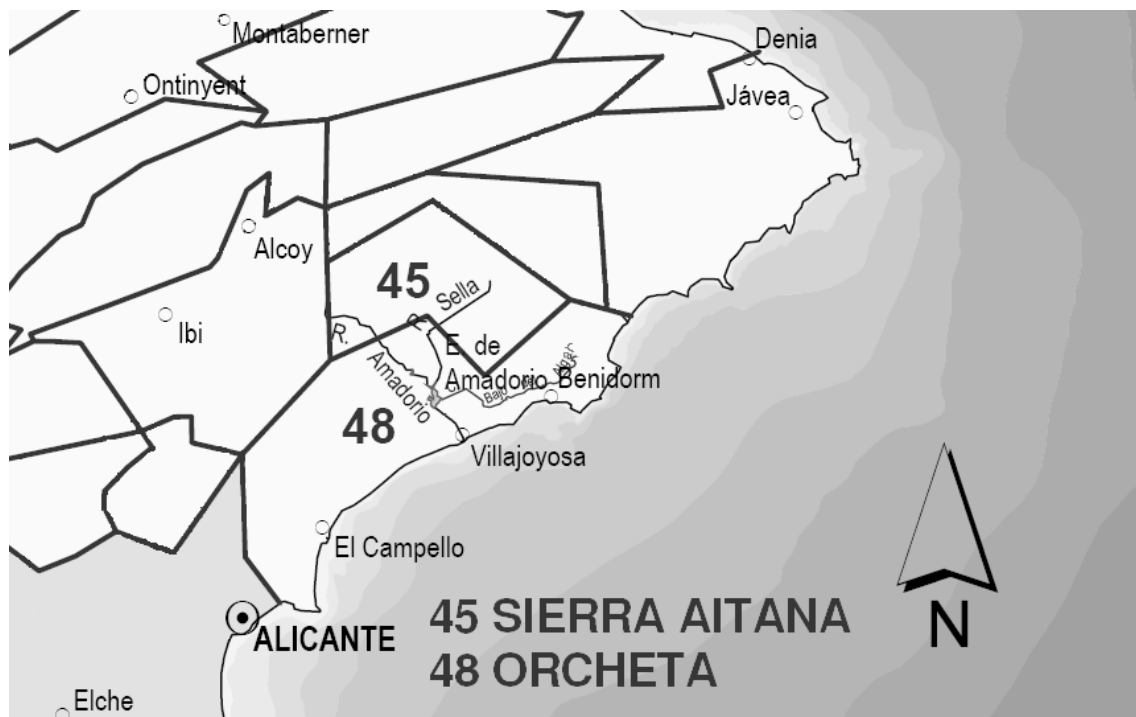


FIGURA 37 Unidades hidrogeológicas asociadas a la Cuenca del Río Amadorio (Fuente: <http://www.chj.es/>)

La cuenca del Río Amadorio se encuentra en una zona en la que la erosión potencial del suelo es alta o muy alta dependiendo de la ubicación, con valores que van desde las 40 t/ha/año a valores por encima de las 100 t/ha/año. La erosión actual en esta cuenca está catalogada como moderada, entre 15 y 40 t/ha/año (Paredes et al., 2006b).

Dentro de la superficie de la cuenca existen dos unidades hidrogeológicas: Orcheta y la de Sierra Aitana (Figura 37) (www.chj.es). La unidad de Orcheta se recarga, en parte, desde el Embalse Amadorio.

## 6.5 DESCRIPCIÓN DEL EMBALSE

El Embalse de propiedad estatal se destina al riego y al abastecimiento, a través del Consorcio de la Marina Baja de las ciudades de Benidorm y Villajoyosa (Paredes et al., 2006b).

La información significativa del embalse se presenta por categorías a continuación:

<p><u>DATOS ADMINISTRATIVOS</u></p> <p>Nombre de la presa: Amadorio</p> <p>Fase según RTSPYE: Explotación</p> <p>Titular de la presa: Estado</p> <p>Proyectista: J. Aura Candela</p> <p>Categoría en función del riesgo potencial: A</p> <p>Fecha de finalización de las obras: 01-01-1957</p> <p><u>USOS DEL EMBALSE</u></p> <p>Usuarios: Consorcio Abast. y San. Marina Baja de Alicante, C.R. de las huertas DFE Villajoyosa</p> <p>Tipos: Abastecimiento, riego</p> <p><u>DATOS DEL EMBALSE</u></p> <p>Superficie del embalse a NMN (ha): 103,00</p> <p>Capacidad a NMN (hm<sup>3</sup>): 16,00</p> <p>Cota del NMN (m): 127,00</p>	<p><u>DATOS GEOGRÁFICOS</u></p> <p>Río en el que se encuentra la presa: Amadorio</p> <p>Municipio: Villajoyosa</p> <p>Cuenca hidrográfica: Júcar</p> <p>Provincia: Alicante</p> <p><u>DATOS HIDROLÓGICOS</u></p> <p>Superficie de la cuenca hidrográfica (km<sup>2</sup>): 205,00</p> <p>Aportación media anual (hm<sup>3</sup>): 6,00</p> <p>Precipitación media anual (mm): 360,00</p> <p>Caudal punta avenida de proyecto (m<sup>3</sup>/s): 400,00</p> <p><u>DATOS DEL DESAGÜE</u></p> <p>Número total de desagües en la presa: 1</p> <p>Capacidad (m<sup>3</sup>/s): 68,00</p>
---	---

<p><u>DATOS DE LA PRESA</u></p> <p>Tipo de presa: Gravedad</p> <p>Cota de coronación (m): 130,00</p> <p>Altura desde cimientos (m): 64,00</p> <p>Longitud de coronación (m): 333,00</p> <p>Cota cimentación (m): 66,00</p> <p>Cota del cauce en la presa (m): 72,00</p> <p>Volumen del cuerpo presa (miles de m<sup>3</sup>): 268,00</p> <p><u>DATOS DEL ALIVIADERO</u></p> <p>Número total de aliviaderos en la presa: 1</p> <p>Tipo: Creager, con compuertas</p> <p>Vanos: Dos, de 15 m cada uno</p> <p>Capacidad a NMN(m<sup>3</sup>/s): 505,6</p> <p>Regulación: Compuertas</p> <p>Compuertas: Taintor estructura abierta, 15 x 4 m, cada una</p> <p>Cota labio: 123,00 m</p>	<p><u>DATOS DE LAS TOMAS</u></p> <p>Toma inferior; margen derecha</p> <p>Diámetro: 0,50 m</p> <p>Número de conductos: 2</p> <p>Cota (m): 90,00</p> <p>Capacidad desagüe ambos conductos (m<sup>3</sup>/s): 2x3</p> <p>Toma superior; margen izquierda</p> <p>Diámetro: 0,50 m</p> <p>Número de conductos: 1</p> <p>Cota (m): 108,50</p> <p>Capacidad desagüe del conducto (m<sup>3</sup>/s): 2,25</p>
---	---

**TABLA 5** Resumen de las características del Embalse Amadorio (Fuente: Paredes Arquiola et al., 2006b)

Los recursos hídricos del Embalse Amadorio proceden del río del mismo nombre y de su afluente el Sella, aunque la mayor aportación tiene su origen en los sobrantes y escorrentías del Río Algar que se recogen junto a las del Río Guadalest en el azud del Mandem. Desde la estación de bombeo que existe en el azud, las aportaciones son transportadas por el Canal bajo del Algar con una capacidad de transporte aproximada de 1,2 m<sup>3</sup>/s, hasta llegar al azud del Río Torres donde son impulsadas junto con las pequeñas aportaciones de esta cuenca, que incluye los vertidos de la EDAR de Finestrat, hasta el Embalse de Amadorio. La capacidad de impulsión en el azud del Río Torres es de 1,5 m<sup>3</sup>/s (Paredes et al., 2006b).

## 6.6 CLIMA

El clima del Sistema de la Marina Baja es mediterráneo semiárido presentándose los valores más bajos, en lo que a precipitación se refiere, en la desembocadura del río

---

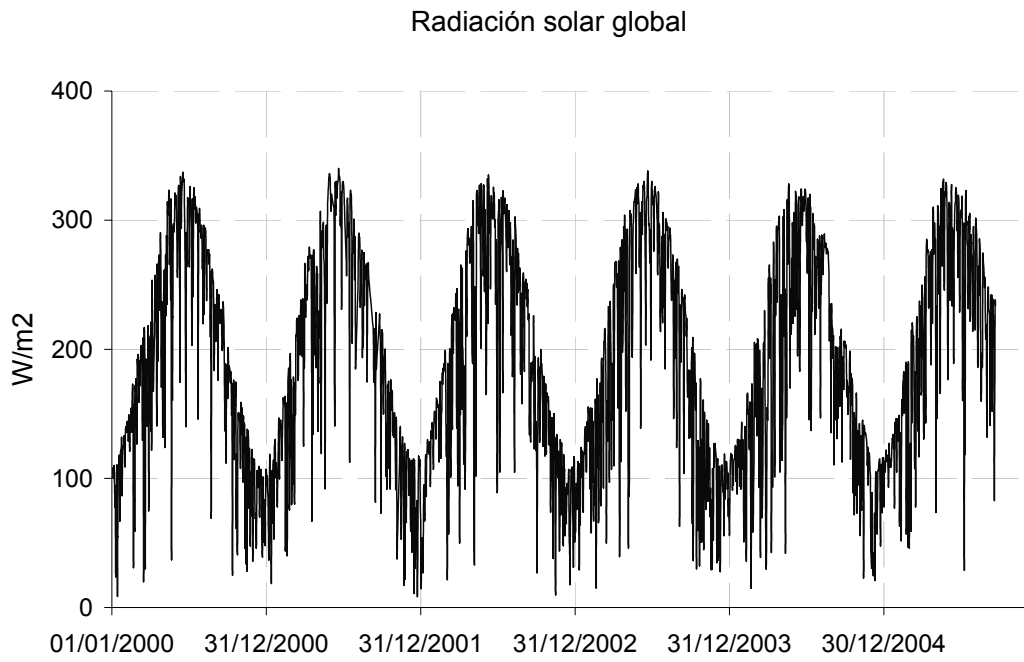
Amadorio. Esta variable aumenta en sentido Norte-Oeste, alcanzándose en Callosa d'Ensarría casi los 600 mm. La precipitación media anual es de 400 mm/año y la temperatura media es de 16°C (Paredes et al., 2006).

La información meteorológica se obtuvo de la estación de Villajoyosa (la estación más cercana al embalse). Los valores de las variables de interés fueron registrados por la estación a escala diaria. Algunas de las variables más significativas medidas por la estación fueron (www. chj.es):

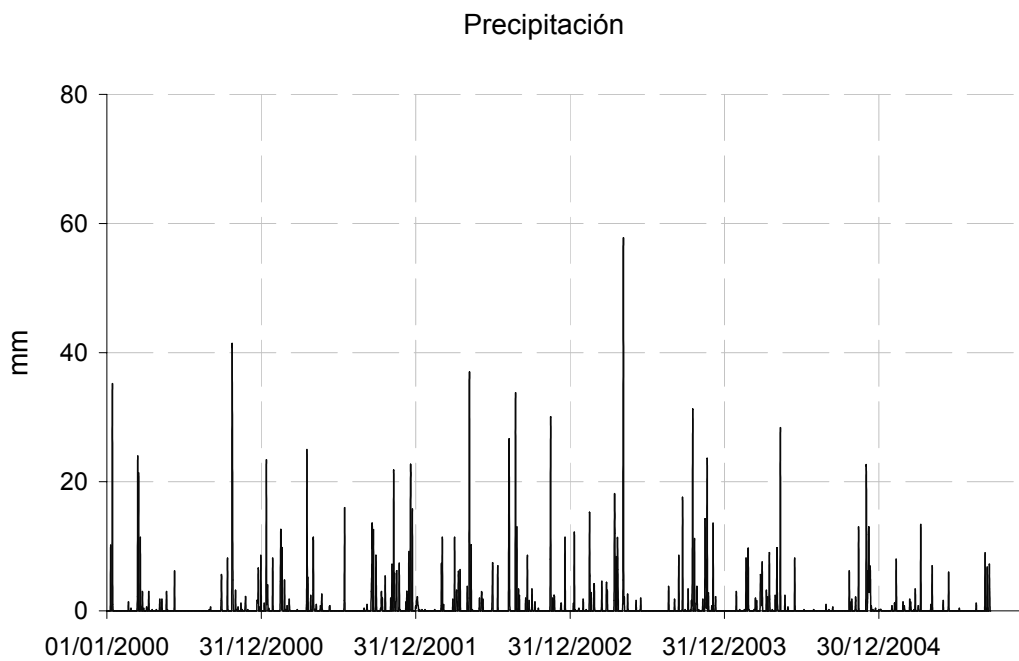
- Velocidad media del viento.
- Dirección media del viento.
- Velocidad máxima del viento.
- Temperatura media del aire.
- Temperatura mínima del aire.
- Temperatura máxima del aire.
- Humedad relativa media del aire.
- Radiación solar global (Figura 38).
- Precipitación diaria total (Figura 39).
- Evapotranspiración de referencia (Figura 40).

Se espera que por la ubicación tan próxima de la estación Villajoyosa al Embalse Amadorio (aproximadamente dos kilómetros) los datos recolectados en la estación sean representativos del embalse. En el modelo mecanicista (sección 5.2) se consideró una constante general llamada Factor por viento ( $\beta$ ) (Hornung, 2002), que es un factor que multiplica a la velocidad del viento; se utiliza para ajustar este parámetro físico en caso de que sus valores sean distintos debido a que la estación meteorológica no se encuentra directamente en el embalse, sino a cierta distancia. Para esta investigación el valor ajustado del Factor por viento fue de 1,10 (sección 8.4), lo cual indica que la velocidad del viento en los cálculos del modelo será incrementada en un 10% para todas las operaciones a efectuar con el modelo mecanicista.

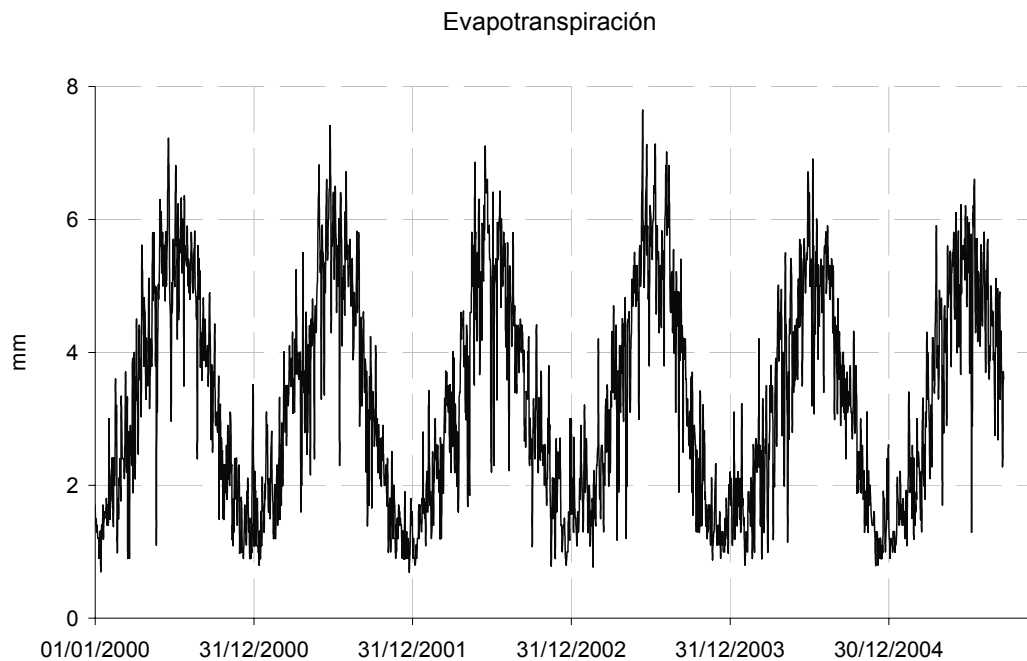




**FIGURA 38** Radiación solar global diaria en Villajoyosa (Fuente: <http://www.chj.es>)



**FIGURA 39** Precipitación diaria en la Estación Villajoyosa (Fuente: <http://www.chj.es>)



**FIGURA 40** Evapotranspiración diaria en la Estación Villajoyosa (Fuente: <http://www.chj.es>)

## 6.7 GEOLOGÍA Y USOS DEL SUELO

El embalse se encuentra dentro del dominio de las cordilleras Béticas. Presentan un estilo estructural propio individualizado por una complicadísima estructura de mantos superpuestos. Esta estructura se debe a que las Cordilleras Béticas se sitúan en una zona de colisión entre la placa Europea y la Africana. Durante la mayor parte de su historia geológica constituyó un dominio de plataforma adyacente al continente, con depósito predominante de materiales carbonatados. La unidad Bética abarca el límite sur de la Provincia de Valencia y gran parte de la de Alicante (Figura 41). Se caracteriza por presentar alineaciones SO-NE, formando tres bandas paralelas, alargadas, denominadas Zona Prebética, Zona Sub-bética y Zona Bética, según la intensidad de la influencia Bética. El embalse está dentro de lo que se ha denominado Prebético Interno (Paredes et al., 2006b).



**FIGURA 41** Zonas del dominio Bético (Fuente: [www.geoiberica.com](http://www.geoiberica.com))

El Prebético interno representa la zona más alejada del continente. Se caracteriza por ser un dominio muy plegado, con frecuentes pliegues cabalgados hacia el Norte o volcados con vergencia norte; a medida que se avanza hacia el Sur los materiales son progresivamente más margosos (medios sedimentarios marinos más profundos) y de mayor potencia (debido a la subsidencia). El Prebético interno comprende las comarcas de L'Alcoia, el Comtat, la Marina Alta, la Marina Baja, el Vinalopó y L'Alacantí (Paredes et al., 2006b).

En cuanto a la litología del suelo destaca principalmente el material calcáreo (Paredes et al., 2006b).

En la cuenca del Río Amadorio, además de las rocas carbonatadas, aparecen también rocas detríticas y evaporíticas. Se encuentran margas, rocas calcáreas y combinaciones de ambas. En las zonas próximas al embalse aparecen nuevamente calcarenitas y también cantos, gravas y limos. En la cuenca del Sella se encuentran combinaciones de arcillas, margas y yesos (Paredes et al., 2006b).

## 6.8 INFORMACIÓN DEMOGRÁFICA

La Comarca de la Marina Baja está integrada por 18 municipios y tiene una población censada de 166.255 habitantes, según los datos del año 2005. En temporada estival y debido al flujo turístico se registra un incremento muy importante de la población, en torno al 200% (Paredes et al., 2006b).

Municipio	Número de habitantes		Incremento 2000 – 2005 (%)
	2000	2005	
L'Alfás de Pí	14.139	16.843	19,1
Altea	15.579	21.154	35,8
Benidorm	54.321	67.492	24,2
Finestrat	2.169	3.797	75,1
La Nucia	8.616	12.573	45,9
Polop	2.136	3.353	57,0
La Villajoyosa/Vila Joiosa	23.704	27.983	18,1
<b>Total</b>	120.664	153.195	27,0

**TABLA 6** Población de los municipios que forman el Consorcio de Abastecimiento de Aguas y Saneamiento “Marina Baja” (Fuente: <http://www.dip-alicante.es/documentación/censo.asp>)

Los municipios más importantes dentro del Consorcio de la Marina Baja, son Benidorm y Villajoyosa, que concentran más del 60 % de la población de este sistema de explotación.

Se puede observar que los municipios que en el año 2000 tenían el menor número de habitantes son los que tuvieron un mayor incremento en su población. En general, el incremento del número de habitantes para todo el Sistema de Explotación de la Marina Baja, que es de un 27,0 %, es significativo y permite pronosticar que la demanda de recursos hídricos para suministro urbano (la que requiere mayor calidad del agua) se incrementará de forma marcada en los próximos años.

## 6.9 VERTIDOS AL EMBALSE Y DEMANDAS

En el Embalse de Amadorio existen tres poblaciones cuyo núcleo se encuentra dentro de su cuenca. Sella y Relleu cuentan, cada una, con una EDAR. La población de Orcheta está conectada a la EDAR de Villajoyosa y por lo tanto su vertido se realiza aguas abajo del embalse, por lo que no afecta a la calidad de éste (Paredes et al., 2006b).

Además de las aguas residuales depuradas que llegan al Embalse de Amadorio de manera directa a través de los cauces a las que son vertidas, aguas depuradas procedentes de otros municipios pueden alcanzar el embalse a través de los diferentes bombeos que se realizan. Es el caso de Bolulla, Tárben y la EDAR “Cuenca del Guadalest” cuyas aguas depuradas son vertidas a cauces que alimentan el bombeo del Mandem. Por otra parte la depuradora de Finestrat vierte una parte de las aguas de la población de forma que pueden llegar al bombeo de Torres, el resto del caudal

---

residual generado en Finestrat es desviado a la depuradora de Benidorm (Paredes et al., 2006b).

El uso que se hace del agua del Embalse Amadorio es principalmente para abastecimiento de la demanda urbana y riego. El aprovechamiento para fines industriales es prácticamente nulo (Paredes et al., 2006b).

La demanda agrícola que se satisface desde el almacenamiento es aquella que no se puede cubrir con la reutilización de las aguas depuradas de las estaciones de Benidorm y Villajoyosa, principalmente. Entre los años 2003 y 2005 la media anual del volumen suministrado para este propósito desde Amadorio es algo superior a 1 hm<sup>3</sup>/año (Paredes et al., 2006b).

Por otra parte, la demanda urbana de los municipios de Benidorm y La Villajoyosa puede suministrarse, de forma parcial, desde el Embalse Amadorio (Paredes et al., 2006b).

Desde el almacenamiento se ha suministrado una magnitud total de 6 hm<sup>3</sup>/año de valor medio (Paredes et al., 2006). Este valor puede aumentar en el futuro como consecuencia del, casi seguro, aumento de la población en los centros urbanos servidos.

DISTRIBUCIÓN MENSUAL DE LA DEMANDA (%)											
OCT	NOV	DIC	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SEP
7	4	0	0	0	9	5	7	11	23	23	11

**TABLA 7** Demanda consuntiva en el ámbito del Plan Hidrológico del Júcar (Fuente: <http://www.chj.es/>)

## 6.10 NORMATIVIDAD APLICABLE AL RECURSO HÍDRICO EN EL EMBALSE

El Plan Hidrológico del Júcar menciona que el tramo correspondiente al Embalse de Amadorio al estar destinado para producción de agua potable tiene como objetivo de calidad el A2, según la nomenclatura del Anexo 1 del Reglamento de la Administración Pública del Agua y de la Planificación Hidrológica (<http://www.chj.es/>).

Para las aguas superficiales susceptibles de ser destinadas al consumo humano clasificadas en el grupo A2 los niveles de calidad para las variables analizadas en este trabajo son:

Parámetro	Límite
Temperatura	25°C
Oxígeno disuelto	50% saturación
pH	5,5 – 9
Conductividad a 20°C	1000 μS/cm

**TABLA 8** Algunos límites de variables de calidad del grupo A2 (Fuente: Real Decreto 1541/1994)

Respecto a la temperatura, las herramientas de predicción podrían permitir conocer, según las condiciones ambientales y la profundidad, a qué nivel se debe efectuar la extracción.

La concentración de saturación para el oxígeno disuelto en agua es la máxima concentración de oxígeno que puede estar disuelto en agua. En general, varios factores ambientales pueden afectar a este valor. Desde el punto de vista de la modelación de la calidad del agua los más importantes son la temperatura, la salinidad y las variaciones de la presión parcial debidas a la elevación del lugar (Chapra, 1997).

La siguiente ecuación puede usarse para establecer la dependencia de la saturación de oxígeno con respecto a la temperatura (Chapra, 1997):

$$\ln o_{sf} = -139,34411 + \frac{1,575701 \times 10^5}{T_a} - \frac{6,642308 \times 10^7}{T_a^2} + \frac{1,243800 \times 10^{10}}{T_a^3} - \frac{8,621949 \times 10^{11}}{T_a^4} \quad (193)$$

Donde  $o_{sf}$  es la concentración de saturación de oxígeno disuelto en agua dulce a una atmósfera (en mg/ℓ) y  $T_a$  es la temperatura absoluta (Chapra, 1997).

La dependencia de la saturación con respecto a la salinidad puede expresarse con la siguiente expresión (Chapra, 1997):

$$\ln o_{ss} = \ln o_{sf} - S \left( 1,7674 \times 10^{-2} - \frac{1,0754 \times 10^1}{T_a} + \frac{2,1407 \times 10^3}{T_a^2} \right) \quad (194)$$

Donde  $o_{ss}$  es la concentración de saturación de oxígeno disuelto en agua salada a una atmósfera (en mg/ℓ) y  $S$  es la salinidad (en g/ℓ) (Chapra, 1997).

Se puede apreciar en las expresiones anteriores que la temperatura del agua tiene una gran influencia sobre la concentración de saturación del oxígeno disuelto.

---

El pH tiene como límites inferior y superior de las observaciones los valores 5,885 y 9,053, respectivamente. Por lo tanto, prácticamente se encuentra siempre cumpliendo los límites 5,5 - 9 de las aguas superficiales del grupo A2.

La conductividad varía con la temperatura y por convención se había establecido una temperatura a la que debían de referirse las medidas de conductividad. Anteriormente la temperatura estándar de referencia era de 25°C; en tiempos más recientes se ha establecido como temperatura estándar la de 20°C. La determinación práctica de la conductividad estándar puede efectuarse de forma instrumental, compensando electrónicamente la temperatura o de forma algebraica (Romero Rojas, 1999).

#### 6.11 DESCRIPCIÓN DE LA SONDA

El régimen de los ríos es muy variable llegando algunos a tener caudales muy bajos en épocas de estiaje e incluso a no tener caudal, poniendo en peligro el uso del recurso en una parte del año (probablemente cuando es más necesario). Este panorama conduce obligadamente a contar con medios de almacenamiento, los embalses, que cumplen las funciones de laminar caudales en periodos de avenidas y la de almacenar agua para periodos durante el año en que los caudales existentes no podrían satisfacer la demanda (Prats i Vime et al., 2004).

Frecuentemente, las fuentes que alimentan de agua a los embalses no garantizan la calidad requerida para los distintos usos solicitados, e inclusive, los caudales que ingresan al embalse pueden estar cargados de nutrientes y otros contaminantes (Prats i Vime et al., 2004).

Debido a lo anteriormente expresado se necesita dar seguimiento a la calidad del agua almacenada en los embalses por medio de diversos mecanismos (Prats i Vime et al., 2004).

El seguimiento real de la calidad del agua embalsada presenta problemas que en ocasiones son de difícil solución. Las campañas de muestreo, de hecho, se efectúan en intervalos de tiempo muy amplios por lo que deben complementarse con acciones adicionales para disponer de un verdadero control de la calidad y tomar acciones oportunas (Prats i Vime et al., 2004).

Para garantizar en todo momento la calidad para los usos requeridos es necesario detectar ciclos que se desarrollan en el embalse, algunos de ellos de duración muy corta en comparación con el tiempo transcurrido entre campañas de muestreo (como

---

por ejemplo, el ciclo de aumento de la cantidad de algas que generalmente ocurre en unos pocos días) (Prats i Vime et al., 2004).

Si se pretende detectar estos ciclos se hace patente que la técnica clásica de muestreo no es la solución más adecuada para estos retos, a menos que se efectúe cuando los fenómenos se producen (algo poco probable) o que la campaña de muestreo sea permanente (lo que implica utilizar gran cantidad de recursos económicos, técnicos y humanos, lo que en la práctica no es factible) (Prats i Vime et al., 2004).

La solución al problema apunta a un tipo de actuación permanente y constante y por tanto automática que informe de las observaciones en tiempo real y que tenga la capacidad de detectar los parámetros que evidencian la aparición de estos ciclos (Prats i Vime et al., 2004).

El seguimiento del estado de la calidad del agua en un embalse se puede abordar desde dos enfoques: i) medir los parámetros directos determinantes de la calidad o ii) los parámetros indirectos que caracterizan cada situación (Prats i Vime et al., 2004).

Para datos continuos y en tiempo real es más adecuado el segundo enfoque (por ser más sencillo y económico, generalmente, obtener los parámetros indirectos que configuran cada ciclo) (Prats i Vime et al., 2004).

Algunos de los parámetros que pueden ayudar a establecer la situación de la calidad del agua embalsada y el riesgo de cambio brusco de esta calidad pueden ser: la temperatura, el oxígeno disuelto y el potencial redox, todos ellos medidos a diferentes profundidades (Prats i Vime et al., 2004).

La meteorología puede causar cambios en la calidad del agua por lo que interesa su registro para efectuar determinadas predicciones e incluso para modelizar el comportamiento del agua embalsada. La información meteorológica más importante es: temperatura, precipitación atmosférica, radiación solar, intensidad y dirección del viento, así como evaporación (Deas et al., 2000; Prats i Vime et al., 2004).

La meteorología presente es determinante para la continuidad en el tiempo de un determinado estado de calidad. Las variaciones de la temperatura del agua en superficie generadas por los cambios en la temperatura del aire o por la lluvia pueden acelerar las corrientes de convección y provocar la rotura de la termoclina, provocando, por ejemplo, la aparición rápida de fitoplancton (Prats i Vime et al., 2004).





**FIGURA 42** Trabajos de instalación de equipos para la sonda multiparamétrica autoposicionable (Fuente: [www.chj.es](http://www.chj.es))

Para efectuar la medición de los parámetros que auxiliarán en el establecimiento del estado de calidad del agua en el embalse se puede utilizar un equipo fijo. Este se monta en un punto crítico del embalse, por ejemplo, en una toma de agua que será potabilizada para suministro a una población (Prats i Vime et al., 2004).

En el año 2003 se efectuó la instalación de una sonda multiparamétrica autoposicionable en el Embalse Amadorio (Figura 42). Esta sonda desarrolla el seguimiento continuo de los parámetros básicos de calidad del agua (como son temperatura, pH, conductividad y oxígeno disuelto) a distintos niveles del embalse, lo cual permite elaborar perfiles verticales de estos constituyentes. Los perfiles verticales pueden servir como auxiliares en el seguimiento ecológico del embalse, estudiar los procesos de mezcla, determinar la estructura de estratificación, etc. ([www.chj.es](http://www.chj.es)).

El sistema está constituido por una unidad de medida fija, medios de transmisión de datos y un centro de recepción de información (Prats i Vime et al., 2004).

La sonda se montó en una plataforma situada en el paramento del embalse, aprovechando la infraestructura de una caseta de accionamiento de válvula de toma (Prats i Vime et al., 2004).



**FIGURA 43** Vista de la caseta donde opera la sonda multiparamétrica (Fuente: [www.chj.es](http://www.chj.es))

La sonda multiparamétrica (Figura 43) aporta una información muy útil para la gestión de la calidad del agua embalsada, en tiempo real y con unos medios relativamente sencillos, con una tasa baja de mantenimiento, debido al sistema automático instalado (Prats i Vime et al., 2004).

Cabe señalar que los avances siguen adelante, ya que existe la posibilidad actualmente de disponer de sondas capaces de medir más parámetros (potencial redox, clorofila a, cianobacterias, etc.), para agregar información a la actualmente disponible y, con esto, mejorar la gestión sobre todo en el aspecto temporal.

---

## CAPÍTULO 7. APLICACIÓN DE LOS MÉTODOS MULTIVARIANTES

### 7.1 INTRODUCCIÓN

Antes de aplicar cualquier método estadístico un paso fundamental es el establecimiento del intervalo de tiempo que se usará para reportar las observaciones, tanto las meteorológicas, las hidrológicas, así como las de calidad del agua del embalse.

Los parámetros meteorológicos fueron obtenidos de la estación más cercana al Embalse Amadorio, la estación Villajoyosa. Las magnitudes medidas fueron las siguientes:

- V, Velocidad media del viento (Km/h).
- DV, Dirección media del viento.
- Vx, Velocidad máxima del viento (Km/h).
- T, Temperatura media (°C).
- Tn, Temperatura mínima (°C).
- Tx, Temperatura máxima del aire (°C).
- Horas frío, Número de horas por debajo de 7 °C.
- HR, Humedad relativa media del aire (%).
- Rad, Radiación solar global (W/m<sup>2</sup>).
- P, Precipitación total (mm).
- ETo, Evapotranspiración de referencia, calculada mediante el método de Penman-Monteith, salvo en el caso de incidencias con el anemómetro donde se utiliza el método de Hargreaves (mm).

Se dispone de estas medidas a escala de tiempo diaria (intervalo de 24 horas representativo del día de referencia).

El intervalo de tiempo del que se cuenta con información va del 01/01/2004 al 20/09/2005. Se observa que en el periodo mencionado la información está completa para todos los parámetros meteorológicos, sin existir interrupciones o datos faltantes.

Las magnitudes hidrológicas y de calidad del agua, calculadas y observadas, fueron las que a continuación se enumeran (Paredes et al., 2006b):

- Volumen de explotación (m<sup>3</sup>).

- 
- Caudal de entrada al embalse (m<sup>3</sup>/día).
  - Caudal de salida del embalse (m<sup>3</sup>/día).
  - Temperatura del agua que entra al embalse (°C).
  - Filtración (m<sup>3</sup>/día).
  - SS, sólidos suspendidos en el caudal de entrada (mg/l).
  - DBO<sub>5</sub>, demanda bioquímica de oxígeno de 5 días en el caudal de entrada (mg/l).
  - O<sub>2</sub>, oxígeno disuelto en el caudal de entrada (mg/l).
  - Norg, nitrógeno orgánico en el caudal de entrada (mg/l).
  - NH<sub>4</sub>, nitrógeno amoniacal en el caudal de entrada (mg/l).
  - NO<sub>3</sub>, nitratos en el caudal de entrada (mg/l).
  - A, fitoplancton en el caudal de entrada (mg/l).
  - Porg, fósforo orgánico en el caudal de entrada (mg/l).
  - Pinorg, fósforo inorgánico en el caudal de entrada (mg/l).

Todas las cantidades anteriores están expresadas a escala diaria (lapso de tiempo de 24 horas).

Los métodos que pueden ser usados para medir las variables químicas fueron los siguientes (AENOR, 2005):

- Para los sólidos suspendidos se empleó el método de filtración por filtro de fibra de vidrio.
- La demanda bioquímica de oxígeno de 5 días se puede determinar o por el Método de dilución y siembra con adición de alil tiourea o por el Método para muestras no diluidas.
- El oxígeno disuelto puede obtenerse por el Método yodométrico o por el Método electroquímico.
- Para el nitrógeno orgánico se usó la determinación de nitrógeno Kjeldahl por el Método de mineralización con selenio.
- La cantidad de nitrógeno amoniacal fue obtenida por análisis de flujo (CFA y FIA) y detección espectrométrica.
- Los nitratos fueron determinados con el análisis por inyección de flujo (CFA y FIA) con detección espectrométrica.
- Para el fitoplancton se empleó el recuento del mismo por microscopía invertida.
- El fósforo orgánico y el inorgánico por medio de las directrices indicadas en la Norma Española para determinación del fósforo mediante el Método

---

espectrométrico de molibdato de amonio.

Para el periodo de tiempo del 01/01/2004 al 20/09/2005, los valores de los distintos parámetros enumerados están completos, sin existir valores faltantes.

Es necesario hacer notar que las mediciones de los parámetros de calidad del agua que ingresa al embalse fueron escasamente monitoreados, por lo que con el fin de efectuar una modelación realizada en el “Tomo III. Estudio de la calidad ambiental de las aguas en el Embalse de Amadorio. Propuesta de actuaciones” basada en ecuaciones en diferencias finitas para dos capas (epilimnion e hipolimnion), se supusieron valores constantes para los parámetros: Temperatura del agua que entra al embalse, SS, DBO<sub>5</sub>, O<sub>2</sub>, Norg, NH<sub>4</sub>, NO<sub>3</sub>, A, Porg y Pinorg.

Una parte de la propuesta de trabajo de tesis está basada en algunos métodos multivariantes, con el objetivo de caracterizar y modelar estados de calidad del agua en el embalse. El fundamento para diferenciar entre estos estados y modelarlos es que cada una de las variables consideradas presenten variaciones, pues los cambios en algunas de estas variables de calidad están correlacionados con las modificaciones en otras. Por lo tanto, una variable que posea un valor constante no será útil para distinguir (diferenciar) entre fases de calidad del agua en el embalse ni para modelar, pues al ser su valor constante, no presenta correlación con los cambios que se presentan en otras variables.

Debido a lo anterior los parámetros Temperatura del agua que entra al embalse, SS, DBO<sub>5</sub>, O<sub>2</sub>, Norg, NH<sub>4</sub>, NO<sub>3</sub>, A, Porg y Pinorg, no serán incluidos en el desarrollo de los métodos multivariantes empleados.

La sonda multiparamétrica midió en el interior del embalse los siguientes valores de parámetros de calidad del agua a distintas profundidades:

- Temperatura (°C).
- Oxígeno disuelto (mg/l).
- pH.
- Conductividad (μS/cm).

Respecto al intervalo de tiempo en el que se efectuaron los registros de la sonda se observa que fueron efectuados en distintos formatos: en ocasiones sólo se obtuvo un perfil al día (en diferentes horarios entre días), en otras se obtuvo el perfil existente

---

durante varios momentos del día.

Debido a que en el procesamiento se debe efectuar para registros (conjuntos de observaciones) que estén aplicados al mismo sistema y representando la misma unidad de tiempo y como los datos meteorológicos e hidrológicos están expresados para una escala de tiempo diaria (24 horas) como unidad temporal, se tomará como unidad temporal para todos los análisis a realizar 24 horas (escala diaria). Para el caso de la sonda multiparamétrica, si durante un día específico sólo se obtuvo un perfil se tomará como representativo del día; por otra parte, si se obtuvieron varios perfiles de concentraciones en el mismo día se obtendrá la media aritmética de ellos para cada profundidad correspondiente. Esa media aritmética será considerada como representativa para ese día específico. Es necesario mencionar que para contar con información de mayor calidad es necesario que los registros que se tomen con la sonda multiparamétrica en el futuro sean recolectados a intervalos regulares de tiempo y todos los días en los mismos horarios, pues esto permitirá que los valores medios del día sean más representativos de las 24 horas que se están analizando. Asimismo, esto permitiría que la resolución temporal del modelo pueda ser mejor (lapsos de tres o cuatro horas, cada hora, etc.).

Es necesario mencionar los días concretos para los que se cuenta con valores medidos en la sonda. Los días que no se muestran en la Tabla (9), entre los periodos que se muestran con información, son en los cuales la sonda no registró información por algún motivo.

En el periodo en el que se efectúa el estudio (01/01/2004 a 20/09/2005) se tienen 153 días con información para el año 2004 y 233 días con información en el año 2005.

De los días con información disponible, se tiene que en el año 2004 los días con información completa para todas las variables y dentro de los rangos normales que se espera encontrar son 138 días. Los días con información faltante o con valores no verosímiles son 15, para el mismo año 2004.

Para el año 2005 se tiene que todos los días en lo que la sonda sí registró información, poseen información verosímil y sin observarse valores perdidos.

Una vez que se conoce la información con la que se cuenta (variables medidas, su disponibilidad y calidad), se procede a la aplicación de los métodos multivariantes convenientes. En principio, para los días con información válida se estructuran los

datos de calidad del agua con respecto a la profundidad (medida desde la superficie del embalse) y los datos meteorológicos e hidrológicos se repiten para cada profundidad registrada el mismo día. Las matrices estructuradas así para cada día son concatenadas una encima de la otra, generándose una nueva matriz de datos aumentada.

Periodos de tiempo en los que existen registros	
Periodo	Observaciones
01/01/2004	Registros de oxígeno, temperatura y conductividad no válidos
31/01/2004	Registro de oxígeno, temperatura y conductividad no válidos; medición sólo en superficie
11/02/2004 – 13/02/2004	Registro de oxígeno y conductividad no válidos
18/02/2004 – 19/02/2004	Registro de oxígeno no válido
24/02/2004 – 02/03/2004	Registro de oxígeno no válido
31/03/2004 – 04/04/2004	
05/07/2004 – 16/08/2004	
28/09/2004 – 07/10/2004	
09/10/2004 – 07/12/2004	
09/12/2004 – 26/12/2004	
28/12/2004 – 29/12/2004	
05/01/2005 – 18/02/2005	
21/02/2005 – 16/03/2005	
23/03/2005	Para algunas horas del día la información está incompleta
24/03/2005 – 25/03/2005	
26/03/2005	Para la hora 0 sólo se midió a una profundidad de 5,83
27/03/2005 – 30/03/2005	
06/04/2005 – 09/04/2005	
13/04/2005	
19/04/2005 – 09/07/2005	
10/07/2005 – 11/07/2005	Sólo se tienen registros de algunas profundidades
15/07/2005 – 19/09/2005	

**TABLA 9** Disponibilidad de información en la sonda

Se iniciará realizando estadística descriptiva de los datos. También se revisará la normalidad univariante de cada variable. Se mencionarán posibles transformaciones de los datos para hacerlos más normales y homocedásticos.

---

Después se efectuará un análisis de componentes principales (PCA) con el propósito de conocer si existen componentes que agrupen a variables o que expliquen el comportamiento de conjuntos de ellas.

También se efectuará el análisis factorial (FA) de los datos para obtener las puntuaciones factoriales de cada una de las observaciones, ya que al estar los factores o variables latentes incorrelacionados entre sí las puntuaciones constituyen el punto de partida de otros métodos (como los de regresión múltiple o los métodos de regresión más avanzados).

El análisis cluster (CA) se aplicará a los datos para formar grupos de datos que sean más homogéneos entre sí y, con esto, auxiliar a que los análisis cumplan más adecuadamente con la suposición de linealidad. Además, se efectuará un análisis cluster de variables para conocer si existen conglomerados de variables que muestren algunas similitudes de comportamiento entre ellas.

A continuación se usará el análisis discriminante (DA), el cual servirá como una confirmación de la existencia de los grupos obtenidos con el análisis cluster.

El modelo lineal de regresión múltiple será aplicado con propósitos exploratorios para conocer la forma posible de las expresiones de regresión y contar con una primera aproximación de la correlación.

El análisis de correlación canónica (CCA) servirá para establecer las relaciones entre variables de manera numérica. Es un método de regresión más avanzado que el anteriormente mencionado.

Por último, se efectuarán aplicaciones del PLSR para obtener expresiones para las variables de salida. Para tener un panorama de los resultados conseguidos se harán comparaciones con los métodos de regresión usados anteriormente y se obtendrán conclusiones.

## 7.2 SOFTWARE

Los paquetes estadísticos proporcionan capacidades extensas para el análisis de datos. Muchos paquetes de análisis estadísticos generales proveen capacidades para la gestión de bases de datos, gráficas y generación de reportes junto con una variedad de cálculos estadísticos, que incluyen estadística descriptiva, varios tipos de pruebas estadísticas, análisis de regresión y correlación y análisis de series temporales. Varios



---

programas estadísticos de propósito general son mostrados en la Tabla (10) (Wurbs, 1994).

Debido a que los datos estadísticos se pueden tratar como matrices se pueden utilizar los paquetes MATLAB y GAUSS, excelentes programas para las operaciones con matrices, por lo que son recomendables para elaborar programas propios y probar análisis nuevos, no incluidos en los paquetes tradicionales. Su gran ventaja es la flexibilidad y tienen el inconveniente de que son menos automáticos para análisis convencionales (Peña, 2002).

Programas de análisis estadístico
Minitab
BMDP
S-PLUS
SAS
SPSS
STATA
StatGraphics

**TABLA 10** Paquetes de análisis estadístico para propósitos en general (Fuente: Wurbs, R. A. 1994; Peña, D., 2002)

Para aplicar el PLSR se tienen diversos programas. Algunos de ellos son:

Statgraphics, que aparte de ser un paquete estadístico de propósito general incluye el ajuste de un modelo lineal por el método PLSR (StatPoint Inc., 2005).

El software libre ParLeS sirve para la modelación multivariante y la predicción. Una de sus funciones es precisamente el desarrollo del PLSR con distintas opciones para el tratamiento de los datos (Viscarra Rossel, 2007).

El SPSS, programa estadístico de propósito general, incluye en su versión 16.0 (no disponible actualmente en países de habla hispana) este método que requiere contar con la Extensión de Programación de SPSS ([www.spss.com](http://www.spss.com), 2008).

Desde la versión 6.11, SAS (paquete estadístico) incluye este método con el procedimiento PROC PLS, así como el CALIS para modelos más complejos ([statisticssolutions.com](http://statisticssolutions.com)).

TOMCAT es un toolbox para MatLab. La interfaz gráfica y sus rutinas son disponibles libremente, están programadas en MatLab 6.5. Entre los métodos que implementa se

---

encuentra el de Mínimos Cuadrados Parciales (PLSR) (Daszykowski et al., 2007).

LIBRA es una librería para MatLab y desarrolla métodos estadísticos robustos. Estos programas son proporcionados de manera libre y para uso no comercial solamente. Algunas de las funciones requieren el toolbox de Estadística de MatLab. Uno de los métodos que desarrolla es el de Mínimos Cuadrados Parciales (PLSR), que en esta librería tiene su denominación como `csimpls`, el cual desarrolla el PLSR clásico. También esta librería dispone del proceso `rsimpls`, en donde la `r` al inicio del nombre significa que se está aplicando una versión robusta del PLSR (versión robusta se refiere a que esa metodología se ve poco afectada por valores anómalos) (Verboven et al., 2005).

Unscrambler es una herramienta de análisis estadístico, en la cual se trata el análisis multivariante. Incluye dentro de sus funciones el método PLSR ([www.camo.com/](http://www.camo.com/)).

Se comenzará utilizando los programas Excel, Statgraphics y SPSS para el desarrollo de la mayoría de los métodos multivariantes.

### 7.3 EXPLORACIÓN INICIAL DE LOS DATOS

Debe tenerse siempre en cuenta que se está tratando con tres grupos o familias de datos: meteorológicos, hidrológicos y de calidad del agua; que para esta investigación serán expresados a escala diaria. Uno de los objetivos más importantes que se persigue lograr es que al tener como información de partida los grupos de datos meteorológicos e hidrológicos se logren efectuar predicciones adecuadas de las variables de calidad del agua.

Los días para los cuales se tiene información están incluidos en los años 2004 y 2005. La información para el año 2004 es de 138 días con registros válidos y 233 para el 2005. El total es de 371 días.

Por lo general en los estudios de modelación se considera una parte de la información para efectuar la calibración del modelo y la parte restante para la validación (como será para el caso que se analiza). El criterio más común es el que toma el conjunto de datos entero y lo divide en dos subconjuntos de igual tamaño, llamándolos conjuntos de datos de calibración y validación.

Entonces el conjunto de calibración será el de los primeros 186 días (por orden cronológico) y el conjunto de validación el de los 185 días restantes.

Por lo tanto la distribución de los días para la calibración y la validación, será, como propuesta de inicio la siguiente:

Proceso	Periodo de tiempo
CALIBRACIÓN (186 días)	31/03/2004 – 04/04/2004
	05/07/2004 – 16/08/2004
	28/09/2004 – 07/10/2004
	09/10/2004 – 07/12/2004
	09/12/2004 – 26/12/2004
	28/12/2004 – 29/12/2004
	05/01/2005 – 18/02/2005
	21/02/2005 – 23/02/2005
VALIDACIÓN (185 días)	24/02/2005 – 16/03/2005
	23/03/2005 – 26/03/2005
	27/03/2005 – 30/03/2005
	06/04/2005 – 09/04/2005
	13/04/2005
	19/04/2005 – 11/07/2005
	15/07/2005 – 19/09/2005

**TABLA 11** Observaciones consideradas para los procesos de calibración y validación

Con el propósito de lograr una mayor claridad, las variables que serán parte del análisis se clasificarán por su procedencia en tres categorías: meteorológicas, hidrológicas o de estado del embalse y de calidad del agua.

Para involucrar de alguna manera en el análisis la temporada del año se agregará dentro de las variables una que represente el día del año en el que se efectuó la observación. Tal variable se denomina día juliano. Esta variable también se utiliza en cálculos relacionados con la radiación solar en programas como el QUAL2K para representar la presencia de ciclos periódicos.

Adicionalmente, los valores de las variables de calidad del agua dependerán de la profundidad, por lo que la profundidad a la cual se presenta la observación será también una variable en el análisis.

En el Anexo 1 se proporcionan los primeros valores de las variables para mostrar como se estructuraron los datos antes de iniciar la aplicación de los métodos estadísticos, iniciando con el primer día del periodo en estudio, el 31/03/2004. Sólo se considerarán en el análisis los días en lo que existen registros válidos en la sonda del Embalse Amadorio.

Como planteamiento inicial las variables que serán analizadas son:

Clasificación de la variable	Variable y sus unidades de medida
Temporal	DJ, Día juliano
Meteorológicas	V, Velocidad media del viento (Km/h)
	DV, Dirección media del viento (°)
	Vx, Velocidad máxima del viento (Km/h)
	T, Temperatura media (°C)
	Tn, Temperatura mínima (°C)
	Tx, Temperatura máxima del aire (°C)
	Horas frío, Hf, Número de horas por debajo de 7 °C
	HR, Humedad relativa media del aire (%)
	Rad, Radiación solar global (W/m <sup>2</sup> )
	P, Precipitación total (mm)
	ETo, Evapotranspiración de referencia (mm)
Hidrológicas	VolExp, Volumen de explotación (m <sup>3</sup> )
	Qent, Caudal de entrada al embalse (m <sup>3</sup> /día)
	Qsal, Caudal de salida del embalse (m <sup>3</sup> /día)
	Filt, Filtración (m <sup>3</sup> /día)
	Profundidad (m)
Calidad del agua	Ta, Temperatura del agua (°C)
	Odis, Oxígeno disuelto (mg/l)
	pH
	Con, Conductividad (μS/cm)

**TABLA 12** Variables con las que se inicia el análisis

Es adecuado señalar en que forma pueden afectar las variables predictivas a las variables respuesta (variables a estimar). La temperatura del agua se ve influenciada por las siguientes variables predictivas:

- Las velocidades media y máxima del viento, pues influyen en los procesos de mezclado (Chapra, 1997) y en su caso rompimiento de la estratificación, incidiendo sobre la distribución de la energía termal en el embalse a todos niveles.
- Las temperaturas media, máxima y mínima del aire, pues al estar en contacto con la superficie del embalse operan directamente sobre el intercambio de energía calorífica entre el embalse y la atmósfera (Hydrologic Engineering Center, 1986; Martin et al., 1999).
- Las horas de frío pueden llegar a influenciar en la temperatura del embalse.

- 
- La radiación solar global es un aporte de energía directo para el almacenamiento (Environmental Laboratory, 1995).
  - La evapotranspiración tiene influencia en los procesos de intercambio de calor (Hydrologic Engineering Center, 1986; Chapra, 1997; Martin et al., 1999).
  - El volumen del embalse proporciona una medida de la inercia térmica que posee el mismo (dificultad para que su temperatura sea modificada).
  - La precipitación, caudales de entrada y salida, además de la filtración constituyen entradas y egresos de agua en el almacenamiento que implican también ganancias o pérdidas de energía calorífica (Environmental Laboratory, 1995; Martin et al., 1999).
  - La profundidad es una variable de primordial importancia en la distribución de la temperatura, pues a menor profundidad el agua del embalse se ve más afectada por la temperatura de la atmósfera adyacente (Hydrologic Engineering Center, 1986; Chapra, 1997, Martin et al., 1999). Asimismo, la radiación de onda corta y de onda larga se absorben hasta determinadas profundidades (Centre for Water Research, 2003).

El oxígeno disuelto es afectado por las siguientes variables predictivas:

- La velocidad media y máxima del viento producen efectos de mezclado que distribuyen el oxígeno a través de la profundidad del almacenamiento (Chapra, 1997).
- Las temperaturas media, máxima y mínima del aire afectan a la temperatura del agua en el embalse, lo que a su vez incide sobre la saturación de oxígeno del agua (Chapra, 1997).
- La precipitación, caudales de entrada y salida, aparte de la filtración, constituyen elementos del balance de oxígeno en el embalse (Environmental Laboratory, 1995).
- La profundidad influye en el oxígeno disuelto, debido a que a mayor profundidad el agua tiene generalmente más dificultad para acceder al oxígeno proveniente del intercambio con la atmósfera (Chapra, 1997). Además las entradas y salidas de oxígeno del sistema representadas por la precipitación, caudales de entrada y salida, así como la filtración suceden a profundidades determinadas (Environmental Laboratory, 1995).

El pH tiene relación con las siguientes variables predictivas:

- 
- La velocidad media y máxima del viento, debido a que inducen el mezclado en el almacenamiento, por lo que pueden ser un factor que ayude a que las concentraciones en perfil sean más homogéneas (Chapra, 1997).
  - La precipitación, los caudales de entradas y salidas, además de la filtración son ingresos y egresos de líquido que puede modificar el valor del pH existente previamente.
  - La profundidad influye, debido a que la variación del pH puede verse afectada por esta variable, por ejemplo debido a los volúmenes que ingresan o egresan del embalse Amadorio a distintas profundidades.

La conductividad se ve influenciada por las siguientes variables predictivas:

- La velocidad media y la máxima del viento produce efectos de mezclado en el embalse, tendiendo a uniformizar las concentraciones principalmente en las partes del almacenamiento más expuestas a su acción (Chapra, 1997).
- Las temperaturas media, mínima y máxima del aire afectan directamente a la temperatura del embalse, haciendo que se puedan presentar fenómenos como el rompimiento de la estratificación, que modifican los valores de conductividad (Chapra, 1997; Martin et al., 1999).
- Las horas de frío y la evapotranspiración afectan a la conductividad debido a que en la temporada del año en que se tienen registros de horas de frío, las temperaturas son más bajas y se puede tener la seguridad de que el mezclado en el embalse es prácticamente completo (Wetzel, 2001). Por otra parte, la presencia de valores altos de evapotranspiración genera que la conductividad se incremente, pues existe un menor volumen de agua y los mismos elementos disueltos.
- La radiación solar global es uno de los factores que al variar puede afectar a los periodos de estratificación y mezcla completa en el almacenamiento (Chapra, 1997; Martin et al., 1999).
- La precipitación, los caudales de entradas y salidas, junto con la filtración constituyen entradas y salidas de volúmenes con sus concentraciones particulares, que pueden modificar la conductividad en el embalse.
- El volumen de explotación también representa un factor en la concentración de la conductividad, pues al ser mayor el volumen almacenado la variación en la conductividad será menor ante una determinada entrada o salida.

- 
- La profundidad también ayuda a determinar la concentración de conductividad, pues la densidad del agua es aproximadamente proporcional a la concentración de conductividad (Chapra, 1997) y en algunas épocas del año se tendrá que a mayor profundidad la conductividad tendrá un valor más alto.

Respecto a la variable Dirección media del viento, esta puede tener influencia debido a la orientación del almacenamiento en los procesos de evaporación y mezclado, pero al ser la orientación una característica particular de cada embalse, las relaciones que se obtengan serán de aplicación exclusiva al Embalse Amadorio.

La información inicial de partida está expresada a escala diaria. Las variables meteorológicas e hidrológicas estaban directamente expresadas en esta escala en la información original. En cambio, para las variables de calidad del agua medidas con la sonda multiparamétrica se tenían días con registros efectuados a diferentes horas y con distinto espaciamiento entre ellos, por lo que se tomó como criterio efectuar el promedio aritmético de los distintos registros existentes para una misma profundidad y día y considerarlos como representativos del día correspondiente.

Un paso preliminar al efectuar un estudio estadístico de los datos es realizar una exploración. Esto se puede hacer por medio de Estadística descriptiva. Los parámetros de Estadística descriptiva de variables iniciales para el periodo de estudio completo (del 31/03/2004 al 19/09/2005) se muestran en la Tabla (13).

Para conocer las características de los datos una de las primeras revisiones que se deben efectuar concierne a la media de una variable y a su mediana. Si la media es muy grande o muy pequeña en comparación con la mediana esto indica que los valores no están distribuidos normalmente (SPSS, 2006).

La media es bastante sensible a los datos anómalos. La presencia de estos puede incrementar (o decrementar) la media, resultando un valor de esta atípicamente grande (o pequeño). Sucede lo mismo si el conjunto de datos es asimétrico. En general, la media es sensible a las observaciones extremas. Debido a esto, se dice que la media no es una medida resistente (robusta) (Chase et al., 2000).

<i>Parámetro</i>	Media	Error típico	Mediana	Moda	Desviación estándar	Varianza de la muestra	Curtosis	Coefficiente de asimetría	Rango	Mínimo	Máximo
<i>Día juliano</i>	186,9	1,030	199	197	97,56	9518	-0,9222	-0,1076	359	5	364
<i>V (Km/h)</i>	1,283	0,0032	1,2	1,1	0,3030	0,0918	2,650	1,085	2,2	0,5	2,7
<i>DV (°)</i>	235,1	0,8434	251	292	79,89	6383	0,1099	-0,8445	359	0	359
<i>Vx (Km/h)</i>	5,168	0,0145	4,9	4,1	1,370	1,876	0,6175	0,9389	7,6	2,5	10,1
<i>T (°C)</i>	19,07	0,0661	20,2	25,6	6,260	39,18	-1,096	-0,4031	25,2	3,9	29,1
<i>Tn (°C)</i>	14,86	0,0639	15,9	17,5	6,052	36,62	-1,047	-0,3981	24,2	-0,1	24,1
<i>Tx (°C)</i>	23,63	0,0666	25,2	25,3	6,311	39,83	-0,9467	-0,4365	27	8,1	35,1
<i>Hf (horas)</i>	1,012	0,0329	0	0	3,119	9,725	12,74	3,537	20	0	20
<i>HR (%)</i>	58,16	0,1353	61	69	12,82	164,4	-0,4131	-0,5587	65	20	85
<i>Rad (W/m<sup>2</sup>)</i>	201,3	0,8595	210	285	81,42	6629	-1,124	-0,2277	310	21	331
<i>P (mm)</i>	0,4601	0,0214	0	0	2,025	4,102	45,59	6,164	22,4	0	22,4
<i>ETo (mm)</i>	3,509	0,0168	3,7	5,2	1,593	2,537	-1,393	-0,1074	6,1	0,8	6,9
<i>VolExp (m<sup>3</sup>)</i>	8853793	15916	9158000	6720000	1507776	2,273E+12	-1,370	-0,1984	4992000	6110000	11102000
<i>Qent (m<sup>3</sup>/día)</i>	22448	482,2	7780	0	45677	2086348622	120,0	9,298	667336	0	667336
<i>Qsal (m<sup>3</sup>/día)</i>	27094	124,0	29376	20304	11742	137878399	-1,027	-0,3040	50630	778	51408
<i>Filt (m<sup>3</sup>/día)</i>	-710,0	43,19	0	0	4091	16739881	265,8	-15,50	72096	-72096	0
<i>Profundidad (m)</i>	8,885	0,0805	6,67	1,67	7,629	58,21	-0,0584	1,026	29,58	0,42	30
<i>Ta (°C)</i>	16,73	0,0671	14,20	13,30	6,361	40,46	-1,295	0,4874	19,29	9,367	28,65
<i>Odis (mg/l)</i>	8,041	0,0445	8,96	0,125	4,215	17,77	-0,7895	-0,6162	15,42	0,0367	15,46
<i>pH</i>	8,237	0,0043	8,32	8,43	0,409	0,1672	1,446	-0,8267	3,268	5,885	9,153
<i>Con (μS/cm)</i>	984,6	0,817	970	970	77,42	5994	1,476	1,364	370	845	1215

**TABLA 13** Estadística descriptiva de los datos iniciales



---

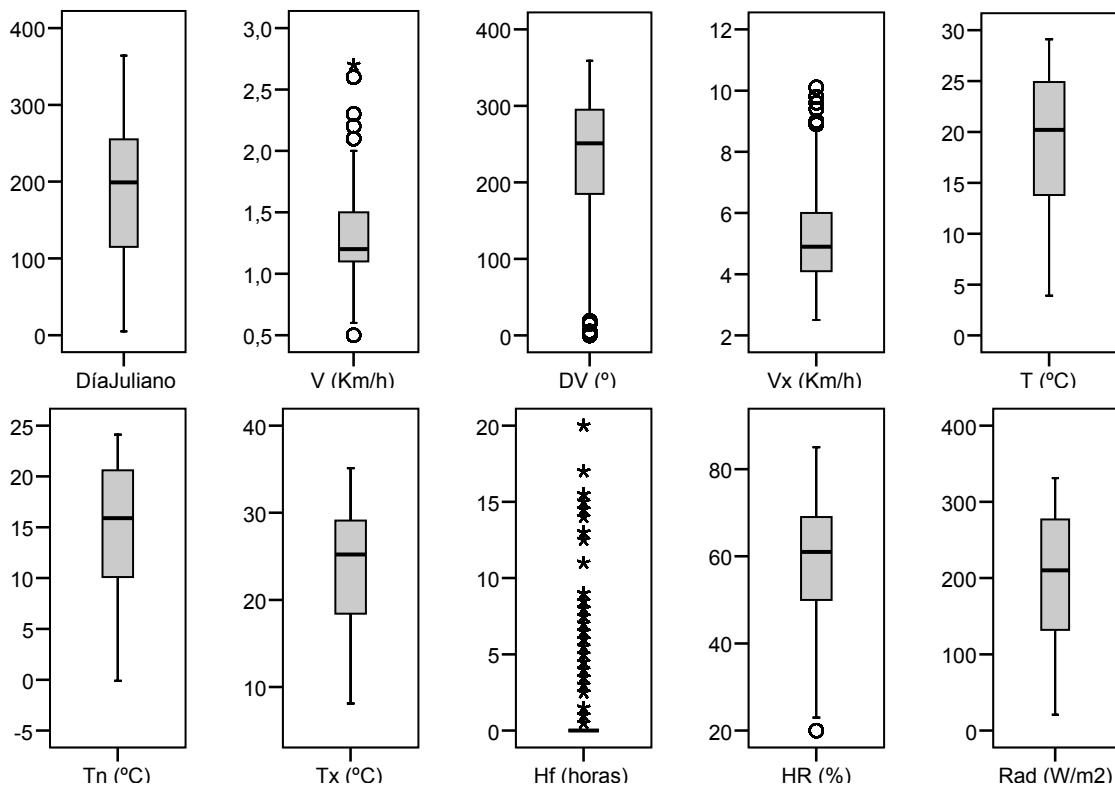
Una importante propiedad de la mediana en contraste con la media es que la mediana es muy insensible al número de valores de datos extremadamente pequeños o extremadamente grandes (datos anormales) (Devore, 1998).

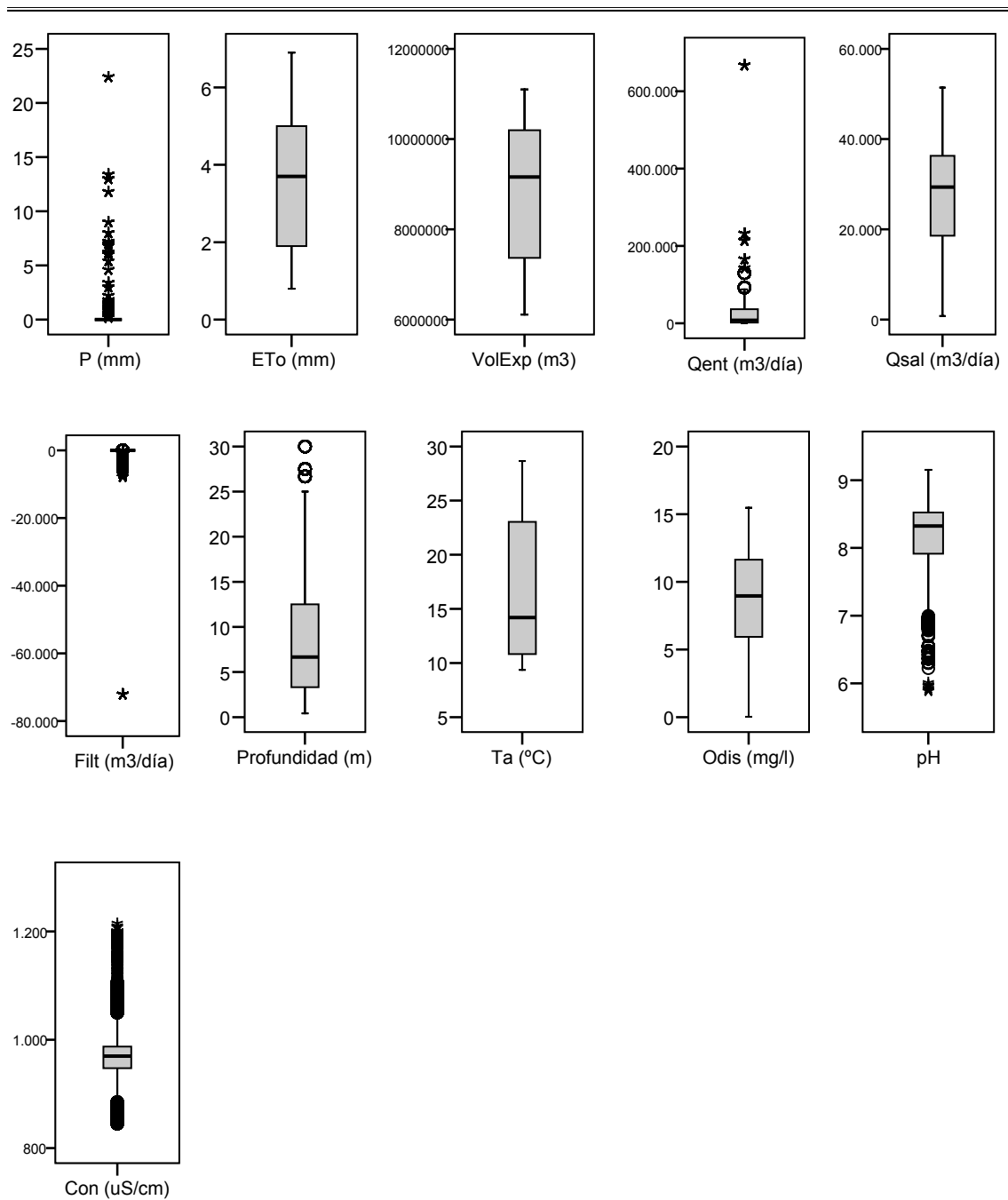
Con el propósito de estudiar de manera inicial la normalidad de las distribuciones de los valores de las variables y además analizar si los periodos de calibración y validación seleccionados son homogéneos (o no lo son), se muestran las mismas estadísticas descriptivas para calibración (31/03/2004 a 23/02/2005) y validación (24/02/2005 a 19/09/2005) en el Anexo 2.

La conclusión de revisar el Anexo 2 y el comportamiento de los análisis que se efectuarán para los dos periodos se expondrá en el Capítulo 9.

Para continuar con la exploración de datos se procede a elaborar diagramas de caja y bigotes de los valores registrados en cada una de las variables para el periodo completo de estudio. Estos se muestran en la Figura (44).

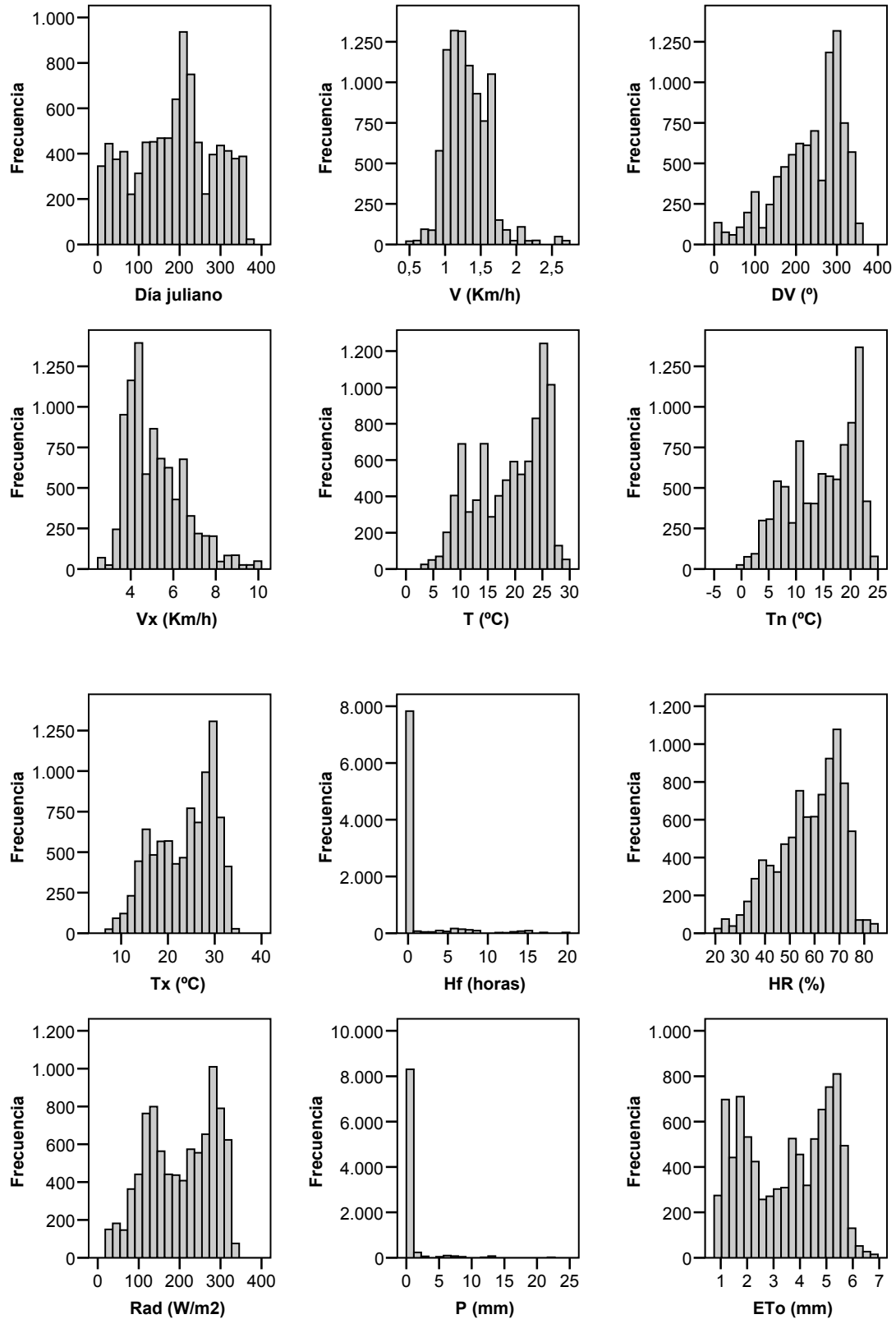
Además, en el Anexo 3 se muestran los diagramas de caja y bigotes para los datos en cada una de las variables para calibración y validación. Se muestran agrupados por variable para los dos periodos con el fin de comparar el comportamiento de cada variable para ambos lapsos de tiempo.

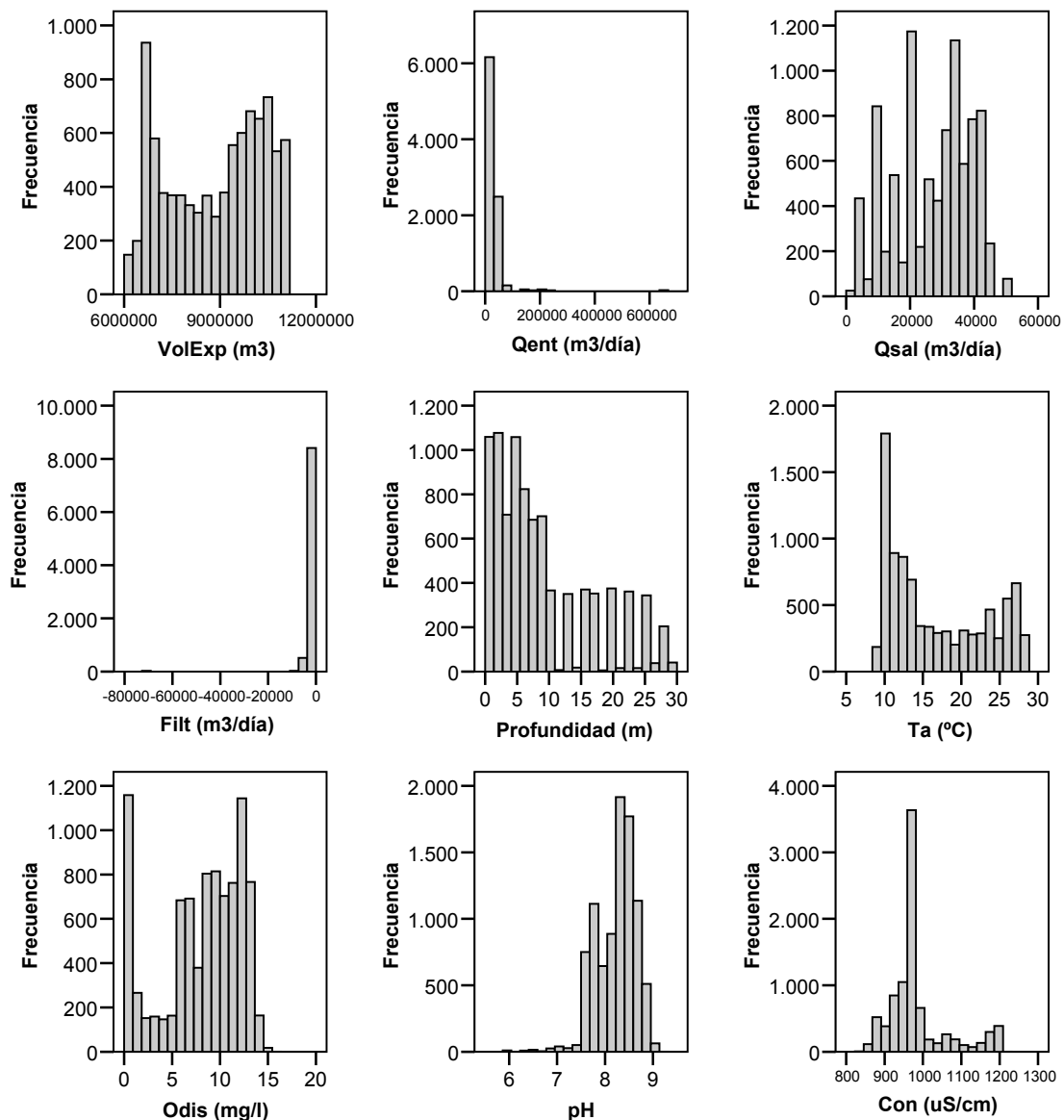




**FIGURA 44** Diagramas de caja y bigotes para las variables con los datos iniciales sin transformar para el periodo completo de estudio (31/03/2004 a 19/09/2005)

Para proseguir con la exploración de los datos enseguida se muestran los histogramas correspondientes a cada variable (Figura 45), tomando en cuenta el lapso de tiempo completo del estudio.





**FIGURA 45** Histogramas de datos iniciales sin transformar para el periodo completo de estudio

Observando los histogramas anteriores se puede apreciar por la forma que presentan, que ninguno se adapta a una distribución normal. Son marcadamente asimétricos los histogramas de las variables Hf, P, Qent y Filt.

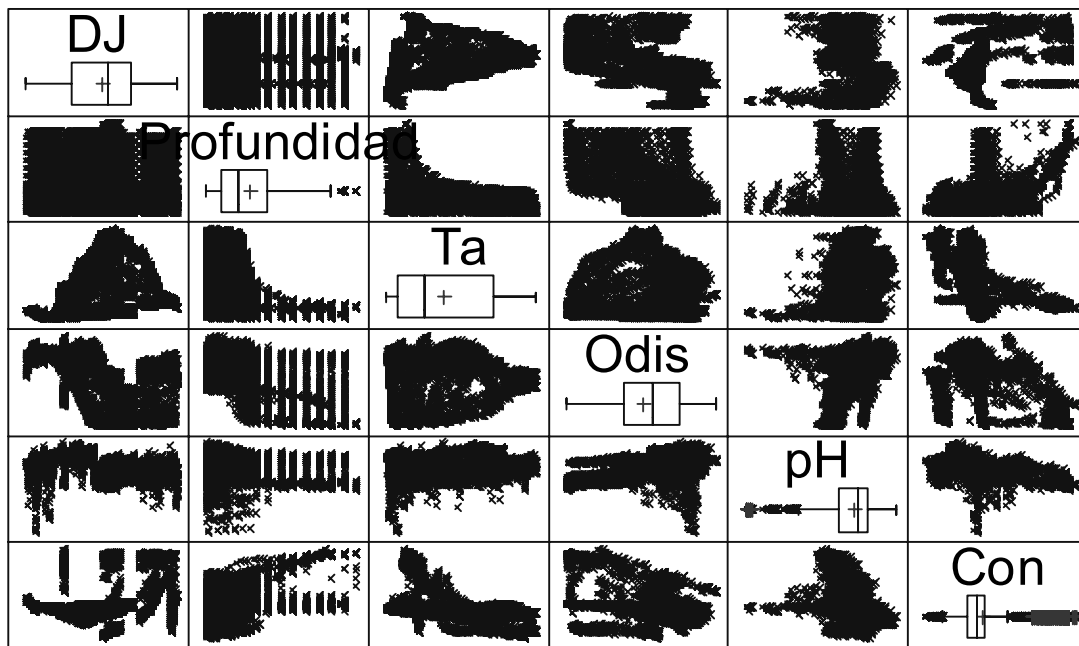
En el Anexo 4 se muestran histogramas de los datos iniciales separados para los periodos de calibración y validación.

Para valorar la normalidad de cada una de las variables se pueden utilizar los gráficos Q-Q. Cuando los puntos están muy próximos a una línea recta trazada se puede establecer que los datos están distribuidos normalmente. En caso contrario para las técnicas estadísticas que así lo exijan, se debe buscar la transformación de datos

adecuada.

En el Anexo 5 se muestran los gráficos Q-Q correspondientes a las variables en el periodo de estudio completo. Estos gráficos muestran los cuantiles reales y teóricos de una distribución normal. Los cuantiles dividen la muestra en una serie de grupos de igual tamaño (Visauta, 2002).

Con el propósito de continuar con la inspección de las variables se efectúan diagramas de dispersión por pares (diagrama de dispersión matricial de variables por pares). Este tipo de gráficos ayudan a saber si existen relaciones, ya sea lineales o no lineales entre los datos (Figura 46). Se presentan en el Anexo 6 los gráficos de dispersión matricial para las variables meteorológicas (con un registro por día), las variables hidrológicas del embalse (un registro por día) y las variables de calidad del agua (por día y profundidad).



**FIGURA 46** Gráfico de dispersión matricial para seis variables

A menos que se utilicen pruebas de normalidad, no se puede hablar de la normalidad de la población de donde fueron obtenidas las muestras de datos (Martín et al., 2008).

Con el propósito de poder afirmar con una cierta probabilidad si una variable es normal o no se requiere una prueba inferencial. Se usará la prueba de normalidad de Kolmogorov-Smirnov (con la corrección de la significación de Lilliefors) (Martín et al., 2008).

La prueba de Kolmogorov-Smirnov es una prueba de bondad de ajuste que se usa para contrastar la hipótesis nula de que la distribución de una variable se ajusta a una distribución normal. Si el  $p$ -valor asociado al estadístico de contraste es inferior que un cierto nivel de significación (que se fija generalmente en 0,05) rechazamos la hipótesis nula a dicho nivel. Cuando el  $p$ -valor asociado al estadístico de contraste es mayor que 0,05 no se puede rechazar la hipótesis nula. En ese último caso se concluye que la distribución de una variable puede ser considerada normal (Ferrán, 2002).

La prueba para contrastar la hipótesis de distribución normal de Kolmogorov-Smirnov-Lilliefors es una particularización de la prueba de Kolmogorov-Smirnov cuando la distribución esperada es una normal. Esta prueba es más precisa que la prueba general de Kolmogorov-Smirnov cuando se aplica al caso de una distribución normal (Ferrán, 1996).

A continuación se presentan las pruebas de normalidad para las variables analizadas:

#### Pruebas de normalidad

	Kolmogorov-Smirnov(a)		
	Estadístico	gl	Sig.
Día juliano	,058	8974	,000
V (Km/h)	,125	8974	,000
DV (°)	,123	8974	,000
Vx (Km/h)	,116	8974	,000
T (°C)	,110	8974	,000
Tn (°C)	,110	8974	,000
Tx (°C)	,106	8974	,000
Hf (horas)	,497	8974	,000
HR (%)	,096	8974	,000
Rad (W/m2)	,104	8974	,000
P (mm)	,467	8974	,000
ETo (mm)	,120	8974	,000
VolExp (m3)	,100	8974	,000
Qent (m3/día)	,312	8974	,000
Qsal (m3/día)	,117	8974	,000
Filt (m3/día)	,431	8974	,000
Profundidad (m)	,174	8974	,000
Ta (°C)	,157	8974	,000
Odis (mg/l)	,099	8974	,000
pH	,094	8974	,000
Con (uS/cm)	,259	8974	,000

a Corrección de la significación de Lilliefors

**TABLA 14** Prueba de normalidad de Kolmogorov-Smirnov para las variables sin transformar

---

Para las 21 variables analizadas el estadístico de Kolmogorov-Smirnov está en el rango de 0,058 hasta 0,497, con 8974 grados de libertad. Debido a que en todas las variables la significación del contraste ( $p$ -valor asociado al estadístico de contraste) es de 0, menor que 0,05 y significativo a cualquier nivel, se rechaza la hipótesis nula de que la distribución univariante de cada una de las variables se ajusta a una distribución normal. Por lo tanto, ninguna de ellas se adapta a una distribución normal univariante.

#### 7.4 TRANSFORMACIÓN DE LOS DATOS

Debido a que algunas técnicas estadísticas requieren que los datos cumplan con ciertos requisitos, siendo uno de los más importantes que las variables sigan una distribución normal, se analizarán en este apartado transformaciones para lograr este objetivo y que los datos parezcan “más normales”.

Se probaron inicialmente con la variable Día juliano (una de las que tiene una distribución no tan alejada de la distribución normal, no como en el caso de otras variables) las transformaciones  $x^2$ ,  $x^3$ ,  $1/x$ ,  $\log(x)$ ,  $\sqrt{x}$ ,  $\sqrt[3]{x}$ ,  $\sqrt[4]{x}$ ,  $\sqrt[5]{x}$ ,  $1/(1+e^x)$ ,  $2,6 \cdot x^{0,84}$ ,  $e^{-x}$ ,  $e^{1+2x} / (1+e^{1+2x})$  y  $\sqrt[3]{2,6 \cdot x^{0,84}}$ , no resultando enteramente satisfactoria ninguna de ellas.

Después se revisó el artículo de Van Albada et al. (2007) donde se sugiere una metodología más efectiva para realizar la transformación, la cual se adoptó en este trabajo. La metodología desarrollada para cada una de las variables fue la siguiente:

1. Se estandarizaron los datos con la fórmula (Mendenhall et al., 1986):

$$z = \frac{x - \mu}{\sigma} \quad (195)$$

2. El menor valor de la variable estandarizada se hace igual con cero por medio de la expresión:

$$z_1 = z + (-\text{MIN}(z)) \quad (196)$$

3. Se hace que el rango de la variable transformada esté entre 0 y 1:

$$z_2 = z_1 / \text{MAX}(z_1) \quad (197)$$

- 
4. Para efectuar la transformación se usará la inversa de la función de distribución acumulada normal estándar ( $\Phi$ ) que se conoce con el nombre de función cuantil normal o función probit, que puede expresarse en términos de la función de error (*erf*) como:

$$\mathbf{y} = \sqrt{2} \operatorname{erf}^{-1}(2\mathbf{z}_2 - 1) \quad (198)$$

La variable  $\mathbf{y}$  es la transformada, cuya distribución lucirá “más normal”.

5. Una vez efectuados los procesos de regresión, para re-expresar los resultados en las unidades originales se usa la expresión:

$$\mathbf{x} = \Phi(\mathbf{y}) = \frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{\mathbf{y}}{\sqrt{2}} \right) \right] \quad (199)$$

Donde  $\Phi$  es la Función de distribución acumulada normal estándar.

En las fórmulas anteriores se tienen funciones a las que se definirá tanto su significado como una manera de calcular sus valores aproximados.

La función de error, *erf*, se define como ([es.wikipedia.org/wiki/](http://es.wikipedia.org/wiki/), 2008):

$$\operatorname{erf}(\mathbf{x}) = \frac{2}{\sqrt{\pi}} \int_0^{\mathbf{x}} e^{-t^2} dt \quad (200)$$

A la función anterior también se le conoce como la Función de error de Gauss ([es.wikipedia.org/wiki/](http://es.wikipedia.org/wiki/), 2008).

La inversa de la función de error se puede expresar mediante la serie ([es.wikipedia.org/wiki/](http://es.wikipedia.org/wiki/), 2008):

$$\operatorname{erf}^{-1}(\mathbf{x}) = \sum_{k=0}^{\infty} \frac{c_k}{2k+1} \left( \frac{\sqrt{\pi}}{2} \mathbf{x} \right)^{2k+1} \quad (201)$$

Donde  $c_0 = 1$  y ([es.wikipedia.org/wiki/](http://es.wikipedia.org/wiki/), 2008):

$$c_k = \sum_{m=0}^{k-1} \frac{c_m c_{k-1-m}}{(m+1)(2m+1)} = \left\{ 1, 1, \frac{7}{6}, \frac{127}{90}, \dots \right\} \quad (202)$$



Para efectos del estudio presente la  $\text{erf}^{-1}$  se calculó mediante los 8 primeros términos de la serie infinita que define a esta función (functions.wolfram.com/):

$$\text{erf}^{-1}(\mathbf{x}) = \frac{1}{2}\sqrt{\pi} \left[ \mathbf{x} + \frac{\pi \mathbf{x}^3}{12} + \frac{7\pi^2 \mathbf{x}^5}{480} + \frac{127\pi^3 \mathbf{x}^7}{40320} + \frac{4369\pi^4 \mathbf{x}^9}{5806080} + \frac{34807\pi^5 \mathbf{x}^{11}}{182476800} + \frac{20036983\pi^6 \mathbf{x}^{13}}{398529331200} + \frac{2280356863\pi^7 \mathbf{x}^{15}}{167382319104000} + \dots \right] \quad (203)$$

La integral de la función de error no puede ser evaluada en términos precisos de funciones elementales, pero se puede expandir el integrando en una Serie de Taylor y se obtiene la serie de Taylor para la función de error así (es.wikipedia.org/wiki/, 2008):

$$\text{erf}(\mathbf{x}) = \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n \mathbf{x}^{2n+1}}{n!(2n+1)} = \frac{2}{\sqrt{\pi}} \left[ \mathbf{x} - \frac{\mathbf{x}^3}{3} + \frac{\mathbf{x}^5}{10} - \frac{\mathbf{x}^7}{42} + \frac{\mathbf{x}^9}{216} - \frac{\mathbf{x}^{11}}{1320} + \frac{\mathbf{x}^{13}}{9360} - \frac{\mathbf{x}^{15}}{75600} + \dots \right] \quad (204)$$

El cálculo de la función de error se efectuó considerando los primeros 8 términos de la serie expresada en (204).

En el caso de que una técnica estadística no requiera para su aplicación que los datos sigan una distribución normal se utilizarán los datos originales. Por el contrario, cuando un supuesto de la técnica sea el de normalidad se usarán los valores que se transformaron con la función probit. Para cada caso se hará mención explícita del tipo de datos con los cuales se opera.

## 7.5 ANÁLISIS DE COMPONENTES PRINCIPALES

En los análisis efectuados con el PCA, FA y CA, se utilizaron los datos experimentales estandarizados con el fin de evitar errores de clasificación debidos a las amplias diferencias en la dimensionalidad de los datos. La estandarización tiende a incrementar la influencia de las variables cuya varianza es pequeña y reduce la influencia de las variables cuya varianza es grande. Además, la estandarización elimina la influencia de las diferentes unidades de medida y homogeniza la dimensionalidad de los datos (Singh et al., 2004).

### 7.5.1 Exploración de datos

Antes de iniciar con el análisis es importante tener en mente los siguientes aspectos

---

---

(Tusell, 2007):

1. El uso de las componentes principales no presupone ningún modelo subyacente.
2. El principal objetivo es la reducción de dimensionalidad de los datos originales.

En este análisis se incluirán de inicio todas las variables consideradas y se efectuará por separado para los periodos de calibración y validación, con el propósito de apreciar las diferencias y las similitudes que existen entre ellos y el comportamiento de las variables meteorológicas, hidrológicas y de calidad del agua.

Luego, se efectuará el PCA solamente para los conjuntos de las variables meteorológicas y las hidrológicas, sin incluir las de calidad del agua, también en los periodos de calibración y validación. Así, se observará si algunas variables muestran comportamientos en conjunto que puedan servir con propósitos de pronóstico de las variables de calidad del agua.

A continuación se muestra el PCA para el periodo de calibración. Para realizar este análisis se estandarizaron todas las variables de manera univariante. Se hacen también varias observaciones breves al respecto. Comentarios más extensos del mismo se efectuarán en el Capítulo 9.

Calibración

#### Resumen del Análisis

Variables: DJ, V (Km/h), DV (°), Vx (Km/h), T (°C), Tn (°C), Tx (°C), Hf (horas), HR (horas), Rad ( $W/m^2$ ), P (mm), ETo (mm), VolExp ( $m^3$ ), Qent ( $m^3/día$ ), Qsal ( $m^3/día$ ), Filt ( $m^3/día$ ), Profundidad (m), Ta (°C), Odis (mg/l), pH y Con ( $\mu S/cm$ ).

Entrada de datos: observaciones

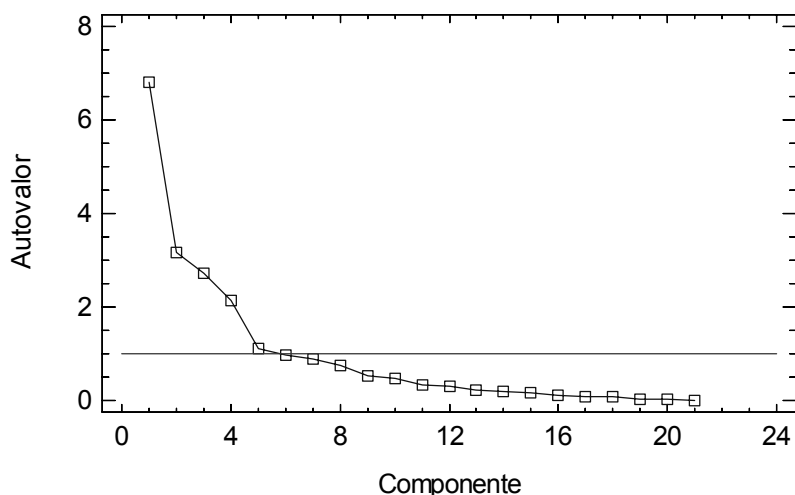
Número de casos completos: 4401

Tratamiento de valor perdido: lista considerada

Estandarizado: sí

Análisis de Componentes Principales			
Componente número	Autovalor	Porcentaje de varianza	Porcentaje acumulado
1	6,79448	32,355	32,355
2	3,16844	15,088	47,442
3	2,73154	13,007	60,450
4	2,12718	10,129	70,579
5	1,11699	5,319	75,898
6	0,975219	4,644	80,542
7	0,889645	4,236	84,779
8	0,744069	3,543	88,322
9	0,532399	2,535	90,857
10	0,457111	2,177	93,034
11	0,323965	1,543	94,576
12	0,289435	1,378	95,955
13	0,226527	1,079	97,033
14	0,179863	0,856	97,890
15	0,153877	0,733	98,623
16	0,100315	0,478	99,100
17	0,0775255	0,369	99,469
18	0,0708559	0,337	99,807
19	0,0251807	0,120	99,927
20	0,0127728	0,061	99,988
21	0,00260514	0,012	100

**TABLA 15** Resumen del análisis de componentes principales para el periodo de calibración



**FIGURA 47** Gráfico de sedimentación para los autovalores en el periodo de calibración

El propósito del análisis es obtener un pequeño número de combinaciones lineales de las 21 variables que explican la mayoría de la variabilidad en los datos. En este caso se tienen 5 componentes con autovalor mayor que la unidad (Figura 47). Juntos explican el 75,9% de la variabilidad en los datos originales.

En el Anexo 7 se incluyen varios elementos generados en el análisis de componentes principales, como son:

- Las Tablas de pesos de los componentes.

- Diagramas de dispersión de los componentes por pares.
- Gráficos de pesos de los componentes.

Al analizar los diagramas de dispersión de los componentes (Anexo 7, los diagramas y los otros elementos del análisis de hicieron con las variables estandarizadas, lo que se estima proporciona mejores resultados que sin estandarizar), se pueden apreciar en algunos de los diagramas conjuntos pequeños de puntos muy alejados del resto. Son los datos posiblemente anómalos. Se analizaron los casos más evidentes (periodo de calibración):

- En el gráfico de dispersión correspondiente a los componentes 5 y 6, los registros 454 a 479 (correspondientes al día 20/07/2004) se observan muy alejados de todo el resto del conjunto (Anexo 7). Al revisar los datos se tiene que la Filtración para ese día tiene un valor de  $-72.000 \text{ m}^3/\text{día}$ , que es el valor más extremo de esta variable para la calibración e inclusive también para la validación. Por lo tanto, se tiene casi la seguridad de que es un valor erróneo. Por ello será sustituido.

Para efectuar esto, primero se debe tomar en cuenta que el balance se tomó de otro estudio previo y está calculado a escala diaria. Para enmendar el dato anómalo se hará el balance entre los días 18 y 20 de julio de 2004, tomando un lapso de tiempo de 48 horas. Los resultados de los datos hidrológicos son (su valor medio diario):

Fecha	Qsal ( $\text{m}^3/\text{día}$ )	Filt ( $\text{m}^3/\text{día}$ )	Qent ( $\text{m}^3/\text{día}$ )
19/07/2004	31968	0	5420
20/07/2004	31968	0	5420

- En el gráfico de dispersión de los componentes 5 y 6 se tienen como datos anómalos los registros 4096 a 4118. Al revisar los datos se observa que la única variable que muestra un cambio radical ese día (el 09/02/2005) es la Dirección media del viento. Los valores anteriores y posteriores a este día están en un rango de variación bien definido. Por lo tanto, es muy probable que este valor sea incorrecto. Para remediar el error este valor será sustituido por el valor medio de los tres días anteriores y los tres días posteriores, esto es, el valor de  $307^\circ$ .
- En el gráfico de dispersión de los componentes 7 y 8 se observan como datos alejados del resto los correspondientes a los registros 2613 al 2632. La posible

---

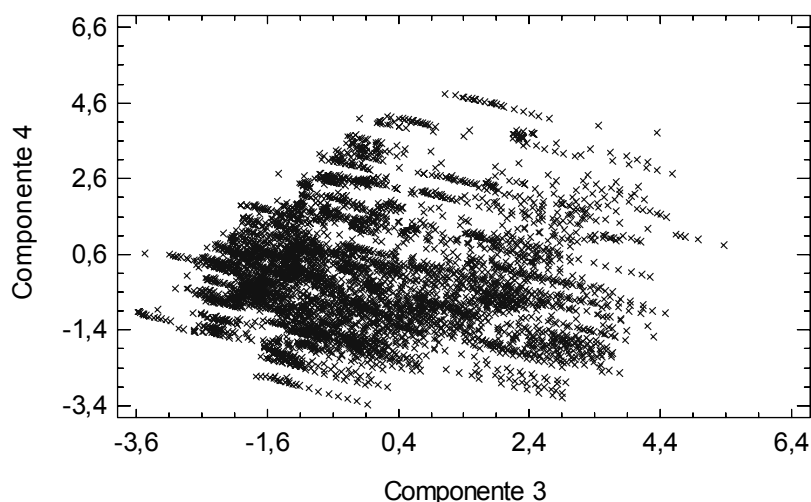
explicación de esto es: esos registros ocurrieron el día 30/11/2004. En esa fecha ocurrió una precipitación (22,4 mm) después de no haberse presentado ninguna en los 18 días anteriores. Este hecho provoca que los puntos de ese día se alejen del resto, pero es por una situación explicable y plausible, por lo que se dejarán tal como están sin hacer modificación alguna.

- Por último, en el gráfico de dispersión de los componentes 11 y 12 se tienen como datos apartados del resto de los registros del 1785 al 1804. La explicación probable es: esos registros acontecieron el día 25/10/2004. En esa fecha ocurrió una precipitación (1,4 mm) y a su vez la Radiación solar global disminuyó hasta  $87 \text{ W/m}^2$ . Y como estos valores están dentro de los rangos que se pueden presentar en la realidad, se dejan así.

Es necesario mencionar que en otros registros que se aprecian como anormales se observa que son los días en que ocurrió precipitación, después de varios días de no haberla. Debido a esto, se considera que dichas observaciones sí son posibles y se cree que no se debe retirarlas del análisis.

Al existir modificaciones en la información se requiere obtener nuevamente:

- Autovalores.
- Gráfico de sedimentación.
- Las Tablas de pesos de los componentes.
- Diagramas de dispersión de los componentes por pares (Figura 48).
- Gráficos de pesos de los componentes.



**FIGURA 48** Diagrama de dispersión de las componentes 3 y 4 en la calibración después de hacer correcciones en los datos

Para obtener estos elementos nuevamente se aplica el PCA. Esto se muestra en el Anexo 8.

Ahora se procede a realizar el análisis sobre los datos originales, pero en el periodo de validación para los datos meteorológicos, hidrológicos y de calidad del agua.

Para efectuar el PCA se estandarizaron las variables. Observaciones en mayor profundidad acerca del mismo se harán en el Capítulo 9.

## Validación

### Resumen del Análisis

Variables:

DJ, V (Km/h), DV (°), Vx (Km/h), T (°C), Tn (°C), Tx (°C), Hf (horas), HR (horas), Rad (W/m<sup>2</sup>), P (mm), ETo (mm), VolExp (m<sup>3</sup>), Qent (m<sup>3</sup>/día), Qsal (m<sup>3</sup>/día), Filt (m<sup>3</sup>/día), Profundidad (m), Ta (°C), Odis (mg/l), pH y Con (μS/cm).  
Entrada de datos: observaciones

Número de casos completos: 4573

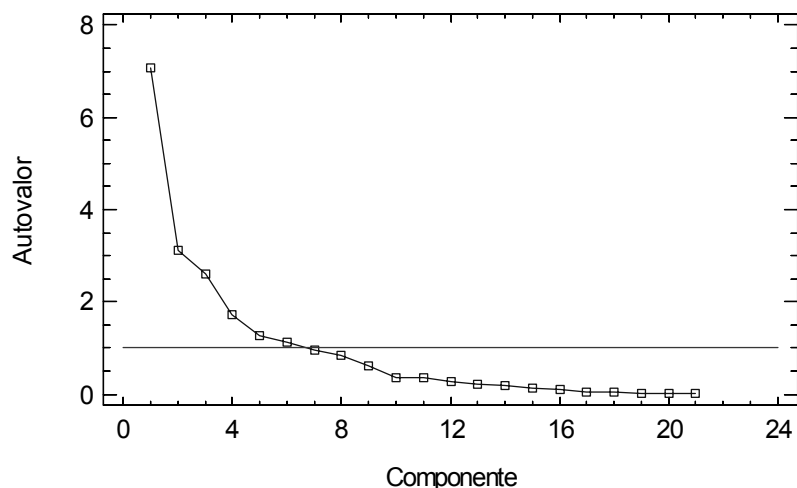
Tratamiento de valor perdido: lista considerada

Estandarizado: sí

Análisis de Componentes Principales			
Componente número	Autovalor	Porcentaje de varianza	Porcentaje acumulado
1	7,06814	33,658	33,658
2	3,10922	14,806	48,464
3	2,61874	12,470	60,934
4	1,71601	8,171	69,105
5	1,26001	6,000	75,105
6	1,11537	5,311	80,417
7	0,946606	4,508	84,924
8	0,834233	3,973	88,897
9	0,611696	2,913	91,810
10	0,357924	1,704	93,514
11	0,350898	1,671	95,185
12	0,278083	1,324	96,509
13	0,201686	0,960	97,470
14	0,182888	0,871	98,341
15	0,127015	0,605	98,945
16	0,103275	0,492	99,437
17	0,0583826	0,278	99,715
18	0,0302146	0,144	99,859
19	0,0163144	0,078	99,937
20	0,0101874	0,049	99,985
21	0,00310108	0,015	100

**TABLA 16** Resumen del análisis de componentes principales para el periodo de validación

El propósito del análisis es obtener un pequeño número de combinaciones lineales de las 21 variables que explican la mayoría de la variabilidad en los datos. En este caso, se extrajeron 6 componentes que poseen un autovalor mayor a 1 (Figura 49). Juntos explican el 80,42% de la variabilidad en los datos originales.



**FIGURA 49** Gráfico de sedimentación para los autovalores en el periodo de validación

En el Anexo 9 (el análisis se efectuó con las variables estandarizadas) se encuentran otras partes del análisis de componentes principales, las cuales son:

- Las Tablas de pesos de los componentes.
- Diagramas de dispersión de los componentes por pares.
- Gráficos de pesos de los componentes.

Se revisaron los diagramas de dispersión de los componentes (Anexo 9). Se tienen algunos casos posibles de datos anómalos (periodo de validación):

- En el gráfico de dispersión de las componentes 3 y 4, los registros 2864 a 2890 (del día 15/07/2005) parecen ser anómalos. Al revisar el conjunto de la información se tiene que el caudal de entrada al embalse, Qent, aumenta hasta un valor extremo de 667.336,3 m<sup>3</sup>/día, el cual está fuera del rango observado y, por tanto, es casi seguro que es un error. Para corregirlo se hace un balance conjunto para los días 14 y 15 de julio de 2005. Una vez efectuado, los nuevos valores para el día 15/07/2005 son:

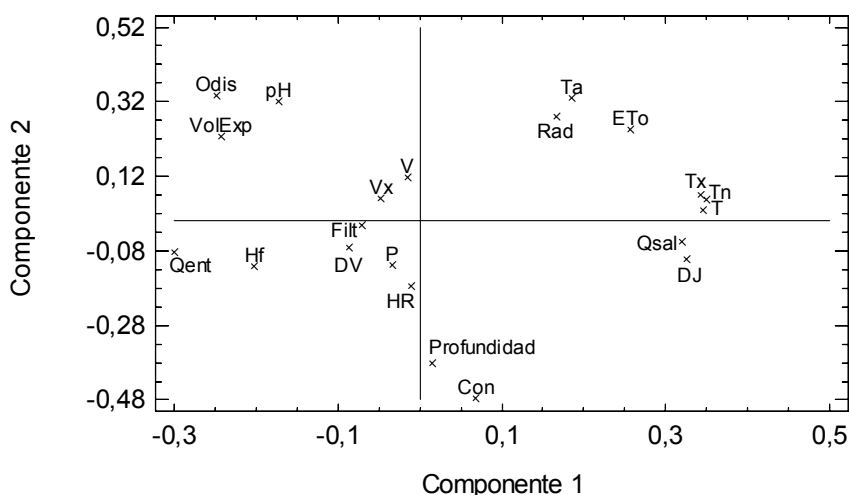
Qsal (m <sup>3</sup> /día)	Filt (m <sup>3</sup> /día)	Qent (m <sup>3</sup> /día)
41040	0	1225,05

- En el gráfico de dispersión de los componentes 19 y 20, los registros 2798 a 2817 (del día 08/07/2005) están alejados del resto de puntos. Para estos registros el valor de la radiación es demasiado bajo ( $\text{Rad} = 29 \text{ W/m}^2$ ). Es un valor muy alejado de los valores que se presentan en esas fechas por lo que con mucha seguridad es un error. Fue sustituido por el valor medio de los tres días anteriores y los tres días posteriores a ese día,  $256 \text{ W/m}^2$ .

En algunos gráficos se observan datos alejados del conjunto principal. Se investigaron todos los más separados del resto y se determinó que esto era así por varias causas: variación de la profundidad, aumento de las horas de frío, cambio de la velocidad media del viento, modificación de la dirección del viento, disminución de la humedad ligada a que ese día, ya no se presentó precipitación, aumento de la filtración, después de varios días sin llover se presenta una precipitación. Todos estos cambios y variaciones se encuentran dentro de rangos comunes por lo que estos registros no deben ser eliminados.

Al haber cambios en la información se modifican los siguientes elementos:

- Autovalores.
- Gráfico de sedimentación.
- Las Tablas de pesos de los componentes.
- Diagramas de dispersión de los componentes por pares.
- Gráficos de pesos de los componentes (Figura 50).



**FIGURA 50** Gráfico de pesos de las componentes 1 y 2 en la validación después de hacer correcciones en los datos

Para obtener los elementos anteriores, se vuelve a realizar el análisis por medio del



---

---

PCA. Todo esto se presenta en el Anexo 10. Comentarios adicionales de estos resultados se efectúan en el Capítulo 9.

### 7.5.2 Puntuaciones

Un paso muy común en el análisis de datos multivariantes es obtener una representación en pocas dimensiones de los datos (Tusell, 2007).

Al ser las componentes principales variables incorrelacionadas es posible aplicar otros métodos multivariantes (por ejemplo, el de regresión múltiple) evitando así efectos de colinealidad. Se consideran las componentes principales como variables explicativas, para así poder efectuar inferencias (Cuadras, 2007).

Otra aplicación del PCA es ordenar (clasificar) elementos de un conjunto por medio de sus valores en las componentes y estudiar su variabilidad (Cuadras, 2007).

Por lo tanto, se hace necesario conocer los valores que toman las componentes en cada una de las observaciones. Esto se puede realizar por medio de la fórmula (35).

Debido a que el propósito del presente trabajo es realizar inferencia, sólo se considerarán las variables (como independientes o explicativas) de la Tabla (17) para el cálculo de las puntuaciones.

Las cuatro variables de calidad del agua de las que se dispone constituyen las variables dependientes o variables respuesta.

Primero se obtendrán las puntuaciones de la calibración y luego las de validación.

Se calculan los valores de las puntuaciones,  $Z_{hi}$ , haciendo uso de los valores de los pesos de los componentes.

En el Anexo 11 se muestran los resultados del proceso aplicado para obtener las puntuaciones. Se incluyen en él:

- Los pesos de los componentes.
- Las primeras puntuaciones de las componentes en cada observación.

La interpretación de componentes principales es más sencilla, en ocasiones, basándose en los valores que adquieren los pesos de las componentes (Tusell, 2007).

Clasificación de la variable	Variable y sus unidades de medida
Temporal	DJ, Día juliano
Meteorológicas	V, Velocidad media del viento (Km/h)
	DV, Dirección media del viento (°)
	Vx, Velocidad máxima del viento (Km/h)
	T, Temperatura media (°C)
	Tn, Temperatura mínima (°C)
	Tx, Temperatura máxima del aire (°C)
	Horas frío, Hf, Número de horas por debajo de 7 °C
	HR, Humedad relativa media del aire (%)
	Rad, Radiación solar global (W/m <sup>2</sup> )
	P, Precipitación total (mm)
	ETo, Evapotranspiración de referencia (mm)
Hidrológicas	VolExp, Volumen de explotación (m <sup>3</sup> )
	Qent, Caudal de entrada al embalse (m <sup>3</sup> /día)
	Qsal, Caudal de salida del embalse (m <sup>3</sup> /día)
	Filt, Filtración (m <sup>3</sup> /día)
	Profundidad (m)

**TABLA 17** Variables (independientes) que se usaron para obtener las puntuaciones de las observaciones en las componentes

Pesos altos, positivos y negativos, indican que esas variables tienen presencia importante en la componente principal considerada (Terrado et al., 2006).

Es muy importante mencionar que este proceso se llevó a cabo sobre los valores tal como se encuentran en los registros, esto es, sin estandarizar. Comentarios acerca de estos resultados se harán en el siguiente capítulo.

En necesario indicar que la revisión de las varianzas y covarianzas muestrales y los coeficientes de correlación muestral se efectuarán en el apartado correspondiente al tema de análisis factorial en los capítulos que corresponda.

## 7.6 ANÁLISIS FACTORIAL

La aplicación del análisis factorial se efectuará con la siguiente secuencia:

- Como fase inicial (exploratoria) se aplicará el análisis factorial (FA) al conjunto de datos para el periodo completo de estudio, con los datos que se corrigieron en la fase de PCA (para no tener que manejar datos erróneos) sin aplicarles ninguna transformación. Luego se aplicará el FA para ese mismo conjunto de datos (sin datos anómalos y sin aplicar transformación alguna) dividido en dos

---

---

partes: calibración y validación. Con esto se comprobará la estabilidad de la solución (ver sección 4.3.5).

- Verificar la normalidad de las variables corregidas y sin transformar, así como de las primeras puntuaciones factoriales (scores). Esto para el periodo de estudio completo.
- Después transformar las variables corregidas con la transformación descrita en la sección 7.4, por medio de la función probit, con el fin de que se acerquen lo más posible hacia la normalidad. A las variables transformadas se les aplica el FA y se verifica si la interpretación obtenida es mejor. Se obtienen las puntuaciones factoriales y se comprueba si éstas se aproximan más a una distribución normal.
- Todos los análisis se harán extrayendo los factores con el método de componentes principales y en algunos, para verificar, con el de máxima verosimilitud. Se hará una rotación con el objetivo de simplificar la interpretación de los factores con el método Varimax (que trata de reducir el número de variables con saturaciones altas en cada factor). La Rotación Varimax permite incrementar la contribución de los valores de las cargas factoriales más altas y la reducción del peso de valores bajos. Esto lleva a una mejor comprensión de la estructura de datos (Simeonov et al., 2002).

#### 7.6.1 Periodo completo de estudio y variables sin transformar

##### a) Análisis factorial aplicado al conjunto de datos corregido:

- Utilizando el Método de componentes principales para extraer los factores se obtienen los resultados que se presentan en el Anexo 12. Una revisión más extensa se realizará en el Capítulo 9. Sólo algunos de los aspectos más importantes se comentan a continuación.

El determinante de la matriz de correlaciones tiene un valor  $5,13 \times 10^{-11}$ , como está próximo a cero, esto indica que las variables están linealmente relacionadas. Por lo tanto, es pertinente desarrollar el FA. La matriz de correlaciones (Anexo 12) tiene elementos fuera de la diagonal principal con valores que aseguran que esta matriz no esta próxima a ser una matriz identidad (si esto ocurriera, el análisis factorial llevaría a una solución deficiente).

La medida de adecuación muestral de Kaiser-Meyer-Olkin (KMO) indica si la

relación entre cada par de variables puede ser explicada por el resto de variables, por lo que llevar a cabo un análisis factorial es una buena solución. Valores de KMO por debajo de 0,5 no son aceptables, considerándose inadecuados los datos para un modelo de FA. Se considera excelente la adecuación de los datos para valores de KMO muy cercanos a 0,9. La medida KMO en este caso tiene un valor de 0,798, que según el criterio de Kaiser se puede considerar como entre aceptable y bueno (Visauta et al., 2003).

**KMO y prueba de Bartlett**

Medida de adecuación muestral de Kaiser-Meyer-Olkin		,798
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	212407,580
	gl	210
	Sig.	,000

**TABLA 18** Medida de adecuación muestral de Kaiser-Meyer-Olkin (KMO) y Prueba de esfericidad de Bartlett para el periodo de estudio completo, los datos sin contener anómalos evidentes y sin transformar

La Prueba de esfericidad de Bartlett contrasta la hipótesis nula de que la matriz de correlaciones es una matriz identidad. Si se diera este caso, no existirían correlaciones significativas entre las variables y no sería adecuado utilizar el modelo factorial.

En este caso la Prueba de Bartlett indicará que se puede seguir con el análisis cuanto mayor sea el valor de la transformación  $\chi^2$  del determinante de la matriz de correlaciones (212407,580 en este caso) y por tanto menor el grado de significación ( $p = 0,000$  para esta situación). Si el nivel crítico es mayor que 0,05 no se podrá rechazar la hipótesis nula de esfericidad y, por lo tanto, no se podrá garantizar que el modelo factorial sea adecuado para explicar los datos.

Por lo valores obtenidos en la Prueba de Bartlett se continúa con el FA.

En la tabla de varianza total explicada se tiene que se extrajeron en total cinco factores utilizando el criterio de extraer todos los factores que tengan autovalor mayor que uno (>1).

Los cinco factores extraídos explican el 75,258% de la varianza total, el cual es un valor que se puede considerar como bueno.

**Varianza total explicada**

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	6,841	32,576	32,576	6,841	32,576	32,576
2	3,342	15,915	48,491	3,342	15,915	48,491
3	2,500	11,903	60,394	2,500	11,903	60,394
4	1,996	9,507	69,901	1,996	9,507	69,901
5	1,125	5,357	75,258	1,125	5,357	75,258
6	,977	4,655	79,912			
7	,791	3,767	83,680			
8	,766	3,648	87,327			
9	,485	2,308	89,635			
10	,396	1,886	91,521			
11	,371	1,767	93,287			
12	,306	1,455	94,742			
13	,262	1,250	95,992			
14	,229	1,091	97,083			
15	,174	,828	97,911			
16	,160	,760	98,671			
17	,123	,588	99,258			
18	,106	,506	99,765			
19	,028	,135	99,899			
20	,018	,088	99,987			
21	,003	,013	100,000			

Método de extracción: Análisis de Componentes principales.

**TABLA 19** Extracción de factores por el Método de componentes principales en el periodo completo de estudio, variables sin anómalos y sin transformar

Luego se obtuvo Matriz de componentes. Para observar las variables que se agrupan en cada factor hay que detectar a las variables cuyas cargas sean altas en un factor y bajas en los otros. Los valores menores que 0,25 se consideran bajos, por lo que han sido eliminados de la Tabla (20) para ganar en claridad.

En la primera componente están representadas las variables T, Tn, Tx y Qent. La segunda componente incorpora, entre otras, a la variable Con. La variable DJ está presente en las componentes 1, 2 y 4. V y HR están representadas en las componentes 3 y 4. DV está presente en las componentes 1 y 4. Vx está añadida en las componentes 3, 4 y 5. Hf está en las componentes 1 y 5, al igual que Filt. Rad está presente en las componentes 1, 2 y 3. P y pH están representadas en las componentes 2,3 y 5. ETo está incorporada en las componentes 1,2 y 3. VolExp aparece en las componentes 2 y 4. Qsal está en las componentes 1 y 2. Profundidad está incorporada a las componentes 2, 3 y

4. Ta está incluida en las componentes 1, 3 y 4. Y Odis aparece en las componentes 1 a 4.

**Matriz de componentes(a)**

	Componente				
	1	2	3	4	5
DJ	,368	-,667		,447	
V			,756	,439	
DV	-,526			,292	
Vx			,594	,435	,499
T	,980				
Tn	,970				
Tx	,967				
Hf	-,567				-,342
HR			-,729	-,398	
Rad	,714	,453	,264		
P		-,267	-,326		,615
ETo	,850	,337	,302		
VolExp		,743		-,486	
Qent	-,679				
Qsal	,787	-,370			
Filt	-,307				-,349
Profundidad		-,434	,445	-,630	
Ta	,670		-,333	,476	
Odis	-,548	,639	-,317	,297	
pH		,585	-,435		,298
Con		-,763			

a 5 componentes extraídos

**TABLA 20** Matriz de componentes de los 5 factores extraídos por el Método de componentes principales, periodo de estudio completo, datos sin anómalos evidentes y sin transformar

Se observa que es difícil agrupar a las variables en factores, con lo que se hace necesaria una rotación. Se aplicará una Rotación Varimax.

b) Extracción de los factores por el Método de componentes principales, Rotación Varimax:

- Una vez efectuada la rotación Varimax, la matriz de componentes es la siguiente (se han eliminado los valores inferiores a 0,25 para tener mayor claridad):

**Matriz de componentes rotados(a)**

	Componente				
	1	2	3	4	5
DJ		,861			
V				,904	
DV	-,599				
Vx				,868	,252
T	,934	,288			
Tn	,909	,321			
Tx	,926	,274			
Hf	-,473	-,331			-,440
HR				-,777	,379
Rad	,841	-,291			
P					,711
ETo	,936				
VolExp		-,887			
Qent	-,719				
Qsal	,630	,654			
Filt	-,338				-,353
Profundidad			-,879		
Ta	,570	,384	,613		
Odis	-,478	-,370	,727		
pH		-,333	,644		,341
Con	-,300	,306	-,737		

Método de rotación: Normalización Varimax con Kaiser

**TABLA 21** Matriz de componentes de los 5 factores extraídos por el Método de componentes principales, periodo de estudio completo, variables sin transformar y Rotación Varimax

Se tiene que después de la rotación de los factores ha mejorado mucho la agrupación de las variables en torno a ellos. En el primer factor quedan representados DV, T, Tn, Tx (estos tres últimos con cargas factoriales mayores que 0,9), Hf, Rad, ETo (con carga factorial mayor que 0,9) y Qent. En el segundo factor quedan incorporados DJ, VolExp y Qsal. El tercer factor comprende las variables Profundidad, Ta, Odis, pH y Con. El cuarto factor contiene a V, Vx y HR. Y el quinto factor representa a P y Filt. Un análisis más a fondo del significado de estos factores se efectuará en el Capítulo 9.

c) Extracción de los factores por el Método de máxima verosimilitud, incluyendo una rotación con el Método Varimax

- Los resultados de utilizar el Método de la máxima verosimilitud se incluyen en el Anexo 12. Como este método se aplica para compararlo con la solución del análisis factorial desarrollado con el Método de componentes principales sólo

se obtendrán las comunalidades, la tabla de varianza total explicada, la matriz de componentes y la matriz de componentes rotados.

**Comunalidades(a)**

Variable	Inicial
DJ	,700
V	,730
DV	,437
Vx	,612
T	,996
Tn	,989
Tx	,986
Hf	,583
HR	,650
Rad	,939
P	,318
ETo	,974
VolExp	,735
Qent	,537
Qsal	,804
Filt	,119
Profundidad	,715
Ta	,783
Odis	,840
pH	,602
Con	,614

Método de extracción: Máxima verosimilitud.

a Se han encontrado una o más estimaciones de comunalidad mayores que 1 durante las iteraciones. La solución resultante deberá ser interpretada con precaución.

**TABLA 22** Comunalidades de las variables al extraerse los factores por el Método de máxima verosimilitud para el periodo completo de estudio, variables sin transformar

Al comparar las comunalidades obtenidas para las variables por los dos métodos se ve que al extraer los factores por el Método de máxima verosimilitud disminuyó la comunalidad de 14 variables (DJ, V, DV, Vx, HR, P, VolExp, Qsal, Filt, Profundidad, Ta, Odis, pH y Con) y aumentó la comunalidad de 7 variables (T, Tn, Tx, Hf, Rad, Eto y Qent). Es necesario hacer notar que todas las variables respuesta (Ta, Odis, pH y Con) disminuyeron su comunalidad (la parte de su variabilidad explicada por los factores) al extraerse los factores por el Método de máxima verosimilitud.

La Tabla de varianza total muestra los resultados siguientes:



**Varianza total explicada**

Factor	Autovalores iniciales		
	Total	% de la varianza	% acumulado
1	6,841	32,576	32,576
2	3,342	15,915	48,491
3	2,500	11,903	60,394
4	1,996	9,507	69,901
5	1,125	5,357	75,258
6	,977	4,655	79,912
7	,791	3,767	83,680
8	,766	3,648	87,327
9	,485	2,308	89,635
10	,396	1,886	91,521
11	,371	1,767	93,287
12	,306	1,455	94,742
13	,262	1,250	95,992
14	,229	1,091	97,083
15	,174	,828	97,911
16	,160	,760	98,671
17	,123	,588	99,258
18	,106	,506	99,765
19	,028	,135	99,899
20	,018	,088	99,987
21	,003	,013	100,000

Método de extracción: Máxima verosimilitud.

**TABLA 23** Varianza total explicada por cada factor para el periodo completo de estudio, sin datos anómalos, variables sin transformar y Método de extracción de máxima verosimilitud

Los factores que se considerarán serán los que tengan autovalor mayor que uno (cinco factores). Se observa que los porcentajes de varianza de cada factor obtenidos por ambos métodos son idénticos.

Al comenzar con el cálculo de la matriz de componentes se observa que el programa SPSS sólo es capaz de proporcionar un resultado de la matriz de componentes hasta cuando se efectúan 5 iteraciones. Con un mayor número de iteraciones no es capaz de efectuar el cálculo. Por lo tanto, la matriz de componentes y la matriz de componentes rotados se calcularán con 5 iteraciones. Esto provoca que la varianza total explicada por cada factor disminuya, pero sin ser demasiado significativo (11,05% de disminución del total de los cinco factores con autovalor mayor a la unidad), como se muestra en la tabla siguiente:

**Varianza total explicada**

Factor	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	6,841	32,576	32,576	6,589	31,374	31,374	6,093	29,014	29,014
2	3,342	15,915	48,491	2,504	11,922	43,296	2,499	11,902	40,916
3	2,500	11,903	60,394	2,418	11,513	54,809	2,407	11,461	52,376
4	1,996	9,507	69,901	1,802	8,581	63,391	1,975	9,403	61,779
5	1,125	5,357	75,258	,746	3,551	66,941	1,084	5,162	66,941
6	,977	4,655	79,912						
7	,791	3,767	83,680						
8	,766	3,648	87,327						
9	,485	2,308	89,635						
10	,396	1,886	91,521						
11	,371	1,767	93,287						
12	,306	1,455	94,742						
13	,262	1,250	95,992						
14	,229	1,091	97,083						
15	,174	,828	97,911						
16	,160	,760	98,671						
17	,123	,588	99,258						
18	,106	,506	99,765						
19	,028	,135	99,899						
20	,018	,088	99,987						
21	,003	,013	100,000						

Método de extracción: Máxima verosimilitud.

**TABLA 24** Varianza total explicada por cada uno de los factores al ser extraídos con el Método de máxima verosimilitud en cinco iteraciones, para el periodo completo de estudio y las variables sin transformar

La matriz factorial que se obtuvo al efectuar las cinco iteraciones con el método de Máxima verosimilitud es la que se muestra en la Tabla (25).

La matriz factorial de la Tabla (25) muestra a la gran mayoría de las variables saturando en dos o más factores. Como el ideal del análisis factorial es hallar un modelo en el que todas las variables saturan en algún factor, esto es, pesos factoriales altos en uno de los factores y reducidos en el resto (Visauta et al., 2003), se efectuará una Rotación Varimax para intentar lograr este objetivo.

**Matriz factorial(a)**

	Factor				
	1	2	3	4	5
DJ	,361	-,562		,323	-,268
V		,399	,256	,724	
DV	-,486				
Vx				,654	,457
T	1,000				
Tn	,986				
Tx	,987				
Hf	-,575				
HR		-,440		-,571	
Rad	,668	,688			
P		-,290			
ETo	,825	,548			
VolExp		,690		-,376	,390
Qent	-,612				
Qsal	,755	-,305			-,272
Filt	-,251				
Profundidad			,785	-,309	
Ta	,633		-,531	,252	-,250
Odis	-,553		-,721		
pH			-,603		
Con		-,363	,647		

Método de extracción: Máxima verosimilitud.

a Se han intentado extraer 5 factores. Se efectuaron 5 iteraciones. (Convergencia=,035). Se ha terminado la extracción.

**TABLA 25** Matriz factorial de los 5 factores extraídos por el Método de máxima verosimilitud con cinco iteraciones efectuadas, periodo de estudio completo, datos sin anómalos y sin transformar

Después de obtener la matriz de factores rotados (Tabla 26) se tienen que los factores están constituidos por las siguientes variables: el factor 1 por DV, T, Tn, Tx, Hf, Rad, ETo, Qent, Qsal y Filt. El factor 2 incluye a DJ y VolExp. En el factor 3 encontramos a las variables Profundidad, Ta, Odis, pH y Con. V, Vx y HR están en el factor 4. Y el factor 5 sólo incluye a la variable P.

Por lo tanto, sólo se tienen estos cambios de pertenencia de las variables a los factores al pasar del método de componentes principales al de máxima verosimilitud: Qsal se transfiere del factor 2 al 1, Filt del factor 5 al 1. Los valores de las saturaciones no presentan cambios altamente significativos. Debido a estas modificaciones mínimas se considera que la solución obtenida con el método de componentes principales es adecuada.

**Matriz de factores rotados(a)**

	Factor				
	1	2	3	4	5
DJ		,752			
V				,866	
DV	-,542				
Vx				,818	
T	,971				
Tn	,953				
Tx	,957				
Hf	-,561				
HR				-,680	-,314
Rad	,690	-,263			,634
P					-,268
ETo	,839				,479
VolExp		-,869			
Qent	-,603				-,264
Qsal	,632	,618			
Filt	-,258				
Profundidad			-,849		
Ta	,566	,363	,582		
Odis	-,473	-,346	,739		
pH		-,336	,565		
Con	-,255	,271	-,663		

Método de extracción: Máxima verosimilitud.

Método de rotación: Normalización Varimax con Kaiser.

a La rotación ha convergido en 8 iteraciones.

**TABLA 26** Matriz de factores rotados de los 5 factores extraídos por el Método de máxima verosimilitud para cinco iteraciones, periodo de estudio completo, variables sin transformar y Rotación Varimax

7.6.2 Periodo de calibración, variables sin transformar

Para verificar la estabilidad de la solución para el periodo de estudio se desarrollará el FA por el Método de componentes principales y se efectuará una rotación por el Método Varimax, tanto para el periodo de calibración como el de validación. Los elementos que se compararán son los siguientes:

- Matriz de correlaciones.
- Comunalidades de las variables.
- Varianza total explicada por los factores.
- Matriz de componentes.
- Matriz de componentes rotados.

Se inicia con el FA para el periodo de calibración. Los cinco elementos anteriores se

---

---

muestran en el Anexo 13. La mayoría de las comunalidades (66,7% de ellas) muestran gran coincidencia en sus valores para las variables, mientras que un 19,0% de ellas muestran una menor correspondencia y 14,3% restante no tienen similitud entre sí.

Las varianzas explicadas por los factores extraídos (cinco en ambos casos) tienen grandes semejanzas y en este aspecto se puede considerar que el FA realizado para todo el periodo de estudio sí es estable.

La matriz de componentes es más fácil de interpretar al realizar una Rotación Varimax. Se observa una gran coincidencia en la pertenencia de las variables a los factores siendo las únicas diferencias encontradas en las variables Hf, Qent y Ta (para las tres variables hay valores muy próximos en varios factores), debido a pequeños cambios en los valores de las saturaciones. En general, la estabilidad en la formación de los factores se puede considerar como buena.

#### 7.6.3 Periodo de validación, variables sin transformar

Se aplicará el FA a los datos de validación y se obtendrá (ver Anexo 14):

- Matriz de correlaciones.
- Comunalidades de las variables.
- Varianza total explicada por los factores.
- Matriz de componentes.
- Matriz de componentes rotados.

Al comparar las comunalidades de las variables se observa que entre el periodo completo de estudio y el periodo de validación hay grandes coincidencias, teniendo en total 61,9% de variables con valores muy próximos entre sí, un 23,8% con valores medianamente próximos entre sí y el restante 14,3% sin coincidencia entre sí.

Existe una diferencia poco marcada entre el porcentaje de varianza total que explica cada factor en los periodos completo y de validación. La mayor diferencia que se aprecia es que en el periodo completo se extraen 5 componentes con sus autovalores mayores que la unidad. Mientras que en el periodo de validación se extrajeron 6 componentes con autovalores mayores que la unidad. En realidad la diferencia entre ambas extracciones no es importante debido a que en el periodo completo de estudio el factor ubicado en la posición número 6 (en orden decreciente de magnitud de su autovalor) tiene un autovalor de 0,977 y para la validación el factor en la posición

---

---

número 6 posee un autovalor de 1,083. Por lo tanto, sus valores están muy próximos.

Las variables que integran los factores se agruparon después de revisar la solución rotada por el Método Varimax, con el fin de tener más claridad de la pertenencia de cada variable a uno de los factores. De entrada se tiene que el factor 1 incorpora a las variables DJ, T, Tn, Tx, VolExp, Qent, Qsal, Odis y pH. El factor 2 contiene a Hf, Rad y ETo. En el factor 3 saturan las variables Profundidad, Ta y Con. En el cuarto factor están V, Vx y HR. En el 5 se encuentran P y Filt. Y en el sexto factor la única variable es DV.

El principio de parsimonia dice que entre varias soluciones admisibles se escoge el modelo factorial que implique un número mínimo de factores comunes (Cuadras, 2007).

Debido a esto, se observa que el sexto factor sólo representa a una variable (DV) lo que no ayuda a encontrar explicaciones comunes para la información. Debido a esto, se obtendrán nuevamente la Matriz de componentes y la Matriz de componentes rotados pero considerando sólo cinco factores (Anexo 14).

Con 5 factores de la Matriz de componentes rotados se tiene que el primer factor contiene a las variables DJ, T, Tn, Tx, Hf, VolExp, Qent, Qsal y Odis. En el factor 2 saturan principalmente Rad, P y ETo. El factor 3 representa a las variables Profundidad, Ta, pH y Con. El cuarto factor incluye a las variables V, Vx y HR. El factor 5 incluye a DV y Filt.

Las diferencias en los factores entre el periodo completo de estudio y el de validación son: DJ, VolExp y Qsal salen del segundo factor y entran todas al primero; Rad y ETo pasan del primer al segundo factor; DV pasa del factor 1 al 5; Odis sale del tercer factor y llega al primero; P se va del quinto factor para ingresar al segundo; las otras variables siguen igual. Por tanto, 8 variables (de 21) cambiaron de factor por lo que existe una estabilidad en la formación de los factores de calidad intermedia, siendo inferior a la encontrada en el caso de la calibración. Esto puede deberse a que en el periodo de validación se incluyeron menos días de invierno que en el de calibración, por lo que la formación de los factores muestra ciertas fluctuaciones.

#### 7.6.4 Normalidad de las puntuaciones factoriales extraídas, periodo completo de estudio

Para los cinco factores extraídos en el periodo completo de estudio se muestran los

estadísticos descriptivos más importantes:

	<i>Factor 1</i>	<i>Factor 2</i>	<i>Factor 3</i>	<i>Factor 4</i>	<i>Factor 5</i>
Media	0,000	0,000	0,000	0,000	0,000
Error típico	0,011	0,011	0,011	0,011	0,011
Mediana	0,254	-0,073	0,333	-0,164	-0,130
Moda	1,275	-0,286	0,748	0,110	-0,131
Desviación estándar	1,000	1,000	1,000	1,000	1,000
Varianza de la muestra	1,000	1,000	1,000	1,000	1,000
Curtosis	-1,32	-1,32	0,00	1,47	9,74
Coficiente de asimetría	-0,31	0,05	-0,99	1,02	1,81
Rango	3,94	4,00	4,60	6,33	10,64
Mínimo	-2,26	-2,18	-3,11	-2,18	-3,98
Máximo	1,69	1,83	1,50	4,15	6,66
Número de casos	8974	8974	8974	8974	8974

**TABLA 27** Estadísticos descriptivos de los 5 factores extraídos en el periodo completo de estudio, variables sin datos anómalos y sin transformar

Y la prueba de normalidad de Kolmogorov-Smirnov aplicada para cada uno de los cinco factores da por resultado:

#### Pruebas de normalidad

	Kolmogorov-Smirnov(a)		
	Estadístico	gl	Sig.
Factor 1	,123	8974	,000
Factor 2	,092	8974	,000
Factor 3	,133	8974	,000
Factor 4	,081	8974	,000
Factor 5	,118	8974	,000

a Corrección de la significación de Lilliefors

**TABLA 28** Pruebas de normalidad de Kolmogorov-Smirnov para los cinco factores obtenidos en el periodo completo de estudio, información sin datos anómalos y variables sin transformar

Se observa de la Tabla (28) que ninguno de los factores se ajusta a una distribución normal.

#### 7.6.5 Periodo completo de estudio, variables transformadas

Se transformarán las 21 variables considerando el periodo completo de estudio. La transformación que se usará en todos los casos es la expuesta en el apartado 5.4, para aproximar su distribución a una normal.

a) Análisis factorial aplicado al conjunto de datos corregido de anómalos:

- Con el Método de componentes principales se obtienen los resultados que se muestran en el Anexo 15. Algunos aspectos se comentan en seguida y otros en el Capítulo 9.

**Varianza total explicada**

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	6,853	32,635	32,635	6,853	32,635	32,635
2	3,250	15,476	48,111	3,250	15,476	48,111
3	2,541	12,100	60,211	2,541	12,100	60,211
4	1,995	9,500	69,711	1,995	9,500	69,711
5	1,115	5,311	75,022	1,115	5,311	75,022
6	,975	4,642	79,664			
7	,797	3,795	83,459			
8	,713	3,394	86,853			
9	,440	2,097	88,950			
10	,434	2,066	91,016			
11	,360	1,716	92,732			
12	,298	1,420	94,153			
13	,269	1,282	95,435			
14	,226	1,076	96,511			
15	,202	,964	97,475			
16	,185	,882	98,357			
17	,159	,756	99,114			
18	,116	,553	99,667			
19	,041	,197	99,864			
20	,024	,115	99,979			
21	,004	,021	100,000			

Método de extracción: Análisis de Componentes principales.

**TABLA 29** Extracción de factores por el Método de componentes principales en el periodo completo de estudio, variables sin anómalos y transformadas

El determinante de la matriz de correlaciones tiene un valor  $2,63 \times 10^{-10}$  que está próximo a cero, lo que señala que las variables están relacionadas linealmente, por lo que es adecuado emplear el FA. La matriz de correlaciones (Anexo 15) posee elementos fuera de la diagonal principal con valores que garantizan que la matriz no está próxima a parecerse a una matriz identidad (una condición para que el FA sea eficiente).

El KMO tiene un valor de 0,808 que se puede considerar como bueno. La Prueba de esfericidad de Bartlett tiene un grado de significación de  $p = 0,000$ ,



por lo que se rechaza la hipótesis nula de esfericidad y se puede continuar con el análisis.

Se extrajeron en total cinco factores con autovalores mayores que la unidad.

Los cinco factores extraídos representan el 75,022% de la varianza total (un valor ligeramente inferior al de las variables sin transformar).

La Matriz de componentes sin los elementos con valor inferior a 0,25 es:

**Matriz de componentes(a)**

	Componente				
	1	2	3	4	5
DJ_T	,335	-,611		,460	
V_T			,762	,405	
DV_T	-,499			,297	
Vx_T			,599	,417	,511
T_T	,972				
Tn_T	,956				
Tx_T	,955				
Hf_T	-,612				-,371
HR_T			-,739	-,357	
Rad_T	,682	,476	,252		
P_T		-,310	-,390		,554
ETo_T	,842	,351	,315		
VolExp_T		,723		-,475	
Qent_T	-,805				
Qsal_T	,762	-,356			
Filt_T	-,359				
Profundidad_T		-,426	,430	-,638	
Ta_T	,671		-,328	,503	
Odis_T	-,502	,647	-,342	,329	
pH_T		,591	-,427		,363
Con_T		-,733	,269		

a 5 componentes extraídos

**TABLA 30** Matriz de componentes de los 5 factores extraídos por el Método de componentes principales, periodo de estudio completo, datos sin anómalos evidentes y transformados

Para que sea más sencilla la agrupación de las variables en factores se aplicará una Rotación Varimax:

- b) Extracción de los factores por el Método de Componentes principales, Rotación Varimax:

- En el primer factor están presentes las variables DV, T, Tn, Tx, Hf, Rad, ETo, Qent, Qsal y Filt. En el segundo factor quedan incorporados DJ y VolExp. En el tercero quedan incluidos la Profundidad, Ta, Odis, pH y Con. Para el cuarto factor se tienen V, Vx y HR. Y en el quinto factor satura principalmente la variable P.

**Matriz de componentes rotados(a)**

	Componente				
	1	2	3	4	5
DJ_T		,780			,258
V_T				,894	
DV_T	-,578				
Vx_T				,880	
T_T	,934	,264			
Tn_T	,912	,292			
Tx_T	,918	,253			
Hf_T	-,586	-,288			-,425
HR_T				-,752	,417
Rad_T	,765	-,317			-,354
P_T					,728
ETo_T	,896				
VolExp_T		-,874			
Qent_T	-,824				
Qsal_T	,626	,621			
Filt_T	-,413				
Profundidad_T			-,884		
Ta_T	,568	,407	,601		
Odis_T	-,450	-,351	,751		
pH_T		-,442	,574		,389
Con_T	-,265	,296	-,733		

Método de rotación: Normalización Varimax con Kaiser.

**TABLA 31** Matriz de componentes de los 5 factores extraídos por el Método de componentes principales, periodo de estudio completo, variables transformadas y Rotación Varimax

Al representar el quinto factor primariamente a una sola variable (P) es deseable que se cumpla con el principio de parsimonia, por lo que se busca que un factor incluya mínimamente dos variables. Por lo tanto, se efectúa nuevamente la Rotación Varimax pero con sólo 4 factores.

Las variables se agruparon así en los factores: en el 1, DV, T, Tn, Tx, Hf, Rad, ETo, Qent, Qsal, Filt y Ta. Para el 2 están DJ y VolExp. En el 3 Profundidad, Odis, pH y Con. Y en el 4, V, Vx, HR y P.

Por lo que, en apariencia, la transformación de las variables genera una reducción en el número de factores que explican la información. El modelo posiblemente se hace más simple.

**Matriz de componentes rotados(a)**

	Componente			
	1	2	3	4
DJ_T		,818		
V_T				,892
DV_T	-,570			
Vx_T				,742
T_T	,932	,275		
Tn_T	,908	,307		
Tx_T	,920	,258		
Hf_T	-,529	-,394		
HR_T				-,852
Rad_T	,798	-,358		
P_T	-,268	,279		-,373
ETo_T	,907			,313
VolExp_T		-,836		
Qent_T	-,833			
Qsal_T	,631	,602		
Filt_T	-,385			
Profundidad_T			-,852	
Ta_T	,580	,447	,550	
Odis_T	-,440	-,288	,786	
pH_T		-,301	,654	
Con_T	-,294	,259	-,737	

Método de rotación: Normalización Varimax con Kaiser.

**TABLA 32** Matriz de componentes de 4 factores extraídos por el Método de componentes principales, periodo de estudio completo, variables transformadas y Rotación Varimax

c) Extracción de los factores por el Método de la máxima verosimilitud, incluyendo una rotación con el Método Varimax

- Los resultados generados se incluyen en el Anexo 12.
- Al utilizar el Método de máxima verosimilitud se pueden desarrollar sólo 14 iteraciones y se observa que las comunalidades iniciales de algunas variables tienen valores medios al comienzo y bajan aún más durante la extracción. Debido a esto, todas las conclusiones que se obtengan deben ser tomadas con precaución.
- En 16 variables es más alta la comunalidad (que es la porción de la varianza

de una variable que recibe la contribución de los factores) obtenida por el Método de componentes principales que por el de máxima verosimilitud. Las únicas excepciones son T, Tn, Tx, Rad y ETo. Por lo tanto, es mejor utilizar el Método de componentes principales pues se explica mejor la varianza de un mayor número de variables.

**Comunalidades(a)**

Variable	Inicial	Extracción
DJ_T	,604	,513
V_T	,695	,514
DV_T	,396	,350
Vx_T	,611	,254
T_T	,994	,995
Tn_T	,981	,993
Tx_T	,980	,999
Hf_T	,601	,440
HR_T	,627	,586
Rad_T	,914	,878
P_T	,341	,144
ETo_T	,967	1,000
VolExp_T	,682	,700
Qent_T	,739	,636
Qsal_T	,742	,694
Filt_T	,186	,106
Profundidad_T	,664	,692
Ta_T	,785	,782
Odis_T	,826	,894
pH_T	,570	,471
Con_T	,583	,569

Método de extracción: Máxima verosimilitud.

a Se han encontrado una o más estimaciones de comunalidad mayores que 1 durante las iteraciones. La solución resultante deberá ser interpretada con precaución.

**TABLA 33** Comunalidades de las variables al extraerse los factores por el Método de máxima verosimilitud, para el periodo completo de estudio, variables sin transformar

En la Tabla (33) se tiene que al efectuar la extracción el tercer factor toma un autovalor inferior a la unidad, lo que muestra la inestabilidad de la solución por el Método de máxima verosimilitud. La varianza total explicada por los cinco factores también disminuye significativamente (más de un 12% de la variabilidad total).

La varianza total explicada por cada factor se muestra en la tabla siguiente:

**Varianza total explicada**

Factor	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	6,853	32,635	32,635	5,586	26,601	26,601
2	3,250	15,476	48,111	3,119	14,851	41,452
3	2,541	12,100	60,211	,451	2,149	43,600
4	1,995	9,500	69,711	2,401	11,432	55,033
5	1,115	5,311	75,022	1,655	7,882	62,915
6	,975	4,642	79,664			
7	,797	3,795	83,459			
8	,713	3,394	86,853			
9	,440	2,097	88,950			
10	,434	2,066	91,016			
11	,360	1,716	92,732			
12	,298	1,420	94,153			
13	,269	1,282	95,435			
14	,226	1,076	96,511			
15	,202	,964	97,475			
16	,185	,882	98,357			
17	,159	,756	99,114			
18	,116	,553	99,667			
19	,041	,197	99,864			
20	,024	,115	99,979			
21	,004	,021	100,000			

Método de extracción: Máxima verosimilitud.

**TABLA 34** Varianza total explicada por cada factor para el periodo completo de estudio, sin datos anómalos, variables sin transformar y Método de extracción de máxima verosimilitud

La matriz factorial que se obtuvo con 14 iteraciones usando el método de máxima verosimilitud y considerando cinco factores en total, es la que se muestra en la Tabla (35). Se extrajeron cinco factores en total y se puede apreciar como se agrupan las variables en cada uno de los factores. Como se puede apreciar, la pertenencia de cada una de las variables a un factor no está bien determinada en algunos casos, pues los valores absolutos son de magnitud cercana entre sí.

**Matriz factorial(a)**

	Factor				
	1	2	3	4	5
DJ_T		,506			,456
V_T		-,473			,498
DV_T	-,435				,272
Vx_T		-,280			,395
T_T	,804	,577			
Tn_T	,748	,594	,283		
Tx_T	,808	,585			
Hf_T	-,372	-,545			
HR_T	-,272	,434			-,498
Rad_T	,900				
P_T	-,316				
Eto_T	1,000				
VolExp_T	,259	-,519			-,555
Qent_T	-,761				
Qsal_T	,466	,569			,330
Filt_T	-,277				
Profundidad_T				-,765	-,316
Ta_T	,476	,397		,530	,323
Odis_T	-,338	-,394		,760	
pH_T				,585	
Con_T	-,307			-,673	

a Se han intentado extraer 5 factores. Requidas más de 14 iteraciones. (Convergencia=,019).

**TABLA 35** Matriz factorial de los 5 factores extraídos por Máxima verosimilitud con cinco iteraciones, periodo de estudio completo, datos sin anómalos y transformados

Para interpretar de forma más sencilla se efectúa una Rotación Varimax:

**Matriz de factores rotados(a)**

	Factor				
	1	2	3	4	5
DJ_T		,690			
V_T				,708	
DV_T	-,520				
Vx_T				,500	
T_T	,952	,262			
Tn_T	,939	,294			
Tx_T	,924				,279
Hf_T	-,535	-,292			
HR_T				-,751	
Rad_T	,765	-,410		,311	
P_T					
ETo_T	,911			,349	
VolExp_T		-,813			
Qent_T	-,789				
Qsal_T	,609	,563			
Filt_T	-,315				
Profundidad_T			-,814		
Ta_T	,582	,367	,552		
Odis_T	-,439	-,309	,778		
pH_T		-,353	,553		
Con_T	-,276	,257	-,652		

**TABLA 36** Matriz de factores rotados de los 5 factores extraídos por Máxima verosimilitud para cinco iteraciones, periodo de estudio completo, variables transformadas y Rotación Varimax

---

Los factores están constituidos así: el 1 por DV, T, Tn, Tx, Hf, Rad, ETo, Qent, Qsal, Filt y Ta. El 2 incluye a DJ y VolExp. El 3 incorpora a la Profundidad, Odis, pH y Con. En el 4 están V, Vx y HR. En el 5 sólo satura débilmente Tx. La variable P no saturó con un valor superior a 0,25 en ninguno de los cinco factores.

Por lo tanto, los Métodos de componentes principales y máxima verosimilitud sólo difirieron al ubicar dos variables (Ta y P), por lo que la formación de los factores con variables transformadas parece ser estable al mostrar alta coincidencia ambos métodos.

#### 7.6.6 Periodo de calibración, variables transformadas

Con el propósito de comprobar la estabilidad de la solución obtenida para el periodo de estudio se aplica el FA por el Método de componentes principales y se aplicará una rotación por el Método Varimax en los periodos de calibración y validación. Las partes del FA que se compararán son:

- Matriz de correlaciones.
- Comunalidades de las variables.
- Varianza total explicada por los factores.
- Matriz de componentes.
- Matriz de componentes rotados.

Para el periodo de calibración las cinco secciones anteriores se incluyen en el Anexo 16. Respecto a las comunalidades el 66,7% muestran una alta correspondencia en sus valores dentro de las variables, el 14,3% muestran una coincidencia regular y el 19,0% restante no tienen similitud entre sí.

La varianza total explicada por los factores extraídos (cinco en los dos casos) muestra grandes similitudes y en este aspecto el FA efectuado para el periodo de estudio completo sí es estable.

Al efectuar la Rotación Varimax se tiene que sólo tres variables cambiaron de factor: Ta, Qsal y Hf. Debido a esto, la estabilidad en la formación de los factores se puede considerar como buena.

**Matriz de componentes rotados(a)**

	Componente				
	1	2	3	4	5
DJ_T			,829		
V_T				,879	
DV_T	-,598				
Vx_T				,871	
T_T	,953				
Tn_T	,935				
Tx_T	,948				
Hf_T	-,474		-,482		-,524
HR_T				-,775	,428
Rad_T	,810		-,336		
P_T	-,255		,259	-,250	,449
ETo_T	,900			,305	
VolExp_T	,388		-,828		
Qent_T	-,748				,344
Qsal_T	,583		,650		
Filt_T	-,387				
Profundidad_T		,850			
Ta_T	,720	-,448	,389		
Odis_T	-,490	-,804			
pH_T		-,585			,547
Con_T		,867			

Método de extracción: Análisis de componentes principales.

Método de rotación: Normalización Varimax con Kaiser.

**TABLA 37** Matriz de factores rotados de los 5 factores extraídos por el Método de componentes principales, periodo de calibración, variables transformadas y Rotación Varimax

### 7.6.7 Periodo de validación, variables transformadas

Se desarrollará el FA en los datos de validación, obteniéndose (Anexo 17):

- Matriz de correlaciones.
- Comunalidades de las variables.
- Varianza total explicada por los factores.
- Matriz de componentes.
- Matriz de componentes rotados.

En la comparación de las comunalidades de los periodos completo y de validación se observa que hay muy buenas correspondencias en el 66,7% de las variables, un 14,3% de valores medianamente cercanos entre sí y el remanente 19,0% sin correspondencia entre sí.



Respecto a la varianza total explicada, en el periodo completo se extrajeron 5 factores, mientras que en la validación fueron 6. Esto se debe a que el sexto factor del periodo completo tiene un autovalor de 0,975 y es rechazado por el criterio, pero se encuentra próximo al sexto factor de la validación con 1,148.

En la validación cada uno de los factores explica más varianza total que en el caso del periodo completo pero la diferencia en realidad no es muy marcada. En este aspecto se puede considerar que el FA es suficientemente estable.

Con el fin de que los factores sean comparables se extraerán sólo 5 factores para comparar objetivamente el periodo completo y el de validación. Los resultados de la Rotación Varimax son:

**Matriz de componentes rotados(a)**

	Componente				
	1	2	3	4	5
DJ_T	,937				
V_T				,934	
DV_T					-,541
Vx_T				,880	
T_T	,873		,361		
Tn_T	,876				
Tx_T	,846		,405		
Hf_T	-,500		-,353	,251	-,291
HR_T			-,444	-,685	
Rad_T			,869		
P_T			-,613		,354
ETo_T	,510		,763		
VolExp_T	-,824		,339		,261
Qent_T	-,693				-,496
Qsal_T	,859				
Filt_T					-,680
Profundidad_T		-,880			
Ta_T	,648	,708			
Odis_T	-,566	,724			
pH_T	-,457	,617			
Con_T		-,844			

Método de extracción: Análisis de componentes principales. Se extrajeron sólo 5 factores.

Método de rotación: Normalización Varimax con Kaiser.

**TABLA 38** Matriz de factores rotados de 5 factores extraídos por el Método de componentes principales, periodo de validación, variables transformadas y Rotación Varimax

Se tiene que al pasar del periodo completo de estudio al de validación en total siete variables cambian de factor: DJ, VolExp, Rad, P, ETo, DV y Filt. Se puede considerar

la estabilidad de la formación de los factores como regular y el uso de los factores en el periodo de validación se debe hacer con precaución, pues los resultados podrían tener variaciones significativas.

#### 7.6.8 Normalidad de las puntuaciones factoriales extraídas, variables transformadas, periodo completo de estudio

Los estadísticos descriptivos de los cinco factores extraídos en el periodo completo de estudio, usando las variables transformadas, son:

	<i>Factor 1</i>	<i>Factor 2</i>	<i>Factor 3</i>	<i>Factor 4</i>	<i>Factor 5</i>
Media	0,000	0,000	0,000	0,000	0,000
Error típico	0,011	0,011	0,011	0,011	0,011
Mediana	0,224	0,128	-0,054	0,012	-0,128
Moda	-0,590	-0,577	-0,615	0,357	-0,470
Desviación estándar	1,000	1,000	1,000	1,000	1,000
Varianza de la muestra	1,000	1,000	1,000	1,000	1,000
Curtosis	-1,04	-0,10	1,13	-0,06	2,84
Coefficiente de asimetría	-0,41	-0,52	0,23	0,03	0,92
Rango	4,44	6,49	9,02	7,08	9,77
Mínimo	-2,63	-4,27	-4,77	-2,98	-4,22
Máximo	1,82	2,21	4,24	4,10	5,55
Número de casos	8974	8974	8974	8974	8974

**TABLA 39** Estadísticos descriptivos de los 5 factores extraídos en el periodo completo de estudio, variables sin datos anómalos y transformadas

Se observa que el efecto de transformar las variables es el de suavizar los picos en la mediana, moda, curtosis (marcadamente) y el coeficiente de asimetría. El rango y el mínimo empeoran sus valores en casi todos los factores.

La prueba de normalidad de Kolmogorov-Smirnov para los factores proporciona los resultados siguientes:

#### Pruebas de normalidad

	Kolmogorov-Smirnov(a)		
	Estadístico	gl	Sig.
Factor 1	,104	8974	,000
Factor 2	,054	8974	,000
Factor 3	,032	8974	,000
Factor 4	,013	8974	,001
Factor 5	,061	8974	,000

a Corrección de la significación de Lilliefors

**TABLA 40** Pruebas de normalidad de Kolmogorov-Smirnov para los cinco factores obtenidos en el periodo completo de estudio, información sin datos anómalos y variables transformadas

---

---

En la Tabla (40) se aprecia que ninguno de los cinco factores se ajusta a una distribución normal, pero en el factor 4 se ve que el nivel de significación llega a un valor de 0,001, el más alto que se ha obtenido hasta el momento. Además, cada uno de los valores del estadístico de Kolmogorov-Smirnov es menor en cada uno de los factores (ver también Tabla 28) y debido a que cuanto más pequeño es el estadístico y más grande por tanto el nivel de significación, mejor es el ajuste de los datos a una ley normal (Visauta, 2002).

Así, se puede asegurar que el efecto de la transformación de las variables es acercar ligeramente la distribución de los factores a una distribución normal.

## 7.7 ANÁLISIS CLUSTER

### 7.7.1 Exploración inicial, datos sin anómalos y sin transformar - Periodo completo de estudio

Se iniciará analizando el periodo completo de estudio estando removidos los datos anómalos y usando las variables sin transformar. Será aplicado el algoritmo de K-medias, pues es un método de agrupación no jerárquica y este tipo de algoritmos son útiles para grandes conjuntos de observaciones (como es el caso) (Pérez, 2005; McNeil et al., 2005) debido a que los métodos no jerárquicos pueden aplicarse a conjuntos de datos mucho mayores que las técnicas jerárquicas (Johnson et al., 1992), además de ser el método más importante dentro de las técnicas no jerárquicas, debido a que es el más utilizado y el que proporciona los mejores resultados (Pérez, 2005). Antes de aplicar el algoritmo de K-medias cada una de las variables será estandarizada.

Además, se revisará si existen variables altamente correlacionadas. Si existe un conjunto de variables con alta correlación sólo se dejará para el estudio la variable más representativa o importante de ese conjunto.

Al observar el gráfico de dispersión matricial (Anexo 6) y la matriz de correlaciones del periodo completo de estudio (Anexo 12) se tiene que las variables T, Tn y Tx están altamente correlacionadas, al igual que ETo y Rad.

Debido a esto, se toma la decisión de eliminar de todos los análisis a las variables Tn (Temperatura mínima del aire) y Tx (Temperatura máxima del aire) y conservar a T (Temperatura media del aire) por ser de entre estos tres parámetros T un valor medio del día y no un valor extremo.

Asimismo, se prescindirá de Rad (Radiación solar global) y se conservará ETo (Evapotranspiración de referencia) por ser esta última una magnitud más común en la medición de parámetros meteorológicos.

Se aplicó el algoritmo de K-medias retirando esas tres variables. El software SPSS sirvió para este propósito. Para seleccionar el número de grupos se efectuó una prueba F de reducción de variabilidad comparando la suma de cuadrados dentro de los grupos con  $K$  grupos y  $K+1$  grupos y calculando la reducción relativa de variabilidad al aumentar un conglomerado. Se usó la ecuación (51) y la regla práctica de Hartigan.

Número de conglomerados	Suma de cuadrados dentro de los grupos	Valor de F
10	10141,61	
11	9292,71	818,78
50	2649,89	
51	2626,66	78,9
200	743,82	
201	740,91	34,5
500	311,84	
501	311,28	15,4

**TABLA 41** Criterio de selección del número de conglomerados para el periodo completo de datos

La aplicación del criterio anterior conlleva dos problemas: 1) el número de conglomerados resultante es muy grande (por lo tanto, difícil de manejar e implica una cantidad muy grande de regresiones) y 2) existen conglomerados que sólo poseen unos pocos elementos.

Además, se observa en los Anexos 2 y 3 que para las variables Hf (Horas frío), P (Precipitación total diaria) y Filt (Filtración) la mayoría de sus valores son iguales a cero, por lo que no sirven para clasificar ni tampoco muestran la variabilidad necesaria para estar incluidos en los análisis multivariantes. Debido a esto, serán removidos en los procedimientos a efectuar en esta sección desde este punto.

Después de remover las tres variables anteriores se efectúa nuevamente la aplicación del algoritmo de K-medias, pero se tiene que para 501 grupos el valor de F es de 19,3. Asimismo, se sigue con el problema de conglomerados muy pequeños (con pocos registros).

Debido a lo anterior, se decide cambiar el criterio: se pondrá como restricción que la

cantidad mínima de observaciones o registros en cada conglomerado sea aproximadamente de 30. Al comparar la distribución de observaciones para 14 y 15 conglomerados se observa que la que satisface mejor el criterio es la correspondiente a 15 conglomerados.

Conglomerado	Número de casos en cada conglomerado	Conglomerado	Número de casos en cada conglomerado
1	1047	1	553
2	<b>67</b>	2	390
3	92	3	990
4	867	4	133
5	112	5	2260
6	380	6	1184
7	2151	7	76
8	613	8	<b>27</b>
9	627	9	763
10	291	10	967
11	1261	11	67
12	291	12	100
13	459	13	740
14	716	14	418
Válidos	8974	15	306
		Válidos	8974

**TABLA 42** Comparación del número de casos para 14 y 15 conglomerados en el periodo completo de estudio

El resultado del CA eligiendo 15 conglomerados se muestra en el Anexo 18.

El resultado del número de conglomerados, 15, no se tomará como un valor fijo sino como un valor orientativo del número de grupos en los análisis que se realizarán en los siguientes apartados.

#### 7.7.2 Formación de conglomerados para predicción de calidad del agua – Periodo de calibración, variables sin transformar

En este apartado se formarán conglomerados con las siguientes variables consideradas como predictivas (11 variables estandarizadas): DJ, V, DV, Vx, T, HR, ETo, VolExp, Qent, Qsal y Profundidad. Para el número de conglomerados se continuará con el criterio propuesto de que el menor número de elementos en un grupo sea alrededor de 30.

Al aplicar el algoritmo de K-medias el número de conglomerados que más se ajustan al criterio es de 19. La distribución de los registros en los conglomerados se muestra en la siguiente tabla:

Conglomerado	Número de casos en cada conglomerado
1	229
2	92
3	24
4	126
5	254
6	243
7	424
8	45
9	273
10	320
11	241
12	507
13	227
14	143
15	50
16	46
17	508
18	516
19	133
Válidos	4401

**TABLA 43** Número de conglomerados y distribución de casos, periodo de calibración, variables sin transformar

Los centros de cada uno de los 19 conglomerados obtenidos se muestran en el Anexo 19.

### 7.7.3 Formación de conglomerados para predicción de calidad del agua – Periodo de validación, variables sin transformar

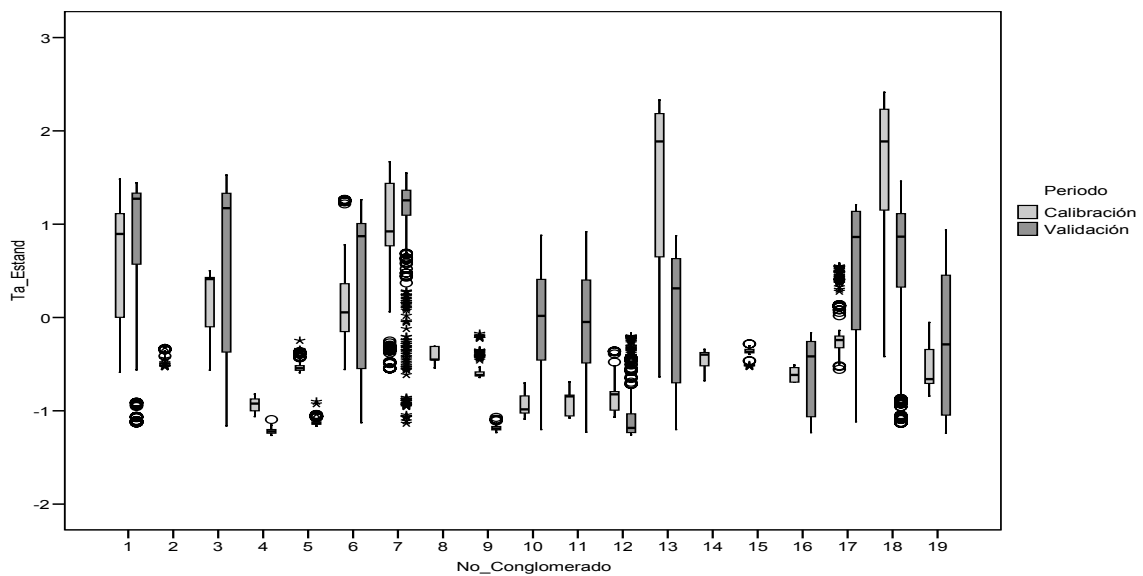
Una vez formados los 19 conglomerados en el apartado anterior y determinados sus centros, se procede a realizar la agrupación de los datos en el periodo de validación con las mismas 11 variables predictivas estandarizadas (DJ, V, DV, Vx, T, HR, ETo, VolExp, Qent, Qsal y Profundidad), utilizando en las opciones del análisis cluster del SPSS 19 conglomerados y como centros de las agrupaciones los calculados en el apartado anterior.

La distribución de observaciones obtenida al aplicar al procedimiento de *K*-medias en los 19 conglomerados es la que se muestra en la Tabla (44).

Se observa en la tabla anterior que para el periodo de validación las observaciones se agrupan sólo en 15 de los 19 conglomerados formados para el periodo de calibración. Todas las observaciones del periodo de validación fueron clasificadas en algún conglomerado.

Conglomerado	Número de casos en cada conglomerado
1	245
2	0
3	334
4	198
5	305
6	271
7	569
8	0
9	217
10	311
11	395
12	446
13	264
14	0
15	0
16	61
17	272
18	475
19	210
Válidos	4573

**TABLA 44** Número de conglomerados y distribución de casos, centros los obtenidos para el periodo de calibración, observaciones del periodo de validación, variables sin transformar



**FIGURA 51** Gráficos de caja y bigotes agrupados por periodos de calibración y validación, para temperatura del agua estandarizada, variables sin transformar y estandarizadas

Por lo tanto, al ser establecidos sólo 15 conglomerados para el periodo de validación se está regresando al número establecido para el periodo completo de estudio. Para este caso el periodo con un menor número de conglomerados establece dicha cantidad también para los otros periodos de análisis.

La comparación de la variable Temperatura del agua (estandarizada) se realizó para cada uno de los conglomerados, tanto para el periodo de calibración como el de validación (Figura 51).

Comparaciones como la anterior para las variables Ta, Odis, pH y Con se realizan para los siguientes casos:

- Variables sin transformar.
- Variables transformadas.
- Puntuaciones en componentes principales de variables sin transformar.
- Puntuaciones factoriales (sólo factores mayores a uno) de variables sin transformar.

Se plantearon estos tratamientos para tener la posibilidad de evaluar la efectividad de cada uno de ellos de manera individual (no combinados). Estas se muestran en los Anexos 21, 22, 23 y 24.

#### 7.7.4 Formación de conglomerados para predicción de calidad del agua – Periodo de calibración, variables transformadas

Conglomerado	Número de casos en cada conglomerado
1	582
2	667
3	389
4	271
5	247
6	122
7	189
8	345
9	<b>37</b>
10	383
11	682
12	75
13	297
14	71
15	44
Válidos	4401

**TABLA 45** Número de conglomerados y distribución de casos, periodo de calibración, variables transformadas y sin estandarizar

Se formarán conglomerados con las siguientes variables predictivas (11 variables, sólo se transformarán sin estandarizar): DJ, V, DV, Vx, T, HR, ETo, VolExp, Qent, Qsal y Profundidad. Se tomará el criterio propuesto de que el menor número de elementos en un grupo sea alrededor de 30.



---

Con el algoritmo de K-medias el número de conglomerados que más se ajustan al criterio es de 15.

Los centros obtenidos de este análisis se usarán en el próximo apartado.

#### 7.7.5 Formación de conglomerados para predicción de calidad del agua – Periodo de validación, variables transformadas

Una vez formados los 15 conglomerados del apartado anterior se procede a realizar la agrupación de los datos en el periodo de validación con las mismas 11 variables predictivas sin estandarizar (DJ, V, DV, Vx, T, HR, ETo, VolExp, Qent, Qsal y Profundidad), utilizando en las opciones del análisis cluster 15 conglomerados y como centros los calculados en el apartado anterior.

Se tiene que las observaciones se clasifican, en el periodo de validación, sólo en 12 de los 15 conglomerados generados en el periodo de calibración. Por el momento, la ventaja que se ve al transformar las variables es que se tienen que procesar un menor número de conglomerados.

La comparación de los valores transformados de las variables respuesta de calidad del agua (Ta, Odis, pH y Con) en los conglomerados, se muestra en el Anexo 22.

La distribución de observaciones obtenida al aplicar al procedimiento de *K*-medias en los 15 conglomerados es la siguiente:

Conglomerado	Número de casos en cada conglomerado
1	334
2	544
3	67
4	189
5	408
6	0
7	0
8	257
9	678
10	724
11	667
12	28
13	327
14	0
15	350
Válidos	4573

**TABLA 46** Número de conglomerados y distribución de casos, centros los obtenidos para calibración, observaciones del periodo de validación, variables transformadas y sin estandarizar

---

---

### 7.7.6 Formación de conglomerados para predicción de calidad del agua – Periodo de calibración, puntuaciones en componentes principales de variables sin transformar

Los conglomerados se formarán con las puntuaciones en componentes principales de las siguientes variables predictivas estandarizadas (11 variables): DJ, V, DV, Vx, T, HR, ETo, VolExp, Qent, Qsal y Profundidad. El criterio propuesto es que el menor número de elementos en un grupo sea alrededor de 30.

Con el algoritmo de K-medias el número de conglomerados que más se ajustan al criterio es de 18.

Conglomerado	Número de casos en cada conglomerado
1	408
2	246
3	50
4	232
5	227
6	342
7	69
8	683
9	92
10	247
11	45
12	575
13	292
14	41
15	386
16	273
17	<b>24</b>
18	169
Válidos	4401

**TABLA 47** Número de conglomerados y distribución de casos, periodo de calibración, puntuaciones en componentes principales de variables sin transformar y estandarizadas

Los centros de los conglomerados se conservan para el apartado siguiente.

### 7.7.7 Formación de conglomerados para predicción de calidad del agua – Periodo de validación, puntuaciones en componentes principales de variables sin transformar

Con los centros de los 18 conglomerados del apartado anterior se procede a realizar la agrupación de las puntuaciones en componentes principales de los datos en el periodo de validación con las mismas 11 variables predictivas sin estandarizar (DJ, V, DV, Vx, T, HR, ETo, VolExp, Qent, Qsal y Profundidad) y utilizando en las opciones del análisis cluster 18 conglomerados y como centros los calculados en el apartado anterior.

La distribución de observaciones al aplicar al procedimiento de *K*-medias en los 18 conglomerados se muestra en la Tabla (48). Se tiene que las observaciones se clasifican en el periodo de validación en 17 de los 18 conglomerados.

Conglomerado	Número de casos en cada conglomerado
1	403
2	198
3	67
4	284
5	323
6	169
7	0
8	452
9	61
10	125
11	269
12	355
13	312
14	438
15	190
16	347
17	349
18	231
Válidos	4573

**TABLA 48** Número de conglomerados y distribución de casos, periodo de calibración, puntuaciones en componentes principales de variables sin transformar y estandarizadas

La comparación de los valores de las variables respuesta de calidad del agua estandarizadas (Ta, Odis, pH y Con) dentro de los conglomerados, se muestra en el Anexo 23.

#### 7.7.8 Formación de conglomerados para predicción de calidad del agua – Periodo de calibración, puntuaciones factoriales de las variables sin transformar

Para agrupar los conglomerados se usarán las puntuaciones factoriales de los factores con autovalor superior a la unidad, para las siguientes variables predictivas estandarizadas (11 variables): DJ, V, DV, Vx, T, HR, ETo, VolExp, Qent, Qsal y Profundidad. Se extrajeron 3 factores. El criterio propuesto es que el menor número de elementos en un grupo sea alrededor de 30.

El análisis factorial fue realizado con el Método de componentes principales, se aplicó una Rotación Varimax y las puntuaciones factoriales se obtuvieron por el Método de regresión.

Al aplicar el algoritmo de K-medias el número de conglomerados que se ajusta más al criterio es 14.

Conglomerado	Número de casos en cada conglomerado
1	154
2	41
3	331
4	397
5	230
6	567
7	146
8	614
9	186
10	466
11	92
12	578
13	46
14	553
Válidos	4401

**TABLA 49** Número de conglomerados y distribución de casos, periodo de calibración, puntuaciones factoriales de variables sin transformar y estandarizadas

Los centros de estos conglomerados se emplearán en el siguiente apartado.

7.7.9 Formación de conglomerados para predicción de calidad del agua – Periodo de validación, puntuaciones factoriales de las variables sin transformar

Conglomerado	Número de casos en cada conglomerado
1	356
2	322
3	345
4	889
5	276
6	258
7	78
8	150
9	327
10	294
11	399
12	392
13	61
14	426
Válidos	4573

**TABLA 50** Número de conglomerados y distribución de casos, periodo de validación, puntuaciones factoriales de variables sin transformar y estandarizadas

Con los 14 centros anteriores se procede a formar los conglomerados para el periodo de validación. Se usarán las puntuaciones factoriales de las variables predictivas DJ, V, DV, Vx, T, HR, ETo, VolExp, Qent, Qsal y Profundidad, obtenidas con los mismos métodos del apartado anterior. Se extrajeron 4 factores con autovalor mayor a la unidad, pero sólo se considerarán los 3 mayores debido a que sólo se tienen 3

---

---

factores en el análisis factorial correspondiente a la calibración.

Aplicando el algoritmo de K-medias se obtiene la distribución de observaciones en los conglomerados que se muestra en la Tabla (50).

Las observaciones se clasifican para la validación en (exactamente) los 14 conglomerados que se formaron en el periodo de calibración.

La comparación de los valores de las variables respuesta de calidad del agua estandarizadas (Ta, Odis, pH y Con) en los conglomerados se muestra en el Anexo 24.

#### 7.7.10 Análisis cluster de variables– Variables estandarizadas y sin transformar

El CA de variables puede auxiliar en la reducción de la dimensión como lo hace el análisis factorial (Peña, 2002), pero el análisis de conglomerados de variables es menos restrictivo porque no exige linealidad (Gil Pascual, 2003).

Atendiendo a los periodos de calibración y validación en conjunto se obtendrán los dendogramas por cada uno de los conglomerados, con el objetivo de poder determinar en cada uno como se agrupan las variables. Para formar los conglomerados se utilizó el Método del promedio entre grupos o vinculación entre grupos como medida de distancia o disimilitud se empleó la distancia euclídea al cuadrado. Los resultados se muestran en el Anexo 25.

Se puede observar que para los distintos conglomerados la forma de asociarse de las variables es diferente en cada agrupamiento.

#### 7.7.11 Análisis cluster de variables – Variables transformadas

Se obtendrán los dendogramas en los periodos de calibración y validación (agrupados conjuntamente) para cada uno de los conglomerados. Como medida de disimilaridad se usó la distancia euclídea al cuadrado y como método para formar los conglomerados el de vinculación inter-grupos. Los dendogramas de los conglomerados se incluyen en el Anexo 26.

Se observa que para los distintos conglomerados las agrupaciones de las variables ocurren para cada uno de ellos de diferentes maneras.

---

---

## 7.8 ANÁLISIS DISCRIMINANTE

En esta sección se aplicará el DA a los registros conforme a la clasificación de conglomerados obtenida con el análisis cluster.

Como una observación importante las estandarizaciones de las variables (en el caso de que se usen) fueron efectuadas para los periodos de calibración y validación con los parámetros del periodo de calibración (media y desviación estándar) y luego agrupadas para el periodo completo de estudio. Esto se efectuó así para que el proceso de estandarización en general sólo dependiera de un conjunto de parámetros (los de calibración).

En todos los apartados siguientes se usó la opción de emplear la matriz de covarianzas intra-grupos en vez de la opción de utilizar la matriz de covarianzas de los grupos separados. Esta última opción proporciona mejores resultados, pero se prefirió usar la primera, por otorgar resultados más conservadores.

### 7.8.1 Periodo completo de estudio – Variables sin transformar

<b>Probabilidades previas para los grupos</b>		
No Grupo	Probabilidad previa (en %)	Casos utilizados en el análisis
1	5,28	474
2	1,03	92
3	3,99	358
4	3,61	324
5	6,23	559
6	5,73	514
7	11,07	993
8	0,50	45
9	5,46	490
10	7,03	631
11	7,09	636
12	10,62	953
13	5,47	491
14	1,59	143
15	0,56	50
16	1,19	107
17	8,69	780
18	11,04	991
19	3,82	343
<b>Total</b>	<b>100,00</b>	<b>8974</b>

**TABLA 51** Grupos y su probabilidad previa de ser clasificados correctamente por mero azar, periodo de estudio completo, variables sin transformar y no estandarizadas

Primero se efectuó el análisis discriminante sobre las variables sin estandarizar. Se

probaron los modos estándar y de inclusión paso a paso (*stepwise*). Debido a que en este último modo fueron incluidas todas las 11 variables predictivas consideradas (DJ, V, DV, Vx, T, HR, ETo, VolExp, Qent, Qsal y Profundidad), ambos modos presentan los mismos resultados en cuanto a clasificación de los registros en los grupos por lo que sólo se presentan los del modo estándar en el Anexo 27.

Los conglomerados considerados para el análisis fueron 19. Las probabilidades previas se calcularon según el tamaño de los grupos.

De inicio las probabilidades de clasificar correctamente un registro por mero azar son las que se muestran en la Tabla (51).

El resultado de la clasificación efectuada con las funciones discriminantes se muestra en la Tabla (52).

**Resultados de la clasificación**

No Grupo	Grupo de pertenencia pronosticado																		Total		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		19	
Recuento	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19		
1	383				8	42	41													474	
2		92																		92	
3			312		2				3				18						23	358	
4				221						80	6	17								324	
5	1		22		356	11	15		51					6			85	12		559	
6	98				10	371											35			514	
7	19		49		16	19	770										81	39		993	
8								45												45	
9					4				392	7	48		13						24	2	490
10				20					6	477	18	89							21		631
11									2	36	373	156							69		636
12										40	109	804									953
13									10				380						88	13	491
14														143							143
15															50						50
16																96				11	107
17					18		147							26			589				780
18									3										988		991
19									6	64		2								271	343

**TABLA 52** Resultados de la clasificación (Matriz de confusión), periodo de estudio completo, variables sin transformar y no estandarizadas

Para los 19 conglomerados los casos clasificados correctamente constituyen el 79,3% de los casos en los grupos originales. En comparación con las probabilidades previas de clasificación correcta este porcentaje se puede considerar como alto, lo que confirma la correcta formación de los conglomerados en el análisis cluster y también se disminuye la importancia del no cumplimiento de la homocedasticidad de las

---

variables independientes (indicada por los resultados de la Prueba M de Box, Anexo 27) (Gil Pascual, 2003).

Los casos clasificados correctamente mediante validación cruzada (opción: dejando uno fuera, Método jackknife) fue del 79,1%, valor que casi coincide con el anterior.

Si sólo se considera los grupos que tienen registros tanto en el periodo de calibración como en el de validación (15 grupos: 1, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 16, 17, 18 y 19), el porcentaje de clasificados correctamente es del 79,0% de los casos, por lo que prácticamente no existe cambio.

Después, al estandarizar las variables con los parámetros del periodo de calibración, se aplicaron nuevamente los modos estándar y de inclusión por pasos. En el modo de inclusión por pasos se incluyeron todas las 11 variables predictivas por lo que los dos modos aportan idénticas clasificaciones. Sólo se presentan los resultados del modo estándar en el Anexo 28 a).

Los grupos en el análisis fueron 19 y las probabilidades previas se calcularon según el tamaño de los grupos:

<b>Probabilidades previas para los grupos</b>		
No grupo	Probabilidades previas (en %)	Casos utilizados en el análisis
1	5,3	474
2	1,0	92
3	4,0	358
4	3,6	324
5	6,2	559
6	5,7	514
7	11,1	993
8	0,5	45
9	5,5	490
10	7,0	631
11	7,1	636
12	10,6	953
13	5,5	491
14	1,6	143
15	0,6	50
16	1,2	107
17	8,7	780
18	11,0	991
19	3,8	343
Total	100,0	8974

**TABLA 53** Grupos y su probabilidad previa de ser clasificados correctamente por mero azar, periodo de estudio completo, variables sin transformar y estandarizadas con los parámetros del periodo de calibración



La clasificación realizada por el DA es:

		Resultados de la clasificación																				
		Número de conglomerado	Grupo de pertenencia pronosticado																	Total		
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17		18	19
Original Recuento	1	382				11	41	40													474	
	2		92																		92	
	3			312		2				3					6				35		358	
	4				238						65	4	17								324	
	5				21	358	11	16		49						5		86	13		559	
	6		101				10	368										35			514	
	7		4		42		16	19	824										49	39		993
	8									45												45
	9						4				394	5	48		13					24	2	490
	10					19					6	487	21	77						21		631
	11										3	33	355	158						75	12	636
	12											39	108	791							15	953
	13										10					380				88	13	491
	14																143					143
	15																	50				50
	16																	96			11	107
	17						22		163							26			569			780
	18										5									986		991
	19										4	61		2							276	343

**TABLA 54** Resultados de la clasificación (Matriz de confusión), periodo de estudio completo, variables sin transformar y estandarizadas con los parámetros del periodo de calibración

Los registros clasificados de manera correcta constituyen un 79,6% de los registros agrupados originalmente. Es un 0,3% superior al caso de cuando las variables están sin estandarizar. Por tanto, se puede afirmar que para el análisis discriminante los resultados serán prácticamente los mismos si las variables se aportan sin estandarizar o estandarizadas para todo el periodo en estudio con los parámetros del periodo de calibración exclusivamente.

Los registros clasificados de manera correcta por validación cruzada (Método jackknife) fue de 79,4 %, valor muy próximo al antes mencionado.

Al tomar en cuenta sólo los conglomerados que se extienden en los dos periodos, el de calibración y validación (15 grupos: 1, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 16, 17, 18 y 19), el porcentaje de clasificaciones correctas es 78,9 %, un valor un poco inferior al obtenido considerando 19 grupos.

Por otra parte, se estandarizaron de forma univariante los periodos de calibración y validación por separado, luego se unieron en un solo conjunto para llevar a cabo el análisis discriminante y se aplicaron nuevamente los modos estándar y de inclusión por pasos. En el modo de inclusión por pasos se incluyeron todas las 11 variables

predictivas por lo que los dos modos aportan idénticas clasificaciones. Sólo se presentan los resultados del modo estándar en el Anexo 28 b).

Los grupos en el análisis fueron 19 y las probabilidades previas se calcularon según el tamaño de los grupos:

Probabilidades previas para los grupos		
No grupo	Probabilidades previas (en %)	Casos utilizados en el análisis
1	5,3	474
2	1,0	92
3	4,0	358
4	3,6	324
5	6,2	559
6	5,7	514
7	11,1	993
8	0,5	45
9	5,5	490
10	7,0	631
11	7,1	636
12	10,6	953
13	5,5	491
14	1,6	143
15	0,6	50
16	1,2	107
17	8,7	780
18	11,0	991
19	3,8	343
Total	100,0	8974

**TABLA 55** Grupos y su probabilidad previa de ser clasificados correctamente por mero azar, periodo de estudio completo, variables sin transformar y estandarizadas por separado en los periodos de calibración y validación

La clasificación realizada por el DA se muestra en la Tabla (56).

Los registros clasificados de manera correcta constituyen un 87,8% de los registros agrupados originalmente. Es un 8,5% superior al caso de cuando las variables están sin estandarizar. Esto puede estar ocurriendo, posiblemente, debido a que los conglomerados antes de ser formados por el análisis cluster son tipificados con el fin de que las diferentes escalas de medida y distintas unidades no influyan en la forma de agrupar. Debido a esto, al estar las variables estandarizadas se ajustan más al esquema tratado por el análisis cluster en su momento. El agrupamiento de los registros se tomó del efectuado para las variables tipificadas en el análisis cluster y este agrupamiento se usó en el DA, tanto para las variables originales como para las tipificadas, acercándose más a la estructura de ordenamiento de estas últimas.

		Resultados de la clasificación																				
	Número de conglomerado	Grupo de pertenencia pronosticado																		Total		
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		19	
Original	1	381				9	64	20													474	
Recuento	2		92																		92	
	3			334				1		1				1						21	358	
	4				308						3									13	324	
	5	2		11		402	12	17		34					15			55	11		559	
	6	38				1	475														514	
	7	36		40		9		908													993	
	8								45												45	
	9									378	17	56		15						22	2	490
	10									3	577	10	18							21	2	631
	11									2	35	455	82							60	2	636
	12										26	49	878									953
	13				11					13					379					88		491
	14															143						143
	15																50					50
	16				11										10		86					107
	17					10		19						43				708				780
	18							13		6										972		991
	19										37										306	343

**TABLA 56** Resultados de la clasificación (Matriz de confusión), periodo de estudio completo, variables sin transformar y estandarizadas por separado para los periodos de calibración y validación

Los registros clasificados de manera correcta por validación cruzada (Método jackknife) fue de 87,6 %, valor muy próximo al antes mencionado.

Al tomar en cuenta sólo los conglomerados que se extienden en los dos periodos, el de calibración y validación (15 grupos: 1, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 16, 17, 18 y 19), el porcentaje de clasificaciones correctas es 87,3 %, un valor un poco inferior al obtenido considerando 19 grupos.

### 7.8.2 Periodo completo de estudio – Variables transformadas

Se utilizaron tanto el modo estándar como el de inclusión por pasos. Debido a que en este último se incluyeron las 11 variables predictivas consideradas los dos métodos proporcionaron los mismos resultados. Por lo tanto, sólo se presentan los del modo estándar en el Anexo 29.

El número de grupos en el análisis fue de 15. Las probabilidades previas de los grupos se muestran en la Tabla (57).

Los casos clasificados correctamente en los 15 grupos es de 84,0% (Tabla 58) de los casos en los grupos originales. Este porcentaje se puede considerar como muy bueno en comparación de las probabilidades previas al azar según el tamaño del grupo.

El único grupo con un porcentaje de clasificación inferior al 50% fue el grupo 5 (42,4% de casos clasificados correctamente), por lo que habrá que tener precauciones al efectuar procesos donde intervenga este.

**Probabilidades previas para los grupos**

No de grupo	Probabilidad previa	Casos en el análisis
1	10,2	916
2	13,5	1211
3	5,1	456
4	5,1	460
5	7,3	655
6	1,4	122
7	2,1	189
8	6,7	602
9	8,0	715
10	12,3	1107
11	15,0	1349
12	1,1	103
13	7,0	624
14	0,8	71
15	4,4	394
Total	100	8974

**TABLA 57** Grupos y su probabilidad previa de ser clasificados correctamente por mero azar, periodo de estudio completo, variables transformadas

La clasificación realizada por el DA es la siguiente:

**Resultados de la clasificación**

Número del grupo	Grupo de pertenencia pronosticado															Total		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15			
1	893				22					1								916
2		1100	48	18							41		4					1211
3		78	362	16														456
4		37	20	403														460
5	194				279			11	27	57			73			14		655
6						107	15											122
7							189											189
8					13			499		90								602
9									585		122		8					715
10	13				65				26	927	66					10		1107
11									73	50	1225		1					1349
12												95				8		103
13			2	4					18	6	53		541					624
14														71				71
15					2				40	60	6	20				266		394

**TABLA 58** Resultados de la clasificación (Matriz de confusión), periodo de estudio completo, variables transformadas

El porcentaje obtenido por validación cruzada (Método jackknife, dejando uno fuera), fue de 83,9%, valor altamente cercano al anterior.

Al considerar sólo los grupos que poseen registros en el periodo de calibración y en el

de validación (12 grupos: 1, 2, 3, 4, 5, 8, 9, 10, 11, 12, 13 15) el porcentaje de clasificaciones correctas es el mismo, observándose una mejoría casi inapreciable en el grupo 5. Se puede concluir que no hay modificación.

### 7.8.3 Periodo completo de estudio – Puntuaciones en componentes principales de las variables sin transformar

Las puntuaciones en componentes principales (con estandarización) se obtuvieron por separado para calibración y validación, se aplicó la técnica del análisis cluster para obtener los conglomerados en el conjunto de calibración y los centros obtenidos fueron empleados para formar los conglomerados en el periodo de validación. Estos resultados serán revisados por medio del DA. La salida de este procedimiento se resume en el Anexo 30.

Los modos estándar y de inclusión por pasos (que incluyó a las 11 componentes) proporcionaron idénticos resultados por lo que sólo se presenta el modo estándar.

La cantidad de grupos en el análisis fue de 18. Las probabilidades previas correspondientes a los grupos se muestran en la Tabla (59).

Probabilidades previas para los grupos		
Número inicial de casos	Probabilidades previas	Casos utilizados en el análisis
1	9,0	811
2	4,9	444
3	1,3	117
4	5,7	516
5	6,1	550
6	5,7	511
7	0,8	69
8	12,6	1135
9	1,7	153
10	4,1	372
11	3,5	314
12	10,4	930
13	6,7	604
14	5,3	479
15	6,4	576
16	6,9	620
17	4,2	373
18	4,5	400
Total	100	8974

**TABLA 59** Grupos y su probabilidad previa de ser clasificados correctamente por mero azar, periodo de estudio completo, puntuaciones en componentes principales (con estandarización)

Los registros clasificados correctamente dentro de los grupos (Tabla 60) constituyen

un 88,1% de los originales. Este porcentaje se puede considerar como excelente. El porcentaje que se obtuvo por validación cruzada tiene un valor de 87,9%.

Si se toman en cuenta sólo los conglomerados que poseen registros tanto en el periodo de calibración como en el de validación (quedaría únicamente fuera el grupo 7) el porcentaje de clasificaciones correctas queda exactamente igual.

**Resultados de la clasificación**

	No inicial de casos	Grupo de pertenencia pronosticado																	Total	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17		18
Original	1	800	11																	811
Recuento	2	12	298			9	33		22			6		33		22			9	444
	3			117																117
	4				437					18			41				20			516
	5	25				502									22			1		550
	6	39			47	20	359		16						10	17		3		511
	7							69												69
	8		1						1079					36	13	6				1135
	9									147						6				153
	10				19						309	6	8				26		4	372
	11		7	11								296								314
	12				43		15		2		4	2	847			16		1		930
	13		28						26					528	22					604
	14								58					14	406			1		479
	15		1				11						28			536				576
	16									7			54				557		2	620
	17	4			8	2			46			3	8	14	6	1		281		373
	18	15	3		25											18			339	400

**TABLA 60** Resultados de la clasificación (Matriz de confusión), periodo de estudio completo, puntuaciones en componentes principales (con estandarización)

#### 7.8.4 Periodo completo de estudio – Puntuaciones factoriales de las variables sin transformar

Las puntuaciones factoriales se obtuvieron por separado para los periodos de calibración y validación. El análisis cluster se usó para formar los conglomerados en la calibración y los centros de estos se emplearon para agrupar a los registros de validación. La agrupación obtenida se analizará con el DA. Los resultados están en el Anexo 31.

Los modos estándar y de inclusión por pasos (que incluyó mediante su algoritmo a las puntuaciones en los factores 1,2 y 3) entregaron salidas de clasificación idénticas. Se tomará en cuenta únicamente el modo estándar.

La cantidad de grupos analizados fue de 14. Las probabilidades previas de una clasificación correcta en los grupos fueron:

**Probabilidades previas para los grupos**

No de grupo	Probabilidades previas (%)	Casos utilizados
1	5,7	510
2	4,0	363
3	7,5	676
4	14,3	1286
5	5,6	506
6	9,2	825
7	2,5	224
8	8,5	764
9	5,7	513
10	8,5	760
11	5,5	491
12	10,8	970
13	1,2	107
14	10,9	979
<b>Total</b>	<b>100</b>	<b>8974</b>

**TABLA 61** Grupos y su probabilidad previa de ser clasificados correctamente por mero azar, periodo de estudio completo, puntuaciones factoriales (tres factores)

Las clasificaciones correctas representan un 88,5% (Tabla 62). Esta proporción, considerando la cantidad de datos involucrados y el número de grupos, se considera excelente.

El porcentaje conseguido mediante validación cruzada (dejando uno fuera) fue de 88,4%. Por lo que, de todos los apartados de clasificación vistos en el análisis discriminante, este es el que proporciona los mejores resultados.

**Resultados de la clasificación**

	Número de conglomerado	Grupo de pertenencia pronosticado													Total		
		1	2	3	4	5	6	7	8	9	10	11	12	13		14	
Original	1	399							59			52					510
Recuento	2		320									43					363
	3			559						117							676
	4	4			1154							128					1286
	5			6		477					23						506
	6						788	14	23								825
	7						34	190									224
	8						12		752								764
	9			82	31					395						5	513
	10			44		44					616					56	760
	11	22										467				2	491
	12				125		20						825				970
	13													107			107
	14						29		21		38					891	979

**TABLA 62** Resultados de la clasificación (Matriz de confusión), periodo de estudio completo, puntuaciones factoriales (tres factores)

Los 14 conglomerados tienen elementos tanto en el periodo de calibración como en el de validación.

## 7.9 MODELO LINEAL DE REGRESIÓN MÚLTIPLE

Con el propósito de tener una referencia de comparación, se aplicará la regresión lineal múltiple (MLR) sobre los datos de calibración completos (usando al arrancar con el proceso todas las 17 variables predictivas), se obtendrá la ecuación correspondiente y se probará en el periodo de validación para contar con una idea del ajuste esperado y el error cometido. Este proceso se llevará a cabo para las 4 variables de calidad del agua de interés: temperatura del agua (Ta), oxígeno disuelto (Odis), el pH y conductividad (Con).

La calibración y la validación son a menudo utilizadas para medir el desempeño de un modelo. El rango de aplicabilidad del modelo está generalmente empatado al rango de condiciones usadas para calibrar, validar o probar el modelo (Deas et al., 2000). Para los distintos métodos multivariantes y el modelo mecanicista que se usarán para efectuar estimaciones es necesario tomar en cuenta que tanto las regresiones como los coeficientes del modelo mecanicista están supeditados al rango que poseen para esta aplicación las variables predictivas. Y por otra parte el hecho de efectuar los procesos de calibración y validación garantiza que para estas condiciones los resultados sean estables, no presentándose estos por mera casualidad.

El método que se usó en la selección de variables fue el de selección hacia delante que permitió obtener un modelo prácticamente con los mismos resultados, pero con menos variables, que es un modelo más simple e interpretable. Los resultados son:

Periodo	Variable	R <sup>2</sup> corregido	Error estándar de estimación
Calibración	Ta	0,746	2,768
	Odis	0,774	2,055
	pH	0,286	0,340
	Con	0,654	59,011
Validación	Ta	0,672	4,124
	Odis	0,658	2,810
	pH	0,380	0,473
	Con	0,592	76,569

**TABLA 63** Resultados del ajuste inicial efectuado para los periodos de calibración y validación de las variables de calidad del agua

Para el periodo de validación el R<sup>2</sup> corregido recibe el nombre de Q<sup>2</sup> (o R<sup>2</sup> corregido de



---

validación cruzada). Este se obtuvo por medio de la ecuación (116).

Es necesario aclarar en este punto que el coeficiente de determinación, tanto para el periodo de calibración como en el de validación, fue calculado en base a valores diarios (cada 24 horas) para cada profundidad correspondiente, por lo que se está evaluando la calidad de las estimaciones hechas a escala temporal diaria y a escala espacial unidimensional (en la dirección vertical hacia abajo). Esta forma de obtener el coeficiente de determinación será aplicada durante todo el desarrollo de la tesis.

El error estándar de estimación,  $S_e$ , cuando interviene más de una variable independiente se obtuvo con la expresión (Nolberto Sifuentes, 2002):

$$S_e = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - k - 1}} \quad (205)$$

Donde (Nolberto Sifuentes, 2002):

$S_e$  = Error estándar de estimación (en las mismas unidades de la variable observada).

$y_i$  = Valor observado de la variable.

$\hat{y}_i$  = Valor estimado de la variable.

$n$  = Número de observaciones.

$k$  = Número de variables predictivas.

Para los grupos formados mediante estandarización de variables (sin transformar), transformación de variables, puntuaciones en componentes principales y puntuaciones factoriales, se aplicará el MLR a los casos del grupo existentes en el periodo de calibración para ajustar el modelo y se comprobará en el de validación. Esto se hará para las 4 variables respuesta (Ta, Odis, pH y Con). En cada caso, se usarán los tratamientos más adecuados para las variables independientes y las variables respuesta. Sólo en el caso de grupos formados mediante las variables transformadas se usarán estas últimas dentro del proceso.

Si se calcula la ecuación de regresión sobre datos a los que previamente se les haya aplicado un tratamiento específico se debe tener en cuenta que las propiedades de los estimadores de los coeficientes de regresión se refieren a la forma tratada, no a los originales (Etxeberria, 1999).

Con el propósito de poder comparar realmente la efectividad de cada tratamiento y tipo de agrupamiento las medidas de adecuación y error cometido se calcularán con los

---

---

valores estimados de la variable respuesta en sus dimensiones y unidades originales. En otras palabras, se le aplicará el tratamiento inverso al que se le haya administrado, para comparar los resultados de la calibración y validación en las unidades originales.

#### 7.9.1 MLR en grupos formados con las variables estandarizadas

Al usar variables independientes en las que la mayoría de sus valores toman valor cero, se observó que algunas estimaciones estaban en valores extremos (arriba de  $1 \times 10^{14}$  °C para la temperatura del agua), por lo que se determinó excluir esas variables de todos los análisis (Hf, P y Filt), con el objetivo de contar con mayor estabilidad en las estimaciones.

Además, como las variables T, Tn y Tx presentan colinealidad, sólo se usará T. Asimismo, Rad y ETo también son colineales por lo que sólo intervendrá en los análisis ETo.

Aquí se consideran los 19 grupos formados en los apartados 7.7.2 (conglomerados formados por medio del algoritmo de K-medias en el análisis cluster para el periodo de calibración, con las variables sin transformar y aproximadamente 30 elementos por grupo como mínimo) y 7.7.3 (agrupamientos formados con el algoritmo de K-medias del análisis cluster para el periodo de validación, con las variables sin transformar y aproximadamente 30 elementos por conglomerado como mínimo). Las variables predictivas y las variables respuesta fueron probadas para un conglomerado en el periodo de calibración sin estandarizar las variables y luego estandarizadas, observándose que el coeficiente de determinación corregido adquirió el mismo valor en ambos casos, por lo que el procedimiento MLR fue llevado a cabo con las variables expresadas en su escala y unidades originales con el fin de que las ecuaciones de regresión sean más fáciles de utilizar al efectuar las sustituciones en ellas.

Para el primer conglomerado en la calibración se ajusta una ecuación por medio de regresión múltiple, la ecuación obtenida se usa para efectuar estimaciones en el mismo conglomerado de los valores del periodo de calibración y también para efectuar predicciones en el periodo de validación para los registros que fueron clasificados por el análisis cluster como del mismo grupo. Se continúa así hasta terminar con todos los conglomerados y una vez completas las estimaciones se comparan con los datos observados, obteniéndose el  $R^2$  corregido y el error estándar de estimación para cada una de las variables a estimar, separados por periodo correspondiente (calibración o validación).

Los resultados de este proceso de regresión por conglomerados fue el siguiente:

Periodo	Variable	R <sup>2</sup> corregido	Error estándar de estimación
Calibración	Ta	0,966	1,048
	Odis	0,944	1,042
	pH	0,722	0,214
	Con	0,928	27,014
Validación	Ta	0,477	6,786
	Odis	0,235	5,312
	pH	0,197	0,716
	Con	0,091	123,523

**TABLA 64** Resultados de la regresión múltiple efectuada para 15 grupos seleccionados (de 19) que poseen registros en los periodos de calibración y validación

Se puede observar en la tabla anterior que respecto a la modelación por periodo entero (calibración), el ajuste por conglomerados mejora en la calibración pero su calidad predictiva para el periodo de validación disminuye bastante.

Como conclusión inicial, el proceso anterior puede proporcionar buenos resultados para efectuar interpolación dentro de un conglomerado (por ejemplo, para recuperar valores faltantes), pero su uso en periodos distintos llevaría a inferencias de baja calidad.

#### 7.9.2 MLR en grupos formados con las variables transformadas

Se usarán los 15 grupos obtenidos en los apartados 7.7.4 (agrupamientos generados con el algoritmo de K-medias del análisis cluster en la calibración, utilizando variables transformadas y aproximadamente 30 elementos por agrupamiento como mínimo) y 7.7.5 (conglomerados formados con el algoritmo de K-medias del análisis cluster en la validación, empleando variables transformadas y aproximadamente 30 elementos por grupo como mínimo). Las variables predictivas serán en total 11: DJ, V, DV, Vx, T, HR, ETo, VolExp, Qent, Qsal y Profundidad. Las variables predictivas y las variables respuesta se usarán transformadas, de los periodos de calibración y validación por separado. Una vez obtenidos los resultados se les aplicará la transformación inversa a las variables respuesta, para obtener sus estadísticas referidas a sus unidades originales. Como se trata de un planteamiento predictivo, para realizar esta transformación inversa se usarán los parámetros obtenidos en el periodo de calibración para pronosticar los valores de las variables de calidad del agua en el otro periodo.

El procedimiento para obtener las estimaciones de las variables de calidad del agua en la calibración y la validación, además del  $R^2$  corregido y el error estándar de estimación, es similar al descrito en el apartado anterior (7.9.1).

Los índices de calidad de la predicción fueron los siguientes:

Periodo	Variable	$R^2$ corregido	Error estándar de estimación
Calibración	Ta	0,951	1,257
	Odis	0,909	2,843
	pH	0,742	0,207
	Con	0,925	27,765
Validación	Ta	0,017	$5,4 \times 10^{14}$
	Odis	0,000	$1,2 \times 10^{15}$
	pH	-0,002	0,8944
	Con	-0,002	$9,0 \times 10^{13}$

**TABLA 65** Resultados de la regresión múltiple efectuada para 12 grupos seleccionados (de 15) que poseen registros en los periodos de calibración y validación

Con respecto al caso de las variables estandarizadas, en el periodo de calibración todas las variables, excepto el pH, empeoran su estimación. Y en el periodo de validación los resultados obtenidos son de muy baja calidad.

### 7.9.3 MLR en grupos formados con puntuaciones en componentes principales

Las puntuaciones se obtendrán por medio de la matriz de correlación, usando las variables independientes (11): DJ, V, DV, Vx, T, HR, ETo, VolExp, Qent, Qsal y Profundidad. Las variables dependientes se expresarán en sus unidades originales.

Se usarán 17 conglomerados (de 18 totales) que poseen información tanto en el periodo de calibración como en el de validación. Las puntuaciones que se usarán en los cálculos provienen de los periodos de calibración y validación, obtenidas por separado para cada uno de los periodos.

Los resultados para el periodo de calibración y validación se muestran en la Tabla (66). Se puede concluir que existe un adecuado poder de interpolación dentro de los grupos en la calibración, pero en la validación la capacidad predictiva es pobre. Por tanto, este planteamiento puede servir para completar información faltante en el periodo para el que se hicieron las regresiones, pero no para pronosticar en otros periodos.

Periodo	Variable	R <sup>2</sup> corregido	Error estándar de estimación
Calibración	Ta	0,964	1,0508
	Odis	0,938	1,0803
	pH	0,720	0,2126
	Con	0,932	26,2916
Validación	Ta	0,105	29,1127
	Odis	0,245	51,3197
	pH	0,166	137,8221
	Con	0,100	1024,7778

**TABLA 66** Resultados de la regresión múltiple efectuada para los periodos de calibración y validación usando puntuaciones en componentes principales

#### 7.9.4 MLR en grupos formados con las puntuaciones factoriales

Las puntuaciones factoriales se obtuvieron para los tres primeros factores (con autovalor mayor a la unidad) usando las variables independientes del apartado anterior. Las variables dependientes se emplearon en sus unidades originales.

El número de conglomerados utilizado es de 14 (que abarcan el total de registros del estudio), los cuales se extienden en los dos periodos de análisis. Las puntuaciones se obtuvieron por separado para los dos periodos.

Los parámetros de calidad del ajuste son:

Periodo	Variable	R <sup>2</sup> corregido	Error estándar de estimación
Calibración	Ta	0,613	3,4163
	Odis	0,452	3,1985
	pH	0,341	0,3270
	Con	0,251	86,9016
Validación	Ta	0,001	25,8111
	Odis	0,027	19,1120
	pH	0,092	1,8434
	Con	0,000	567,4418

**TABLA 67** Resultados de la regresión múltiple efectuada para los periodos de calibración y validación usando puntuaciones factoriales (considerando tres factores)

De la tabla anterior se puede concluir que este planteamiento no es el más adecuado tanto para interpolación en el periodo de calibración, así como para pronóstico en el de validación.

---



---

## 7.10 ANÁLISIS DE CORRELACIÓN CANÓNICA

La correlación canónica es una generalización de la regresión lineal múltiple que considera las relaciones “espaciales” entre las variables que se requiere pronosticar. Se busca que dadas unas variables predictivas  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_{p_1})$  y unas variables respuesta  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_{p_2})$ , se requiere estimar el valor del vector respuesta de forma conjunta a partir de los patrones espaciales que mayor correlación presenten con el vector predictivo (Gutiérrez et al., 2004).

En este trabajo más que conocer como se relaciona el conjunto de variables respuesta con el conjunto de variables predictivas lo que se pretende es predecir los valores que adquiere cada una de las variables respuesta dados valores del conjunto de las variables predictivas.

El planteamiento sería que dadas dos muestras simultáneas de tamaño  $n$  (Gutiérrez et al., 2004):

$$\mathbf{X}^T = (\mathbf{x}_1, \dots, \mathbf{x}_{p_1})^T \qquad \mathbf{Y}^T = (\mathbf{y}_1, \dots, \mathbf{y}_{p_2})^T \qquad (206)$$

Se trata de encontrar dos subespacios proyectores  $(\{\mathbf{f}_1, \dots, \mathbf{f}_r\}$  y  $\{\mathbf{g}_1, \dots, \mathbf{g}_r\})$ , respectivamente, con  $r \leq \min\{p_1, p_2\}$ , para los dos espacios de datos:

$$\mathbf{u} = \mathbf{F}^T \mathbf{X}^T = \begin{pmatrix} \mathbf{f}_{11} & \dots & \mathbf{f}_{1p_1} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ \mathbf{f}_{r1} & \dots & \mathbf{f}_{rp_1} \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_{p_1} \end{pmatrix} \qquad \mathbf{v} = \mathbf{G}^T \mathbf{Y}^T \qquad (207)$$

De manera tal que las proyecciones  $\mathbf{u}_i$  y  $\mathbf{v}_i$  tengan correlación máxima para cada pareja  $i = j$  de vectores y estén incorrelacionadas para  $i \neq j$  (Gutiérrez et al., 2004).

Cuando ya se han obtenido los vectores canónicos en cada espacio, conociéndose un valor de  $\mathbf{X}$  para el vector predictivo se puede obtener una estimación de  $\mathbf{Y}$  procediendo así: se proyecta  $\mathbf{X}$  sobre el espacio de vectores canónicos de acuerdo con la ecuación (207) y se obtiene  $\mathbf{u}$ . Cada componente del vector proyectado está linealmente relacionada con las componentes de  $\mathbf{Y}$ . Después, se efectúan las regresiones entre estos parámetros con fundamento en los valores disponibles (Gutiérrez et al., 2004):

$$\mathbf{v}_i = \mathbf{a}_i \mathbf{u}_i + b_i, \quad i = 1, \dots, r \quad (208)$$

Así, se obtiene el valor estimado de  $\mathbf{v}$  y se invierte la proyección para calcular el vector de estimaciones en el espacio original:

$$\hat{\mathbf{y}} = (\mathbf{G}^T)^{-1} \hat{\mathbf{v}} \quad (209)$$

La inversa anterior sólo es posible calcularla cuando el número de variables respuesta es menor que el número de variables predictivas (Gutiérrez et al., 2004).

#### 7.10.1 CCA sobre periodos de calibración y validación completos

Para comenzar se efectuó el ajuste de datos por medio del CCA para el periodo de calibración completo usando todas las variables predictivas (17 en total), expresando las variables en su escala original. El análisis se presenta en el Anexo 32. Los resultados pronosticados en la validación son:

	Q <sup>2</sup> corregido	Error típico de la estimación
Ta	0,019	7,3105
Odis	0,438	4,0510
pH	0,226	0,4436
Con	0,215	54,5706

**TABLA 68** Resultados del CCA aplicado para el periodo de validación usando 17 variables predictivas expresadas en su escala original

Se observa que los resultados no son adecuados.

Ahora se efectuará el mismo proceso sobre el periodo de calibración completo, pero con sólo 11 variables independientes (DJ, V, DV, Vx, T, HR, ETo, VolExp, Qent, Qsal y Profundidad). El análisis se incluye en el Anexo 33. Los resultados del periodo de validación fueron:

	Q <sup>2</sup> corregido	Error típico de la estimación
Ta	0,151	52,9007
Odis	0,430	29,4936
pH	0,254	6,5348
Con	0,215	87,9793

**TABLA 69** Resultados del CCA aplicado para el periodo de validación usando 11 variables predictivas expresadas en su escala original

---

---

Los resultados son menos satisfactorios que en el caso anterior, lo cual se ve reflejado en el error típico.

Al desarrollar los dos procesos anteriores se detectó que en el cálculo de variables canónicas influyen más unas variables predictivas que otras. Por lo que se procederá a estandarizarlas antes de efectuar el CCA. En el Anexo 34 se incluye el análisis. Los resultados de validación (con 17 variables) son:

	Q <sup>2</sup> corregido	Error típico de la estimación
Ta	0,640	4,9211
Odis	0,677	2,5213
pH	0,383	0,3900
Con	0,555	86,0564

**TABLA 70** Resultados del CCA aplicado para el periodo de validación usando 17 variables predictivas estandarizadas

Con el propósito de saber si los valores de estandarización influyen al usar las ecuaciones ajustadas se procedió a efectuar la tipificación de las 17 variables del periodo de validación, pero con los valores correspondientes a la media y desviación estándar del periodo de calibración. Los resultados de validación fueron (se usó el mismo CCA del Anexo 34):

	Q <sup>2</sup> corregido	Error típico de la estimación
Ta	0,679	4,0730
Odis	0,682	2,6199
pH	0,434	0,4431
Con	0,604	74,9468

**TABLA 71** Resultados del CCA aplicado para el periodo de validación usando 17 variables predictivas, estandarizadas con los parámetros del periodo de calibración

Se observa que los valores del coeficiente de determinación en el periodo de validación y el error típico correspondiente han mejorado bastante.

A continuación se efectúa en los mismos términos el análisis anterior, excepto que se removerá de los cálculos la única variable que no se considera dentro de las variables meteorológicas o hidrológicas, el día juliano (DJ). El Anexo 35 contiene el análisis efectuado. Los valores que representan la calidad de la estimación para el periodo de validación quedan así:



	Q <sup>2</sup> corregido	Error típico de la estimación
Ta	0,664	4,1044
Odis	0,738	2,6086
pH	0,474	0,4459
Con	0,603	79,4605

**TABLA 72** Resultados del CCA aplicado para el periodo de validación usando 16 variables predictivas, estandarizadas con los parámetros del periodo de calibración

Se observa una mejora en la predicción de las variables Odis y pH y una disminución ligera de Con. Pero como es un modelo con menos variables (más sencillo y que se desempeña mejor que el anterior), se prefiere este.

Así, es preferible para cumplir con el Principio de Parsimonia (simplificar el modelo sin perder poder de predicción) intentar reducir el número de variables. Para esto, en el Anexo 36 fueron seleccionadas de entre las dos primeras variables canónicas (por ser las que explican más del 50% de la variabilidad) los pesos de las variables independientes con valor mayor a 0,10, eliminado del análisis a las variables que no cumplían con este requisito. Se procedió así en dos ocasiones hasta que todas las variables predictivas tenían cuando menos un peso mayor a 0,10. Las variables que resultaron satisfactorias en este sentido fueron nueve: T, Tn, Tx, Hf, HR, ETo, VolExp, Qsal y Profundidad. Los resultados del ajuste se muestran en la tabla siguiente:

	Q <sup>2</sup> corregido	Error típico de la estimación
Ta	0,657	4,1295
Odis	0,738	2,5785
pH	0,475	0,4412
Con	0,607	77,9733

**TABLA 73** Resultados del CCA aplicado para el periodo de validación usando 9 variables predictivas, estandarizadas con los parámetros del periodo de calibración

Se observa que los resultados son muy similares a los de la Tabla (72), pero con sólo nueve variables, por lo que este modelo se prefiere.

#### 7.10.2 CCA sobre las variables originales estandarizadas

En base a los resultados obtenidos en el apartado anterior se efectuarán varias regresiones sobre los conglomerados obtenidos en la sección 7.7.2 y 7.7.3 por medio del CCA, con las características siguientes: se usarán 16 variables independientes (se excluye DJ), las estandarizaciones se efectuarán con la media y la desviación estándar del periodo de calibración entero.

---

---

Al efectuar el procedimiento por conglomerados se tiene que las variables del bloque independiente son linealmente dependientes, por lo que no se puede aplicar el análisis.

#### 7.10.3 CCA sobre las variables transformadas

Se realizarán análisis en los conglomerados formados en las secciones 7.7.4 y 7.7.5 empleando 16 variables independientes (no se incluye DJ). Las transformaciones de las variables se efectuarán con los parámetros obtenidos del periodo de calibración.

Cuando se realiza el proceso se detecta que las variables son linealmente dependientes en varios de los grupos, por lo que no se genera una salida válida. No se puede aplicar el análisis para obtener los resultados del periodo de validación completo.

#### 7.10.4 CCA sobre las puntuaciones en componentes principales

Usando los conglomerados generados en las secciones 7.7.6 y 7.7.7 y empleando en el análisis 16 variables predictivas (la variable DJ no se usa), se calcularán 16 componentes principales y sus correspondientes puntuaciones para ser manejadas en el análisis.

Para realizar el CCA el segundo grupo de variables estará integrado por las puntuaciones en componentes principales de las variables respuesta.

Después de aplicado el procedimiento se observa que los errores típicos de estimación son demasiado grandes en los conglomerados (de valor igual o mayor a la media de las variables que se requiere predecir). Por lo que este análisis no se tomará en cuenta.

#### 7.10.5 CCA sobre las puntuaciones factoriales

Con los conglomerados formados en las secciones 7.7.8 y 7.7.9 y usando 16 variables predictivas (DJ no se empleará) se calcularán las puntuaciones factoriales para los autovalores mayores que uno, para formar el primer conjunto de variables para el análisis de correlación canónica.

Para formar el segundo conjunto se obtendrán las puntuaciones factoriales de las 4 variables dependientes, también para autovalores mayores a la unidad.

---

---

En todos los casos los factores se obtendrán por medio de una Rotación Varimax y el método de extracción será el de componentes principales.

En las variables predictivas se extrajeron cuatro factores (pues se obtuvo la misma cantidad de autovalores mayores a la unidad). También para las variables respuesta se calcularon cuatro factores, pero en este caso se hizo para poder obtener por medio de las puntuaciones factoriales estimadas los valores de las variables respuesta para el periodo de validación.

El coeficiente de determinación corregido ( $R^2$  corregido) es cercano a cero para las 4 variables respuesta, por lo que no se considera satisfactorio este análisis y no se presenta.

### 7.11 PLSR

Su aplicación está basada sobre la suposición de homogeneidad, esto es, el sistema analizado debe encontrarse en un estado similar a través de toda la investigación (Wold et al., 2001b).

En todas las aplicaciones que se efectúen del método las variables **X** y **Y** serán estandarizadas (centradas y escaladas –en varianza– por columnas) antes de iniciar con el mismo (Baffi et al., 1999).

De inicio se aplicará el algoritmo PLS1, esto es, las variables **Y** predichas de manera individual (una a la vez) (Ergon, 2006). En el caso donde se tenga evidencia referente a un buen proceso de predicción se desarrollarán variantes en el modelado, con el objetivo de intentar mejorar la estimación.

En cada caso se probará con varias componentes para saber cuántas utilizar para conseguir la mejor predicción posible en el periodo de validación. Este será el único criterio utilizado para seleccionar el número de componentes.

#### 7.11.1 Periodos de calibración y validación completos

Se tomarán las 17 variables independientes iniciales y las 4 variables dependientes, se estandarizarán de forma univariante (por cada variable) antes de llevar a cabo el procedimiento de regresión con el PLSR.

Los resultados de este análisis se incluyen en el Anexo 37. Los resultados sucintos

---

---

son:

Periodo	Variable	Número de componentes extraídas	R <sup>2</sup>	Error estándar de estimación
Calibración	Ta	7	0,7448	2,77
	Odis	14	0,7744	2,05
	pH	11	0,2882	0,34
	Con	8	0,6533	59,10
Validación	Ta	7	0,6965	4,02
	Odis	14	0,6862	2,61
	pH	11	0,4342	0,44
	Con	8	0,6083	72,36

**TABLA 74** Resultados del PLSR aplicado a los periodos de calibración y validación, algoritmo PLS1

Estos resultados son ligeramente superiores en poder de predicción para Ta y Con, siendo un poco menos efectivos en Odis y pH que los mejores obtenidos con el CCA para 9 variables predictivas (Tabla 73), mostrándose el PLSR más flexible.

#### 7.11.2 Variables originales estandarizadas

Se aplicará el PLSR a los conglomerados de las secciones 7.7.2 y 7.7.3. Se usarán 17 variables predictivas. Las estandarizaciones en las variables se harán para el periodo de calibración completo.

Los valores del coeficiente de determinación y el error típico de validación por grupos son peores que los que se muestran en la Tabla (74) para el periodo de validación completo (la variable más afectada fue pH). Por lo tanto, al no ser mejor este análisis no se muestra.

#### 7.11.3 Variables transformadas

En este caso se considerarán los conglomerados formados en las secciones 7.7.4 y 7.7.5 para emplear el PLSR. Las estandarizaciones utilizadas serán las del periodo de calibración entero.

Los valores de los dos estadísticos que miden la calidad del ajuste indican que por este procedimiento la validación no arroja buenos resultados (no es adecuada), por lo que no se incluyen aquí.

---

---

#### 7.11.4 Puntuaciones en componentes principales

Se emplearán los conglomerados generados en las secciones 7.7.6 y 7.7.7 para la aplicación del PLSR. Las puntuaciones en componentes principales corresponden a los periodos de calibración y validación completos.

Se usaron 17 variables independientes obteniéndose 17 componentes principales. De las 4 variables respuesta se obtuvieron 4 componentes principales. El PLSR se aplicó al periodo de calibración. Se comprobó en el periodo de validación.

Al realizar la validación se observa que el coeficiente de determinación toma valores bajos y el error típico se dispara enormemente. Por lo tanto, la estimación de los valores no es adecuada y no se presenta.

#### 7.11.5 Puntuaciones factoriales

En este caso se usarán los conglomerados obtenidos en las secciones 7.7.8 y 7.7.9 para la regresión con el PLSR. Las puntuaciones factoriales para realizar la regresión fueron obtenidas con los periodos de calibración y validación enteros.

En el periodo de calibración fueron empleadas 17 variables independientes, obteniéndose 4 factores con autovalor mayor que la unidad, que fueron los utilizados en el análisis. El método de extracción de factores fue el de componentes principales, se efectuó una rotación Varimax y las puntuaciones factoriales se obtuvieron por el procedimiento de regresión. Para las 4 variables respuesta también se obtuvieron 4 factores con las mismas opciones antes mencionadas; esto se efectuó así debido a que para obtener las puntuaciones factoriales usando las variables de calidad del agua es necesario multiplicar matricialmente las variables respuesta del periodo de calibración por una matriz de coeficientes. Luego se usa el PLSR para obtener las estimaciones de los puntuaciones factoriales y después para calcular las variables respuesta del periodo de validación es necesario utilizar la matriz inversa de la matriz de coeficientes, que debe ser cuadrada para poder invertirla; por este motivo se obtiene el mismo número de factores que la cantidad de variables respuesta. Los resultados de la calibración arrojan un  $R^2$  con valores bajos, lo cual augura una mala predicción.

Para efectuar el pronóstico se obtuvieron 4 factores de los valores correspondientes a 17 variables independientes para sustituirlos en las expresiones de regresión.

---

---

Los resultados muestran un coeficiente de determinación muy bajo (cercano a cero) para las 4 variables respuesta. Por lo tanto, este análisis no se tomará en cuenta.

## 7.12 MEJORA DE LOS RESULTADOS

Considerando que dentro de los métodos revisados el PLSR es uno de los que proporciona los mejores resultados y es una técnica que no impone suposiciones a cumplir de inicio, se intentará mejorar los resultados de la regresión sobre las variables de calidad del agua proponiendo variantes dentro de la aplicación del PLSR.

### 7.12.1 Variables predictivas físicas

Con el objetivo de darle un sentido estrictamente físico a las variables empleadas en la estimación de parámetros de calidad del agua la variable Día Juliano (DJ) no será incluida en los análisis subsiguientes.

Los resultados empleando 16 variables predictivas quedan así (por constituir la auténtica comprobación de la efectividad del proceso sólo se muestran los valores de validación):

Periodo	Variable	Número de componentes extraídas	R <sup>2</sup>	Error estándar de estimación
Validación	Ta	7	0,6749	4,04
	Odis	14	0,7410	2,59
	pH	12	0,4757	0,44
	Con	12	0,6061	78,76

**TABLA 75** Resultados del PLSR aplicado al periodo de validación, algoritmo PLS1, 16 variables predictivas

Disminuye ligeramente el R<sup>2</sup> de la variable Ta, pero aumenta significativamente en Odis y pH. El error estándar de la variable Con creció ligeramente.

Este análisis globalmente es de calidad similar al mostrado en la Tabla (74), pero con una variable independiente menos por lo que cumple con el principio de parsimonia y por lo tanto es superior.

### 7.12.2 Recortado de datos

Para realizar el recortado sobre los datos primero se ordenó por cada variable y se reemplazaron todos los datos mayores que el percentil 99 y los datos menores que el

percentil 1, por los valores de estos límites. El recortado se aplicó tanto a las variables independientes como a las dependientes.

Es muy importante hacer notar que el proceso de recortado sólo se aplicó al periodo de calibración. El periodo de validación se utilizó con sus valores originales. Los resultados son:

Periodo	Variable	Número de componentes extraídas	R <sup>2</sup>	Error estándar de estimación
Validación	Ta	4	0,6935	4,03
	Odis	13	0,7419	2,61
	pH	10	0,4723	0,43
	Con	8	0,6096	77,86

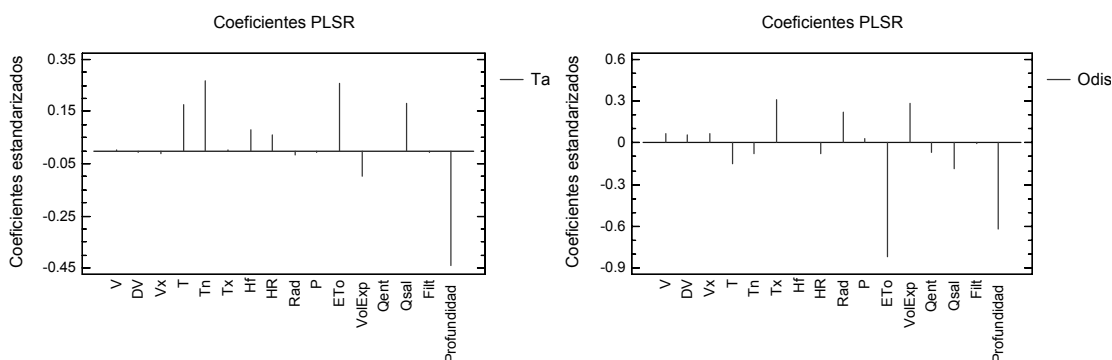
**TABLA 76** Resultados del PLSR aplicado al periodo de validación, algoritmo PLS1 y recortado de los datos, 16 variables predictivas

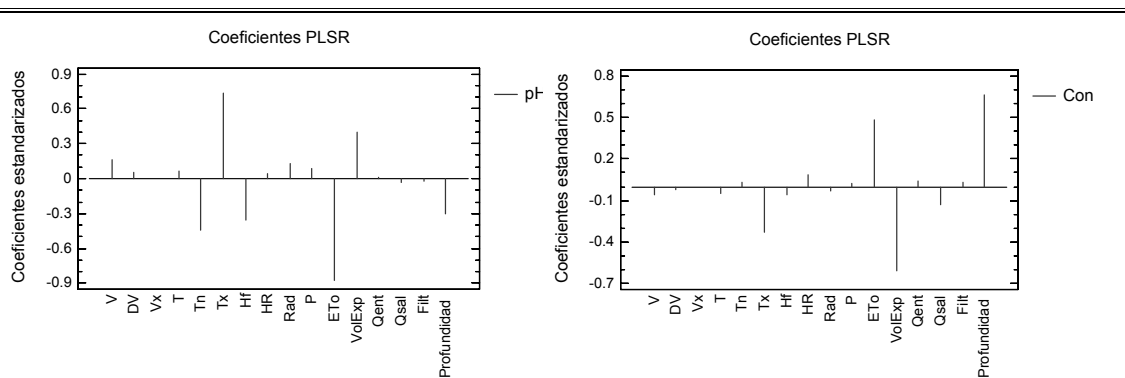
Al comparar los resultados obtenidos con los de la Tabla (75) (idéntica cantidad de variables independientes) se observa que las modificaciones observadas son muy pequeñas. Debido a esto, en este trabajo no se empleará el recortado.

La aplicación del proceso de recortado ha servido para tener la seguridad de que los datos de calibración son estables en sus predicciones y no poseen valores extremos importantes. También se puede concluir que el proceso inicial de depuración de los datos fue hecho correctamente.

### 7.12.3 Modelo reducido – Términos lineales

Fundamentalmente el principio de parsimonia indica que un modelo debe ser útil y lo más sencillo posible.





**FIGURA 52** Coeficientes estandarizados del modelo PLSR, 16 variables predictivas

Haciendo uso de gráficos de coeficientes PLSR para cada variable de calidad del agua obtenidos en el apartado 7.12.1 y que se muestran en la Figura 52, se proponen modelos reducidos con un mínimo de pérdida de información.

El resumen de los modelos obtenidos es el siguiente:

Periodo	Variable	Número de variables predictivas	Número de componentes extraídas	R <sup>2</sup>	Error estándar de estimación
Validación	Ta	5	2	0,6892	4,07
	Odis	4	3	0,7491	2,32
	pH	6	6	0,4843	0,42
	Con	3	3	0,6304	71,44

**TABLA 77** Resultados del PLSR aplicado al periodo de validación, algoritmo PLS1, modelo reducido en términos lineales

Se puede apreciar (al comparar con la Tabla 75) que con la excepción de un muy ligero aumento en el error de la variable Ta, todos los valores restantes que miden la calidad del ajuste mejoran.

Por lo tanto, al existir mejor estimación de las variables de calidad del agua con una cantidad reducida de variables, se concluye que esta es la mejor regresión efectuada hasta este punto del trabajo. Los resultados de la regresión efectuada se presentan en el Anexo 38. Las ecuaciones de estimación son:

$$Ta\_Est = 8,3202 - 0,310551 \cdot Profundidad + 0,000112235 \cdot Qsal + 0,56017 \cdot ETo + 0,191481 \cdot Tn + 0,180964 \cdot T \quad (210)$$

$$Odis\_Est = 11,1268 - 0,752148 \cdot ETo - 0,342047 \cdot Profundidad - 0,238237 \cdot Tx + (8,68778 \times 10^{-7}) \cdot VolExp \quad (211)$$



$$\text{pH\_Est} = 7,43143 - 0,144645 \cdot \text{ETo} + 0,0305618 \cdot \text{Tx} - 0,0259479 \cdot \text{Tn} + (1,11747 \times 10^{-7}) \cdot \text{VolExp} - 0,0408722 \cdot \text{Hf} - 0,0154395 \cdot \text{Profundidad} \quad (212)$$

$$\text{Con\_Est} = 1220,88 + 8,71871 \cdot \text{Profundidad} - 0,0000336155 \cdot \text{VolExp} - 0,854586 \cdot \text{ETo} \quad (213)$$

#### 7.12.4 Matriz extendida de entradas

Con el propósito de perfeccionar la regresión se extenderá la matriz de entradas incluyendo términos no lineales de las variables originales **X** (logaritmos, valores cuadráticos, valores cúbicos, recíprocos). A continuación se desarrollará un PLSR sobre la matriz extendida de entradas (variables predictivas) y cada una de las variables de calidad del agua.

Esta variante del PLSR es denominada INLR (implicit non-linear PLS regression) y se ha mostrado que permite la modelación de no linealidades suaves (de segundo o tercer grado polinómico) en las relaciones **X-Y** (Berglund et al., 1997; Berglund et al., 2001).

Los valores obtenidos son:

Periodo	Variable	Número de términos utilizados	Número de componentes extraídas	R <sup>2</sup>	Error estándar de estimación
Validación	Ta	7	6	0,6945	4,04
	Odis	6	5	0,7190	2,71
	pH	7	4	0,4723	0,41
	Con	4	4	0,6386	73,57

**TABLA 78** Resultados del PLSR aplicado al periodo de validación, algoritmo PLS1, usando la matriz extendida de entradas

Al comparar con la Tabla (77) (modelo reducido en términos lineales) se observa que los resultados prácticamente son los mismos a excepción del parámetro oxígeno disuelto, que aparece mejor modelado en la Tabla (77). Por lo tanto, se preferirá de entre estos dos modelos el reducido en términos lineales.

Siguiendo con la propuesta de modelación no lineal se investigará la siguiente línea: considerando los resultados obtenidos con el modelo reducido en términos lineales se intentará efectuar una regresión no lineal, considerando como única variable independiente (**x**) a los valores estimados de una variable de calidad del agua y como variable dependiente (**y**) a sus correspondientes valores observados. Las expresiones

resultantes para obtener las estimaciones de las variables de calidad del agua son las siguientes:

$$Ta = 11,23160366 - 0,0608831 \cdot (Ta\_Est) - 0,0206138 \cdot (Ta\_Est)^2 + 0,00223407 \cdot (Ta\_Est)^3 \quad (214)$$

$$Odis = \frac{11,915435}{1 + 54,990249 \cdot e^{-0,659802 \cdot (Odis\_Est)}} \quad (215)$$

$$pH = -0,0695477 + 1,017231 \cdot (pH\_Est) - 0,001066627 \cdot (pH\_Est)^2 \quad (216)$$

$$Con = -21,609260 + 1,0422319 \cdot (Con\_Est) - (2,05126 \times 10^{-5}) \cdot (Con\_Est)^2 \quad (217)$$

En las ecuaciones (214) a (217) las variables con terminación \_Est hacen referencia a las ecuaciones (210) a (213).

Los parámetros que miden el ajuste de las observaciones al modelo toman los siguientes valores:

Periodo	Variable	R <sup>2</sup>	Error estándar de estimación
Calibración	Ta	0,8635	2,03
	Odis	0,7668	2,09
	pH	0,2656	0,35
	Con	0,6115	62,56
Validación	Ta	0,7426	3,54
	Odis	0,7554	2,38
	pH	0,4844	0,42
	Con	0,6310	71,48

**TABLA 79** Resultados del PLSR aplicado a los periodos de calibración y validación, modelación no lineal, usando los resultados del modelo reducido en términos lineales

En el periodo de calibración se observa que las variables Temperatura del agua y Oxígeno disuelto son muy bien modeladas; asimismo, la estimación de la Conductividad es regular y el pH es mal simulado.

Respecto al periodo de validación, tanto Ta como Odis son bien modeladas (mejor que en el modelo reducido en términos lineales, Tabla 77), el modelado de la Con sigue siendo regular (de igual calidad que en el modelo reducido en términos lineales) y la variable pH presenta un comportamiento errático al aumentar su R<sup>2</sup> (prácticamente con los mismos valores de la Tabla 77).

Se concluye que la modelación no lineal es adecuada para obtener buenas

---

---

estimaciones de la variable  $T_a$ , las variables  $O_{dis}$  y  $Con$  serán estimadas con un ajuste regular por medio del modelo reducido en términos lineales y para  $pH$  es necesario plantear otras opciones para efectuar su estimación.

#### 7.12.5 Modelación usando la correlación entre variables respuesta

Al examinar el Anexo 12 se observa que existe correlación entre las variables respuesta. En esta sección se intentará hacer uso de esta correlación para mejorar la estimación de las variables de calidad del agua.

Es importante mencionar que las expresiones matemáticas a obtener contienen variables respuesta, por lo que al ser empleadas para calcular variables de ese mismo grupo indican que existe una correlación entre ellas, pero no necesariamente dependencia.

Se efectuará un PLS1 lineal para cada una de las variables respuesta en el periodo de calibración incluyendo en cada ocasión como variables predictivas a las restantes variables dependientes.

Para el periodo de validación se tiene que los parámetros que miden la calidad de la regresión oscilan en sus valores y no tienden hacia un valor fijo. Se concluye que al contar sólo con valores estimados de las variables respuesta no es posible mejorar el pronóstico de éstas, pues los estimados poseen un error (no determinado exactamente para cada caso en la validación) que impide su uso como predictores confiables.



---

---

## CAPÍTULO 8. APLICACIÓN DEL MODELO MECANICISTA

### 8.1 REVISIÓN Y TRATAMIENTO DE LOS DATOS

En la aplicación de un modelo la gestión de información conlleva la revisión de los datos no procesados con anterioridad para detectar valores faltantes, anómalos, fuera de rango (Kettaneh et al., 2005) y tomar las medidas pertinentes en cada caso.

El modelo mecanicista propuesto fue programado especialmente en el lenguaje Visual Basic 2005.

En el caso de las concentraciones iniciales es necesario para el desarrollo del modelo proporcionar, al menos, 2 lecturas de temperatura del agua en el embalse a diferentes profundidades al inicio de la simulación. Las lecturas no utilizadas son rellenadas con ceros.

Cuando en un día del periodo de simulación no se hubiera extraído caudal por una toma o el vertedor en el dato de caudal se escribe cero, para no dar lugar a errores en la interpretación del dato.

Dentro de las características de las salidas (tomas y vertedor) un valor fundamental es la cota del fondo del embalse. Contando con ella y con las cotas al centro de toma y la altura de la cresta del vertedor se obtienen inmediatamente las alturas sobre el fondo de cada toma y el vertedor.

Los datos diarios de volumen en el embalse, evaporación, precipitación, filtración y caudal de entrada por el río se ajustaron para que el balance fuera satisfactorio.

En la tabla calado-volumen-superficie se efectuaron varios ajustes. Primero, para el área correspondiente al calado nulo (cero) se modificó el valor original de cero. Para esto, se efectuó un ajuste tomando los valores desde el calado 0,5 m con incrementos de 0,5 m hasta llegar a 20,0 m. De las funciones que se probaron con ayuda del programa CurveExpert 1.3, la que proporcionó el mejor ajuste fue:

$$A = 8306,12 + 16566,928 \cdot (\text{Calado}) - 63.764653 \cdot (\text{Calado})^2 + 7.9689324 \cdot (\text{Calado})^3 \quad (218)$$

Empleando esta expresión el error estándar es de 5503,3086 y el coeficiente de correlación tiene un valor de 0,9988245. El área correspondiente al calado cero tiene un valor de  $A_0 = 8306,12 \text{ m}^2$ . Este valor tiene su importancia dentro del cálculo del flujo

---

neto de sedimentos en la capa del fondo.

Otros valores que fueron ajustados en la tabla calado-volumen-superficie para que correspondieran a los valores de sus conjuntos de datos vecinos fueron:

Calado	Volumen original	Volumen ajustado
2,1	39052	69012,5

Calado	Área original	Área ajustada
2,5	51340	51330
2,6	51335	51330
10,5	181906,6667	185240
10,6	183573,3333	185240

Los valores de latitud y longitud del embalse (localización geográfica) fueron obtenidos introduciendo los valores de las coordenadas UTM en la Calculadora Geodésica (<http://www.ign.es/ign/home/calculadora/UTM2LL.jsp>). Los resultados son los siguientes:

Latitud	38° 34' 3,97349" N
Longitud	0° 15' 50,0174" W

En la altitud del embalse se reportó, en metros, la cota de coronación al considerarse este como un valor representativo del almacenamiento.

Los datos meteorológicos, como son la velocidad del viento, la radiación solar, la temperatura del aire y la humedad relativa fueron registrados completos a escala diaria para todo el periodo de simulación.

A continuación, se efectúa una mención especial del tratamiento dado a las cantidades temperatura superficial del agua del embalse y la cobertura de nubes.

## 8.2 COBERTURA DE NUBES Y TEMPERATURA SUPERFICIAL DEL EMBALSE

La cobertura de nubes media diaria es un parámetro necesario para el cálculo del flujo de energía de radiación atmosférica neta de onda larga que penetra en el agua (Environmental Laboratory, 1995; Martin et al., 1999; Deas et al., 2000).

---

---

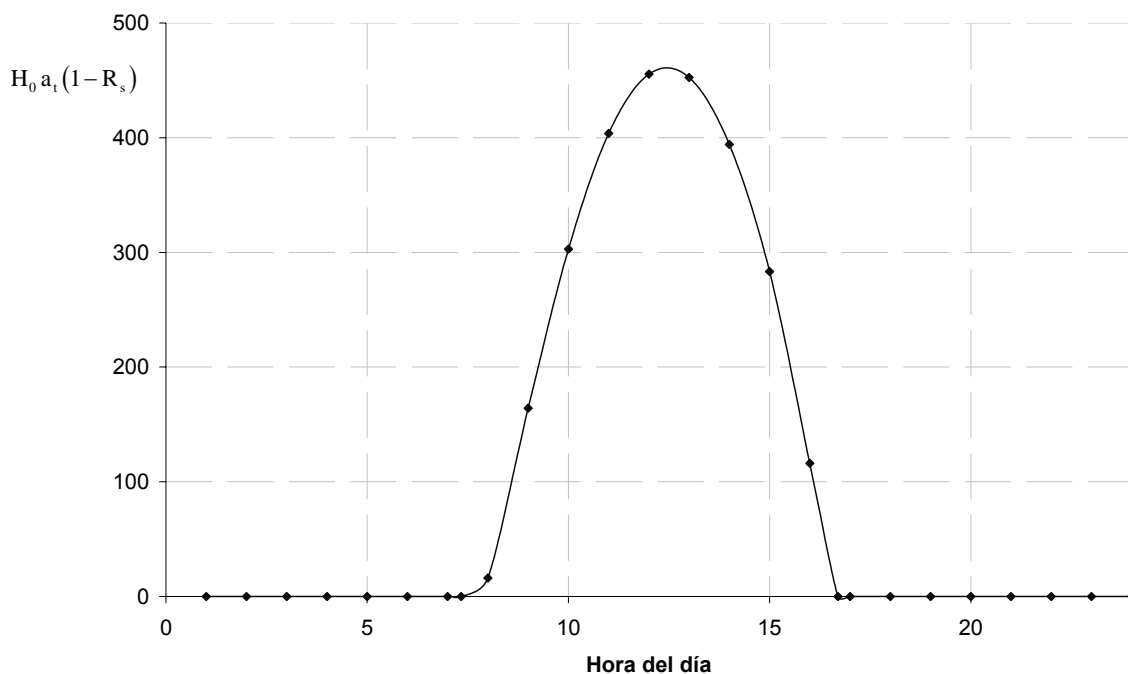
La estación de Villajoyosa, al ser una estación automática, no registra la cobertura de nubes (Deas et al., 2000). Es necesario contar con este dato a escala diaria para que los cálculos de flujos de energía térmica (Environmental Laboratory, 1995; Martin et al., 1999; Deas et al., 2000) y en consecuencia los cálculos de la temperatura en las capas del almacenamiento sean efectuados de manera correcta.

Una propuesta para estimar el valor de la cobertura de nubes, que puede variar día a día, es obtenerla en base a la siguiente metodología empleando los registros de radiación solar media diaria:

Cálculo de la cobertura de nubes media diaria en base a la radiación solar media diaria

1.- Para un día determinado se calculan los valores de las variables  $H_0$ ,  $a_t$  y  $R_s$  (siguiendo la metodología expresada en Martin y McCutcheon, 1999) para cada hora del día, en donde:

- $H_0$  = Flujo de radiación de onda corta que llega a la parte externa de la atmósfera terrestre.
- $a_t$  = Fracción de radiación que alcanza la superficie del agua después de la reducción por difusión y absorción.
- $R_s$  = Albedo o fracción de la radiación solar reflejada en la superficie del agua.



**FIGURA 53** Variación horaria de la radiación solar de onda corta después de difusión, absorción y reflexión en la superficie del agua

---

---

2.- Se elabora la gráfica que muestre la variación de  $H_0 \cdot a_t \cdot (1 - R_s)$  a lo largo del día considerado (Figura 53).

3.- Se obtiene el Área bajo la gráfica y se divide entre 24 (número de horas del día) para obtener el valor medio del día.

$$\text{Valor medio de } \sum H_0 \cdot a_t \cdot (1 - R_s) = \frac{\text{Área}}{24} \quad (219)$$

4.- Por medio de la expresión siguiente:

$$H_{sw} = \left\{ \text{Valor medio de } \sum H_0 \cdot a_t \cdot (1 - R_s) \right\} C_a \quad (220)$$

Donde (Martin et al., 1999):

$H_{sw}$  = Radiación solar media registrada del día.

$C_a$  = Es la fracción de la radiación pasando a través de las nubes.

Como  $C_a = 1 - 0,65 C_L^2$ , se puede despejar el valor de  $C_L$ :

$$C_L = 1,2403 \sqrt{1 - \frac{H_{sw}}{\text{Valor medio de } \sum H_0 \cdot a_t \cdot (1 - R_s)}} \quad (221)$$

Y con la ecuación (221) se obtiene el valor de  $C_L$  para el día en análisis.

Debido a la gran cantidad de datos y operaciones involucradas la obtención de la cobertura de nubes para cada día del periodo de simulación se efectuó con un programa elaborado ex profeso, Nubosidad, programado en Microsoft Visual Basic 2005.

#### Temperatura superficial del embalse

La temperatura superficial del embalse se midió sólo en una parte del periodo de estudio. A continuación se proporcionan los intervalos para los que sí existen mediciones:



---

Intervalo de tiempo
11/02/2004 a 13/02/2004
18/02/2004 a 19/02/2004
24/02/2004 a 28/02/2004
01/03/2004 a 02/03/2004
31/03/2004 a 04/04/2004
05/07/2004 a 16/08/2004
28/09/2004 a 07/10/2004
09/10/2004 a 29/10/2004
31/10/2004 a 06/12/2004
09/12/2004 a 24/12/2004
26/12/2004
28/12/2004 a 29/12/2004
05/01/2005 a 18/02/2005
21/02/2005 a 16/03/2005
23/03/2005 a 30/03/2005
06/04/2005 a 09/04/2005
13/04/2005
19/04/2005 a 06/07/2005
08/07/2005 a 11/07/2005
15/07/2005 a 19/09/2005

Esta variable es importante debido a que interviene en la obtención de la radiación de onda larga que sale del agua, el calor perdido por conducción y convección, así como la evaporación. Por lo que es imprescindible contar con la temperatura media diaria del agua del almacenamiento.

---

---

Con el fin de generar los datos faltantes se eligió efectuar el análisis de regresión. Para elegir las variables de las que dependería la temperatura superficial del agua en el embalse se procesaron junto con los datos existentes de esta variable y para los mismos días las siguientes cantidades:

- Día Juliano (inicia cada año con valor uno).
- Radiación, en  $W/m^2$ .
- Evaporación, en mm.
- Precipitación, en mm.
- Temperatura del aire, en  $^{\circ}C$ .
- Humedad relativa, en %.
- Cobertura de nubes (fracción del cielo cubierto).

En un modelo de regresión una variable respuesta (en este caso la temperatura del agua de la superficie del depósito) es expresada como una función de una o más variables predictivas más la incertidumbre. En muchos casos (pero no en todos) la forma funcional es lineal con coeficientes indeterminados (StatPoint, 2005).

El objetivo del análisis de regresión es construir un modelo que (StatPoint, 2005):

- Describa las relaciones que existen entre las variables de una manera que permita predecir la variable respuesta proporcionando los valores de las variables predictivas.
- No contenga más variables que las necesarias para generar una buena predicción.

La última consideración es en ocasiones referida como parsimonia. Típicamente los modelos que involucran un pequeño conjunto de predictores bien seleccionados tienen el mejor desempeño en la práctica (StatPoint, 2005).

El primer paso para comenzar cuando se inicia la construcción de un modelo de regresión es con el procedimiento de análisis múltiple de variables.

La matriz de gráficas (Figura 54) para cada par de variables se realizó con el software estadístico STATGRAPHICS.

La tabla de correlaciones muestra una matriz de coeficientes de correlación estimados para cada par de variables en el análisis, el número de observaciones utilizado para

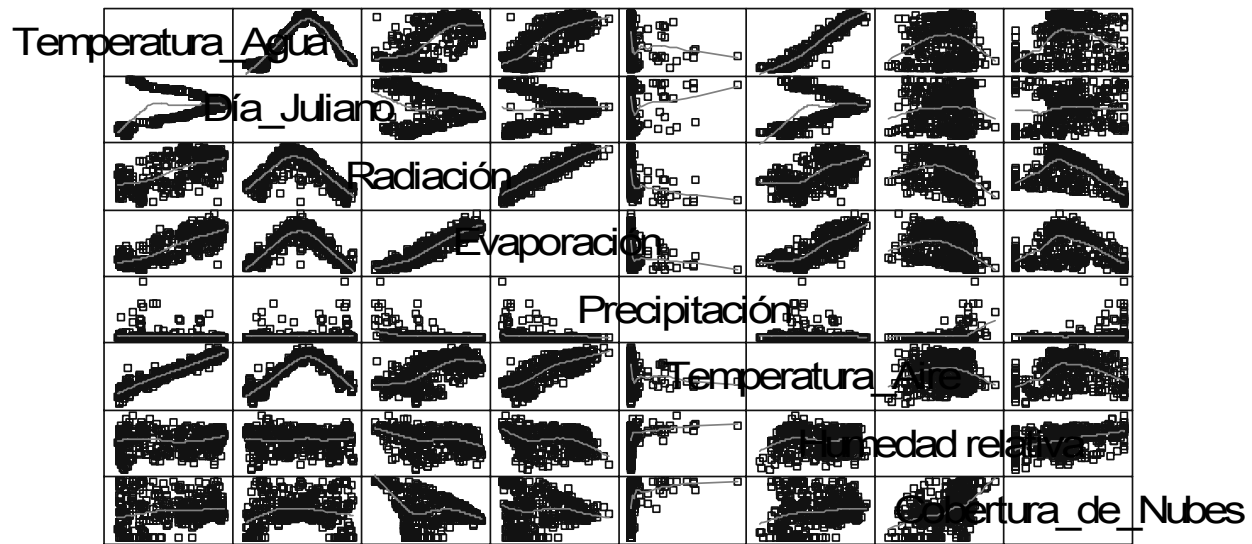


FIGURA 54 Matriz de gráficas para variables del embalse

obtener la estimación y el valor P. El coeficiente de correlación  $r$  es un número entre el rango -1 y 1, el cual mide la solidez de la relación lineal entre dos variables. Entre más cercana es la correlación a -1 o +1, más sólida es la relación. El signo de la correlación indica la dirección de la relación. Un valor positivo significa que la variable respuesta aumenta cuando la variable predictiva aumenta. Un valor negativo significa que la variable respuesta disminuye cuando la predictiva aumenta (StatPoint, 2005).

Para determinar si dos variables están o no significativamente relacionadas entre sí un valor P es calculado por cada coeficiente de correlación. Cualquier par de variables para el cual el valor P es menor que 0,05 exhibe una correlación lineal estadísticamente significativa a un nivel de confianza del 5% (StatPoint, 2005). A continuación se presenta la tabla de correlaciones obtenida:

#### Correlations

	Temperatura_Agua	Día_Juliano	Radiación	Evaporación	Precipitación
Temperatura_Agua		0.3981	0.6675	0.8226	-0.1203
$p$		0.0000	0.0000	0.0000	0.0192
Día_Juliano	0.3981		-0.1058	0.0546	0.1243
	0.0000		0.0395	0.2892	0.0155
Radiación	0.6675	-0.1058		0.9327	-0.3407
	0.0000	0.0395		0.0000	0.0000
Evaporación	0.8226	0.0546	0.9327		-0.2580
	0.0000	0.2892	0.0000		0.0000
Precipitación	-0.1203	0.1243	-0.3407	-0.2580	
	0.0192	0.0155	0.0000	0.0000	
Temperatura_Aire	0.9674	0.3984	0.6527	0.8210	-0.1339
	0.0000	0.0000	0.0000	0.0000	0.0091
Humedad relativa	0.0663	0.0482	-0.2204	-0.2526	0.2718
	0.1978	0.3491	0.0000	0.0000	0.0000
Cobertura_de_Nubes	0.1640	0.0617	-0.2731	-0.0650	0.3900
	0.0014	0.2307	0.0000	0.2069	0.0000

	Temperatura_Aire	Humedad relativa	Cobertura_de_Nubes
Temperatura_Agua	0.9674	0.0663	0.1640
	0.0000	0.1978	0.0014
Día_Juliano	0.3984	0.0482	0.0617
	0.0000	0.3491	0.2307
Radiación	0.6527	-0.2204	-0.2731
	0.0000	0.0000	0.0000
Evaporación	0.8210	-0.2526	-0.0650
	0.0000	0.0000	0.2069
Precipitación	-0.1339	0.2718	0.3900
	0.0091	0.0000	0.0000
Temperatura_Aire		0.0456	0.1418
		0.3756	0.0057
Humedad relativa	0.0456		0.4282
	0.3756		0.0000
Cobertura_de_Nubes	0.1418	0.4282	
	0.0057	0.0000	

**TABLA 80** Correlaciones para variables del embalse

De la matriz de gráficas, Figura (54) y la tabla de correlaciones, Tabla (80), se observa

que la temperatura superficial del agua está más correlacionada con la temperatura del aire y con la evaporación; claro, esto dentro de un modelo de regresión lineal múltiple.

El resultado del análisis de regresión es el siguiente:

**Multiple Regression - Temperatura Agua**

Dependent variable: Temperatura\_Agua (En grados Celsius)

Independent variables:

Día\_Juliano (número de día del año)

Radiación (en W/m<sup>2</sup>)

Evaporación (en mm)

Precipitación (en mm)

Temperatura\_Aire (en ° C)

Humedad relativa (en %)

Cobertura\_de\_Nubes (C<sub>L</sub>)

		<i>Standard</i>	<i>T</i>	
<i>Parameter</i>	<i>Estimate</i>	<i>Error</i>	<i>Statistic</i>	<i>P-Value</i>
CONSTANT	-1.58508	0.48992	-3.23538	0.0013
Día_Juliano	0.00661735	0.0010828	6.11132	0.0000
Radiación	0.00651944	0.00524063	1.24402	0.2143
Evaporación	0.651283	0.355346	1.83281	0.0676
Precipitación	-0.0233754	0.0420644	-0.555706	0.5787
Temperatura_Aire	0.752368	0.0404088	18.6189	0.0000
Humedad relativa	0.030639	0.0104191	2.94067	0.0035
Cobertura_de_Nubes	1.56056	0.501051	3.11457	0.0020

**Analysis of Variance**

<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-Ratio</i>	<i>P-Value</i>
Model	15344.8	7	2192.11	956.40	0.0000
Residual	850.348	371	2.29204		
Total (Corr.)	16195.1	378			

**TABLA 81** Análisis de varianza con todas las variables

R-squared = 94.7494 percent

R-squared (adjusted for d.f.) = 94.6503 percent

Standard Error of Est. = 1.51395

Mean absolute error = 1.14945

Durbin-Watson statistic = 0.759042 (P=0.0000)

Lag 1 residual autocorrelation = 0.614895

**The StatAdvisor**

The output shows the results of fitting a multiple linear regression model to describe the relationship between Temperatura\_Agua and 7 independent variables. The equation of the fitted model is:

$$\begin{aligned} \text{Temperatura\_Agua} = & -1.58508 + 0.00661735*\text{Día\_Juliano} \\ & + 0.00651944*\text{Radiación} + 0.651283*\text{Evaporación} \\ & - 0.0233754*\text{Precipitación} + 0.752368*\text{Temp\_Aire} \\ & + 0.030639*\text{Humedad relativa} + 1.56056*\text{Cobert\_Nubes} \end{aligned} \quad (222)$$

Since the P-value in the ANOVA table is less than 0.05, there is a statistically significant relationship between the variables at the 95.0% confidence level.

El coeficiente de correlación es de 0,947 el cual se puede considerar como un valor de ajuste muy bueno. El error estándar de la estimación es de 1,51. La ecuación ajustada al modelo es la (222).

Para revisar de una forma rápida cuales son las variables más significativas en el análisis se realizó una regresión del tipo Stepwise (por pasos), para intentar mejorar el ajuste. Los resultados fueron los siguientes:

#### **Multiple Regression - Temperatura Agua**

Dependent variable: Temperatura\_Agua (En grados Celsius)

Independent variables:

- Día\_Juliano (número de día del año)
- Radiación (en W sobre m2)
- Evaporación (en mm)
- Precipitación (en mm)
- Temperatura\_Aire (en grados Celsius)
- Humedad relativa (en %)
- Cobertura\_de\_Nubes (C<sub>L</sub>)

<i>Parameter</i>	<i>Estimate</i>	<i>Standard Error</i>	<i>T Statistic</i>	<i>P-Value</i>
CONSTANT	-1.41082	0.475045	-2.96987	0.0032
Día_Juliano	0.00630129	0.00105641	5.96483	0.0000
Evaporación	1.06312	0.124331	8.55074	0.0000
Temperatura_Aire	0.727498	0.0330735	21.9964	0.0000
Humedad relativa	0.0387098	0.00763211	5.07196	0.0000
Cobertura_de_Nubes	1.02514	0.331658	3.09096	0.0021

#### **Analysis of Variance**

<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-Ratio</i>	<i>P-Value</i>
Model	15339.9	5	3067.99	1338.08	0.0000
Residual	855.224	373	2.29282		
Total (Corr.)	16195.1	378			

**TABLA 82** Análisis de varianza con sólo cinco variables

R-squared = 94.7193 percent

R-squared (adjusted for d.f.) = 94.6485 percent

Standard Error of Est. = 1.51421

---

Mean absolute error = 1.15216

Durbin-Watson statistic = 0.778448 (P=0.0000)

Lag 1 residual autocorrelation = 0.605243

**The StatAdvisor**

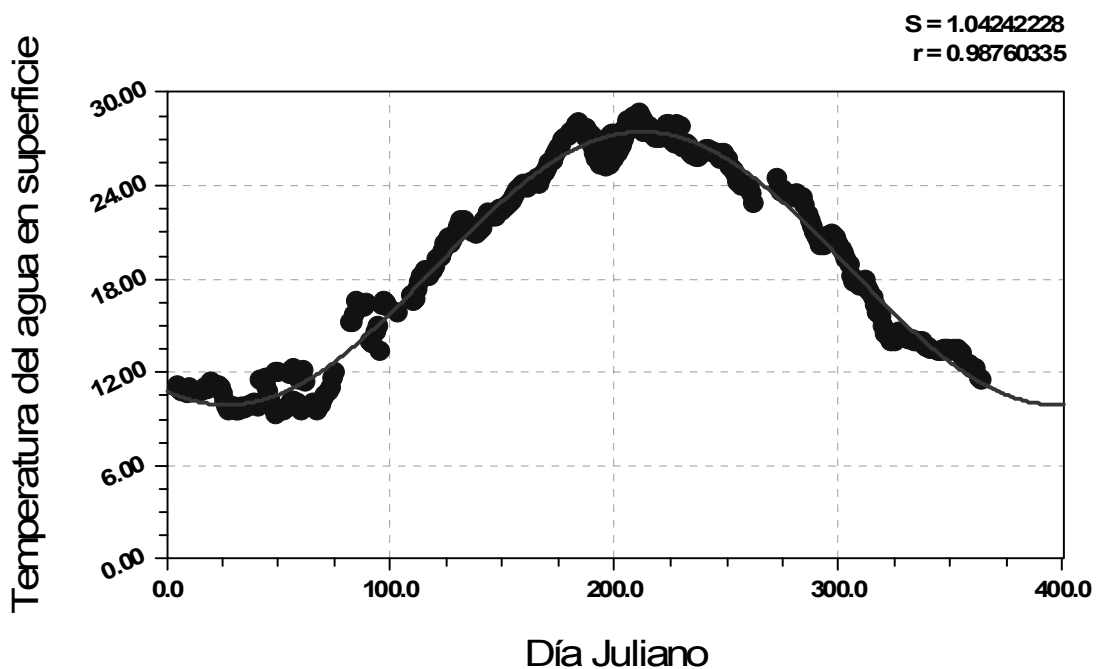
The output shows the results of fitting a multiple linear regression model to describe the relationship between Temperatura\_Agua and 7 independent variables. The equation of the fitted model is

$$\begin{aligned} \text{Temperatura\_Agua} = & -1.41082 + 0.00630129 \cdot \text{Día\_Juliano} + 1.06312 \cdot \text{Evaporación} \\ & + 0.727498 \cdot \text{Temperatura\_Aire} + 0.0387098 \cdot \text{Humedad relativa} \\ & + 1.02514 \cdot \text{Cobertura\_de\_Nubes} \end{aligned} \quad (223)$$

Since the P-value in the ANOVA table is less than 0.05, there is a statistically significant relationship between the variables at the 95.0% confidence level.

Se observa que el coeficiente de correlación y el error estándar son iguales al resultado anterior, por lo que no se aprecia ninguna mejora al realizar este análisis, salvo que se mantiene la misma precisión pero utilizando menos variables (se tiene parsimonia).

Observando la Figura (54) se tiene que la temperatura del agua posee gráficas con poca dispersión de los datos (los datos aproximadamente se ubican sobre una línea de tendencia bien definida) para las variables día Juliano y temperatura del aire. Se harán análisis de regresión simple con cada una de ellas.



**FIGURA 55** Gráfico de ajuste para temperatura del agua y día juliano

---

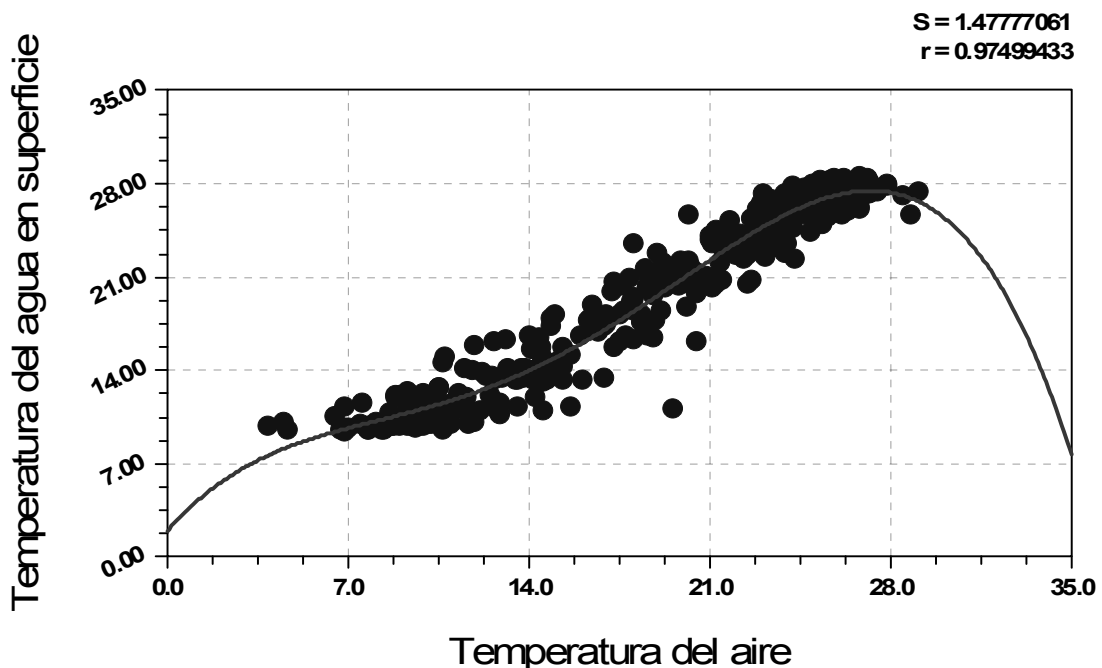
El análisis de regresión para la temperatura del agua y el día juliano se muestra en la Figura 55.

La ecuación ajustada a los datos es la siguiente (que tiene una forma funcional similar a la del radio normalizado de la órbita de la Tierra, Chapra, S., Pelletier, G., 2003. "QUAL2K: A modeling framework for simulating river and stream water quality: Documentation and users manual"):

$$\text{Temp\_del\_agua} = 18,688504 + 8,7790295 \cdot \cos(0,01694358 \cdot \text{Día\_Juliano} - 3,5888365) \quad (224)$$

El coeficiente de correlación es de 0,988 el cual es un valor de ajuste excelente (muy cercano a 1). El error estándar de la estimación es de 1,04. Por lo tanto, este ajuste mejora con mucho a todos los ajustes realizados con anterioridad.

Ahora se realiza el análisis de regresión simple para la temperatura del agua en la superficie del embalse y temperatura del aire (Figura 56). El análisis arroja los siguientes resultados:



**FIGURA 56** Gráfico de ajuste para temperatura del agua en función de la temperatura del aire

La ecuación obtenida es:



---

---

$$\begin{aligned} \text{Temp\_del\_Agua} = & 1,9818128 + 2,2039583 \cdot \text{Temp\_Aire} - 0,243991787 \cdot (\text{Temp\_Aire})^2 \\ & + 0,014066741 \cdot (\text{Temp\_Aire})^3 - 0,000250405 \cdot (\text{Temp\_Aire})^4 \end{aligned} \quad (225)$$

El coeficiente de correlación es de 0,975 el cual es un valor de ajuste excelente, pero el obtenido en la regresión anterior (día Juliano y temperatura del agua en superficie) ofrece un coeficiente aún mejor. El error estándar de la estimación es de 1,48 que es mucho mayor que el de la regresión anterior (el error que se comete al aplicar la presente regresión es mucho mayor que el cometido con la regresión anterior). Por lo tanto, este ajuste es bueno pero el anterior es el mejor de todos los efectuados.

Conclusión.- La expresión que se utilizará para completar los datos faltantes de la temperatura del agua en la superficie del embalse es la ecuación (224) la cual proporciona el mejor ajuste a los datos existentes, tiene parsimonia y su gráfica (Figura 55) muestra una buena representación de la estacionalidad que posee esta variable.

### 8.3 INICIO DE LA SIMULACIÓN

La simulación se inicia con el cálculo del calado en el embalse para cada día del tiempo a simular. Se obtiene con el volumen diario en el embalse y la tabla calados-volumenes. Esta variable es una de las más importantes para todos los procesos a realizar.

Para el primer paso de tiempo (primer día) el embalse se divide en capas de espesor igual a un metro, iniciando del fondo hacia arriba. La última capa (la de superficie) puede tener espesor de un metro o más, sin superar los dos metros. Esto se hace para que los procesos a efectuar en las capas tengan una configuración de inicio estable.

Se obtiene la altura media de cada capa. Haciendo uso de las lecturas de concentraciones iniciales se obtienen las concentraciones (temperatura del agua) del primer día en cada capa por medio de una interpolación lineal. Esto se hizo así pues al requerirse como mínimo dos lecturas no es posible efectuar interpolaciones de mayor orden.

El siguiente proceso efectuado fue obtener la densidad para cada una de las capas.

---

---

Teniendo ya estos valores de inicio se puede comenzar con los procesos y el orden marcado en la sección 5.3, e iniciar con el proceso iterativo de resolución para cada paso de tiempo.

#### 8.4 CALIBRACIÓN

La calibración del modelo es una noción fundamental de la modelación en general y significa “buscar la mejor concordancia entre las variables de estado observadas y las calculadas mediante la variación de un número de parámetros” (Angelini et al., 2000; Logofet et al., 2002).

La calibración se define como las primeras pruebas o afinación del modelo a un conjunto de datos de campo de un sistema. Tal afinación incluye un conjunto consistente y racional de parámetros y constantes teóricamente defendibles. Durante la calibración estos parámetros y constantes se ajustan (Martín et al., 1999; Martin et al., 1999).

El objetivo de la calibración es ajustar los parámetros y constantes de manera que los resultados del modelo se ajusten a las observaciones de una manera óptima (Chapra, 1997).

Existen dos enfoques generales para evaluar la calidad de la calibración: el subjetivo y el objetivo (Chapra, 1997).

La valuación subjetiva está basada en una comparación visual de la simulación con los datos. Estos usualmente consisten en gráficas de series temporales individuales por cada una de las variables de estado. El modelador ajusta parámetros y constantes hasta que estima que la concordancia entre los resultados del modelo y los datos luce adecuada (Chapra, 1997).

En contraste, el enfoque objetivo depende del desarrollo de algunas medidas cuantitativas de la calidad del ajuste (usualmente alguna medida del error cometido). Una vez que tales medidas son adoptadas los parámetros son ajustados hasta que el mejor valor (usualmente un mínimo) es alcanzado (Chapra, 1997).

Hay numerosas medidas que pueden desarrollarse para valorar el ajuste. Una muy empleada es la sumatoria de las diferencias cuadráticas (Chapra, 1997):

---

---

$$S_r = \sum_{i=1}^n (c_{p,i} - c_{m,i})^2 \quad (226)$$

Donde (Chapra, 1997):

$c_{p,i}$  = Predicción de concentración del modelo para el tiempo  $i$ .

$c_{m,i}$  = Concentración medida en el tiempo  $i$ .

La suma de los cuadrados proporciona un valor total que debe ser mínimo cuando las predicciones y las medidas están en concordancia. Esta medida proporciona un valor único que puede monitorearse mientras se ajustan los parámetros (Chapra, 1997).

La calibración para embalses usualmente involucra examinar tanto los perfiles de las concentraciones dentro del almacenamiento (al menos un conjunto de observaciones representativas de la “red interna”), así como las concentraciones en los caudales de salida. Típicamente cuando se analizan perfiles de embalses el comienzo y el rompimiento de la estratificación termal, la localización de la termoclina y las concentraciones del agua en el hipolimnion y el epilimnion, son factores de primera importancia (Martin et al., 1999; Deas et al., 2000).

La evaluación de la calibración puede incluir todos los datos medidos en el perfil vertical o sólo ubicaciones seleccionadas (Environmental Laboratory, 1995; Deas et al., 2000).

Para el caso de la temperatura del agua la comparación de temperaturas cercanas a la superficie es algunas veces pasada por alto en la modelación de grandes embalses, debido a que el agua en la superficie tiende a responder en periodos cortos de tiempo a las condiciones meteorológicas mientras que el agua más profunda responde a variaciones de largo plazo (estacionales) (Martin et al., 1999; Deas et al., 2000).

Cuando los parámetros y constantes del modelo no puedan ser medidos directamente en el campo o el laboratorio frecuentemente son efectuadas estimaciones preliminares de los valores encontrados en la literatura o de relaciones aproximadas. La mayoría de estas estimaciones iniciales serán adecuadas para la calibración final. Sin embargo, un puñado de parámetros críticos deberá normalmente ser cambiado iterativamente, hasta que las simulaciones del modelo concuerden lo mejor posible con el conjunto de datos medidos seleccionado para la calibración (Martín et al., 1999; Paredes et al., 2006a).

El conjunto de valores observados que se usará para calibrar, constituido por los días del periodo y sus profundidades correspondientes, será el mismo que se utilizó en los métodos multivariantes (del 31/03/2004 al 23/02/2005).

Las constantes generales a controlar son las siguientes:

- Ke, coeficiente de extinción lumínica.
- Factor por viento.
- $\beta$ , proporción de la radiación absorbida en la superficie de la radiación entrante neta.
- KeSS, coeficiente de extinción lumínica para los sólidos suspendidos.
- a, coeficiente para calcular la difusión turbulenta vertical.
- c, coeficiente para calcular la difusión turbulenta vertical.

Las constantes se fueron modificando una a una en distintos porcentajes y se midió el error cuadrático de estimación de la temperatura.

Los valores “base” (de inicio para el análisis) de las constantes y parámetro serán los siguientes (Bowie et al., 1985; Chapra, 1997; Hornung, 2002; Centre for Water Research, 2003):

Ke (1/m)	Factor por viento	$\beta$	a	c
0,40	1,00	0,55	0,01	0,005

**TABLA 83** Valores iniciales para la calibración

Considerando lo anterior se inicia con el proceso de calibración de las constantes para las cuales sus valores poseen los rangos de variación aceptados siguientes (Bowie et al., 1985; Hornung, 2002; Centre for Water Research, 2003):

Constante o parámetro	Rango de variación
Ke	0,03 a 4,0 (1/m)
Factor por viento	1,0 a 1,6
$\beta$	0,55 (fijo)

**TABLA 84** Rangos de variación típicos de algunas constantes y parámetros

El orden de calibración será el siguiente:

1. Ke.

- 
- 
2. Factor por viento.
  3. a.
  4. c.

Este orden fue elegido así debido a que se inicio con la calibración de las constantes más generales (que afectan a más procesos) y luego con las más específicas (que involucran menos procesos) (Martín et al., 1999), así como también atendiendo a la estructura del programa estructurado en Visual Basic 2005. La calibración fue efectuada de manera manual y atendiendo como medida de efectividad de la calibración al error cuadrático medio. El número de simulaciones efectuado fue muy numeroso (más de cincuenta), para tener la seguridad de elegir correctamente los coeficientes. Con cada coeficiente analizado (la calibración fue realizada optimizando un coeficiente a la vez) se elaboraron gráficos de valor propuesto del coeficiente contra el error cuadrático medio para contar con elementos detallados que aseguren la elección correcta del coeficiente en cada caso.

Los valores obtenidos en la calibración son los siguientes:

Constante o parámetro	Valor
Ke	0,30
Factor por viento	1,10
a	0,04
c	0,004

**TABLA 85** Valores obtenidos en la calibración

Y los resultados asociados a la modelación de la variable de salida, temperatura del agua, se muestran a continuación como perfiles verticales promedio mensuales de los meses de marzo, abril, julio, agosto, septiembre, octubre, noviembre y diciembre del año 2004 (Figuras 57 a 66); además los meses de enero y febrero del año 2005.

---

---

Gráficos promedio mensuales

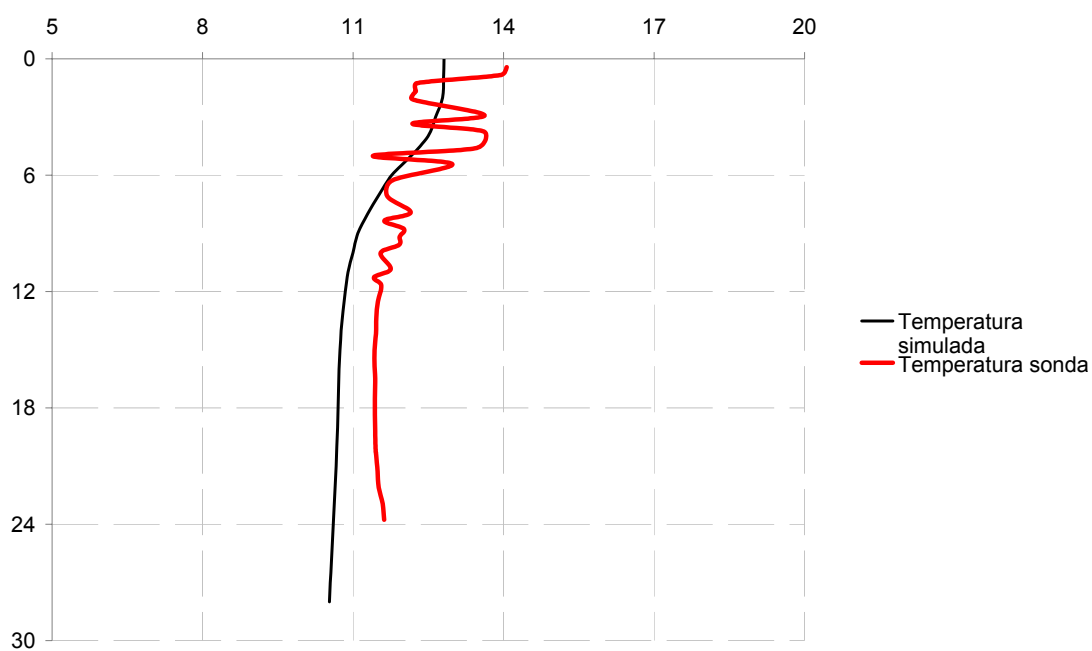


FIGURA 57 Perfil de temperaturas simulada y observada (sonda) para el mes de marzo de 2004

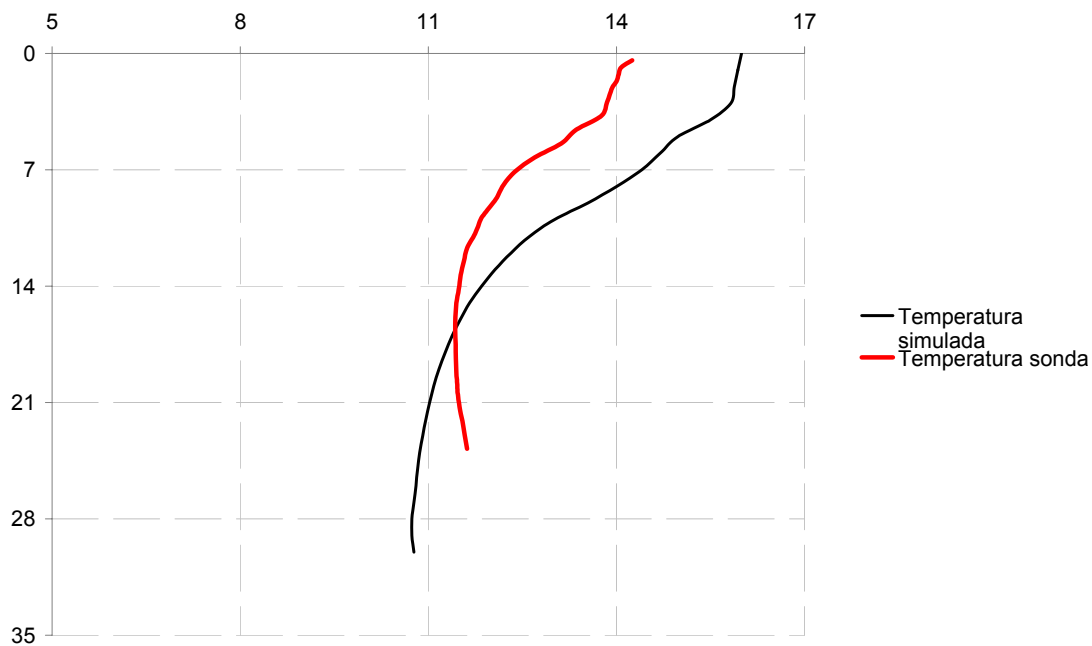
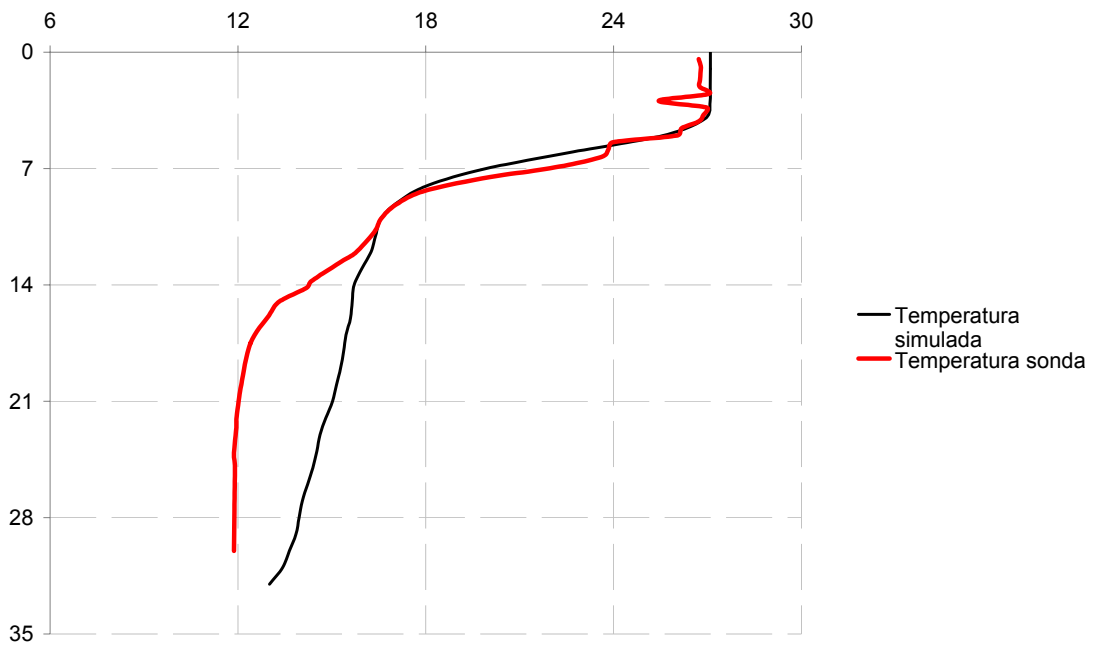
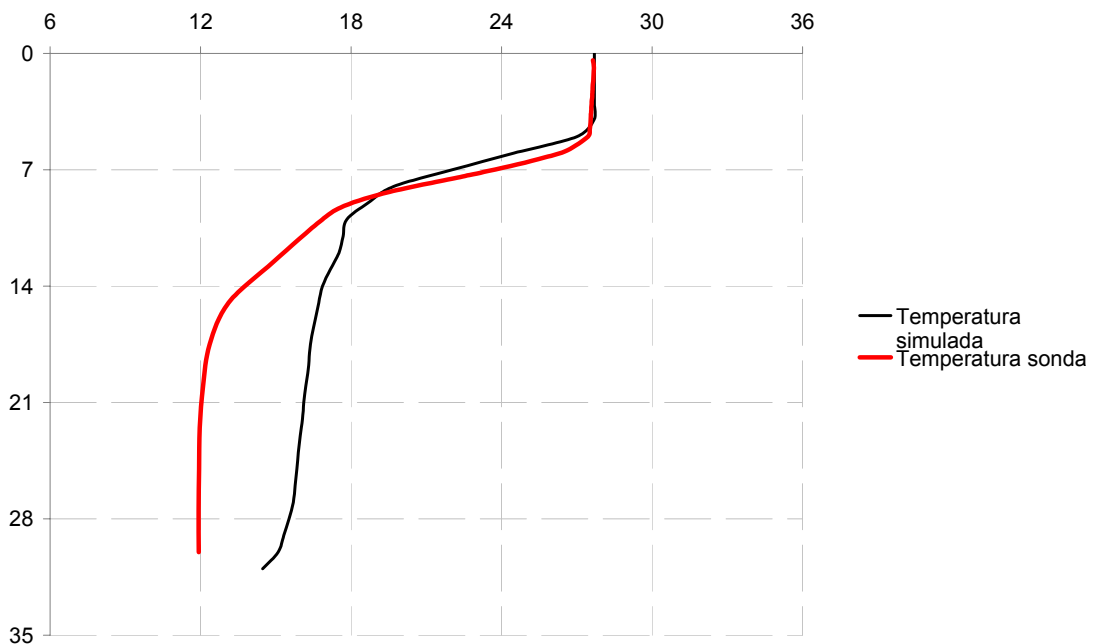


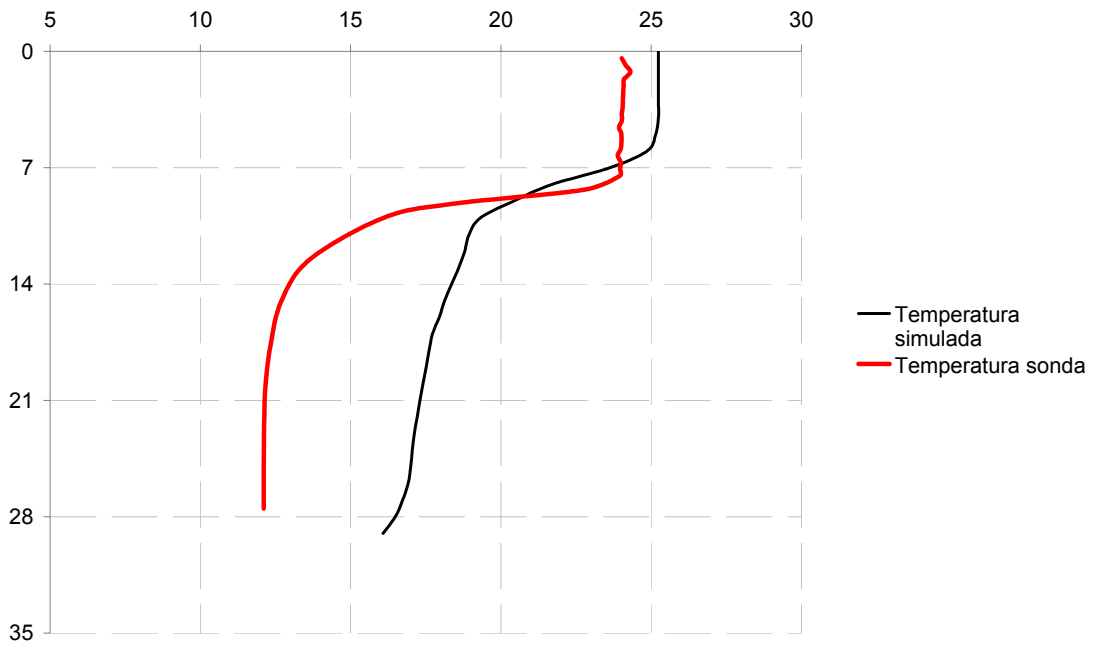
FIGURA 58 Perfil de temperaturas simulada y observada (sonda) para el mes de abril de 2004



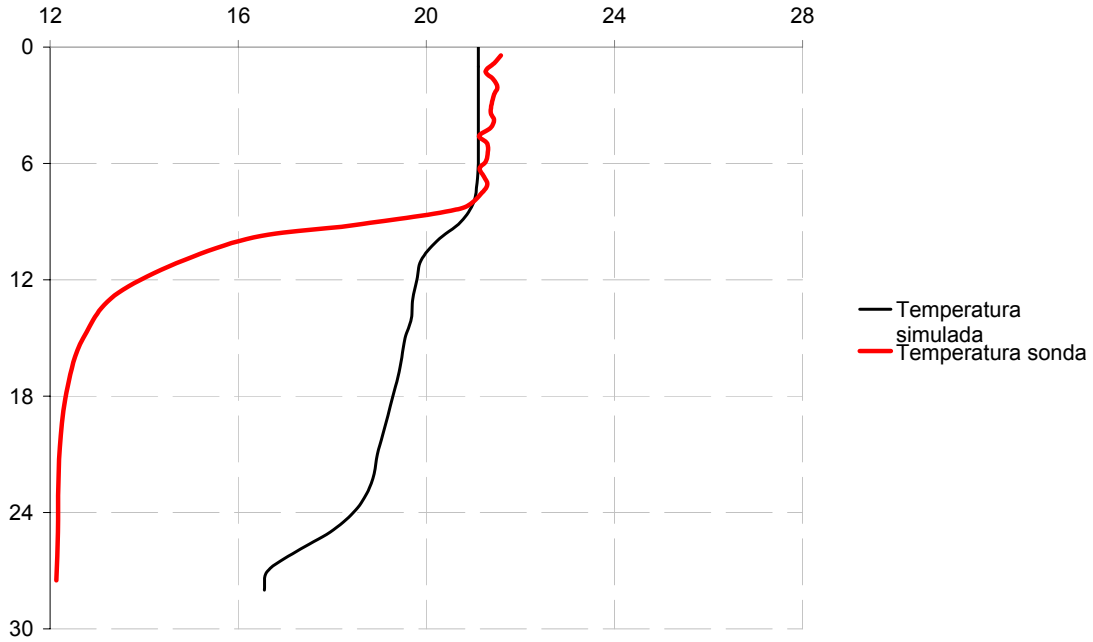
**FIGURA 59** Perfil de temperaturas simulada y observada (sonda) para el mes de julio de 2004



**FIGURA 60** Perfil de temperaturas simulada y observada (sonda) para el mes de agosto de 2004

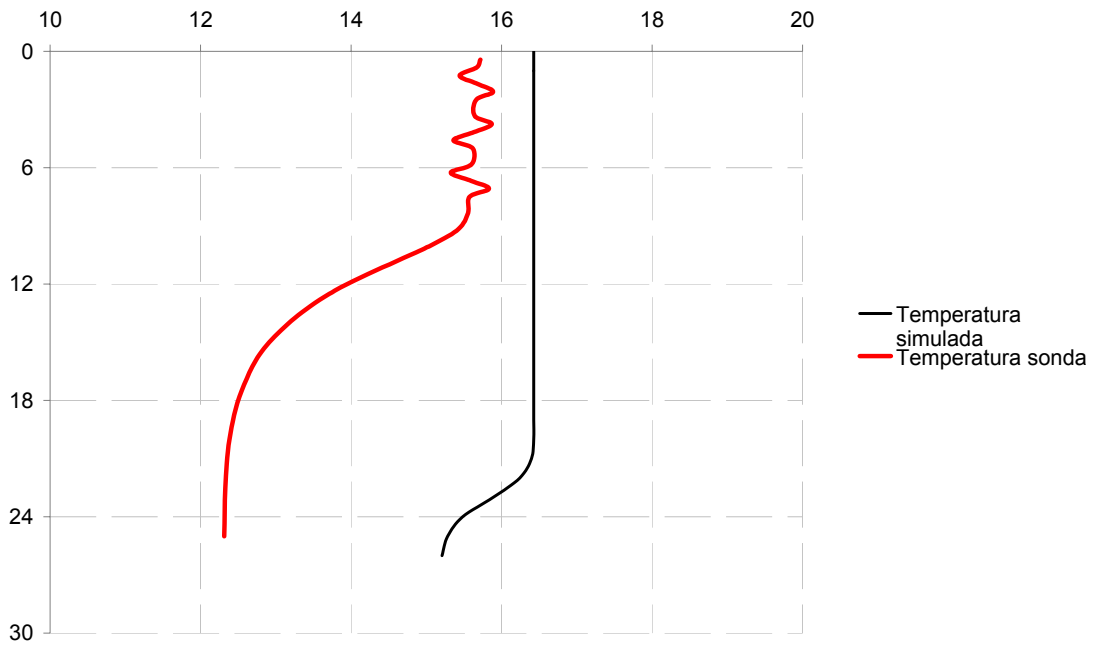


**FIGURA 61** Perfil de temperaturas simulada y observada (sonda) para el mes de septiembre de 2004

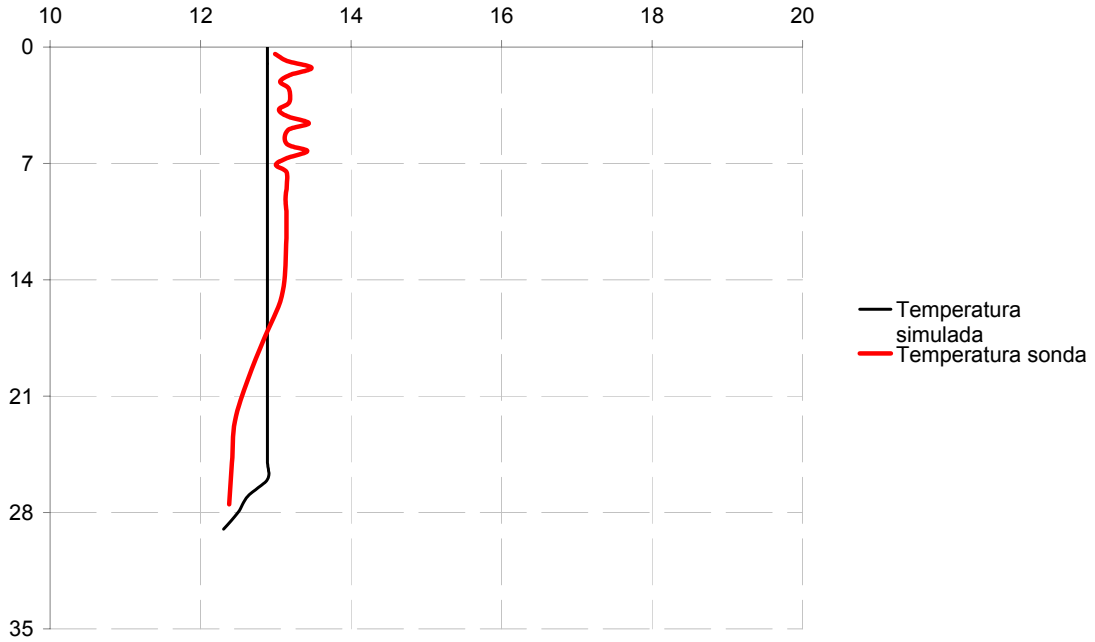


**FIGURA 62** Perfil de temperaturas simulada y observada (sonda) para el mes de octubre de 2004

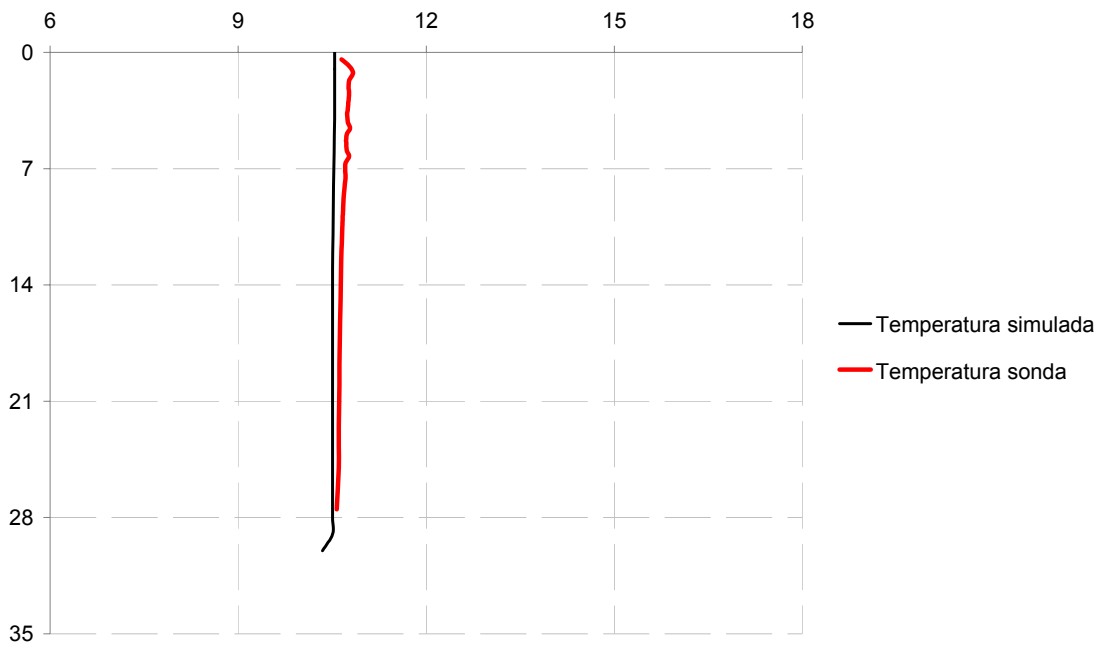




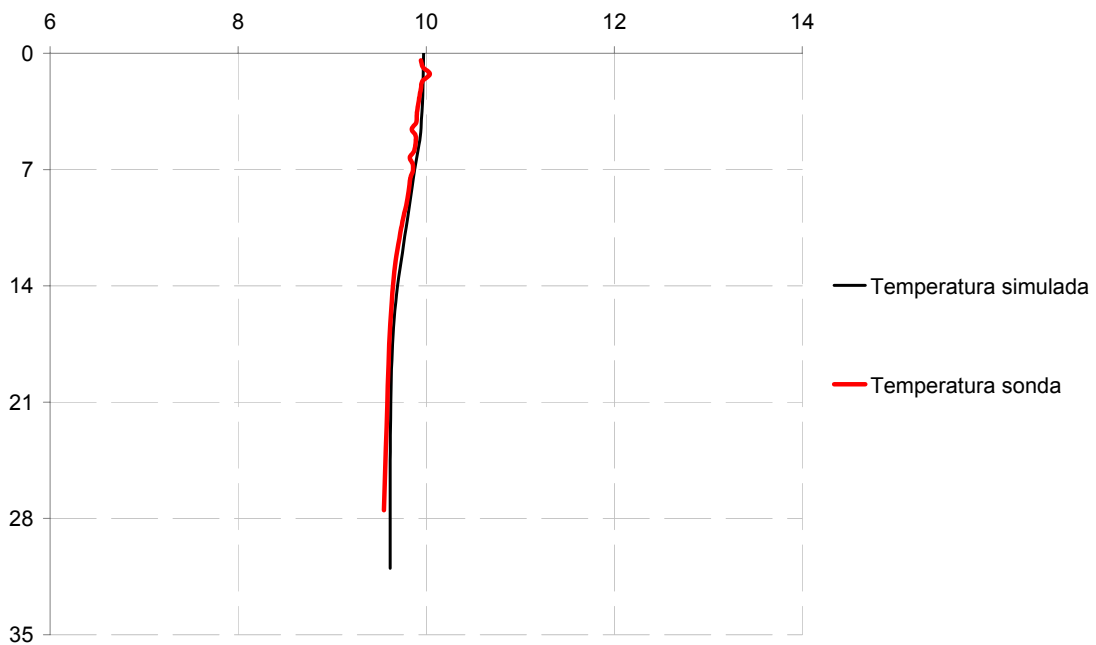
**FIGURA 63** Perfil de temperaturas simulada y observada (sonda) para el mes de noviembre de 2004



**FIGURA 64** Perfil de temperaturas simulada y observada (sonda) para el mes de diciembre de 2004



**FIGURA 65** Perfil de temperaturas simulada y observada (sonda) para el mes de enero de 2005



**FIGURA 66** Perfil de temperaturas simulada y observada (sonda) para el mes de febrero de 2005

---

---

## 8.5 VALIDACIÓN

La validación o verificación del modelo consiste en usar o probar el modelo calibrado en un conjunto diferente de datos de calidad del agua a los utilizados durante el proceso de calibración preferiblemente bajo diferentes condiciones externas (condiciones iniciales y de frontera) para examinar la validez del modelo (Schladow et al., 1997; Hornung, 2002). Una vez que se ha completado esta fase de verificación o auditoria el modelo puede considerarse verificado (Martin et al., 1999).

En el proceso de validación los parámetros del modelo están fijos y no se permiten ajustes adicionales después del paso de calibración (Martín et al., 1999; Deas et al., 2000).

La simulación que se efectúa con el modelo durante el proceso de validación es comparada con los datos registrados (variables respuesta) para valuar el grado de fiabilidad del modelo para hacer simulaciones de tipo seguro. La experiencia, pruebas estadísticas de variables importantes y otros métodos son usados para decidir si el modelo es fiable para los propósitos que se tienen (Martin et al., 1999).

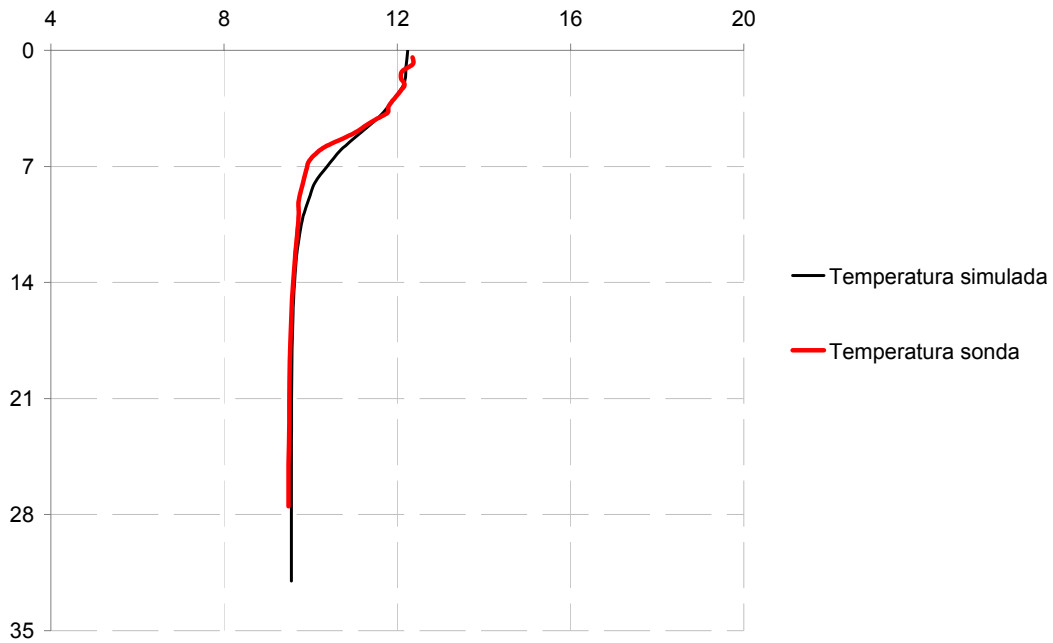
La validación no implica una amplia aplicabilidad del modelo. Si la nueva simulación con el modelo coincide con los nuevos datos recolectados, el modelo ha sido confirmado como una herramienta de predicción efectiva para el rango de condiciones definida por la calibración original y el conjunto de datos de validación (Martin et al., 1999; Deas et al., 2000).

Un modelo calibrado debe seguirse validando para asegurarnos de que proporcione buenos resultados. La calibración y la validación son actualmente consideradas como pasos iniciales de un proceso iterativo que puede involucrar calibraciones y validaciones repetidas cada cierto periodo de años con el fin de actualizar y diseñar un modelo operacional (Chapra, 1997). Además, la calidad, cantidad y tipo de datos para validar puede aumentar en el tiempo.

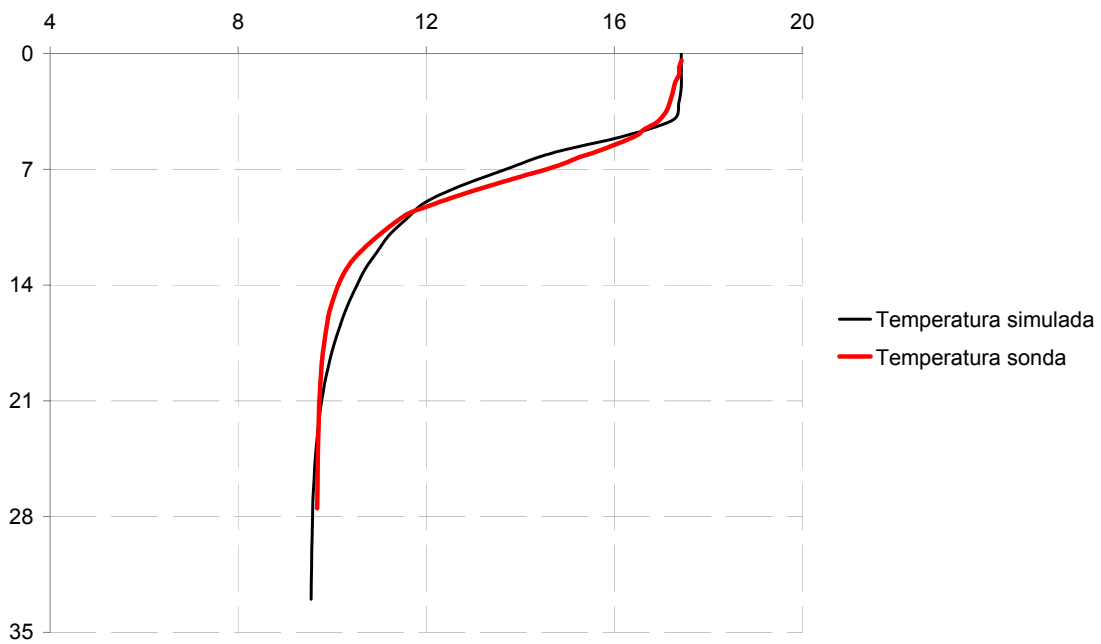
La validación se efectuó con los mismos días y sus profundidades correspondientes utilizados con los métodos multivariantes, del 24/02/2005 al 19/09/2005, en el cual las condiciones dentro del embalse fueron significativamente similares con respecto al periodo elegido para la calibración.

Para la variable de salida temperatura se elaborarán gráficas comparativas de los valores registrados (observados) con la sonda multiparamétrica y los simulados con el

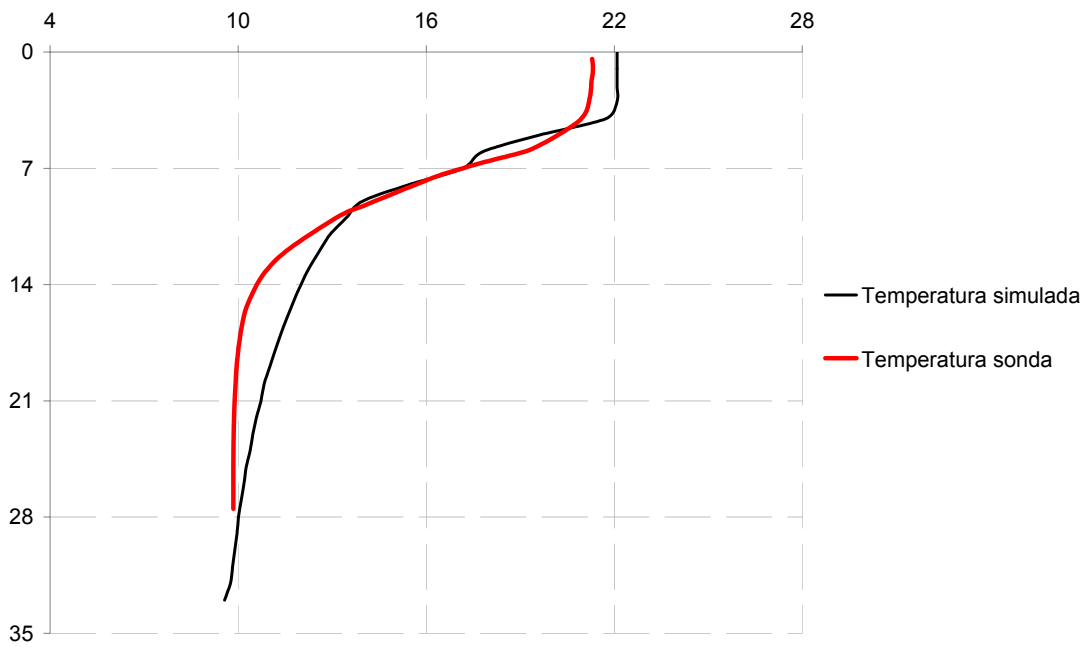
modelo. Se muestran las gráficas del perfil medio mensual a continuación (Figuras 67 a 73) para los meses en los que se efectúa la validación, que son para el año 2005 los meses de: marzo, abril, mayo, junio, julio, agosto y en septiembre se representa el periodo comprendido desde el día primero hasta el día veinte del mismo mes.



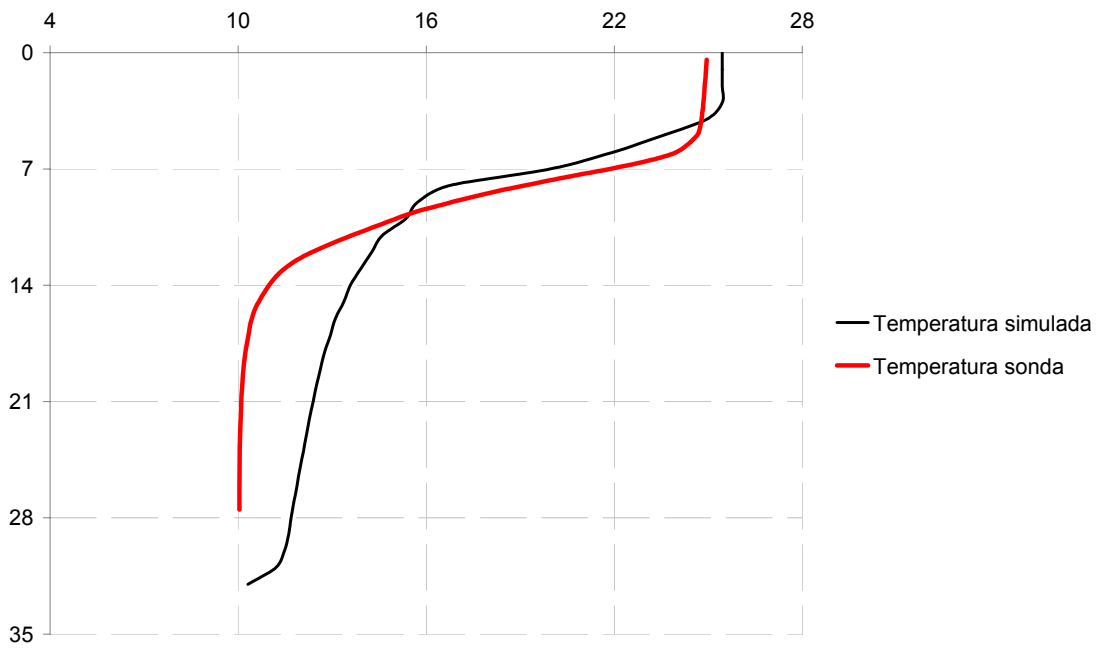
**FIGURA 67** Perfil de temperaturas simulada y observada (sonda) para el mes de marzo de 2005



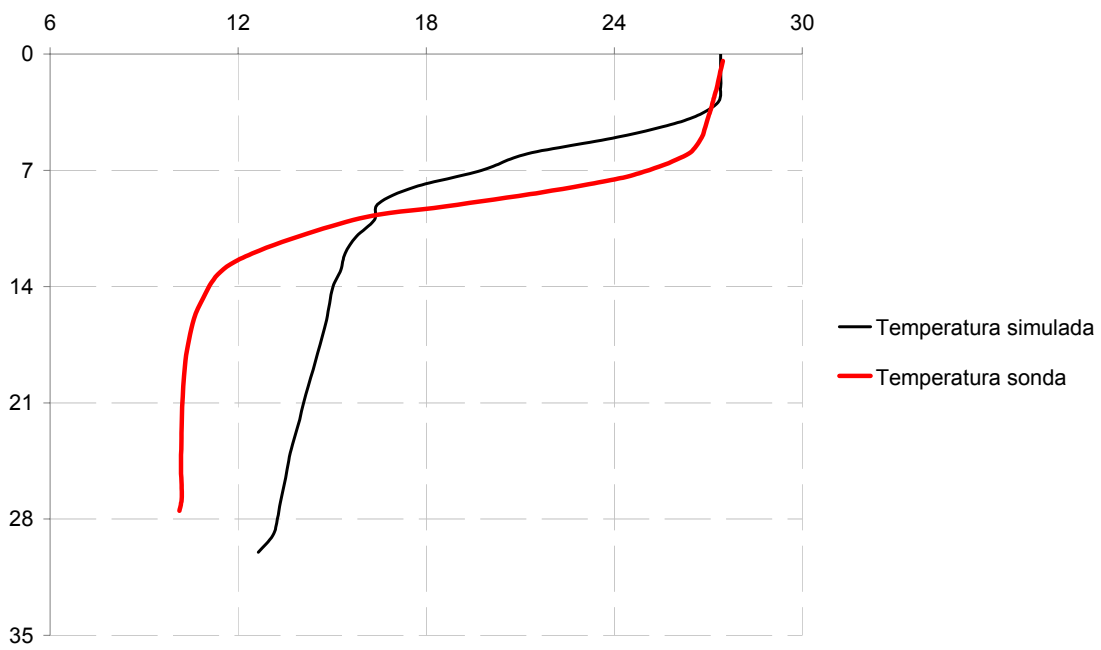
**FIGURA 68** Perfil de temperaturas simulada y observada (sonda) para el mes de abril de 2005



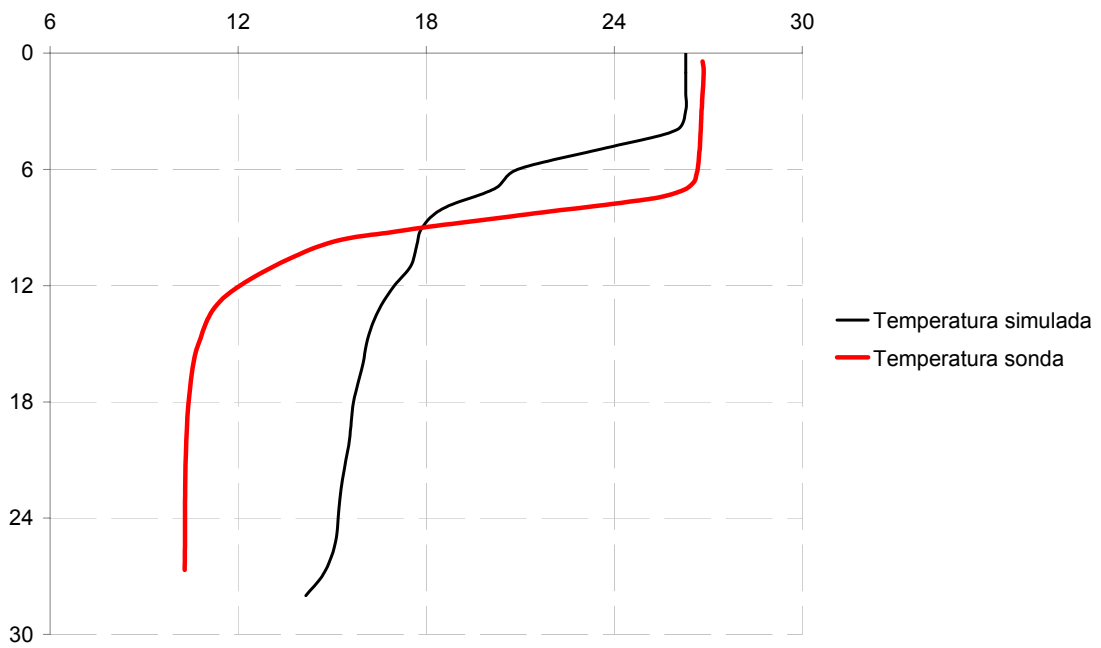
**FIGURA 69** Perfil de temperaturas simulada y observada (sonda) para el mes de mayo de 2005



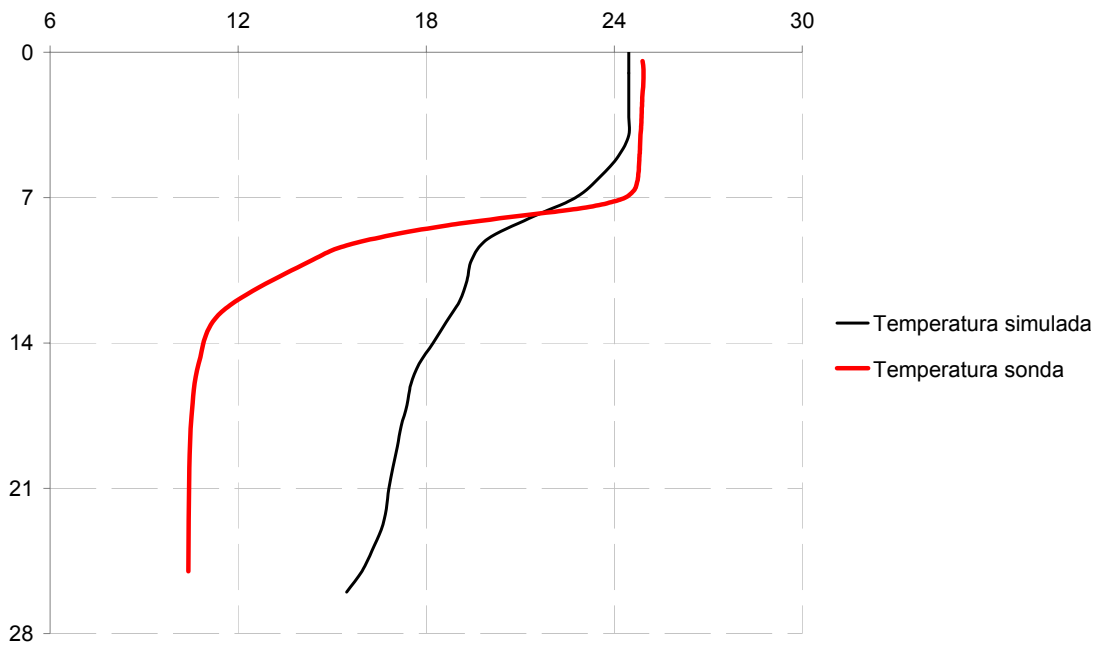
**FIGURA 70** Perfil de temperaturas simulada y observada (sonda) para el mes de junio de 2005



**FIGURA 71** Perfil de temperaturas simulada y observada (sonda) para el mes de julio de 2005



**FIGURA 72** Perfil de temperaturas simulada y observada (sonda) para el mes de agosto de 2005



**FIGURA 73** Perfil de temperaturas simulada y observada (sonda) para el mes de septiembre de 2005

---



---

---

## CAPÍTULO 9. ANÁLISIS DE RESULTADOS Y DISCUSIÓN

### 9.1 OBTENCIÓN DE VARIABLES PREDICTIVAS

#### 9.1.1 De operación del embalse

De las variables correspondientes al embalse las más importantes en las técnicas CCA y PLSR fueron VolExp (Volumen de explotación, m<sup>3</sup>), Qsal (Caudal de salida del embalse, m<sup>3</sup>/día) y Profundidad (m).

En el método mecanicista las variables de estado del embalse que se requieren son los Caudales de entrada y salida del embalse, Qent y Qsal, el Volumen en el embalse, VolExp y la Filtración, Filt.

Para determinar los valores más probables de VolExp y Qsal (la Profundidad se puede tomar a varios niveles y la profundidad máxima depende del volumen embalsado a través de la tabla calados-volúmenes) se pueden seguir varias metodologías.

Una de ellas sería en base a consultar a los Especialistas encargados de la gestión del embalse por medio del llamado “Juicio de experto” para contar con valores estimados de las variables del almacenamiento.

Otra metodología sería el empleo de series temporales, pues estas variables han sido medidas a lo largo de muchos años y se aprecia que los registros están completos (una serie temporal proviene de registrar los valores de una variable a través del tiempo en intervalos regulares) (Peña, 2005a). Si estos modelos se basan exclusivamente en la historia de la propia serie se denominan univariantes. Los pronósticos obtenidos con este tipo de modelos se basan en la suposición de que las condiciones futuras serán semejantes a las del pasado y son muy útiles para predicciones a corto plazo. Por otra parte, al tomar en cuenta además de las variables del embalse otras cantidades como son la evolución de la población servida residente, afluencia de turistas, demanda agrícola o superficie cultivada, disponibilidad de otras fuentes del recurso dentro de la zona de aprovechamiento del almacenamiento y las variables meteorológicas, se puede hacer uso de los modelos multivariantes de series temporales. Con estos existe la posibilidad de mejorar las estimaciones de las variables de estado del embalse (Peña, 2005a).

El caudal de entrada se puede obtener de los registros de bombeos hacia el embalse, así como los volúmenes diarios aportados por las depuradoras en la cuenca que

---

---

constituyen en este caso las principales contribuciones. Así que se deben revisar las proyecciones disponibles del cambio estimado de estas cantidades.

La filtración se puede obtener por balance de las entradas y salidas del almacenamiento, por medio de las estimaciones de las cuales se disponga de estas últimas cantidades.

### 9.1.2 Meteorológicas

Las variables meteorológicas más importantes para los métodos CCA y PLSR fueron T (Temperatura media del aire, °C), Tn (Temperatura mínima del aire, °C), Tx (Temperatura máxima del aire, °C), Hf (Horas de frío, h), HR (Humedad relativa media del aire, %) y ETo (Evapotranspiración de referencia, mm).

Además de las variables ya mencionadas en el párrafo anterior el método mecanicista requiere la velocidad media del viento, la radiación y la precipitación.

Una forma de estimar estas variables y que sirvan de base a la predicción de los parámetros de calidad del agua es por medio del “Juicio de experto” emitido por un profesional de la meteorología que conozca la evolución temporal de las variables de interés.

Otro método sería el de series temporales univariantes (para periodos de predicción cortos). Para los periodos de predicción más amplios se puede hacer uso de los métodos multivariantes de series temporales involucrando información de otras variables (Peña, 2005a) meteorológicas no incluidas y de estaciones vecinas a la estación meteorológica de Villajoyosa, en su caso.

También existen los llamados modelos del clima. Estos sirven para efectuar predicciones en las cuales se trabaja con anomalías (estimaciones de desviaciones mensuales o trimestrales, con referencia a la climatología), o con su carácter (positivo, normal o negativo) y no con los valores específicos de las variables (Gutiérrez et al., 2004).

Existen los modelos acoplados océano-atmósfera de pronóstico estacional que en la actualidad permiten pronosticar con cierto grado de certidumbre anomalías mensuales o estacionales durante ciertas épocas y para algunas regiones del mundo (asociadas primordialmente a las anomalías en la ENSO, El Niño-Oscilación del Sur) (Gutiérrez et al., 2004).

---

---

El problema de algunos modelos globales del clima es su resolución espacial, pues generan predicciones para superficies muy extensas ( $80 \times 80$  km o mayores), con lo cual aspectos locales como la orografía, vegetación, cuerpos de agua, etc., no son considerados para una región más reducida (momac.uclm.es/).

Un modelo que parte de la salida del modelo global AVN (NCEP-NOAA) y que busca superar el problema de la resolución espacial, presentando un mayor nivel en el detalle es el PROMES desarrollado en España por el Grupo de Modelización para el Medio Ambiente y el Clima. Al ser el PROMES un modelo de área limitada, el dominio de aplicación está centrado sobre la Península Ibérica. Durante la simulación es necesario aportarle de manera continua los valores de las variables atmosféricas en los contornos laterales de la región de dominio del modelo. Estos valores provienen de la salida del modelo AVN (NCEP-NOAA). La resolución horizontal del PROMES es de  $15 \times 15$  km. El mismo grupo de investigación desarrolla un modelo del clima a escala regional para simular el clima presente y futuro en Europa y la cuenca Mediterránea (momac.uclm.es/). Este modelo puede ayudar con los valores meteorológicos requeridos por las técnicas CCA y PLSR.

Como datos especiales que también requiere el desarrollo del modelo mecanicista están la temperatura del agua superficial en el embalse y la cobertura de nubes. Para el caso de la temperatura del agua superficial en el embalse se efectuaron varias regresiones y se determinó que la más adecuada era una en función de la variable Día Juliano por medio de una función de forma cosenoidal, aplicable para todo el año.

Las estaciones automáticas como la de Villajoyosa no registran la cobertura de nubes (Deas et al., 2000). Con el fin de obtener la cobertura de nubes media diaria, se planteó una metodología que utiliza el flujo de radiación de onda corta teórica que llega a la parte externa de la atmósfera terrestre, la fracción de radiación que alcanza la superficie del agua después de la reducción por difusión y absorción, el albedo de la superficie del agua y la radiación solar media registrada del día. Debido a la gran cantidad de cálculos involucrados se elaboró un programa en Visual Basic 2005 que desarrollara este planteamiento.

## 9.2 ESTUDIO INICIAL DE LOS DATOS POR MÉTODOS MULTIVARIANTES

Uno de los aspectos que atrae primero la atención en la gestión de grandes volúmenes de datos es la consideración de su confiabilidad. Según Kettaneh et al. (2005), todos los grandes conjuntos de datos poseen datos anómalos.

---

---

Es un hecho que al contar con amplios grupos de datos habrá algunas partes dentro del conjunto que presenten ausencias, omisiones o errores (debidos a varias fuentes), variaciones forzadas o inconsistencias. Por ejemplo, en la información proveniente de la sonda multiparamétrica se observa que en algunos días se midieron las variables a diferentes profundidades en vez de las acostumbradas comúnmente o que en otros sólo se contaba con un perfil diario, pero en estos casos no se tomaba a la misma hora del día (habrá variación en las mediciones al ser registradas a las 3:00 horas o a las 14:00, sólo analícese el caso de la temperatura del agua para distintas horas del día); en ocasiones si se tenían varios perfiles para obtener un promedio de ellos todos eran de un mismo periodo del día (mañana, tarde o noche). Por otra parte, tanto los datos meteorológicos como los hidrológicos también pueden poseer errores inherentes.

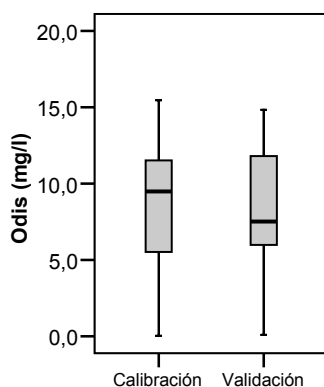
El objetivo es tener conciencia de estos aspectos (que siempre se presentan en diversas formas en la práctica del análisis de datos) para intentar por medio de diversas técnicas minimizar sus efectos adversos en el estudio y modelación de las variables de interés.

Para que los procedimientos estadísticos pudieran ser aplicados de manera confiable de los datos iniciales fueron removidos valores a todas luces erróneos o días que exhibían valores ausentes. Después de que la información pasó por este tamiz, con el propósito de realizar inferencias se consideró la división de ésta en dos periodos: calibración y validación. Para esto, los días existentes se dividieron aproximadamente en dos partes iguales y la primera de ellas (con respecto al tiempo), fue considerada como datos de calibración y la otra se usó para validación.

Pero con el propósito de que las conclusiones de cualquier análisis, cálculo o modelo obtenidas en los datos de calibración puedan ser aplicables a los datos de validación es necesario revisar si las variables en ambos periodos toman valores (aproximadamente) dentro de los mismos límites (Figura 74). Para efectuar esto se obtuvieron estadísticas descriptivas de las variables para los dos periodos (Anexo 2). La conclusión es que en la mayoría de los casos los límites que toma cada una de las variables en los dos periodos son muy similares; para las variables  $H_f$ ,  $P$ ,  $Q_{ent}$ ,  $Q_{sal}$  y  $Con$ , que son las que presentan las diferencias más significativas, los límites se encuentran dentro de los mismos órdenes de magnitud. Por lo tanto, haciendo referencia a estadísticas descriptivas de las variables disponibles se puede concluir que en el periodo de calibración y el de validación las variables toman valores aproximadamente dentro de los mismos límites. Se puede llegar a la misma conclusión

---

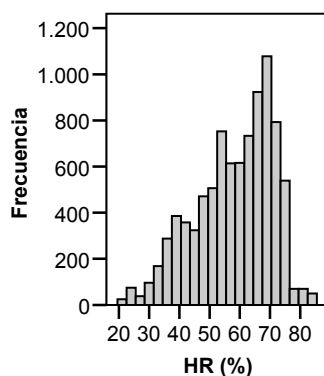
al revisar los diagramas de caja y bigotes en el Anexo 3.



**FIGURA 74** Diagrama de caja y bigotes de oxígeno disuelto para los periodos de calibración y validación

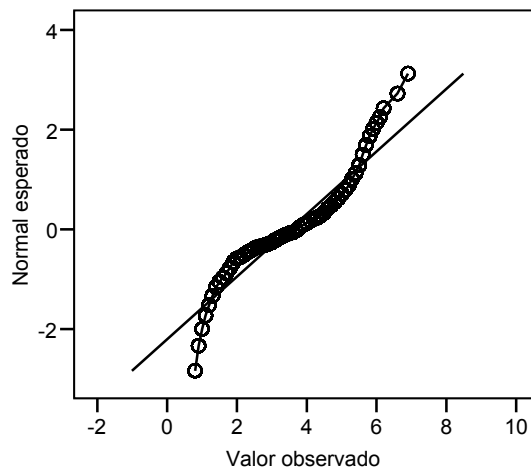
En la Tabla (12) se encuentran las variables utilizadas en esta sección junto con sus abreviaturas.

Respecto a la normalidad de las variables se puede observar en los histogramas de la Figura (45) que ninguna de las variables exhibe una distribución normal (la formas de sus histogramas no se asemejan al gráfico de una distribución normal, como se observa en la Figura 75).



**FIGURA 75** Histograma de humedad relativa (%) para el periodo completo de estudio

Esta apreciación es confirmada al revisar los diagramas Q-Q de los datos iniciales (Anexo 5 y Figura 76) y la prueba de normalidad de Kolmogorov-Smirnov (con la corrección de Lilliefors) para cada una de las variables (Tabla 14).



**FIGURA 76** Diagrama Q-Q de evapotranspiración para el periodo de estudio completo

Se intentó solucionar la situación de falta de normalidad en los datos probando para esto distintos tipos de transformaciones (sección 7.4). Debido a la estructura adoptada para los datos ninguna de las transformaciones intentadas sobre los valores logró que cumplieran con la prueba de normalidad.

### 9.3 APLICACIÓN DEL PCA

Con el objetivo de proporcionar la misma influencia a las variables en este análisis los datos se estandarizaron. Se realizó primero el análisis de componentes principales del periodo de calibración para todas las variables (meteorológicas, hidrológicas, calidad del agua y Día Juliano). Se obtuvieron cinco componentes con autovalores mayores que la unidad (hubo una sexta componente con un autovalor de 0,98 pero al no satisfacer el criterio no se consideró) y que en conjunto explican un 75,9% de la variabilidad contenida en los datos originales.

Análisis de Componentes Principales			
Componente número	Autovalor	Porcentaje de varianza	Porcentaje acumulado
1	6,79448	32,355	32,355
2	3,16844	15,088	47,442
3	2,73154	13,007	60,450
4	2,12718	10,129	70,579
5	1,11699	5,319	75,898
6	0,975219	4,644	80,542

**TABLA 86** Porcentaje de varianza explicada por las primeras componentes principales para el periodo de calibración, variables estandarizadas

La primera componente explica un 32,4% de la varianza. En el Anexo 7 se observa que las variables que poseen mayores pesos ( $|\text{peso}| > 0,25$ ) en esta primera componente son T, Tn, Tx, Rad, Eto, Qsal y Ta. Se puede interpretar que esta

componente representa la energía térmica en el sistema y su balance.

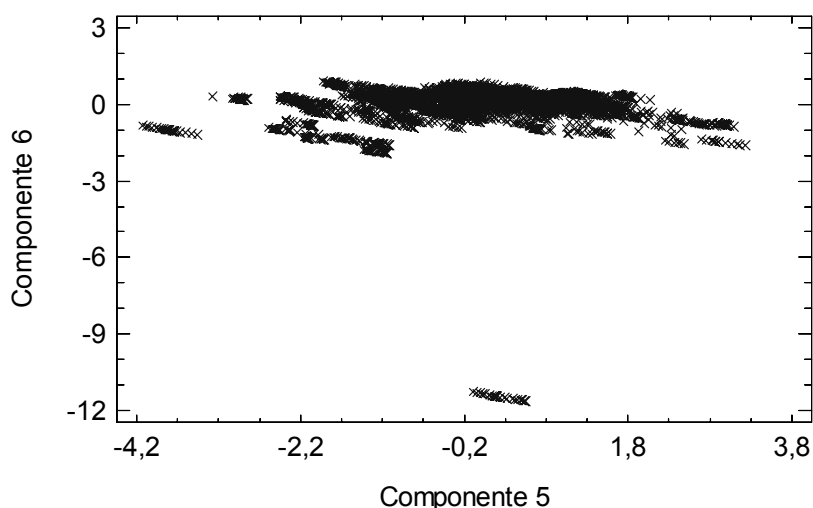
La segunda componente representa un 15,1% de la varianza. Las variables con pesos altos en esta componente son DJ, Hf, Rad, VolExp, Odis y Con. Esta componente se relaciona con la política de gestión del embalse, la variación estacional de energía térmica y los parámetros de calidad del agua más afectados por estas influencias.

Variable	Componentes						
	1	2	3	4	5	6	7
DJ	0,134344	0,371215	-0,069300	-0,332976	-0,001983	-0,077019	0,078531
V	-0,054194	-0,258480	0,274217	-0,404803	-0,135416	-0,081706	0,058346
DV	-0,206907	0,065572	0,028698	-0,260413	0,331011	0,079670	0,045238
Vx	-0,024976	-0,162526	0,206584	-0,408870	-0,485762	-0,134045	-0,112845
T	0,375898	0,018683	-0,023962	-0,010627	-0,062084	0,037994	0,002141
Tn	0,372289	0,048966	-0,039415	-0,003146	-0,053546	0,023360	-0,027970
Tx	0,371293	0,003470	-0,024354	-0,003343	-0,031251	0,069052	0,042544
Hf	-0,198504	-0,251637	0,136428	0,106961	0,369485	-0,111238	-0,266187
HR	0,018820	0,247806	-0,270606	0,426459	-0,119554	0,011712	-0,017541
Rad	0,290548	-0,261377	0,135362	0,108942	0,010109	0,103157	0,038903
P	-0,061232	0,229358	-0,117962	0,003346	-0,293545	-0,243116	-0,786717
Eto	0,336220	-0,205651	0,127691	0,004569	-0,077207	0,037001	-0,070170
VolExp	0,082030	-0,404050	0,086351	0,353197	-0,175572	-0,002178	-0,092796
Qent	-0,216803	0,116647	-0,064483	0,034106	-0,401812	0,099364	0,324327
Qsal	0,285674	0,200727	-0,066627	-0,217541	0,231297	-0,041039	-0,040710
Filt	-0,063614	0,018968	0,022872	-0,136527	-0,013260	0,912368	-0,333095
Profundidad	0,013624	0,175023	0,432751	0,216761	-0,143239	0,079461	0,072626
Ta	0,282310	-0,080991	-0,299427	-0,172867	0,090590	-0,031996	-0,052403
Odis	-0,227576	-0,254965	-0,358857	-0,098802	0,004540	-0,002016	-0,005006
pH	-0,060446	-0,098104	-0,407869	-0,054493	-0,327100	0,142024	0,198437
Con	-0,046205	0,380305	0,380468	0,058782	-0,060759	0,034680	0,051922

**TABLA 87** Tabla de pesos de los primeros siete componentes en el periodo de calibración

La tercera componente contiene un 13,0% de la varianza total del conjunto. Esta componente se encuentra más influida por las variables V, HR, Profundidad, Ta, Odis, pH y Con. La tercera componente representa la influencia del mezclado por el viento y la variación de los parámetros de calidad del agua con respecto a la profundidad. Las otras componentes no se analizarán debido a que representan porcentajes de varianza menores.

Asimismo, continuando con el PCA mostrado en el Anexo 7 los diagramas de dispersión (Figura 77) fueron de gran ayuda para estudiar valores sospechosos de ser erróneos. El proceso completo para su detección y su corrección en los casos donde la revisión demostró que eran con alta probabilidad datos incorrectos se muestra en la sección 7.5.1.



**FIGURA 77** Diagrama de dispersión de las componentes 5 y 6 en el periodo de calibración mostrando algunos probables datos atípicos

A continuación se aplicó el PCA al periodo de validación obteniéndose seis componentes con autovalores mayores a la unidad, que agrupadas explican un 80,4% de la varianza de los datos en este periodo.

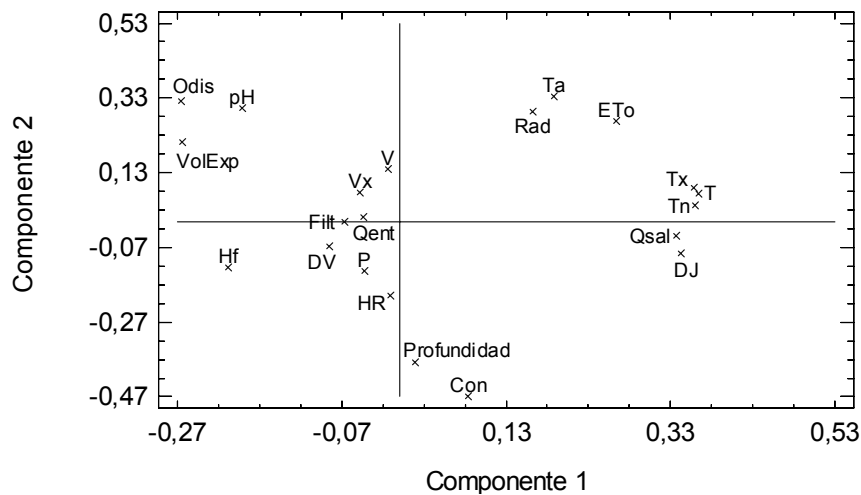
La primera componente explica un 33,7% de la varianza. En el Anexo 9 se muestra que las variables con mayores pesos en esta componente son DJ, T, Tn, Tx, ETo, VolExp, Qsal y Odis (Figura 78). Se podría interpretar como una componente asociada con la operación en el tiempo del embalse, la temperatura ambiente y el oxígeno disuelto.

Análisis de Componentes Principales			
Componente número	Autovalor	Porcentaje de varianza	Porcentaje acumulado
1	7,06814	33,658	33,658
2	3,10922	14,806	48,464
3	2,61874	12,470	60,934
4	1,71601	8,171	69,105
5	1,26001	6,000	75,105
6	1,11537	5,311	80,417

**TABLA 88** Porcentaje de varianza explicada por las primeras componentes principales para el periodo de validación, variables estandarizadas

La segunda componente representa un 14,8% de la variabilidad total. Las variables que más influyen en esta componente son Rad, ETo, Profundidad, Ta, Odis, pH y Con (Figura 78). La componente se asocia con el balance de energía en el almacenamiento, así como el cambio de las variables de calidad del agua con la profundidad.



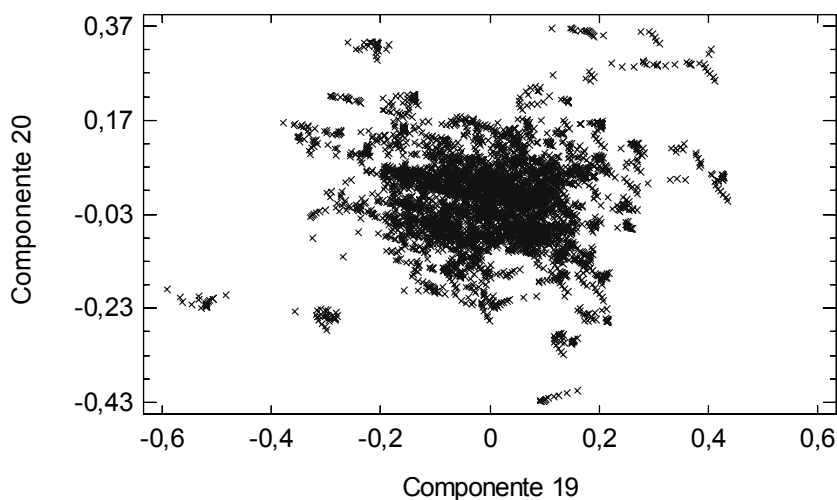


**FIGURA 78** Gráfico de pesos de los componentes 1 y 2 en la validación, datos sin corregir

La tercera componente contiene un 12,5% de la varianza. Esta más influida por las variables V, Vx y HR; se puede relacionar al efecto del viento y la humedad.

Con base en los diagramas de dispersión (Figura 79) del Anexo 9 se revisaron los datos que podrían ser incorrectos. Los pasos seguidos para sustituir esos valores se presentan en la sección 7.5.1.

Atendiendo solamente al porcentaje de variabilidad explicado por las componentes tomadas individualmente o en conjuntos, se puede decir que los periodos de calibración y validación presentan valores muy similares.



**FIGURA 79** Diagrama de dispersión de las componentes 19 y 20 en el periodo de validación mostrando algunos probables datos atípicos

---

---

## 9.4 RESULTADOS DEL ANÁLISIS FACTORIAL

El análisis factorial se utilizará esencialmente como elemento de análisis exploratorio del conjunto de datos.

### 9.4.1 Periodo completo de estudio, variables sin transformar

En el Anexo 12 se muestran los resultados del análisis factorial realizados para el periodo entero de estudio estando las variables sin transformar. Aparte de los comentarios vertidos en la sección 7.6.1 se efectuará a continuación la revisión de algunos elementos del análisis efectuado.

De entrada la matriz de correlaciones mostrada en el Anexo 12 parece adecuada, pues muchos de sus elementos fuera de su diagonal principal son distintos de cero. Una matriz de correlaciones cercana a una matriz identidad señala que el análisis factorial desembocará en una solución deficiente. En la parte inferior de la matriz de correlaciones aparece el nivel crítico unilateral (Sig. unilateral) asociado a cada coeficiente de correlación. Un nivel crítico inferior a 0,05 señala que la correlación entre el par de variables correspondientes puede ser considerada significativamente distinta de cero. Debido a esto lo mejor es encontrar muchos niveles críticos de valor reducido, como sucede en este caso. Una excepción a esto ocurre con la Profundidad que muestra muchos valores superiores al nivel 0,05, pero esto no debe alarmar pues esta variable se establece de acuerdo a consideraciones operativas (en qué niveles se quiere efectuar las lecturas) y por lo tanto no presenta correlación con la mayoría de las variables (excepto con las que tienen relación con ella, como por ejemplo el volumen almacenado).

Al examinar las variables de calidad del agua se observa que  $T_a$  posee altas correlaciones con:  $T$ ,  $T_n$ ,  $T_x$ ,  $ETo$ ,  $Qsal$  y Profundidad; esto sugiere que la temperatura del agua depende principalmente de la temperatura del aire y sus picos, ya que al aumentar estas también aumenta  $T_a$  debido a la conducción del calor en la superficie; así como de la evapotranspiración que al crecer genera un incremento positivo en  $T_a$ , esto indica que cuando existen las condiciones para alta evaporación estas también producen el calentamiento del agua; depende también de las salidas de caudal del almacenamiento ya que al aumentar estas últimas también lo hace  $T_a$ , esto sucede debido a que al aumentar las salidas del embalse disminuye el volumen almacenado y se necesita menos energía para aumentar la temperatura de la masa de agua; además de la profundidad de medición, de manera que a mayor profundidad menor

será Ta pues la energía por conducción incide directamente sobre las capas adyacentes a la superficie y la energía absorbida por radiación disminuye al aumentar la profundidad.

**Matriz de correlaciones**

	DJ	V	DV	Vx	T	Tn	Tx	Hf	HR	Rad	P	Eto	VolExp	Qent
DJ	1,000	-0,062	0,017	-0,014	0,362	0,389	0,336	-0,421	0,036	-0,141	0,145	0,022	-0,670	0,014
V	-0,062	1,000	0,126	0,689	-0,158	-0,179	-0,153	0,208	-0,646	0,068	-0,188	0,122	0,032	0,010
DV	0,017	0,126	1,000	0,006	-0,486	-0,494	-0,454	0,199	-0,202	-0,428	0,108	-0,485	-0,274	0,319
Vx	-0,014	0,689	0,006	1,000	-0,072	-0,101	-0,068	0,065	-0,463	-0,047	0,045	0,069	0,038	0,073
T	0,362	-0,158	-0,486	-0,072	1,000	0,986	0,987	-0,576	0,041	0,668	-0,132	0,825	-0,085	-0,612
Tn	0,389	-0,179	-0,494	-0,101	0,986	1,000	0,956	-0,585	0,110	0,610	-0,090	0,777	-0,118	-0,606
Tx	0,336	-0,153	-0,454	-0,068	0,987	0,956	1,000	-0,562	0,002	0,684	-0,159	0,830	-0,085	-0,606
Hf	-0,421	0,208	0,199	0,065	-0,576	-0,585	-0,562	1,000	-0,194	-0,238	-0,059	-0,349	0,167	0,196
HR	0,036	-0,646	-0,202	-0,463	0,041	0,110	0,002	-0,194	1,000	-0,191	0,245	-0,242	-0,050	0,064
Rad	-0,141	0,068	-0,428	-0,047	0,668	0,610	0,684	-0,238	-0,191	1,000	-0,331	0,930	0,398	-0,567
P	0,145	-0,188	0,108	0,045	-0,132	-0,090	-0,159	-0,059	0,245	-0,331	1,000	-0,248	-0,149	0,087
ETo	0,022	0,122	-0,485	0,069	0,825	0,777	0,830	-0,349	-0,242	0,930	-0,248	1,000	0,277	-0,637
VolExp	-0,670	0,032	-0,274	0,038	-0,085	-0,118	-0,085	0,167	-0,050	0,398	-0,149	0,277	1,000	-0,099
Qent	0,014	0,010	0,319	0,073	-0,612	-0,606	-0,606	0,196	0,064	-0,567	0,087	-0,637	-0,099	1,000
Qsal	0,622	-0,094	-0,270	-0,101	0,755	0,773	0,737	-0,415	0,035	0,303	-0,023	0,478	-0,476	-0,485
Filt	-0,015	0,057	0,196	0,042	-0,251	-0,263	-0,237	0,117	-0,126	-0,216	-0,035	-0,240	-0,060	0,255
Profund.	-0,024	0,004	-0,010	0,006	-0,002	-0,002	-0,002	0,010	0,000	0,015	-0,008	0,015	0,036	0,002
Ta	0,319	-0,099	-0,287	-0,076	0,633	0,636	0,617	-0,336	0,040	0,395	-0,059	0,499	-0,161	-0,416
Odis	-0,449	0,072	0,197	0,070	-0,553	-0,562	-0,536	0,307	-0,053	-0,248	0,020	-0,372	0,342	0,293
pH	-0,249	-0,100	-0,033	-0,016	-0,099	-0,115	-0,086	-0,179	0,111	0,064	0,051	-0,016	0,323	0,025
Con	0,309	0,000	0,230	0,000	-0,188	-0,167	-0,193	-0,048	0,044	-0,336	0,142	-0,315	-0,383	0,264

Determinante = 5,13E-011

**TABLA 89** Parte de la matriz de correlaciones, periodo completo de estudio, variables sin transformar

Odis presenta altas correlaciones con T, Tn, Tx, Qsal y Profundidad; por lo que la concentración de oxígeno disuelto obedece mayormente a: la temperatura del aire y sus picos, de manera que al aumentar la temperatura del aire disminuye el Odis posiblemente porque al aumentar la temperatura del aire esto origina el aumento de la temperatura del agua y al subir esta, genera que la concentración de oxígeno disuelto en el agua sea menor; además con las salidas de caudal del embalse ya que al incrementarse las salidas disminuye el Odis, esto puede ocurrir porque el agua que se extrae es la que posee la mayor concentración de oxígeno lo que representa una pérdida para la masa de agua; asimismo de la profundidad, ya que al aumentar la profundidad disminuye el Odis pues el agua en la superficie está sujeta a la reaireación y cuando ocurre la estratificación térmica es difícil que este aporte de oxígeno sea compartido eficientemente con las capas del fondo, además las capas más profundas están sujetas a la demanda de oxígeno por parte de los sedimentos. Es notorio que tanto Ta como Odis prácticamente dependen de las mismas variables.

---

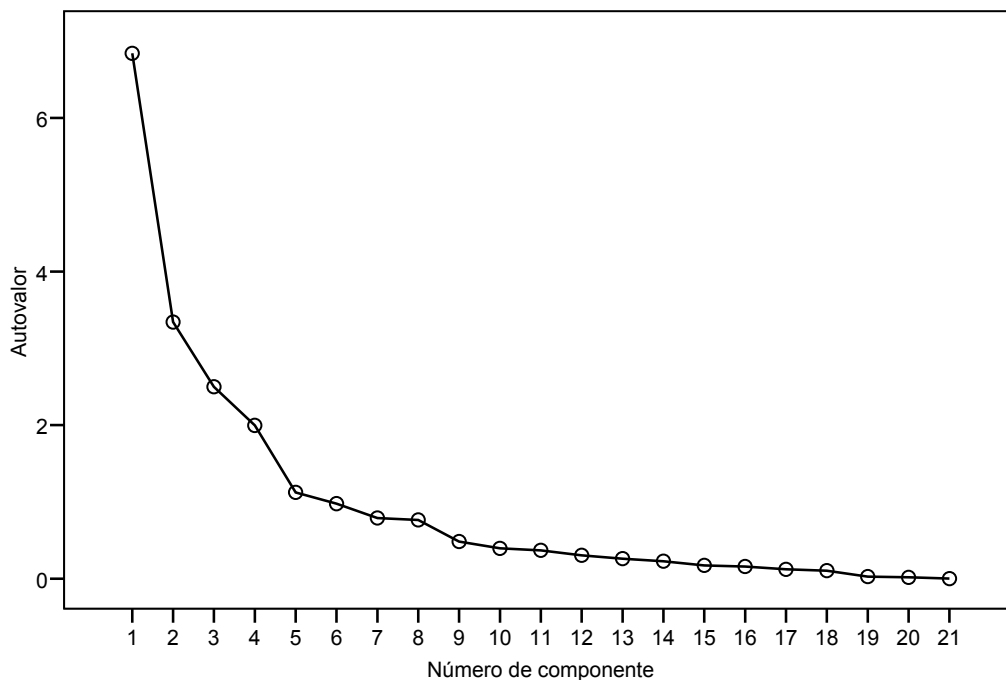
---

pH tiene correlaciones medianamente significativas con: VolExp, ya que al aumentar el volumen también lo hace el pH, esto posiblemente debido a que al ingresar agua al embalse y aumentar su volumen se hacen menos propicias las condiciones para el aumento de la acidez; así también con la profundidad ya que al aumentar la profundidad disminuye el pH, esto ocurre probablemente porque al ir descendiendo las condiciones generales para el desarrollo de la acidez son más favorables.

Y Con sólo posee alta correlación con la profundidad ya que al aumentar la profundidad también lo hace la conductividad, esto se debe a que al aumentar la conductividad del agua esta se hace más densa y tiende a ubicarse en el fondo, en cambio el agua con menor conductividad es menos densa y por flotación tiende a estar cercana a la superficie.

La siguiente parte del análisis es la matriz anti-imagen integrada por los coeficientes de correlación parcial entre cada par de variables, con signo cambiado. Los elementos de la diagonal principal son similares al estadístico KMO e interesa que estén próximos a la unidad. Para este caso la mayoría de los elementos cumplen esta condición.

**Gráfico de sedimentación**



**FIGURA 80** Gráfico de sedimentación para el periodo completo de estudio

En la tabla de la varianza total explicada (Tabla 90) se observa que existen cinco

factores incluidos en el modelo con autovalores mayores a la unidad. La primera componente explica un 32,6% de la varianza total, la segunda un 15,9%, la tercera un 11,9%, la cuarta un 9,5% y la quinta un 5,4%. El gráfico de sedimentación (Figura 80) muestra estas componentes.

#### Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	6,841	32,576	32,576	6,841	32,576	32,576
2	3,342	15,915	48,491	3,342	15,915	48,491
3	2,500	11,903	60,394	2,500	11,903	60,394
4	1,996	9,507	69,901	1,996	9,507	69,901
5	1,125	5,357	75,258	1,125	5,357	75,258

**TABLA 90** Varianza total explicada por los factores incluidos en el modelo, periodo completo de estudio

También se tiene la tabla de comunalidades. La comunalidad es la fracción de variabilidad en cada variable explicada por los factores. Antes de la extracción de los factores su valor para cada variable es la unidad e interesa que una vez efectuada la extracción continúe siendo alta. Las únicas variables que no cumplen con esto son DV y Filt, por lo que el modelo factorial no las reproduce adecuadamente.

#### Comunalidades

	Inicial	Extracción
DJ	1,000	,794
V	1,000	,840
DV	1,000	,457
Vx	1,000	,819
T	1,000	,962
Tn	1,000	,950
Tx	1,000	,936
Hf	1,000	,544
HR	1,000	,748
Rad	1,000	,844
P	1,000	,585
ETo	1,000	,936
VolExp	1,000	,831
Qent	1,000	,526
Qsal	1,000	,831
Filt	1,000	,252
Profundidad	1,000	,814
Ta	1,000	,851
Odis	1,000	,898
Ph	1,000	,647
Con	1,000	,737

**TABLA 91** Comunalidades de las variables para el periodo completo de estudio

En la matriz de componentes rotados se muestra que las variables DV, T, Tn, Tx, Hf, Rad, ETo y Qent, se sitúan en la primera componente que puede ser denominada como “energía térmica externa al embalse y entradas de la misma debido a cambio en el volumen”. En la segunda componente se pueden ubicar las variables DJ, VolExp y Qsal (“variación temporal del volumen almacenado y caudal de salida programado”). En la tercera componente se encuentran Profundidad, Ta, Odis, pH y Con (“variación de la calidad del agua con la profundidad”). En la cuarta componente están V, Vx y HR (“velocidad y humedad del aire”). Y la quinta componente incluye a las variables P y Filt (“precipitación y filtración”).

Por otra parte, al comparar la extracción de factores por el Método de componentes principales y por el de Máxima verosimilitud, los pesos en los factores cambian poco, sólo en algunos casos significativamente pero la configuración de las componentes es prácticamente idéntica, solamente cambia la variable Filt de estar incluida en la quinta componente a ubicarse en la primera y Qsal pasa del factor 2 hacia el uno. Debido a esto se puede considerar que la solución del FA es confiable.

**Matriz de factores rotados**

	Factor				
	1	2	3	4	5
DJ		,752			
V				,866	
DV	-,542				
Vx				,818	
T	,971				
Tn	,953				
Tx	,957				
Hf	-,561				
HR				-,680	-,314
Rad	,690	-,263			,634
P					-,268
ETo	,839				,479
VolExp		-,869			
Qent	-,603				-,264
Qsal	,632	,618			
Filt	-,258				
Profundidad			-,849		
Ta	,566	,363	,582		
Odis	-,473	-,346	,739		
pH		-,336	,565		
Con	-,255	,271	-,663		

Método de extracción: Máxima verosimilitud.  
Método de rotación: Normalización Varimax con Kaiser.

**TABLA 92** Matriz de componentes de los 5 factores extraídos por Máxima verosimilitud, periodo completo de estudio, variables sin transformar y Rotación Varimax

---

---

#### 9.4.2 Periodo de calibración, variables sin transformar

En el Anexo 13 se puede observar que se extrajeron 5 componentes con autovalores mayores que la unidad y en conjunto explican un 76,3% de la varianza total en este periodo. Al comparar con el tiempo completo de estudio, también se extrajeron 5 componentes que cumplen con el criterio del autovalor mayor a la unidad y explican un 75,3% de variabilidad del conjunto de datos. Por lo tanto el FA es confiable, pues coinciden prácticamente los porcentajes de varianza acumulada en ambos casos analizados.

#### Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	6,932	33,010	33,010	6,932	33,010	33,010
2	3,180	15,144	48,154	3,180	15,144	48,154
3	2,730	12,999	61,153	2,730	12,999	61,153
4	2,101	10,006	71,160	2,101	10,006	71,160
5	1,080	5,144	76,303	1,080	5,144	76,303

**TABLA 93** Varianza total explicada por los factores incluidos en el modelo, periodo de calibración

Además, como se menciona en la sección 7.6.2, la formación de los factores es muy similar lo que indica que el comportamiento del sistema y las relaciones entre variables dentro de él son estables a través del tiempo.

Revisando la matriz de correlaciones se tiene que las variables de calidad del agua muestran las siguientes correlaciones relevantes: Ta con las variables T, Tn, Tx, ETo y Qsa; Odis con T, Tn, Tx y Profundidad; asimismo, pH con Profundidad; asimismo Con respecto también a Profundidad.

**Matriz de componentes rotados**

	Componente				
	1	2	3	4	5
DJ		,869			
V				,880	
DV	-,675				
Vx				,874	
T	,931	,291			
Tn	,911	,326			
Tx	,919	,268			
Hf	-,484	-,498			-,474
HR				-,792	,408
Rad	,823	-,256			-,336
P					,451
ETo	,915				-,259
VolExp	,440	-,811			
Qent	-,477				,579
Qsal	,531	,718			
Filt	-,370				
Profundidad			-,841		
Ta	,663	,374	,523		
Odis	-,494	-,267	,776		
pH			,652		,482
Con			-,870		

Método de extracción: Análisis de componentes principales.  
Método de rotación: Normalización Varimax con Kaiser.

**TABLA 94** Matriz de componentes rotados con cinco factores, periodo de calibración, variables sin transformar

9.4.3 Periodo de validación, variables sin transformar

El Anexo 14 muestra que las primeras 5 componentes con autovalores superiores a la unidad explican un 78,4% de la variabilidad total de los datos en este lapso de tiempo. Si se compara este valor con el correspondiente al periodo de calibración se tiene que son muy similares, constituyendo esto una primera indicación de que los dos periodos se pueden considerar, aproximadamente, como homogéneos entre sí.

**Varianza total explicada**

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	7,729	36,807	36,807	7,729	36,807	36,807
2	3,135	14,928	51,735	3,135	14,928	51,735
3	2,578	12,277	64,012	2,578	12,277	64,012
4	1,720	8,188	72,200	1,720	8,188	72,200
5	1,295	6,165	78,366	1,295	6,165	78,366

**TABLA 95** Varianza total explicada por los factores incluidos en el modelo, periodo de validación



Las diferencias observadas en la formación de los componentes entre los periodos de calibración y validación son las siguientes: DJ se ubica en el segundo componente en la calibración y en el primero en la validación; DV está en el primer componente durante la calibración y en el quinto en la validación; en la calibración Hf se ubica en el segundo componente y en la validación en el en el primero; Rad está incluido en el primer componente en la calibración y en el segundo componente en la validación; P está en la quinta componente en la calibración y en la segunda en la validación; ETo pertenece a la primera componente durante la calibración y a la segunda en la validación; VolExp se incluye en la segunda componente en la calibración y en primera para la validación; Qent está incluida en la quinta componente en la calibración y en la primera durante la validación; Qsal pertenece a la segunda componente en la calibración y a la primera en la validación; Filt está ubicada dentro de la primera componente en la calibración y en la quinta durante la validación; Ta se incluye en la primera componente en la calibración y en la tercera para la validación; Odis está en la tercera componente durante la calibración y en la primera en la validación.

**Matriz de componentes rotados – 5 Factores**

	Componente				
	1	2	3	4	5
DJ	,972				
V				,931	
DV					-,635
Vx				,894	
T	,863	,389			
Tn	,872	,297			,278
Tx	,841	,425			
Hf	-,406	-,363		,270	-,320
HR		-,461		-,653	,276
Rad		,894			
P		-,558			,399
ETo	,466	,814			
VolExp	-,886	,300			
Qent	-,690	-,339			-,427
Qsal	,897				
Filt					-,596
Profundidad			-,913		
Ta	,584		,759		
Odis	-,671		,640		
pH	-,516		,589		
Con	,267		-,804		

Método de extracción: Análisis de componentes principales.  
Método de rotación: Normalización Varimax con Kaiser.

**TABLA 96** Matriz de componentes rotados considerando 5 factores, periodo de validación, variables sin transformar

---

---

Estas modificaciones en las componentes, que se producen en varios casos debido a pequeñas variaciones en los valores de los pesos, indican que a pesar de que los dos periodos poseen límites para las variables muy similares, internamente tienen algunas diferencias entre ellos. Esto garantiza que habrá variabilidad entre los periodos haciendo más confiables los resultados obtenidos pues no estaremos reproduciendo un periodo a través de otro exactamente igual.

Analizando la matriz de correlaciones las variables de calidad del agua presentan las siguientes correlaciones de importancia: Ta con DJ, T, Tn, Tx, Qsal y Profundidad; Odis respecto a DJ, T, Tn y Tx, VolExp, Qsal y Profundidad; pH tiene con DJ, T, Tn, Tx, VolExp, Qsal y Profundidad; Con respecto sólo a Profundidad. Al comparar con el periodo de calibración se tiene que la principal diferencia es que en el periodo de validación se incluyeron más días de verano por lo que las temperaturas son más altas (en promedio) y la explotación del almacenamiento cambia, lo cual influye notablemente sobre las variables Ta, Odis y pH; menos se observa esta influencia sobre Con.

#### 9.4.4 Normalidad de las puntuaciones factoriales

Al revisar la Tabla (28) de las pruebas de normalidad de los cinco factores para el periodo completo de estudio se tiene que al considerar los cinco factores como representativos del conjunto de datos en lugar de las variables originales, no se resuelve el problema de la falta de normalidad.

Este resultado va enfocando la atención hacia la selección de técnicas de inferencia que no se vean afectadas por la falta de adecuación de las variables a una distribución normal.

#### 9.4.5 Variables transformadas

Primero se compara la varianza total explicada de las variables en su estado ordinario y luego transformadas. De la Tabla (19) y la Tabla (29) se tiene que la varianza explicada en el primer caso es de 75,3% y después de efectuar las transformaciones y aplicar el FA la varianza es 75,0%. La varianza total disminuye al efectuar las transformaciones.

**Varianza total explicada**

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	6,853	32,635	32,635	6,853	32,635	32,635
2	3,250	15,476	48,111	3,250	15,476	48,111
3	2,541	12,100	60,211	2,541	12,100	60,211
4	1,995	9,500	69,711	1,995	9,500	69,711
5	1,115	5,311	75,022	1,115	5,311	75,022

**TABLA 97** Varianza total explicada considerando cinco factores, periodo completo de estudio, variables transformadas

La formación de las componentes para los dos casos (Tablas 20 y 30) es prácticamente la misma (a excepción de la variable Filt). Esto se puede verificar en las Tablas (98) y (99) que se muestran a continuación, donde se puede apreciar que las saturaciones principales de las variables ocurren en los mismos factores para las variables sin transformar y luego transformadas.

**Matriz de componentes**

	Componente				
	1	2	3	4	5
DJ		-,667			
V			,756		
DV	-,526				
Vx			,594		
T	,980				
Tn	,970				
Tx	,967				
Hf	-,567				
HR			-,729		
Rad	,714				
P					,615
Eto	,850				
VolExp		,743			
Qent	-,679				
Qsal	,787				
Filt					-,349
Profundidad				-,630	
Ta	,670				
Odis		,639			
pH		,585			
Con		-,763			

**TABLA 98** Formación de las componentes por el Método de componentes principales, periodo de estudio completo, variables sin transformar

Se extrajeron puntuaciones factoriales para cinco factores de las variables transformadas. Al efectuar las pruebas de normalidad (Tabla 40) se tiene que ninguna de estas series de puntuaciones se ajusta a una distribución normal. Por lo que se puede concluir que en este caso la transformación aplicada no resuelve la falta de normalidad.

**Matriz de componentes**

	Componente				
	1	2	3	4	5
DJ_T		-,611			
V_T			,762		
DV_T	-,499				
Vx_T			,599		
T_T	,972				
Tn_T	,956				
Tx_T	,955				
Hf_T	-,612				
HR_T			-,739		
Rad_T	,682				
P_T					,554
Eto_T	,842				
VolExp_T		,723			
Qent_T	-,805				
Qsal_T	,762				
Filt_T	-,359				
Profundidad_T				-,638	
Ta_T	,671				
Odis_T		,647			
pH_T		,591			
Con_T		-,733			

**TABLA 99** Formación de las componentes por el Método de componentes principales, periodo de estudio completo, datos transformados

## 9.5 RESULTADOS DEL ANÁLISIS CLUSTER

Al considerar las variables sin transformar y el periodo entero de estudio (Tabla 42) se toman 15 conglomerados para dividir los registros (al considerar el criterio de que el conglomerado más pequeño posea alrededor de 30 registros) y se observa que la conformación de estos no es uniforme: el conglomerado número ocho sólo posee 27 elementos, mientras que el quinto conglomerado tiene 2260.

El motivo por el cual se tomó el criterio de alrededor de 30 registros por conglomerado es el siguiente: Mendenhall et al. (1986) mencionan que un valor de  $n$  mayor o igual a 30 garantizará, generalmente, que su distribución esta bien definida.

---

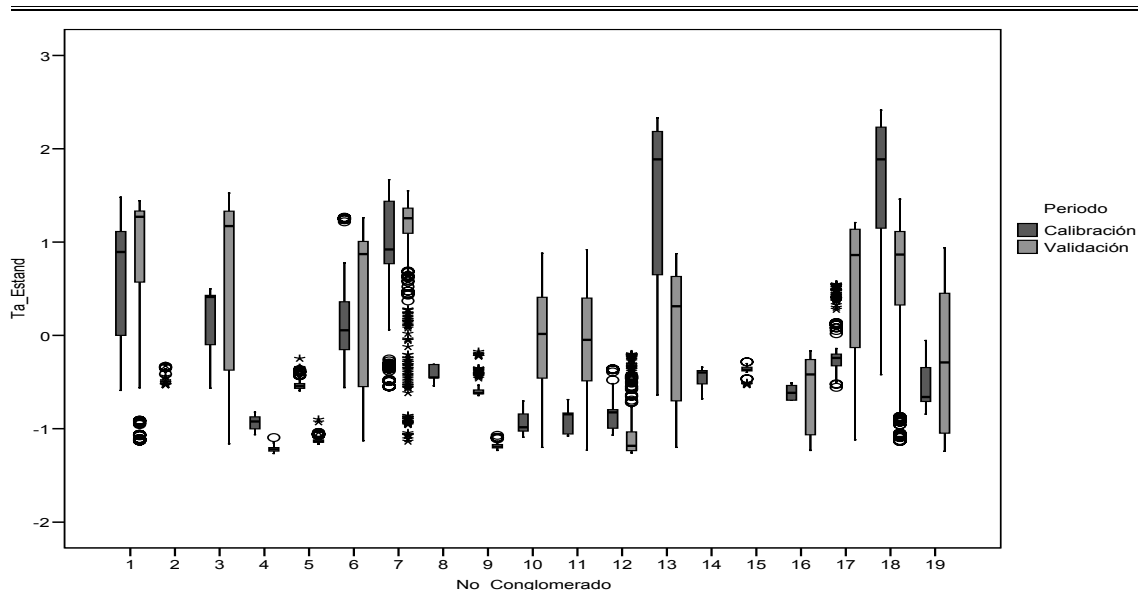
---

En el periodo de calibración se formaron 19 conglomerados (tomando en cuenta el criterio anterior, Tabla 43). Al utilizar los centros de estos 19 grupos obtenidos en el periodo de calibración para ordenar los registros del periodo de validación (Tabla 44), sólo se clasificaron en 15 de los 19 agrupamientos, debido a que los registros se clasifican en el conglomerado cuyo centro está más cercano a ellos, y en este caso 4 conglomerados que se formaron para el periodo de calibración (de un total de 19) no agruparon a ningún registro del periodo de validación, pues esos registros se congregaron en otros conglomerados que desde el punto de vista espacial les quedaban más cercanos.

Cuando se emplearon las variables transformadas fueron seleccionados 15 conglomerados para el periodo de calibración. Al utilizar los centros de esos conglomerados para agrupar las observaciones en el periodo de validación sólo se utilizaron 12 de los 15 conglomerados, debido a que los registros del periodo de validación se encontraron más cercanos a 12 de los 15 agrupamientos y más alejados de los restantes tres, por lo que sólo se tuvieron registros en 12 conglomerados para el periodo de validación. Por lo tanto, la transformación de las variables origina que el número de conglomerados disminuya.

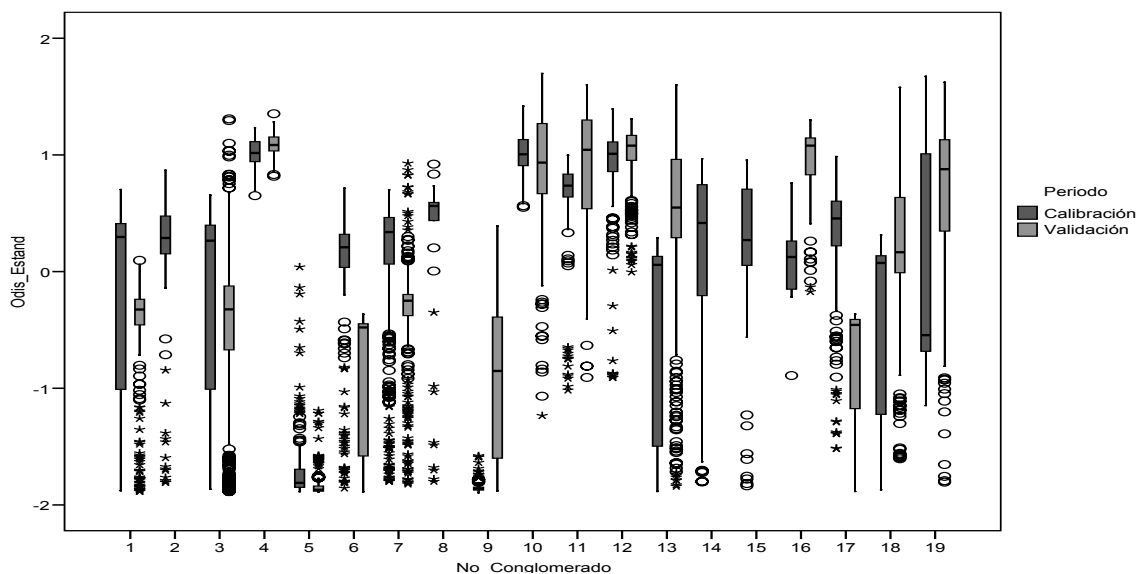
Al aplicar el análisis cluster a las puntuaciones en componentes principales de las variables estandarizadas se generaron 18 conglomerados en la calibración. Al usar los centros obtenidos de estos conglomerados para formar los grupos en la validación se tiene que las observaciones se clasifican en 17 de los 18 conglomerados. Uno de los conglomerados formado en el periodo de calibración no agrupó a ninguna de las observaciones de la validación debido a que desde el punto de vista espacial está más alejado de los registros de la validación que el resto de conglomerados.

Cuando en la calibración se consideraron las puntuaciones factoriales de las variables ordinarias (se extrajeron tres factores) se generan 14 conglomerados. Usando los centros de estos, en el periodo de validación las observaciones se clasifican exactamente en los 14 conglomerados. Todos los agrupamientos formados para el periodo de calibración agruparon a los registros del periodo de validación. De esto se puede deducir que las puntuaciones factoriales son un poco más homogéneas al distinguir entre periodos de tiempo.



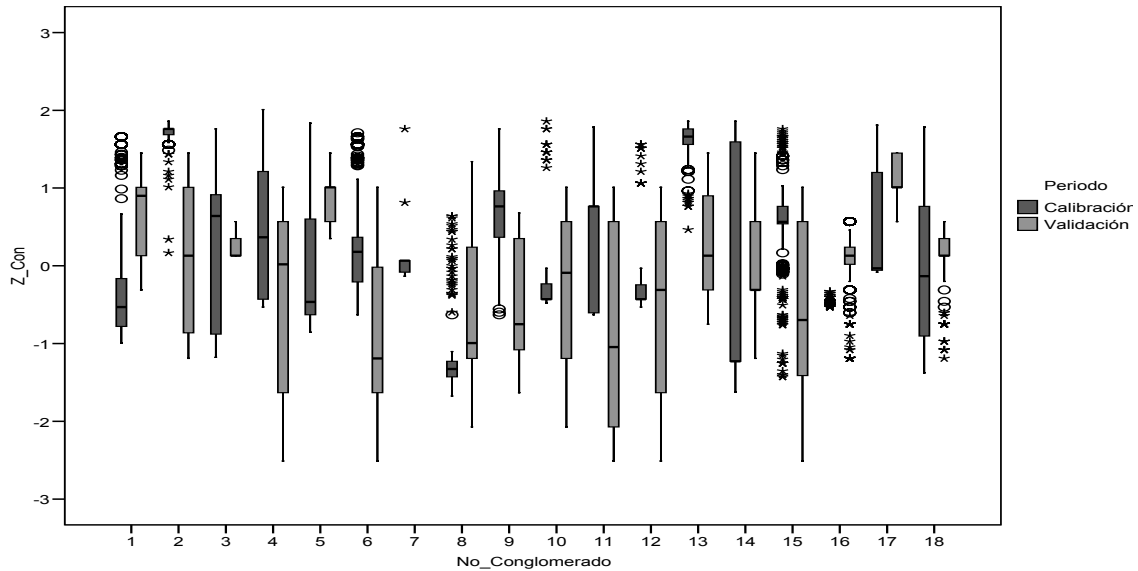
**FIGURA 81** Diagramas de caja y bigotes para comparación de temperatura del agua estandarizada en distintos conglomerados formados con las variables estandarizadas y sin transformar, para calibración y validación

Con el objetivo de conocer si el análisis cluster ha logrado formar conglomerados que sean aproximadamente uniformes en los valores de las variables de calidad del agua se elaboró la Figura (81) para la temperatura del agua, Ta. En ésta figura se puede apreciar que al comparar para un mismo conglomerado los valores correspondientes a la calibración y la validación, en la gran mayoría de los casos los límites superior e inferior, así como las medianas presentan diferencias muy amplias. El mismo fenómeno se presenta para las otras tres variables de calidad del agua: Odis (Figura 82), pH y Con (Figura 83), Anexo 21.



**FIGURA 82** Diagramas de caja y bigotes para comparación de oxígeno disuelto estandarizado en distintos conglomerados formados con las variables estandarizadas y sin transformar, para calibración y validación

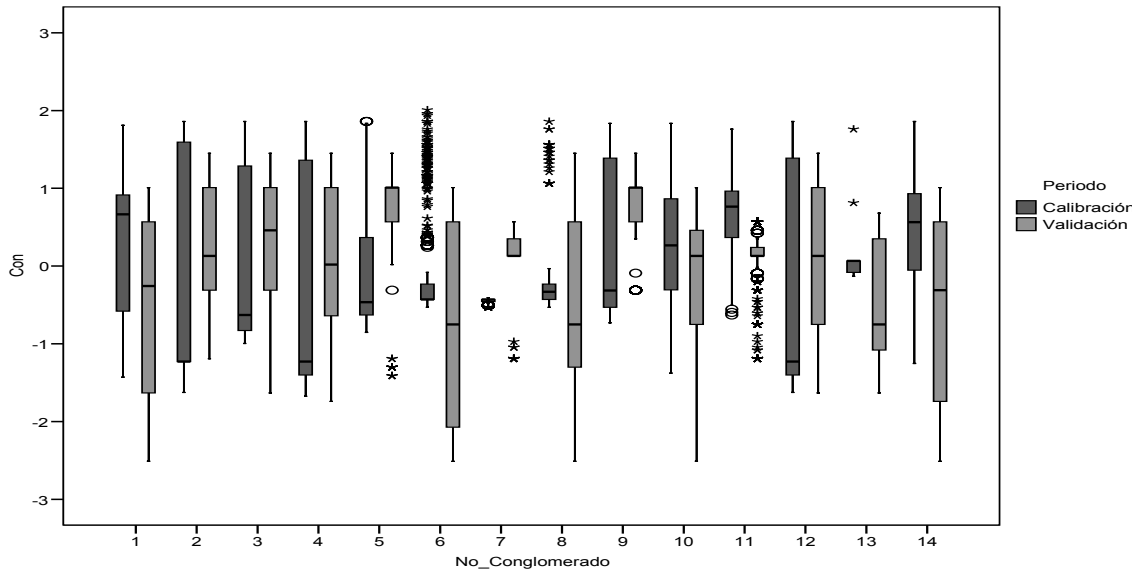
El mismo tipo de gráficos de caja y bigotes se elaboraron para las cuatro variables de calidad del agua considerando variables transformadas (Anexo 22), puntuaciones en componentes principales (Figura 83 y Anexo 23) y puntuaciones factoriales (Figura 84 y Anexo 24).



**FIGURA 83** Diagramas de caja y bigotes para comparación de conductividad estandarizada en grupos formados con puntuaciones en componentes principales, calibración y validación

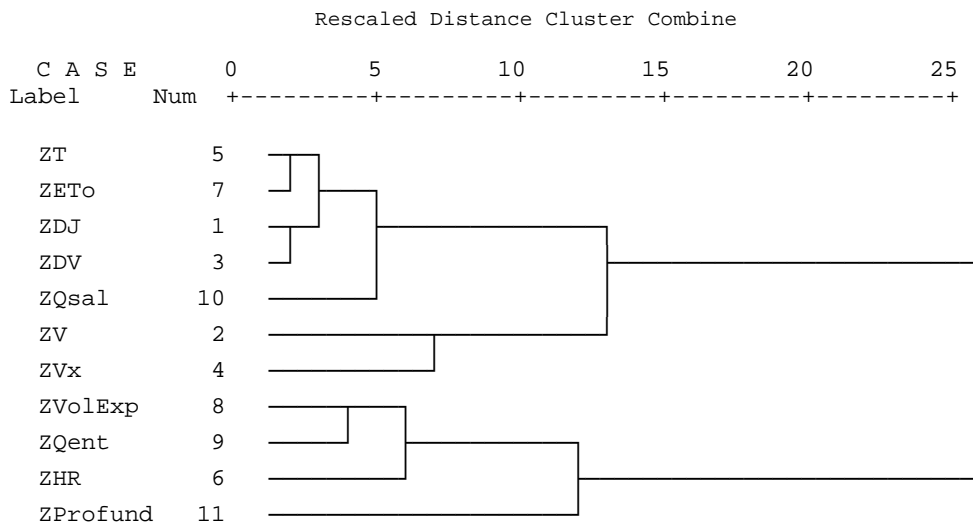
Se observó un comportamiento muy similar al anteriormente descrito para la gran mayoría de los conglomerados, pues los límites muestran grandes diferencias al igual que las medianas, al analizar un mismo conglomerado dividido en sus registros correspondientes a calibración y validación. Esto es inconveniente, como se verá más adelante, pues al ajustar una expresión (mediante regresión) dentro de un conglomerado para el periodo de calibración esta ecuación sólo ajusta bien para ese periodo pero no predice bien los valores que se presentarán en la validación mediante conglomerados

En la sección 7.7.10 se efectuó un análisis cluster jerárquico de variables ordinarias para cada uno de los conglomerados (19 en total), estandarizadas de forma univariante para cada uno de los periodos de tiempo considerados (calibración y validación) por separado y a continuación uniéndolos para efectuar el análisis para cada agrupamiento, usando el método de vinculación inter-grupos para el tiempo total de estudio (Anexo 25).



**FIGURA 84** Diagramas de caja y bigotes para comparación de conductividad estandarizada en distintos conglomerados formados con las puntuaciones factoriales de variables predictivas, para calibración y validación

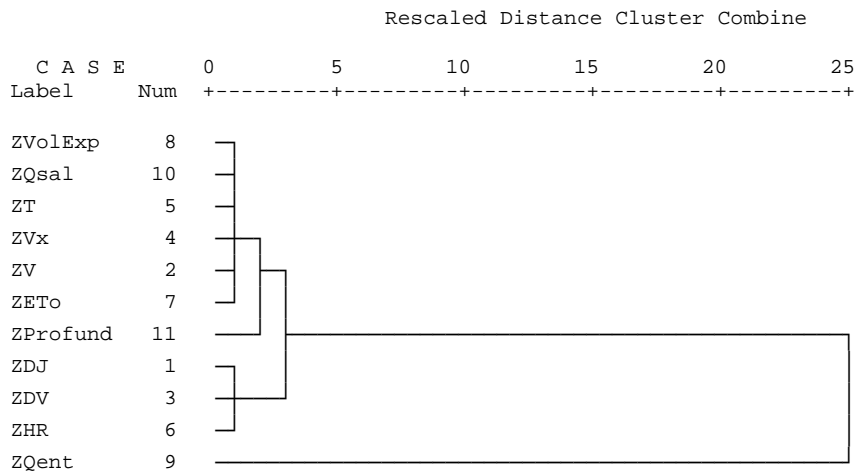
Por ejemplo, en el primer conglomerado (Figura 85) se unieron primeramente T y Eto, luego DJ y Qsal, a continuación V y Vx. A un nivel superior se unen T, ETo, DJ y Qsal. Después DV y Qent. Más adelante se agrupan DV, Qent, V, Vx, VolExp y Profundidad. Luego T, ETo, DJ, Qsal y HR. En la agrupación final se incluyen a las 11 variables.



**FIGURA 85** Dendrograma del primer conglomerado usando el método del promedio entre grupos para conglomerados formados con las variables estandarizadas y sin transformar, periodo completo de estudio

Para cada uno de los conglomerados (Figura 86) la forma en que las variables se asociaron fue distinta, indicando que las relaciones entre las variables dentro de los conglomerados son diferentes para cada uno de ellos.

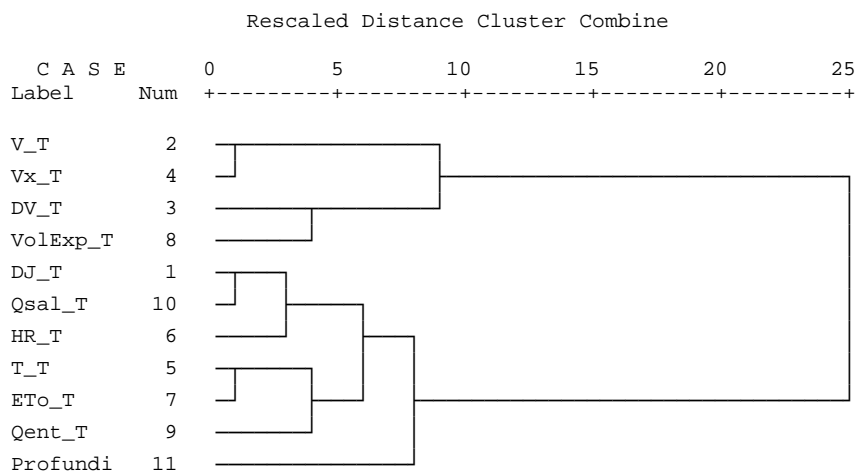




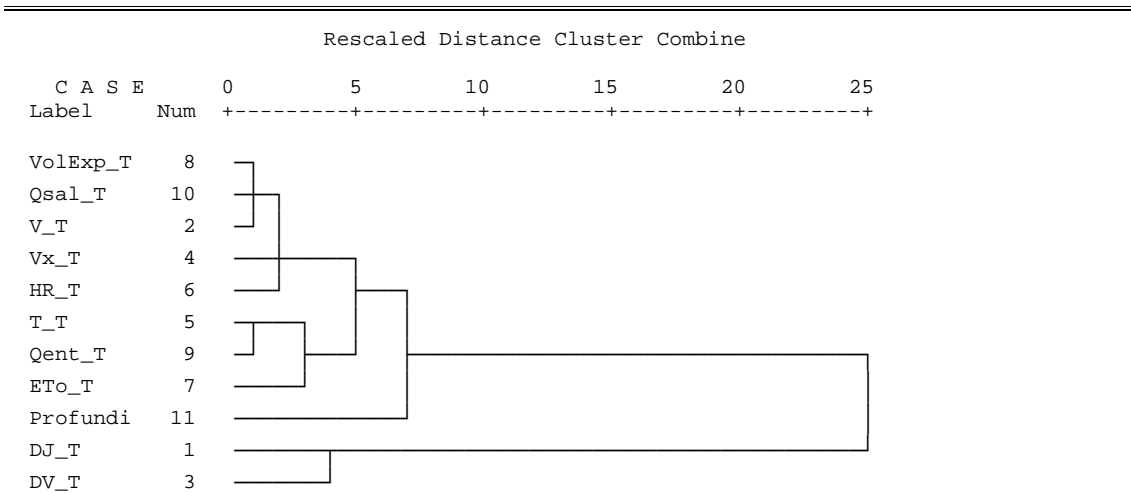
**FIGURA 86** Dendrograma del segundo conglomerado usando el método del promedio entre grupos para conglomerados formados con las variables estandarizadas y sin transformar, periodo completo de estudio

Un análisis cluster jerárquico de variables transformadas fue realizado en la sección 7.7.11.

Como medida de disimilaridad se usó la distancia euclídea al cuadrado y para formar los conglomerados el método de vinculación inter-grupos (Figuras 87 y 88), para el periodo completo de estudio (Anexo 26).



**FIGURA 87** Dendrograma del cuarto conglomerado usando el método del promedio entre grupos para conglomerados formados con las variables transformadas, periodo completo de estudio



**FIGURA 88** Dendrograma del séptimo conglomerado usando el método del promedio entre grupos para conglomerados formados con las variables transformadas, periodo completo de estudio

Se observa el mismo efecto que en el caso anterior: en cada uno de los conglomerados las variables se van asociando de distinta manera. Esto indica que las relaciones entre las variables para cada uno de los conglomerados son distintas.

La conclusión que se obtiene después de aplicar el análisis cluster es que al formar los conglomerados al interior de cada uno se tiene cierta homogeneidad global considerando todas las variables; pero al analizar el comportamiento de una variable en específico para los grupos formados se observa que los valores que toma son, generalmente, distintos en los periodos de calibración y validación. Por lo tanto, la formación de conglomerados propuesta en este caso no será de gran ayuda con el propósito de efectuar predicciones.

## 9.6 ANÁLISIS DISCRIMINANTE

El análisis discriminante se utilizó sobre los conglomerados formados mediante el análisis cluster con el propósito de verificar la pertenencia de las observaciones a los conglomerados.

Se inicio con el periodo de estudio completo utilizando todas las 11 variables consideradas. Estas se emplearon sin estandarizar. Las probabilidades previas de clasificación correcta se obtuvieron de acuerdo al tamaño de los grupos. En los 19 grupos la cantidad de clasificaciones correctas fue de un 79,3% (Anexo 27), resultado que se puede considerar como bueno en comparación a las probabilidades previas de clasificación correcta (Tabla 51).

Resultados de la clasificación																				
No Grupo	Grupo de pertenencia pronosticado																		Total	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		19
% 1	80,8				1,7	8,9	8,6												100	
2		100																	100	
3			87		0,6				0,8				5					6,4	100	
4				68						25	1,9	5,2							100	
5	0,2		3,9		63,7	2,0	2,7		9,1					1,1			15,2	2,1	100	
6	19,1		0,0		1,9	72,2											6,8		100	
7	1,9		4,9		1,6	1,9	77,5										8,2	3,9	100	
8								100											100	
9					0,8				80	1,4	9,8		2,7					4,9	0,4	100
10				3,2					1	76	2,9	14						3,3		100
11									0,3	5,7	59	25						11		100
12										4,2	11	84								100
13									2				77					18	2,6	100
14														100						100
15															100					100
16																90				100
17					2,3		19							3,3			76		10	100
18										0,3								100		100
19									1,7	19		0,6						79		100

Clasificados correctamente el 79,3% de los casos agrupados originales.

**TABLA 100** Resultados de la clasificación, periodo completo de estudio, variables sin transformar y no estandarizadas

A continuación se aplicó nuevamente el DA al periodo de estudio completo con las 11 variables, pero esta vez estandarizadas con la media y la desviación estándar del periodo de calibración exclusivamente. Los registros clasificados correctamente constituyeron un 79,6% del total de registros (Anexo 28 a)), valor que se puede interpretar como bueno y para este caso es un valor muy próximo al conseguido al aplicar el DA con las variables sin estandarizar.

Resultados de la clasificación																				
Número de conglomerado	Grupo de pertenencia pronosticado																		Total	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		19
% 1	80,6				2,3	8,6	8,4												100	
2		100																	100	
3			87,2		0,6				0,8				1,7					9,8	100	
4				73,5						20,1	1,2	5,2							100	
5			3,8		64,0	2,0	2,9		8,8					0,9			15,4	2,3	100	
6	19,6				1,9	71,6											6,8		100	
7	0,4		4,2		1,6	1,9	83,0										4,9	3,9	100	
8								100											100	
9					0,8				80,4	1,0	9,8		2,7					4,9	0,4	100
10				3,0					1,0	77,2	3,3	12,2						3,3		100
11									0,5	5,2	55,8	24,8						11,8	1,9	100
12										4,1	11,3	83,0							1,6	100
13									2,0				77,4					17,9	2,6	100
14														100						100
15															100					100
16																89,7				100
17					2,8		20,9							3,3			72,9		10,3	100
18										0,5								99,5		100
19									1,2	17,8		0,6						80,5		100

Clasificados correctamente el 79,6% de los casos agrupados originales.

**TABLA 101** Resultados de la clasificación, periodo completo de estudio, variables sin transformar y estandarizadas con los parámetros del periodo de calibración

En seguida se utilizó de nuevo el DA en el periodo de estudio completo con las 11 variables, pero en esta ocasión se estandarizaron los valores de las variables correspondientes a los periodos de calibración y validación por separado, para luego unirlos y formar un solo bloque y entonces aplicar el DA. Los registros clasificados correctamente constituyeron un 87,8% del total de registros (Anexo 28 b)), valor que se puede interpretar como muy bueno y está significativamente por arriba del conseguido al aplicar el DA con las variables sin estandarizar.

### Resultados de la clasificación

Número de conglomerado	Grupo de pertenencia pronosticado																			Total	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19		
1	80,4				1,9	13,5	4,2													100	
2		100																		100	
3			93,3				0,3		0,3				0,3						5,9	100	
4				95,1						0,9									4,0	100	
5	0,4		2,0		71,9	2,1	3,0		6,1					2,7			9,8	2,0		100	
6	7,4				0,2	92,4														100	
7	3,6		4,0		0,9		91,4													100	
8								100												100	
9									77,1	3,5	11,4		3,1						4,5	0,4	100
10									0,5	91,4	1,6	2,9							3,3	0,3	100
11									0,3	5,5	71,5	12,9							9,4	0,3	100
12										2,7	5,1	92,1									100
13				2,2					2,6					77,2					17,9		100
14															100						100
15																100					100
16				10,3											9,3		80,4				100
17					1,3		2,4								5,5			90,8			100
18							1,3		0,6										98,1		100
19										10,8										89,2	100

Clasificados correctamente el 87,8% de los casos agrupados originales.

**TABLA 102** Resultados de la clasificación, periodo completo de estudio, variables sin transformar y los periodos de calibración y validación estandarizados por separado y luego reunidos para efectuar el análisis

Se puede decir que en este caso y para los conglomerados formados en la etapa de análisis cluster, la mejor alternativa fue estandarizar las variables en la calibración y la validación por separado, basándonos exclusivamente en que los mejores resultados de clasificación para el método DA fueron obtenidas de esta forma.

Cuando se utilizaron variables transformadas el número de conglomerados fue de 15. Los resultados de emplear la clasificación por medio del DA fueron de un 84,0% de registros clasificados correctamente (Tabla 58). Este resultado se puede evaluar como adecuado, pues es superior al porcentaje de clasificación obtenido para las variables ordinarias a las que solamente se les aplicó como tratamiento la estandarización por columna (estandarización univariante). Los resultados completos de este proceso se aprecian en el Anexo 29.

**Resultados de la clasificación (b)**

Número del grupo	Grupo de pertenencia pronosticado															Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1	97,5				2,4					0,1						100
2		90,8	4,0	1,5							3,4		0,3			100
3		17,1	79,4	3,5												100
4		8,0	4,3	87,6												100
5	29,6				42,6			1,7	4,1	8,7			11,1		2,1	100
6						87,7	12,3									100
7							100									100
8					2,2			82,9		15,0						100
9									81,8		17,1		1,1			100
10	1,2				5,9				2,3	83,7	6,0				0,9	100
11									5,4	3,7	90,8		0,1			100
12												92,2			7,8	100
13			0,3	0,6					2,9	1,0	8,5		86,7			100
14														100		100
15					0,5				10,2	15,2	1,5	5,1			67,5	100

b Clasificados correctamente el 84,0% de los casos agrupados originales.

**TABLA 103** Resultados de la clasificación, periodo completo de estudio, variables transformadas

Al utilizar las puntuaciones en componentes principales obtenidas utilizando once componentes (con estandarización) la cantidad de grupos en el análisis fue de 18 y al aplicar la técnica DA clasificó correctamente un 88,1% de los datos. El porcentaje de clasificaciones correctas se considera como muy bueno y se hace notar que prácticamente es idéntico al obtenido en la clasificación de las variables ordinarias estandarizadas. El análisis se muestra en el Anexo 30.

En esta tesis, para muchos de los análisis que se efectuarán con el fin de captar las relaciones entre variables y los procesos de causa-efecto se prevé que resultará más adecuado el desarrollo de las técnicas estadísticas multivariantes empleando las variables sin transformar y estandarizadas, debido a que su comportamiento estadístico (como se ha visto hasta ahora) es el más adecuado para los resultados en este trabajo.

**Resultados de la clasificación (b)**

No inicial de casos	Grupo de pertenencia pronosticado																		Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
1	98,6	1,4																	100
2	2,7	67,1			2,0	7,4		5,0			1,4		7,4		5,0		2,0		100
3			100																100
4				84,7						3,5		7,9				3,9			100
5	4,5				91,3									4				0,2	100
6	7,6			9,2	3,9	70,3		3,1						2,0	3,3		0,6		100
7							100												100
8		0,1						95,1				3,2	1,1	0,5					100
9									96,1						3,9				100
10				5,1						83,1	1,6	2,2				7,0		1,1	100
11		2,2	3,5								94,3								100
12				4,6		1,6		0,2		0,4	0,2	91,1			1,7		0,1		100
13		4,6						4,3						87,4	3,6				100
14								12,1						2,9	84,8			0,2	100
15		0,2				1,9							4,9			93,1			100
16									1,1				8,7				89,8	0,3	100
17	1,1			2,1	0,5			12,3			0,8	2,1	3,8	1,6	0,3		75,3		100
18	3,8	0,8		6,3											4,5			84,8	100

b Clasificados correctamente el 88,1% de los casos agrupados originales.

**TABLA 104** Resultados de la clasificación, periodo completo de estudio, grupos formados con las puntuaciones en componentes principales

En la comprobación de los 14 conglomerados formados con las puntuaciones factoriales (donde se consideraron únicamente los factores 1,2 y 3) mediante el análisis cluster, el DA clasificó adecuadamente un 88,5% de los registros, porcentaje que es muy alto (los resultados se muestran en el Anexo 31). Se tiene en cuenta que al ser mayores los porcentajes de clasificaciones correctas en este caso, se preferiría desarrollar las técnicas multivariantes con puntuaciones factoriales si consideramos como único aspecto importante la formación de grupos; pero se ve al desarrollar las técnicas de regresión que se exponen en este trabajo que no es la mejor opción para lograr estimaciones de los parámetros de calidad del agua analizados.

**Resultados de la clasificación (b)**

Número de conglomerado	Grupo de pertenencia pronosticado														Total				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14					
1	78,2								11,6			10,2							100
2		88,2											11,8						100
3			82,7									17,3							100
4		0,3		89,7										10,0					100
5				1,2				94,3					4,5						100
6								95,5	1,7	2,8									100
7								15,2	84,8										100
8								1,6		98,4									100
9				16,0	6,0							77,0						1,0	100
10				5,8		5,8							81,1					7,4	100
11		4,5												95,1				0,4	100
12					12,9			2,1							85,1				100
13																100			100
14								3,0		2,1			3,9					91,0	100

b Clasificados correctamente el 88,5% de los casos agrupados originales.

**TABLA 105** Resultados de la clasificación, periodo completo de estudio, grupos formados con las puntuaciones factoriales

---

---

## 9.7 MODELO LINEAL DE REGRESIÓN MÚLTIPLE

Existen cuatro categorías para los objetivos de un análisis de regresión (Burnham et al., 2001):

1. Predicción.
2. Especificación del modelo (interpretación del sistema).
3. Procesamiento o cribado de variables.
4. Estimación de parámetros.

El modelo lineal de regresión múltiple (MLR) fue desarrollado dentro de la estructura de este trabajo fundamentalmente como herramienta exploratoria dentro de los procesos de regresión (ajuste de ecuaciones) y no fue visto, desde el inicio, como una técnica para obtener predicciones.

Esta perspectiva de uso del MLR se debe a que las hipótesis que exige este modelo, como son que los términos de error no presenten autocorrelación, las variables **X** deben ser linealmente independientes y la normalidad de los residuos son condiciones que difícilmente se pueden cumplir para grandes conjuntos de observaciones en problemas prácticos, por lo que al tener la seguridad de su no observancia, los resultados obtenidos mediante esta técnica serán vistos como una referencia (no formal) de los resultados que se pueden esperar al utilizar métodos con menos restricciones.

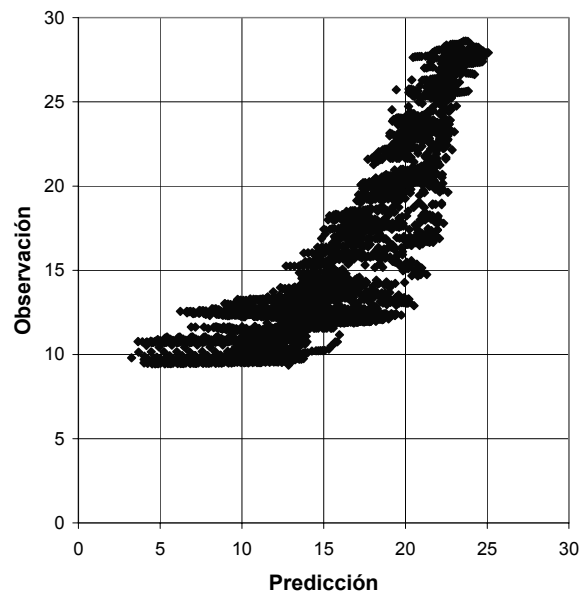
Al aplicar el MLR así como los restantes métodos quimiométricos con el fin de realizar regresiones multivariantes se procedió a dividir los datos existentes para efectuar calibración y validación sobre los datos observados, para garantizar en cierta medida la estabilidad de las relaciones encontradas.

Al aplicar el MLR se utilizaron 17 variables predictivas y 4 variables respuesta (las variables de calidad: temperatura del agua ( $T_a$ , Figura 89), oxígeno disuelto (Odis, Figura 90), pH (Figura 91) y conductividad (Con, Figura 92). El método utilizado para incluir a las variables en el modelo fue el de selección hacia delante. Así, se espera obtener los mejores resultados usando el mínimo de variables posible.

---

---

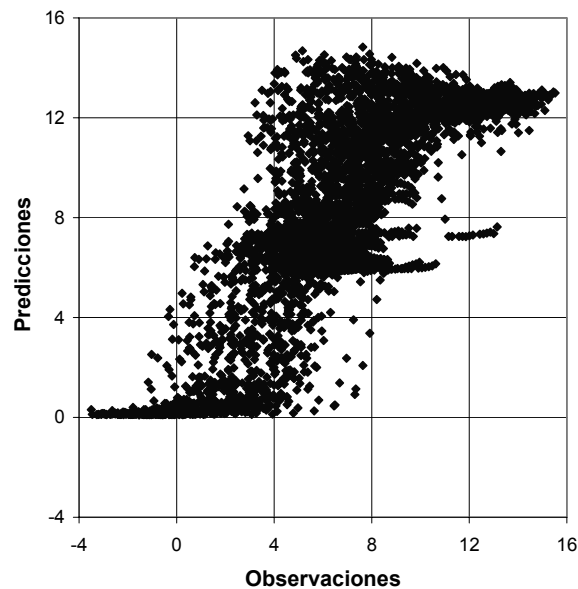
### Temperatura del agua



**FIGURA 89** Gráfico de observaciones y predicciones para la temperatura del agua, periodo de calibración

Los indicadores de la calidad del ajuste efectuado por medio de regresión lineal múltiple usando en todo el proceso las variables en su escala ordinaria de medición se muestran en la Tabla (63).

### Oxígeno disuelto



**FIGURA 90** Gráfico de observaciones y predicciones para oxígeno disuelto, periodo de validación

Se observa en la Figura (90) que algunas de las estimaciones para el oxígeno disuelto presentan valores negativos. Una explicación a este fenómeno es que debido a que el



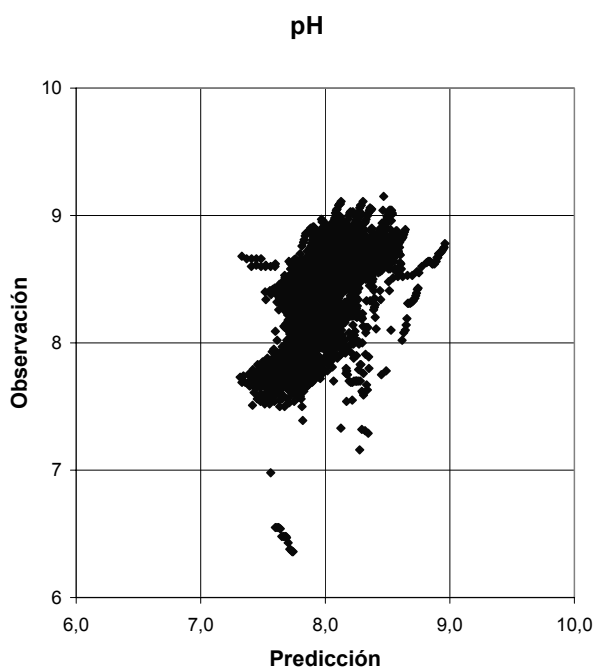
---

---

error de predicción puede ser tanto de signo positivo como negativo, si se tiene el caso de intentar predecir valores del oxígeno disuelto de magnitud muy reducida (siendo estos positivos) y el error en algunas estimaciones fuera de signo negativo, el valor predicho puede llegar a ser negativo. Esto es una evidencia de que el error de estimación en ocasiones posee signo negativo y en ningún caso se debe interpretar como la posibilidad de medir concentraciones negativas de oxígeno disuelto.

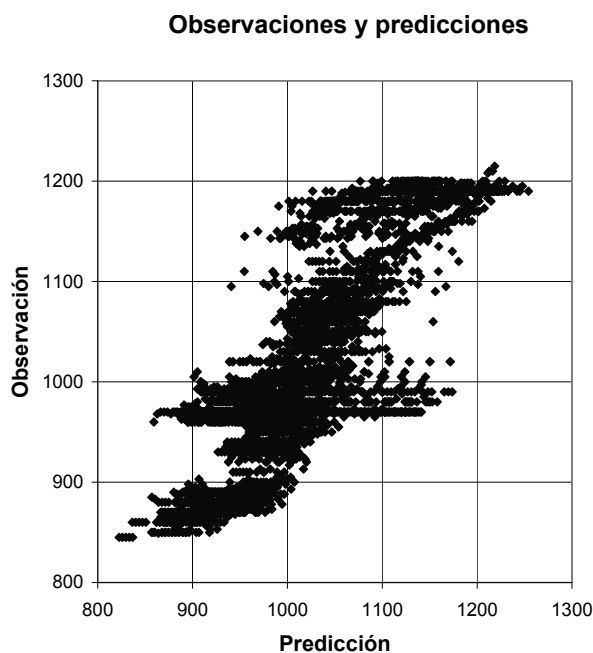
Teniendo en cuenta la Tabla (63) se puede pronosticar que las variables que serán mejor estimadas serán Ta ( $R^2 = 0,672$  de validación) y Odis ( $R^2 = 0,658$  de validación).

La calidad de la predicción será un poco inferior para la variable Con ( $R^2 = 0,592$  de validación) y la variable pH ( $R^2 = 0,381$  para validación) no será bien estimada por los métodos.



**FIGURA 91** Gráfico de observaciones y predicciones para el pH, periodo de validación

Al revisar en el Anexo 2 la estadística descriptiva de la variable pH se observa que su variabilidad medida por el Coeficiente de variación = Desviación típica/Media, es para el periodo de calibración de 0,0497 y en la validación de 0,0449, los cuales son valores muy reducidos (menos del 5%), lo cual indica que su valor cambia muy poco al modificarse otras variables. Esto genera que su modelación no proporcione buenos resultados.

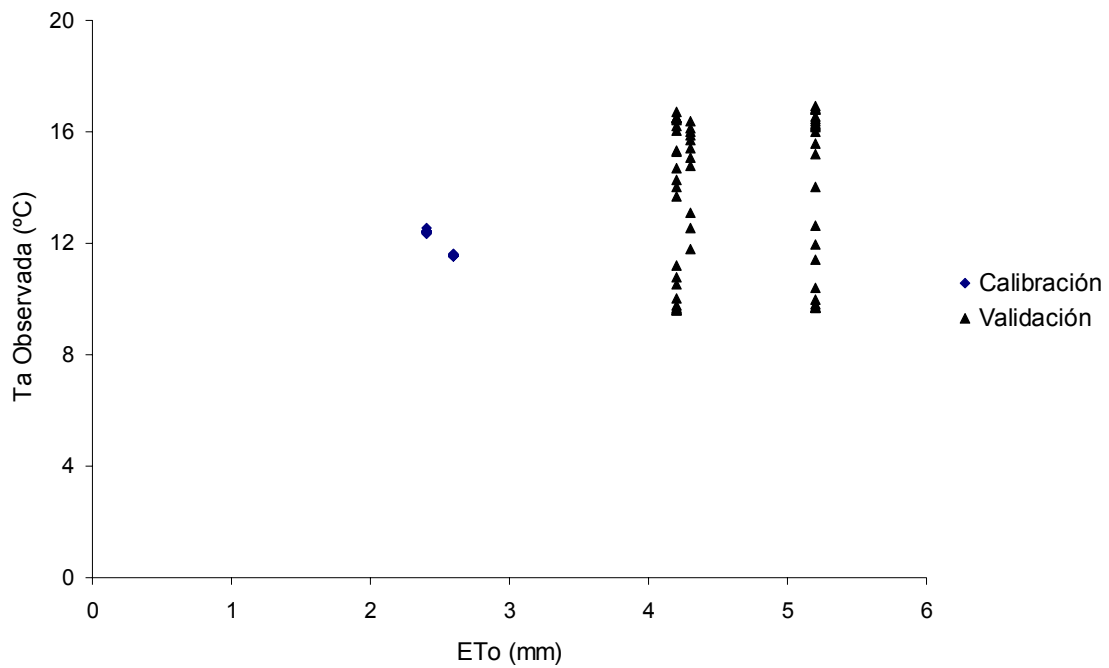


**FIGURA 92** Gráfico de observaciones y predicciones para la conductividad, periodo de calibración

Al aplicar la técnica MLR a los conglomerados formados por medio de las variables estandarizadas, las variables transformadas y las puntuaciones en componentes principales, se observa que tanto el coeficiente de determinación como el error estándar mejoran notablemente su desempeño durante la calibración. El problema estriba en que al utilizar las expresiones ajustadas a cada conglomerado (obtenidas con los datos de calibración) y aplicarlas con las observaciones del periodo de validación (al verificarlas), donde la calidad del  $R^2$  y el error estándar disminuyen enormemente.

Con el propósito de conocer qué sucede al interior de los conglomerados se elaborará un gráfico de las siguientes características: se considerará el conglomerado número 16 formado con las variables estandarizadas y se mostrará el gráfico de variación de una variable (por ejemplo, la temperatura del agua) con respecto a la variable de mayor influencia incluida en su análisis (ETo), para los periodos de calibración y validación (Figura 93).

En la Figura (93) se pueden apreciar varias características relevantes: en primer lugar los valores de la variable independiente más importante para ese conglomerado (ETo) varían en rangos muy distintos. Para el caso de la calibración los valores de ETo van de 2,4 a 2,6; para el periodo de validación de 4,2 a 5,2. Se puede concluir que el rango de variación de la variable independiente es bastante distinto para los dos periodos.



**FIGURA 93** Gráfico de la temperatura observada (Ta) contra la evapotranspiración de referencia (ETo) para el conglomerado 16 formado con variables estandarizadas

Ahora cabe preguntarse, ¿la diferencia de valores tan marcada en la variable predictiva ETo afecta a los valores que se puedan observar en la variable dependiente (Ta)? La misma figura anterior responde esta cuestión. Mientras que en la calibración efectuada por medio de conglomerados formados con las variables estandarizadas la temperatura observada del agua en este conglomerado varía en el intervalo 11,54 a 12,56 °C, en la validación su rango varía entre 9,61 y 16,93 °C. Por lo tanto, los valores de la temperatura del agua en la calibración realizada mediante conglomerados formados con las variables estandarizadas no cubren todos los valores que se presentan en la validación y esto va en contra de los principios generales de la modelación, los cuales indican que un modelo se debe elaborar con todas las condiciones (valores) que se puedan presentar en los periodos para los que se busca hacer predicciones. Se aprecia que por este motivo será mejor no dividir los periodos de calibración y validación en conglomerados, para que cubran un mayor rango de valores y la modelación pueda ser mejor realizada.

Y por último en el sentido gráfico de la regresión, si ajustásemos una recta a los puntos del periodo de calibración se tendría que esa recta (de pendiente negativa en este caso) pasaría bastante lejos de los puntos correspondientes a la validación.

En este punto surge una cuestión: ¿cuál es el motivo por el que la formación de conglomerados por medio del análisis cluster (y verificados con análisis discriminante)

---

---

con el propósito de efectuar análisis de regresión no proporciona buenos resultados? La evidencia que se tiene apunta hacia esta respuesta: en la formación de los conglomerados se consideraron 11 variables independientes con influencia equilibrada en la formación de los agrupamientos. Pero al efectuar la regresión dentro de cada conglomerado mediante métodos de selección de variables para dejar solamente a las que tuvieran más influencia en la variabilidad de la respuesta se tiene que en la mayoría de los casos (en nueve del total de 15 conglomerados), las variables independientes que tienen influencia y que por tanto aparecen en la expresión de regresión son sólo unas pocas (incluso en algunos conglomerados sólo una) y entonces se tiene un conglomerado formado según la influencia de 11 variables que no es el óptimo cuando sólo se consideran unas pocas de ellas con peso sobre la variable a pronosticar.

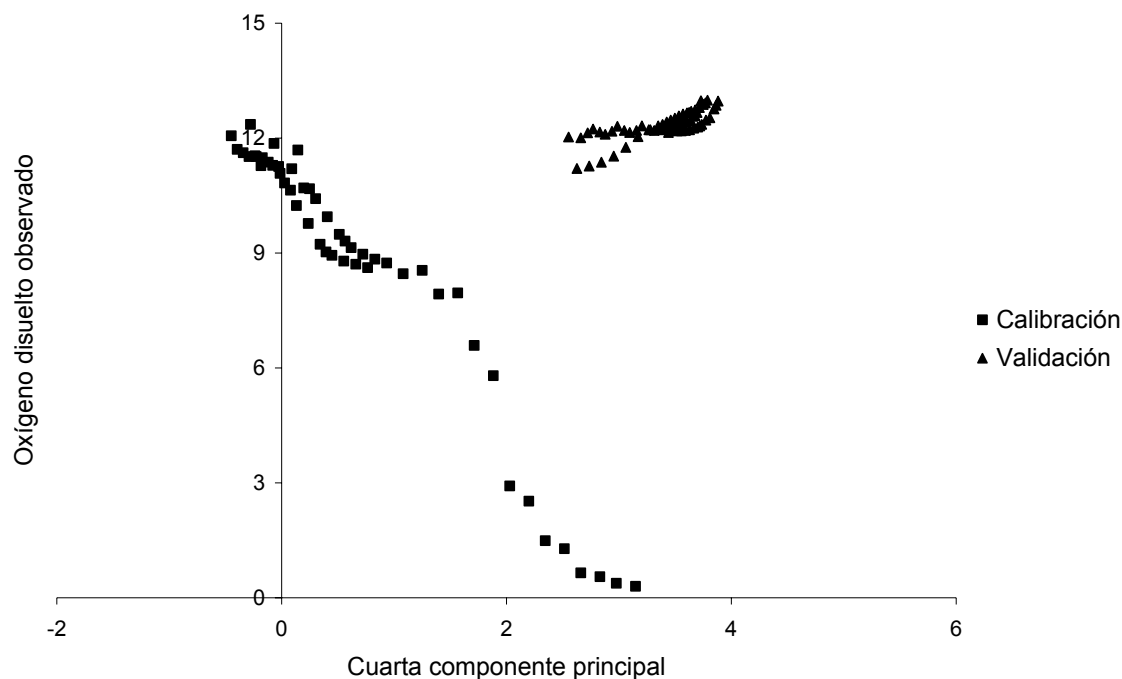
Es lógico pensar, siendo muy probable que así suceda, que al formar los conglomerados considerando menos variables los casos se agrupan de forma distinta (en otro conglomerado y con diferente cantidad de agrupaciones). Pero no es posible cambiar la formación de conglomerados en forma simple pues en la mayoría de los casos se tiene que mientras en unos conglomerados influyen principalmente unas variables, en otros influyen otras variables distintas de ellas. Por lo tanto, se tienen algunos elementos para afirmar que la formación de conglomerados para la mejora de cualquier tipo de regresión no es, en general, un planteamiento que lleve a obtener resultados óptimos. Para intentar mejorar la predicción mediante regresión será necesario buscar o proponer otras metodologías.

En la Figura (94) se muestra a modo de otro ejemplo el conglomerado número 3 creado por medio de las puntuaciones en componentes principales para la variable respuesta Odis, con respecto a la puntuación de mayor peso obtenida en el análisis con MLR, para calibración y validación.

Se puede afirmar para la Figura (94) que aproximadamente más del 50% de los valores que toma la variable independiente (puntuaciones en la cuarta componente principal) en la validación se encuentran fuera del rango o recorrido que esa misma variable tiene en el periodo de calibración, recordando que para la Figura (94) la calibración fue realizada por conglomerados formados con puntuaciones en componentes principales. La diferencia entre los dos periodos (en el mismo conglomerado) se aprecia fácilmente en la forma que adquieren las nubes de puntos, que para cada periodo son muy distintas. Se puede concluir que es más conveniente no dividir los periodos de calibración y validación en conglomerados, con el propósito

---

de que abarquen un mayor rango de valores y que la modelación a realizar proporcione los mejores resultados posibles.



**FIGURA 94** Gráfico de oxígeno disuelto observado (Odis) contra la cuarta componente principal para el tercer conglomerado formado con puntuaciones en componentes principales

Los probables motivos de esto son muy similares a los enunciados para la Figura (93): para formar los conglomerados se usaron 11 componentes principales. Al momento de realizar la regresión al interior de cada conglomerado, sólo considerando las puntuaciones que realmente tienen influencia en la variable respuesta, se tiene que (en la mayoría de los conglomerados) sólo unas pocas de las componentes son necesarias para efectuar la regresión dentro de cada conglomerado. Por lo que los conglomerados se formaron con 11 componentes pero su utilización no es adecuada cuando se consideran menos componentes (muy probablemente los conglomerados formados con menor número de componentes sean distintos). La conclusión es que la regresión por medio de conglomerados disminuye la calidad de la predicción en el periodo de validación.

El mismo fenómeno comentado anteriormente se vió para los conglomerados formados con variables transformadas y con puntuaciones factoriales.

Igualmente cuando se aplicaron los métodos de análisis de correlación canónica (CCA) y regresión en mínimos cuadrados parciales (PLSR), los resultados obtenidos dividiendo la regresión en conglomerados no fueron adecuados sino que fueron de

---

---

menor calidad predictiva que considerando sólo dos bloques de datos: calibración y validación. Debido a esto, en los apartados que prosiguen no se discutirán los resultados obtenidos por medio de conglomerados.

Por lo tanto, sólo se tomarán como valores de referencia de esta sección (MLR) los bloques de calibración y validación, los mostrados en la Tabla (63).

## 9.8 RESULTADOS DEL ANÁLISIS DE CORRELACIÓN CANÓNICA

El objetivo del CCA es encontrar combinaciones lineales de los conjuntos de variables  $X$  y  $Y$  que estén correlacionados al máximo (Jobson, 1992). Los pares de combinaciones lineales son llamados variables canónicas y sus correlaciones son llamadas correlaciones canónicas (Johnson et al., 1992). El CCA puede identificar los modos de covariabilidad lineal dominantes entre los vectores  $X$  e  $Y$  y permite hacer inferencias acerca del vector  $Y$ , conociendo el vector  $X$ .

Se observó que para poder alcanzar los mejores resultados es necesario estandarizar tanto las variables predictivas como las variables respuesta, con el fin de asegurar que todas posean la misma influencia en el análisis.

Un aspecto que se puede apreciar en las Tablas (70) y (71) es precisamente que al operar con valores estandarizados los parámetros de estandarización (la media y la desviación estándar) influyen en los resultados, pues para una misma variable aunque estos valores son relativamente similares para los periodos de calibración y validación, no son exactamente iguales. Por los resultados obtenidos se concluye que para estandarizar las variables independientes, tanto para calibración como validación, es más adecuado usar la media y la desviación estándar obtenidas para las variables en la calibración.

En la Tabla (71) se pueden consultar los valores de ajuste para validación empleando 17 variables predictivas (el análisis corresponde al Anexo 34). Se puede comparar con los mejores resultados obtenidos con MLR (Tabla 63) comprobándose que los resultados del análisis de correlación canónica son mejores en todos los aspectos, al MLR.

Número	Autovalor	Correlación canónica	Lambda Wilks	Chi-Cuadrado	p-valor
1	0,904494	0,951049	0,012279	19310,9	0,0000
2	0,704042	0,839072	0,128571	9003,04	0,0000
3	0,434364	0,659063	0,434423	3659,27	0,0000
4	0,231973	0,481636	0,768027	1158,39	0,0000

**TABLA 106** Análisis de correlación canónica con 17 variables predictivas estandarizadas

En el Anexo 34 se tiene que el  $p$ -valor está cercano a cero en los cuatro casos y por tanto se acepta la hipótesis de que los dos bloques de variables están relacionados (no son independientes entre sí). La máxima correlación canónica corresponde al primer par de variables canónicas (esta es la mayor correlación posible entre las combinaciones lineales de las variables de los conjuntos  $X$  y  $Y$ ) con un valor de 0,95 (Figura 95), el cual se puede considerar como muy alto; esta máxima correlación entre variables predictivas y variables respuesta explica un  $0,90/(0,90+0,70+0,43+0,23) = 40\%$  de la variabilidad y corresponde a la relación entre las variables canónicas  $u_1$  (en donde influyen más T y ETo) que mide la energía térmica en el ambiente y  $v_1$  (que pondera principalmente a las variables Ta –con signo positivo– y Odis –con signo negativo–); se puede interpretar que a mayor energía térmica ambiente habrá mayor temperatura del agua y menos oxígeno disuelto.

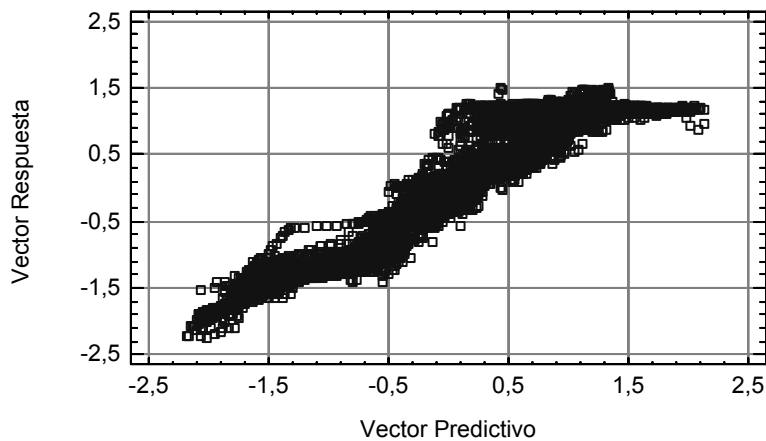
La segunda correlación canónica alcanza la cantidad de 0,84 (valor alto) y que representa un 31% de la variabilidad, contraponiendo a la variable independiente Profundidad, con la variable respuesta Con. Se puede interpretar que al revisar el embalse un día en perfil entre mayor sea la profundidad los valores de conductividad serán mayores.

La tercera correlación tiene un valor de 0,66 (cuantía regular), explica un 19% de variabilidad y manifiesta la relación entre la variable independiente VolExp y Con, de manera que al aumentar el volumen almacenado disminuirá a todos niveles la conductividad.

Y la cuarta correlación es 0,48 la cual se puede clasificar como baja (en comparación con las tres anteriores) y representa el 10% de variabilidad, donde la influencia de algunas de las variables predictivas se manifiesta sobre la variable respuesta pH.

Los coeficientes de las variables canónicas en el primer y segundo bloque indican como se constituyen las variables  $u_i$  y  $v_i$ ; a continuación se muestra el gráfico del primer par de variables canónicas ( $u_1, v_1$ ).

### Variables Canónicas



**FIGURA 95** Gráfico del primer par de variables canónicas usando 17 variables predictivas estandarizadas

A continuación en un nuevo análisis se tomó la decisión de no utilizar a la variable DJ para sólo emplear en el bloque de variables predictivas a las de naturaleza meteorológica e hidrológica. Los resultados se presentan en la Tabla (72). El análisis se incluye en el Anexo 35.

Número	Autovalor	Correlación canónica	Lambda de Wilks	Chi-Cuadrado	p-valor
1	0,888754	0,942737	0,014428	18605,2	0,0000
2	0,702735	0,838293	0,129695	8965,85	0,0000
3	0,432815	0,657887	0,436296	3640,8	0,0000
4	0,23077	0,480385	0,76923	1151,65	0,0000

**TABLA 107** Análisis de correlación canónica con 16 variables predictivas estandarizadas

Los resultados son muy similares a los obtenidos con valores muy próximos y la misma interpretación de las variables, pero usando una variable predictiva menos. Por lo tanto este análisis se considera mejor.

En el Anexo 36 se eligieron las variables independientes que puntuaran en las dos primeras variables canónicas (las que representan la mayoría de la variabilidad) con coeficientes mayores a 0,10.

Nuevamente se aplicó este criterio hasta que todas las variables cumplieron con él, quedando en total nueve variables independientes: T, Tn, Tx, Hf, HR, ETo, VolExp, Qsal y Profundidad.



Variable canónica:	<b>u<sub>1</sub></b>	<b>u<sub>2</sub></b>	<b>u<sub>3</sub></b>	<b>u<sub>4</sub></b>
ZT	<b>0,887309</b>	<b>-0,229647</b>	-0,301244	0,297692
ZTn	-0,029928	<b>0,418870</b>	-0,199611	-0,948600
ZTx	<b>-0,536910</b>	<b>0,156156</b>	0,546816	0,960649
ZHf	-0,015831	<b>0,269957</b>	-0,142189	-0,985355
ZHR	<b>0,166896</b>	<b>-0,104247</b>	-0,171385	0,067276
ZETo	<b>0,521978</b>	-0,015602	-0,584138	-1,08298
ZVolExp	<b>-0,345362</b>	<b>0,288191</b>	1,29268	0,699026
Zqsal	<b>0,240610</b>	<b>0,149963</b>	0,432042	0,081627
ZProfundidad	<b>0,252772</b>	<b>-0,871791</b>	0,345578	-0,220588

**TABLA 108** Coeficientes para las variables canónicas de la primera serie después de seleccionar sólo 9 variables

La evaluación de los resultados obtenidos está en la Tabla (73) que muestra los resultados del CCA aplicado al periodo de validación usando 9 variables predictivas. Este es el mejor ajuste logrado por medio de la técnica de CCA, pues los valores del coeficiente de determinación y los errores típicos de estimación son de los mejores alcanzados con esta técnica, utilizando sólo nueve variables.

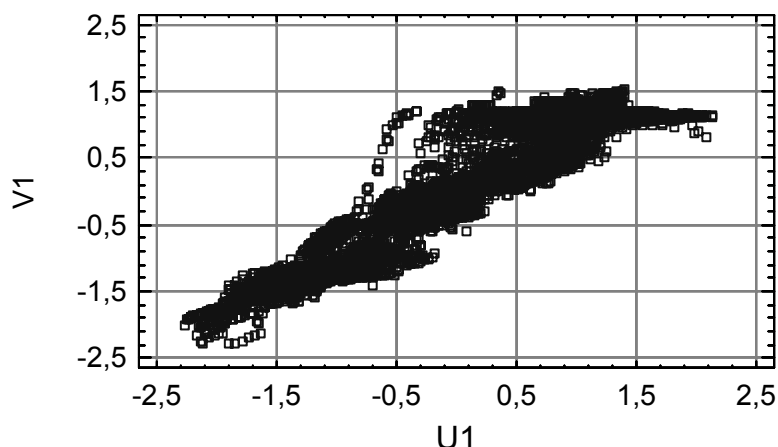
Número	Autovalor	Correlación canónica	Lambda de Wilks	Chi-Cuadrado	p-valor
1	0,878535	0,937302	0,016546	18018,4	0,0000
2	0,698269	0,835625	0,13622	8757,38	0,0000
3	0,426065	0,652736	0,451461	3493,61	0,0000
4	0,213395	0,461947	0,786605	1054,45	0,0000

**TABLA 109** Análisis de correlación canónica con 9 variables predictivas estandarizadas

Los *p*-valores están próximos a cero lo que confirma la hipótesis de que los dos bloques de variables están relacionados entre sí. La mayor correlación canónica tiene un valor de 0,94 (Figura 96), explicando un  $0,88/(0,88+0,70+0,43+0,21) = 40\%$  de la variabilidad y representa la relación entre las variables T y ETo (energía térmica en el ambiente) que son las que más influyen en **u<sub>1</sub>**, además de Ta y Odis (con signos contrarios las dos últimas) que son las que poseen mayor peso en **v<sub>1</sub>**; esto indica lo anteriormente ya expresado, a mayor energía térmica en el ambiente se tendrá mayor temperatura del agua y menor oxígeno disuelto.

La segunda correlación canónica tiene un valor de 0,84 y representa un 31% de variabilidad. La tercera correlación es de 0,65 y engloba un 19% de la varianza del conjunto. Y la cuarta correlación tiene un valor de 0,46, capturando un porcentaje de variabilidad de 10%. Las relaciones entre las variables, tanto las predictivas como las respuesta, que más influencia tienen en las tres restantes variables canónicas son

idénticas a las ya mencionadas para el Anexo 34.



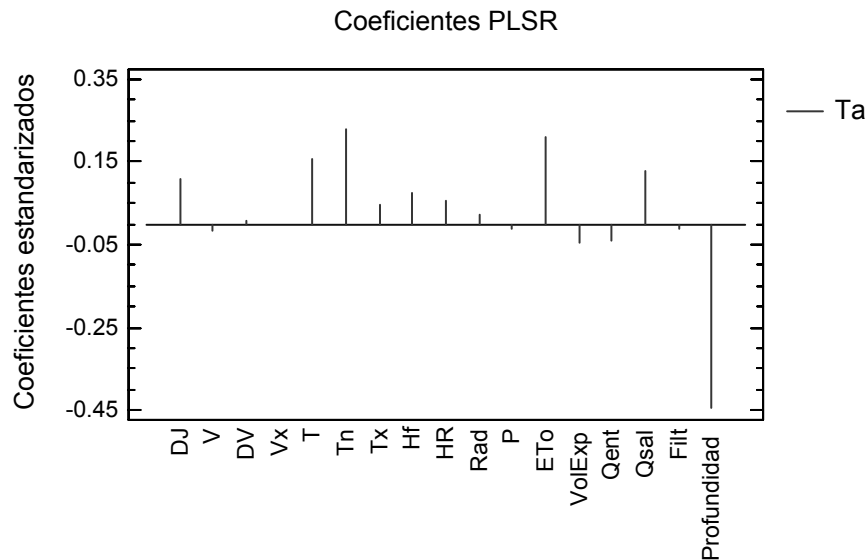
**FIGURA 96** Gráfico del primer par de variables canónicas usando 9 variables predictivas estandarizadas

### 9.9 RESULTADOS DEL PLSR

Con el objetivo de obtener en cada caso el mejor ajuste a los datos se efectuó dentro del periodo de calibración la técnica de validación cruzada para obtener el mejor número de componentes para la regresión. Luego se comprobó que este número de componentes es óptimo para los datos de validación intentando ajustes con distinta cantidad de componentes. Es necesario proceder así, pues los coeficientes de cada una de las variables (Figura 97) dentro de la regresión son distintos según las componentes consideradas.

	Ta
Constante	8,21165
DJ	0,004911
V	-0,275504
DV	0,00055768
Vx	-0,00793173
T	0,132786
Tn	0,198842
Tx	0,0369386
Hf	0,109653
HR	0,0236497
Rad	0,00163213
P	-0,0274099
ETo	0,758365
VolExp	-1,85832E-7
Qent	-0,00000603205
Qsal	0,00006267
Filt	-0,0000645576
Profundidad	-0,312097

**TABLA 110** Coeficientes para las variables sin estandarizar en la expresión de temperatura del agua con 17 variables predictivas



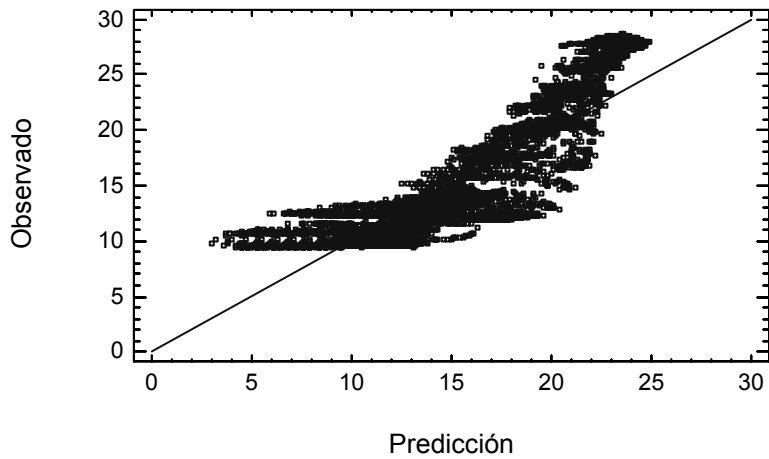
**FIGURA 97** Coeficientes estandarizados del PLSR para la temperatura del agua usando 17 variables predictivas

### 9.9.1 Usando 17 variables predictivas

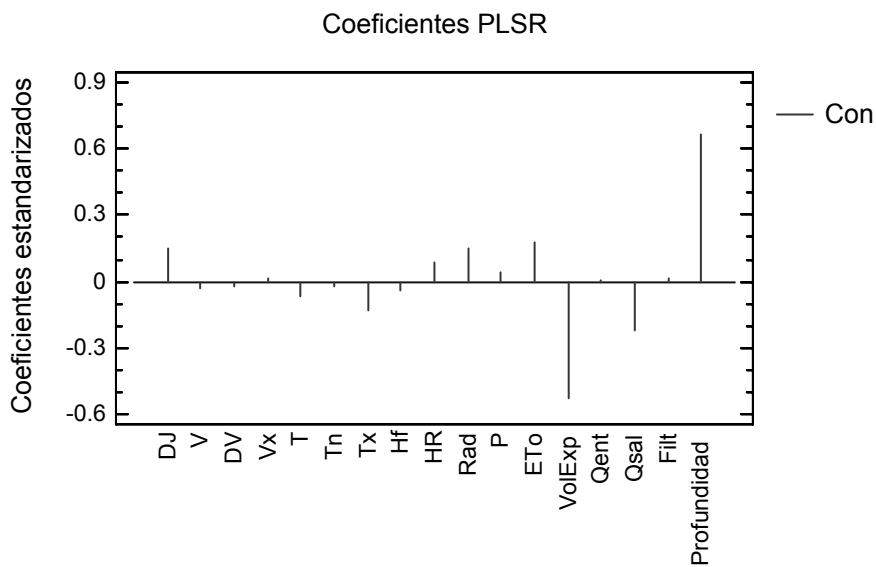
Para iniciar con la aplicación de esta técnica de regresión (Juricek et al., 2005) y verificar su desempeño en distintas situaciones se comenzó utilizando todas las 17 variables predictivas y aplicando el algoritmo PLS1 (una variable respuesta en cada simulación).

Los resultados del ajuste se muestran en la Tabla (74). Al compararlos con los obtenidos en el CCA usando igualmente 17 variables predictivas (Tabla 71), se pueden emitir las siguientes observaciones:

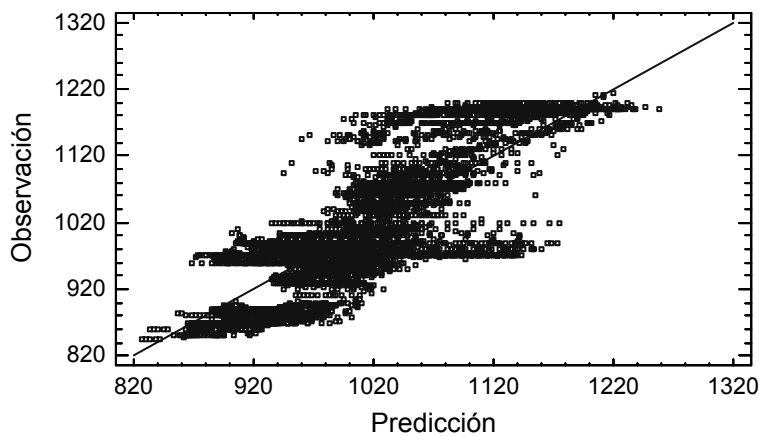
- El ajuste efectuado por medio del PLSR proporciona resultados ligeramente mejores a los del CCA.
- Con el PLSR no existió la necesidad de usar los valores de la media y la desviación estándar de las variables en la calibración para estandarizar en la validación. Debido a esto, se percibe que el PLSR muestra una mayor capacidad de generalización comparado con el CCA.
- Las variables que continúan siendo mejor pronosticadas son Ta (Figura 98) y Odis, la variable Con (Figuras 99 y 100) tiene un ajuste regular y pH muestra un ajuste no adecuado en los casos.



**FIGURA 98** Gráfico de observaciones y predicciones para la temperatura del agua usando 17 variables predictivas



**FIGURA 99** Coeficientes estandarizados del PLSR para la conductividad usando 17 variables predictivas



**FIGURA 100** Gráfico de observaciones y predicciones para la conductividad del agua usando 17 variables predictivas

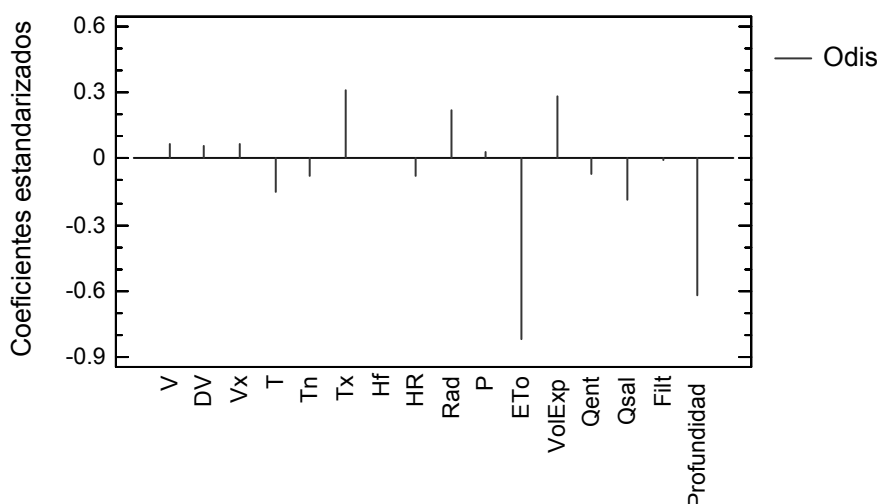
### 9.9.2 Empleando 16 variables predictivas

Al utilizar solamente las variables meteorológicas e hidrológicas (16 variables predictivas) para el pronóstico de las cuatro variables de calidad del agua se generan los resultados de la Tabla (75).

	<i>Odis</i>
Constante	0.0
V	0.0668062
DV	0.0526179
Vx	0.0639692
T	-0.149585
Tn	-0.0759311
Tx	0.313499
Hf	-0.00106866
HR	-0.0787602
Rad	0.215195
P	0.0310925
ETo	-0.819897
VolExp	0.281137
Qent	-0.0731177
Qsal	-0.182441
Filt	-0.00854808
Profundidad	-0.615212

**TABLA 111** Coeficientes para las variables estandarizadas en la expresión de oxígeno disuelto con 16 variables predictivas

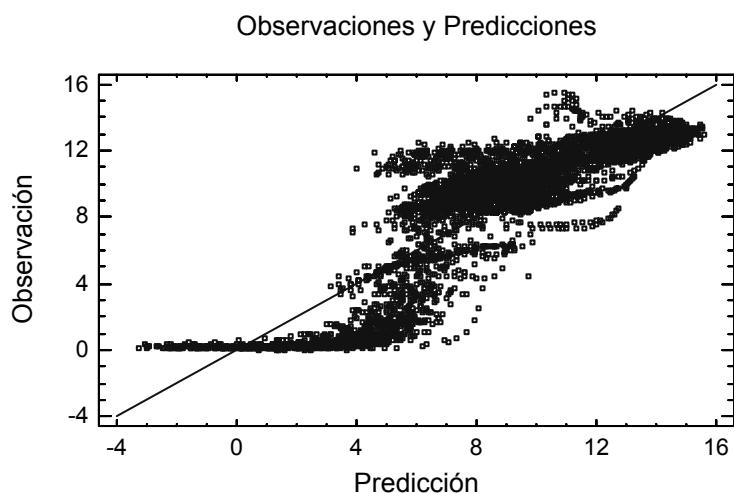
Al comparar con la Tabla (74) (que usa 17 variables predictivas) se tiene que disminuye levemente el coeficiente de determinación de la variable Ta en la validación y aumentó un poco el error estándar en la validación de la variable Con. Fuera de estas dos observaciones, todos los demás parámetros (seis de ellos) que miden la calidad del ajuste mejoraron.



**FIGURA 101** Coeficientes estandarizados del PLSR para el oxígeno disuelto usando 16 variables predictivas

---

Por lo que se puede concluir que el uso de una variable menos en el análisis, DJ, trae aparejado una mejoría en general del ajuste (Figuras 101 y 102) aparte de producir un modelo más sencillo, el cual es siempre más adecuado.



**FIGURA 102** Gráfico de observaciones y predicciones para el oxígeno disuelto usando 16 variables predictivas

En la Figura (102) se puede apreciar que algunos valores de predicción del oxígeno disuelto toman valores negativos. Esto se presenta debido a que el error asociado a la técnica de regresión puede tomar tanto signo positivo como negativo. Para el caso en que el valor observado sea de magnitud reducida (cercano al valor cero, pero con signo positivo siempre) y el error de predicción sea de signo negativo, se puede dar el caso de que la estimación efectuada tenga también signo negativo, pero esto se debe de interpretar como la ocurrencia de un error de signo negativo en una estimación y no como la presencia real de concentraciones de oxígeno disuelto negativas.

### 9.9.3 Recortado de datos

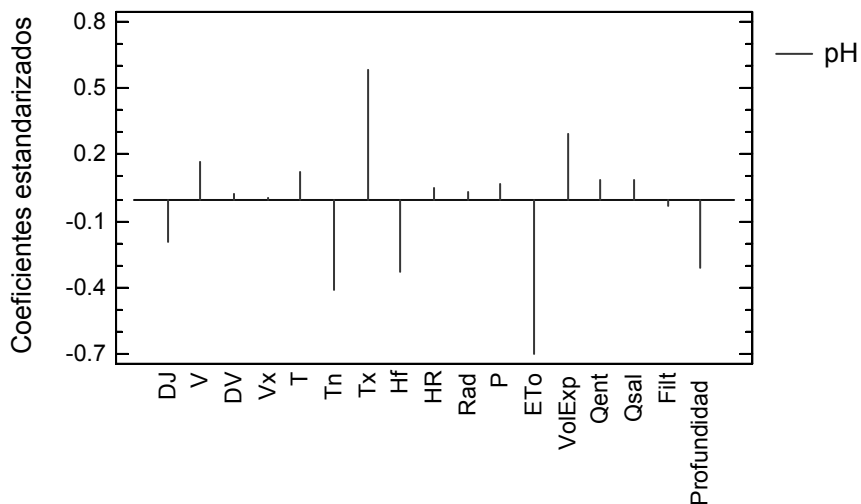
En el proceso de recortado se sustituyeron para todas las variables (predictivas y respuesta) los valores de los extremos y fueron sustituidos por el percentil 1 y el percentil 99. Esto se efectuó sólo para calibración, pues se espera que un modelo bien definido pueda desempeñarse de manera adecuada en la mayoría de los casos (un 98% aproximadamente) cuando se manejen nuevos datos.

Los resultados de este proceder se muestran en la Tabla (76) y al compararlos con el proceso sin recortado (Tabla 75) se tiene que los parámetros de ajuste son prácticamente iguales. Esto puede ser un indicativo acerca de los cuidados y procesos seguidos para la depuración de los datos originales, pues se percibe que no existen

valores atípicos extremos que modifiquen apreciablemente el proceso de regresión.

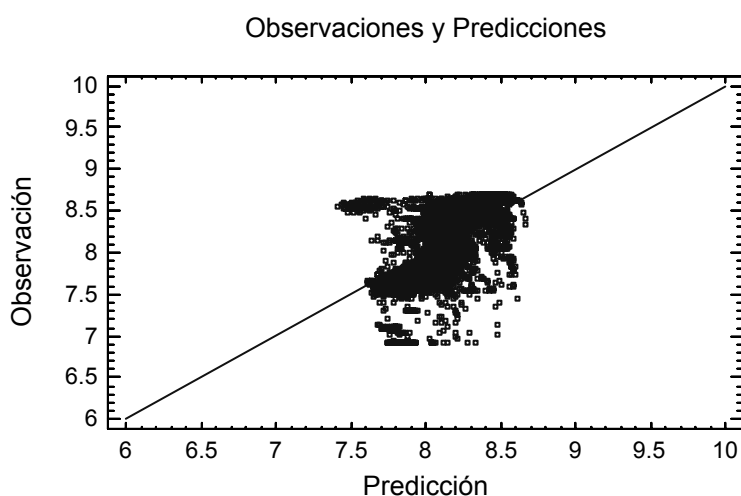
	pH
Constante	0.0
DJ	-0.160396
V	0.174546
DV	0.0315953
Vx	0.00335869
T	0.496119
Tn	-0.443548
Tx	0.37001
Hf	-0.322485
HR	0.0189803
Rad	0.155479
P	0.0789855
ETo	-0.936876
VolExp	0.336707
Qent	0.0595838
Qsal	0.0514418
Filt	-0.0325172
Profundidad	-0.312261

**TABLA 112** Coeficientes para las variables estandarizadas en la expresión del pH usando el recortado



**FIGURA 103** Coeficientes estandarizados del PLSR para el pH obtenidos con las variables recortadas

Pero es necesario manifestar que aunque en este caso el recortado de datos como proceso no significó un cambio en los resultados (Figuras 103 y 104), en otras aplicaciones o estudios donde puedan existir valores extremos no detectados sí puede constituirse en una herramienta indispensable para evitar los problemas de ajuste generados por esos valores.



**FIGURA 104** Gráfico de observaciones y predicciones para el pH usando las variables del recortado

#### 9.9.4 Modelo reducido en términos lineales

Para reducir el modelo, sin perder capacidad predictiva, se utilizaron los coeficientes estandarizados obtenidos con el PLSR (Figura 105). Claramente estos coeficientes son una gran ayuda para lograr la reducción del modelo.

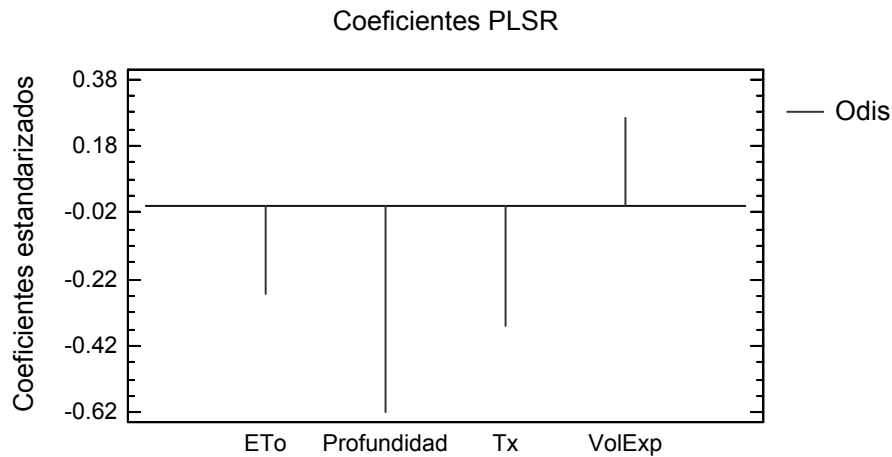
	Odis
Constante	11,1268
ETo	-0,752148
Profundidad	-0,342047
Tx	-0,238237
VolExp	8,68778E-7

**TABLA 113** Coeficientes para las variables sin estandarizar en la expresión del oxígeno disuelto con 4 variables predictivas

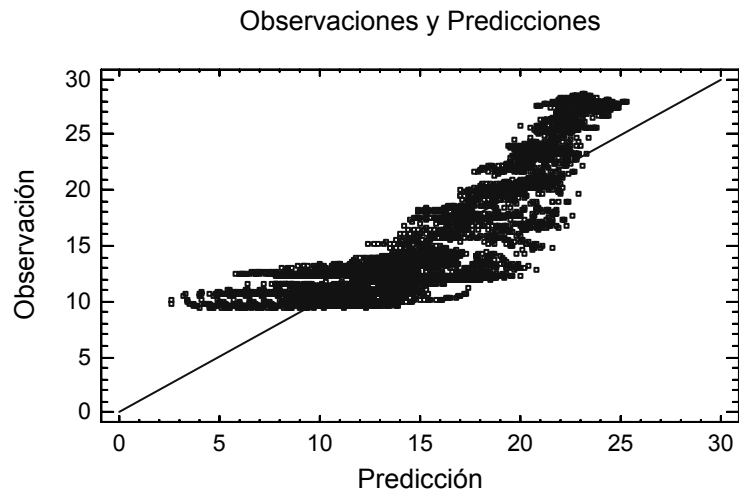
Los parámetros de ajuste que se generan al proponer los modelos reducidos se muestran en la Tabla (77).

Al comparar con el modelo sin reducir de 16 variables (Tabla 75) se tiene que los coeficientes de determinación y los errores de predicción muestran que el ajuste conseguido es superior (Figuras 106 y 107).





**FIGURA 105** Coeficientes estandarizados más significativos del PLSR para el oxígeno disuelto (modelo reducido)



**FIGURA 106** Gráfico de observaciones y predicciones para la temperatura del agua, modelo reducido en términos lineales

Además, las expresiones para estimar las variables de calidad del agua son marcadamente más simples (ecuaciones 210, 211, 212 y 213).

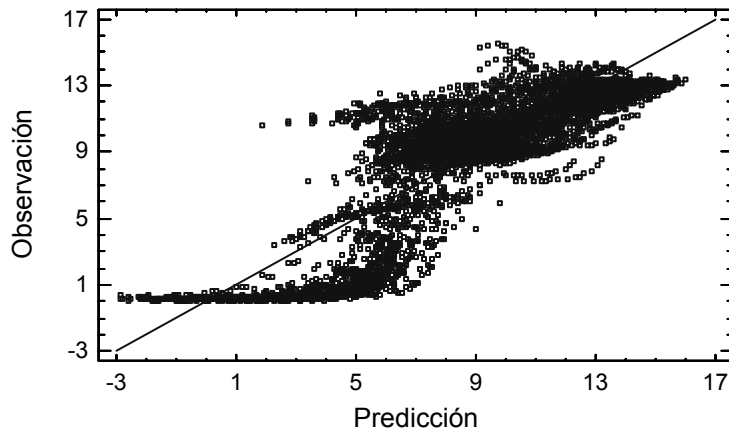
$$Ta\_Est = 8,3202 - 0,310551 \cdot Profundidad + 0,000112235 \cdot Qsal + 0,56017 \cdot ETo + 0,191481 \cdot Tn + 0,180964 \cdot T \quad (221)$$

$$Odis\_Est = 11,1268 - 0,752148 \cdot ETo - 0,342047 \cdot Profundidad - 0,238237 \cdot Tx + (8,68778 \times 10^{-7}) \cdot VolExp \quad (222)$$

$$pH\_Est = 7,43143 - 0,144645 \cdot ETo + 0,0305618 \cdot Tx - 0,0259479 \cdot Tn + (1,11747 \times 10^{-7}) \cdot VolExp - 0,0408722 \cdot Hf - 0,0154395 \cdot Profundidad \quad (223)$$

$$Con\_Est = 1220,88 + 8,71871 \cdot Profundidad - 0,0000336155 \cdot VolExp - 0,854586 \cdot ETo \quad (224)$$

Observaciones y Predicciones



**FIGURA 107** Gráfico de observaciones y predicciones para el oxígeno disuelto, modelo reducido

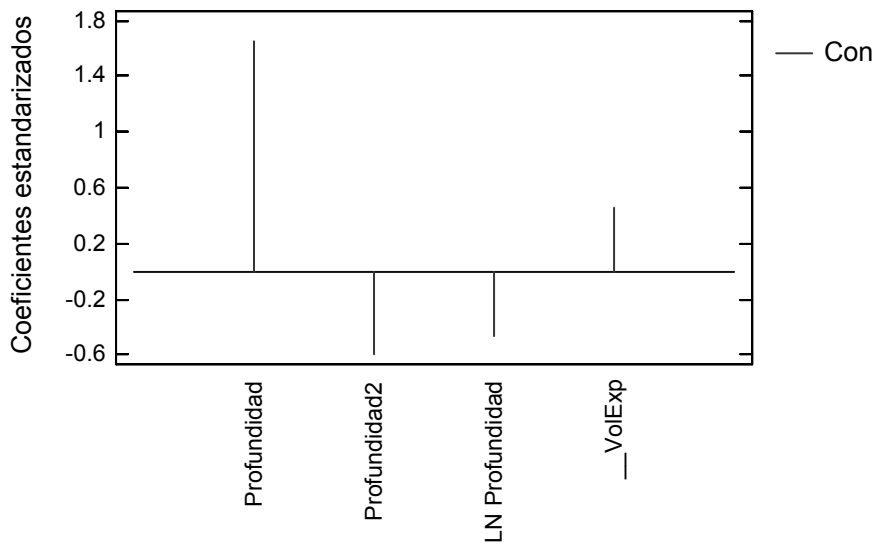
9.9.5 Matriz extendida de entradas - INLR

Al incluirse en la regresión términos no lineales de las variables predictivas (Figura 108) se busca que el método PLSR pueda en un momento dado hacer frente a las posibles faltas de linealidad en las relaciones entre las variables (Figura 109).

	Con
Constant	0,0
Profundidad	1,65575
Profundidad2	-0,597087
LN Profundidad	-0,462874
1/VolExp	0,449577

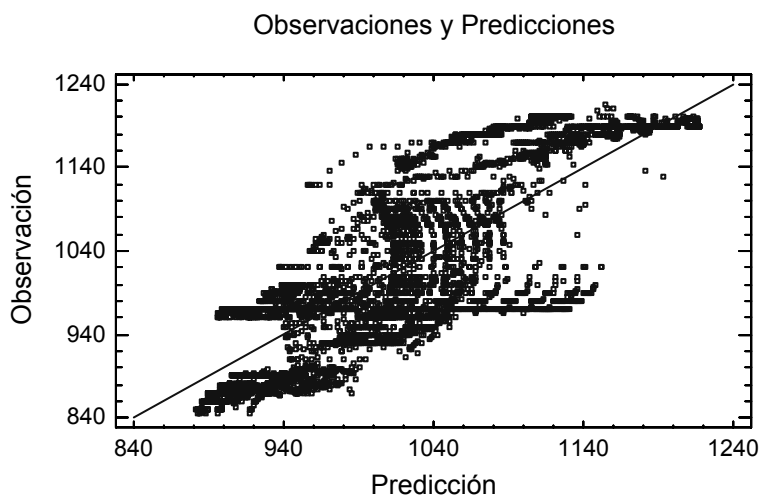
**TABLA 114** Coeficientes estandarizados en el PLSR para determinar conductividad usando términos no lineales relevantes

Coeficientes PLSR



**FIGURA 108** Coeficientes estandarizados más significativos del PLSR para la conductividad considerando términos no lineales

Aplicando este procedimiento a los datos se obtiene la Tabla (78). Al comparar con el modelo reducido lineal (Tabla 77) se tiene, de hecho, que el ajuste lineal proporciona mejores salidas, principalmente se puede apreciar el mejor desempeño del modelo lineal en el oxígeno disuelto.



**FIGURA 109** Gráfico de observaciones y predicciones para la conductividad considerando términos no lineales

Por otra parte, al efectuar regresiones de tipo no lineal sobre las estimaciones efectuadas con el modelo reducido lineal y los valores observados de variables de calidad del agua se obtuvieron las ecuaciones (214), (215), (216) y (217). Los parámetros que evalúan el ajuste se muestran en la Tabla (79).

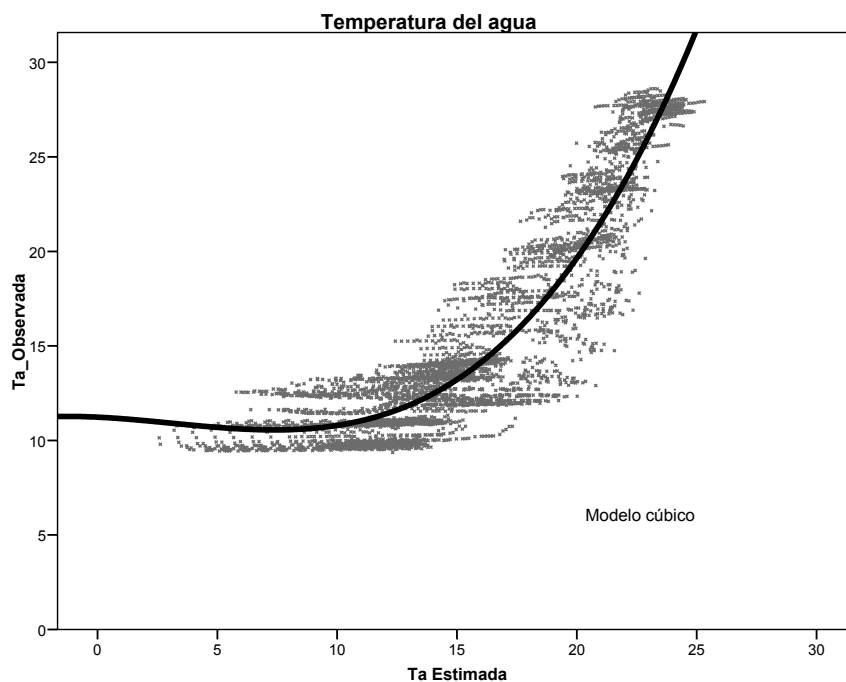
Variable dependiente: Ta\_Obs

Ecuación	Resumen del modelo		Estimaciones de los parámetros			
	R cuadrado	Sig.	Constante	b1	b2	b3
Cúbica	,8635	,000	11,232	-,061	-,021	,002

La variable independiente es Ta\_Est.

**TABLA 115** Resumen del modelo para estimar la temperatura del agua usando como entrada las estimaciones del modelo reducido lineal

Se puede apreciar una mejora significativa de la estimación de la temperatura del agua (al comparar con la Tabla 77), manifestándose esto por el aumento del  $R^2$  y la disminución del error de estimación (Figura 110); el oxígeno disuelto mejora muy poco en  $R^2$  y aumenta levemente su error de estimación.



**FIGURA 110** Gráfico de observaciones y predicciones para la temperatura al efectuar la regresión sobre el modelo reducido lineal y las observaciones

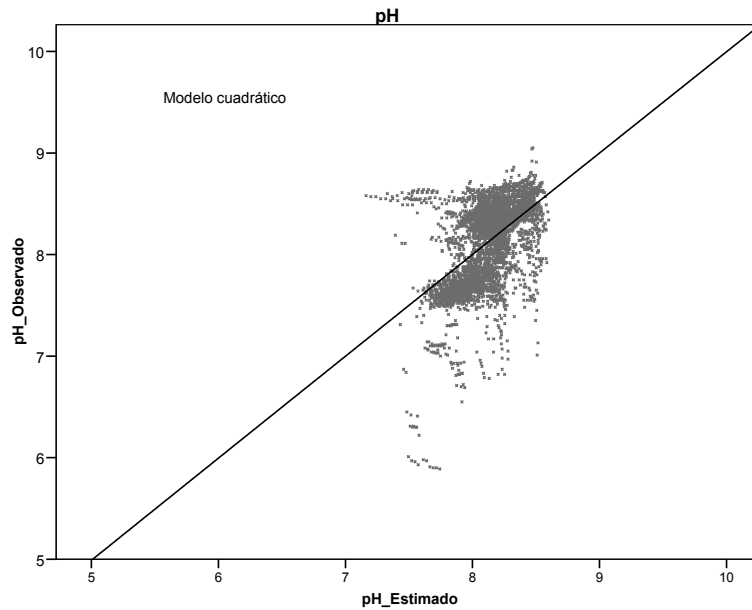
Las variables pH (Figura 111) y Con prácticamente no se ven modificadas en forma alguna.

Variable dependiente: pH\_Obs

Ecuación	Resumen del modelo		Estimaciones de los parámetros		
	R <sup>2</sup>	Sig.	Constante	b1	b2
Cuadrático	,266	,000	-,070	1,017	-,001

La variable independiente es pH\_Est.

**TABLA 116** Resumen del modelo para estimar el pH usando como entrada las estimaciones del modelo reducido lineal



**FIGURA 111** Gráfico de observaciones y predicciones para el pH al efectuar la regresión sobre el modelo reducido lineal y las observaciones

#### 9.10 PREDICCIÓN DE VARIABLES DE CALIDAD DEL AGUA CON MÉTODOS MULTIVARIANTES

En base a los apartados anteriores se tiene que para efectuar pronósticos las mejores expresiones encontradas en este trabajo usando métodos multivariantes son: para temperatura del agua (Ta) la ecuación (214); para oxígeno disuelto (Odis) la ecuación (211); la variable pH obtuvo sus mejores estimaciones con la ecuación (212); en el caso de la conductividad (Con) la mejor ecuación es la (213).

$$\begin{aligned} \text{Ta} = & 11,23160366 - 0,0608831 \cdot (\text{Ta\_Est}) - 0,0206138 \cdot (\text{Ta\_Est})^2 \\ & + 0,00223407 \cdot (\text{Ta\_Est})^3 \end{aligned} \quad (214)$$

$$\begin{aligned} \text{Odis\_Est} = & 11,1268 - 0,752148 \cdot \text{ETo} - 0,342047 \cdot \text{Profundidad} \\ & - 0,238237 \cdot \text{Tx} + (8,68778 \times 10^{-7}) \cdot \text{VolExp} \end{aligned} \quad (211)$$

$$\begin{aligned} \text{pH\_Est} = & 7,43143 - 0,144645 \cdot \text{ETo} + 0,0305618 \cdot \text{Tx} - 0,0259479 \cdot \text{Tn} \\ & + (1,11747 \times 10^{-7}) \cdot \text{VolExp} - 0,0408722 \cdot \text{Hf} - 0,0154395 \cdot \text{Profundidad} \end{aligned} \quad (212)$$

$$\begin{aligned} \text{Con\_Est} = & 1220,88 + 8,71871 \cdot \text{Profundidad} - 0,0000336155 \cdot \text{VolExp} \\ & - 0,854586 \cdot \text{ETo} \end{aligned} \quad (213)$$

La única ecuación no lineal es la correspondiente a Ta y se propone debido a su mejor ajuste. Por otra parte, las expresiones de las otras tres variables de calidad provienen del modelo reducido lineal y se estima que para la mayoría de las situaciones de pronóstico que se puedan presentar tendrán un comportamiento más predecible y adecuado que sus contrapartes no lineales.

Periodo	Variable	R	R <sup>2</sup>	Error estándar de estimación
Calibración	Ta	0,9292	0,8635	2,03
	Odis	0,8420	0,7090	2,33
	pH	0,5154	0,2656	0,35
	Con	0,7820	0,6115	62,56
Validación	Ta	0,8617	0,7426	3,54
	Odis	0,8655	0,7491	2,32
	pH	0,6959	0,4843	0,42
	Con	0,7940	0,6304	71,44

**TABLA 117** Resumen de resultados de los mejores ajustes obtenidos

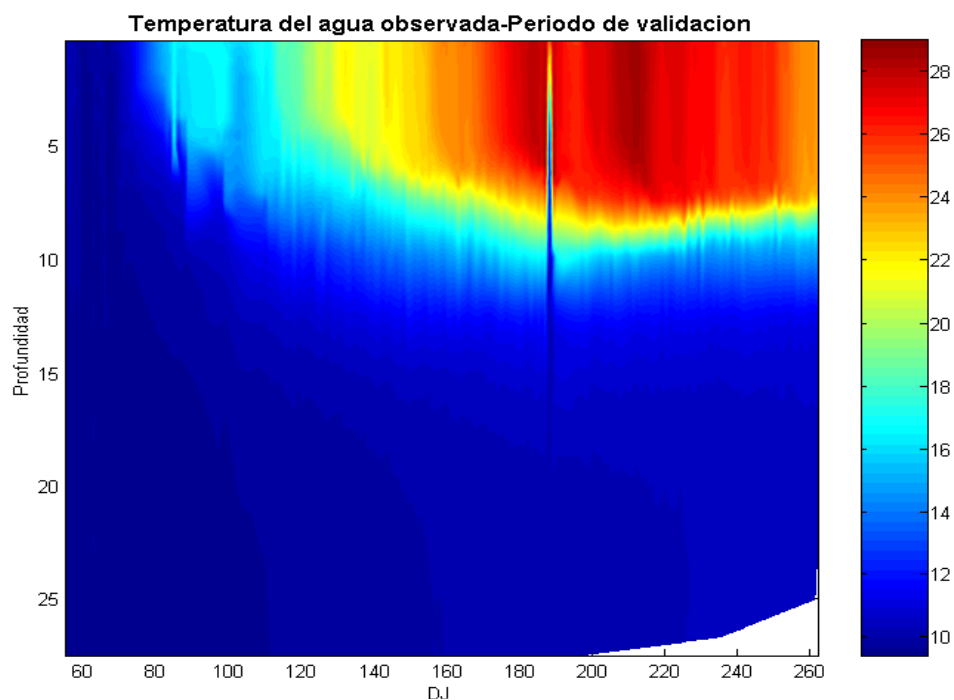
Se compararán los resultados de la tabla anterior con los de otros estudios relacionados con los recursos hídricos efectuados con métodos multivariantes.

En el texto de Gutiérrez et al. (2004), página 103, interesa conocer la relación entre las temperaturas superficiales del mar en Chicama y Paita (Perú) tomándose como predictando la temperatura de Paita y como variable respuesta la de Chicama; se utilizó regresión lineal, consiguiéndose una recta de regresión con  $R^2 = 0,649$ . Además, también se efectúa una regresión lineal para el caudal del Río Piura y la precipitación en Piura, obteniéndose un  $R^2 = 0,781$ .

En el artículo de Petersen et al. (2005) se efectúa un estudio para conocer si las interacciones entre variables de calidad del agua durante un periodo de 5 años en varias localizaciones a lo largo del Río Elba pueden ser recuperadas como patrones de covarianza estadísticamente significativos. Para este propósito se aplicó el PCA y se mostró que el 60% de la variabilidad en el conjunto de datos puede explicarse a través de dos combinaciones lineales de las ocho variables originales.

En el artículo de Ouarda et al (2008) se utilizan varios métodos basados en el análisis de correlación canónica y análisis cluster jerárquico para estimar la frecuencia de caudales en sitios no aforados. Los resultados que se obtienen para varios cuantiles de la descarga específica proporcionan valores del coeficiente de determinación,  $R^2$ , que van desde 0,5419 hasta 0,8383 para los cinco métodos considerados.

En seguida, con el propósito de comprender como se comportan las cuatro variables de calidad del agua y evaluar visualmente las estimaciones conseguidas por medio de las ecuaciones propuestas se muestran gráficos (interpolados) de las cuatro variables respuesta para la validación, mostrando valores observados y valores estimados:



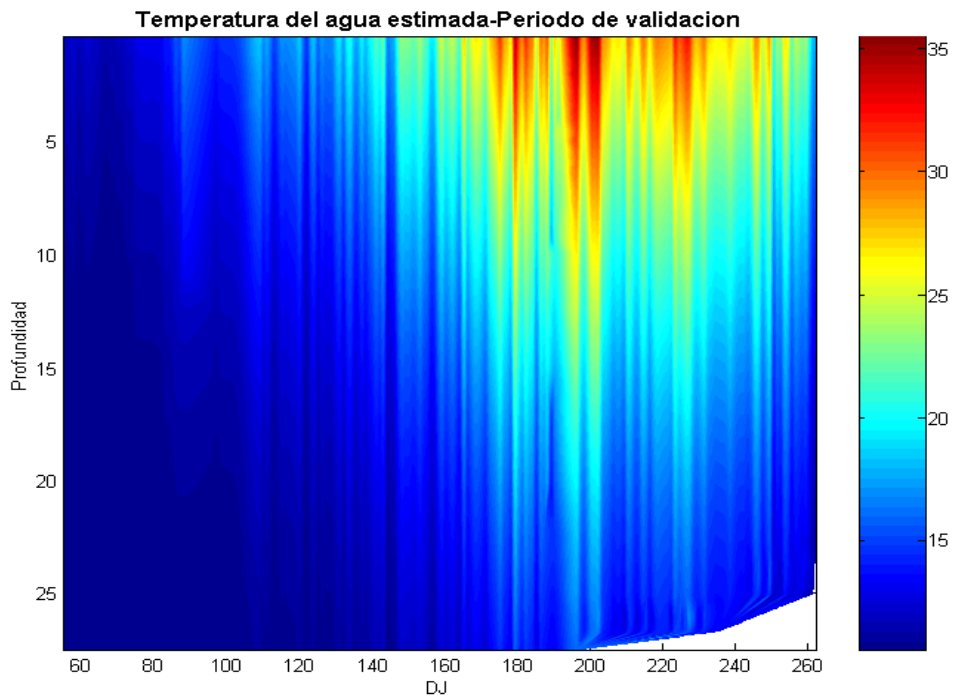
**FIGURA 112** Gráfico de temperatura del agua observada durante el periodo de validación

Se inicia mostrando la temperatura del agua observada para la validación (Figura 112). Como los días con información válida no son continuos, en los días que no hay información se realizó una interpolación para dibujar el gráfico; en este caso se empleó una interpolación cúbica.

Se aprecia en la Figura (112) que la estratificación ocurre aproximadamente del día 120 al 260 (los meses de mayo, junio, julio, agosto y septiembre). Se puede considerar que el resto del tiempo el embalse presenta una temperatura casi uniforme (la temperatura presenta pocas variaciones en perfil vertical).

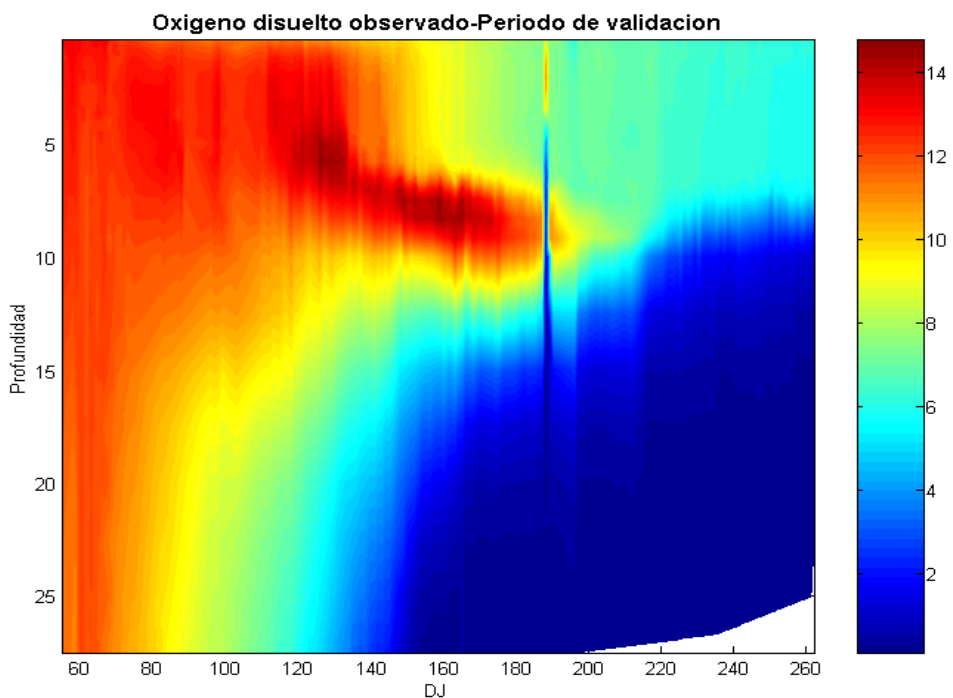
La estimación de la temperatura del agua para validación es la que se muestra en la Figura (113).

Al comparar las Figuras (112) y (113) se puede decir que la forma general de la gráfica está bien representada. La “suavidad” de los valores observados se hecha de menos en el gráfico de estimados. Asimismo, el límite superior que alcanzan los valores estimados es mayor al de los observados, esto se puede ver en la barra de escala.



**FIGURA 113** Gráfico de temperatura del agua estimada durante el periodo de validación

El oxígeno disuelto alcanza valores mayores cuando la temperatura es menor en el embalse. Al aumentar la temperatura del agua con la época del año el oxígeno disminuye. En la Figura (114) se puede observar que la estratificación térmica influye notablemente en la concentración de oxígeno siendo ésta menor en el fondo del almacenamiento.

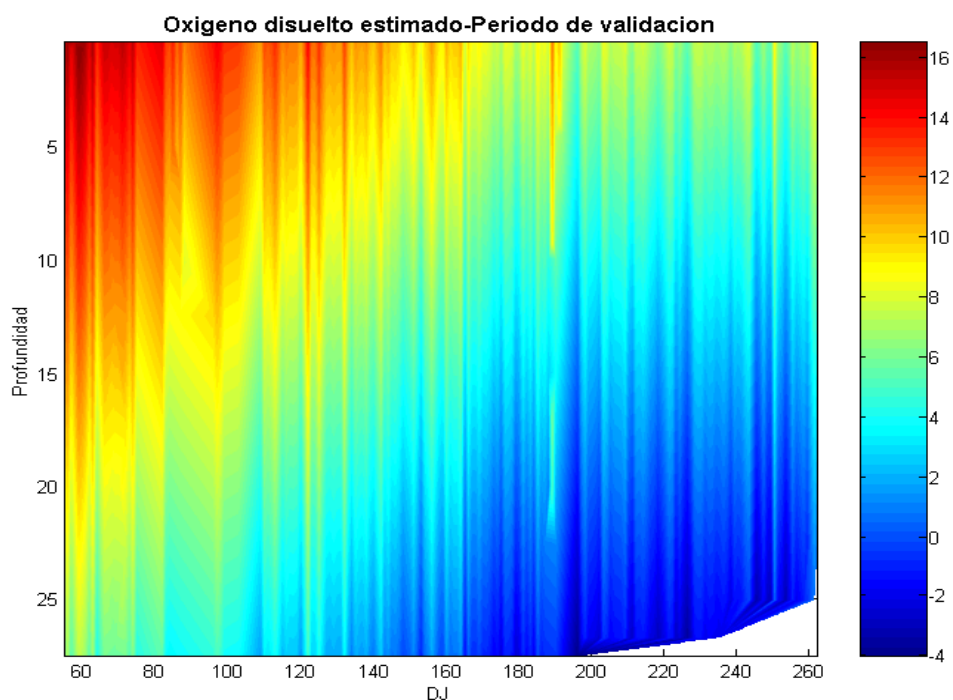


**FIGURA 114** Gráfico de oxígeno disuelto observado durante el periodo de validación



---

Los valores calculados de oxígeno disuelto se muestran en la Figura (115).

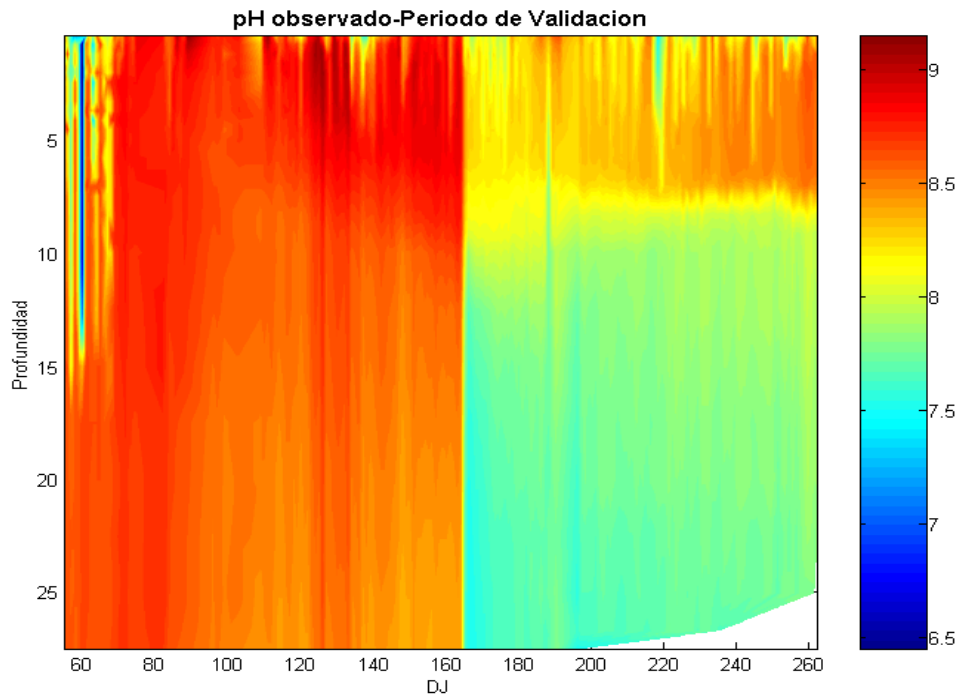


**FIGURA 115** Gráfico de oxígeno disuelto estimado durante el periodo de validación

La forma del gráfico se conserva aproximadamente teniendo que el rango en el que se mueven los valores estimados es más amplio (tanto en el límite superior como en el inferior) que el de los valores observados, llegándose inclusive a estimar unos cuantos valores negativos en el fondo del embalse cuando está presente la estratificación térmica.

Es necesario mencionar que al requerir hacer estimaciones de la variable oxígeno disuelto y los valores observados son muy reducidos (valores positivos, pero próximos a cero), si el error aleatorio de predicción fuera de signo negativo, lo que ocurriría es que el valor pronosticado sería de signo negativo. Esto no quiere decir que se puedan presentar valores negativos en el embalse, sólo es una evidencia de algunos fallos que puede presentar el modelo planteado.

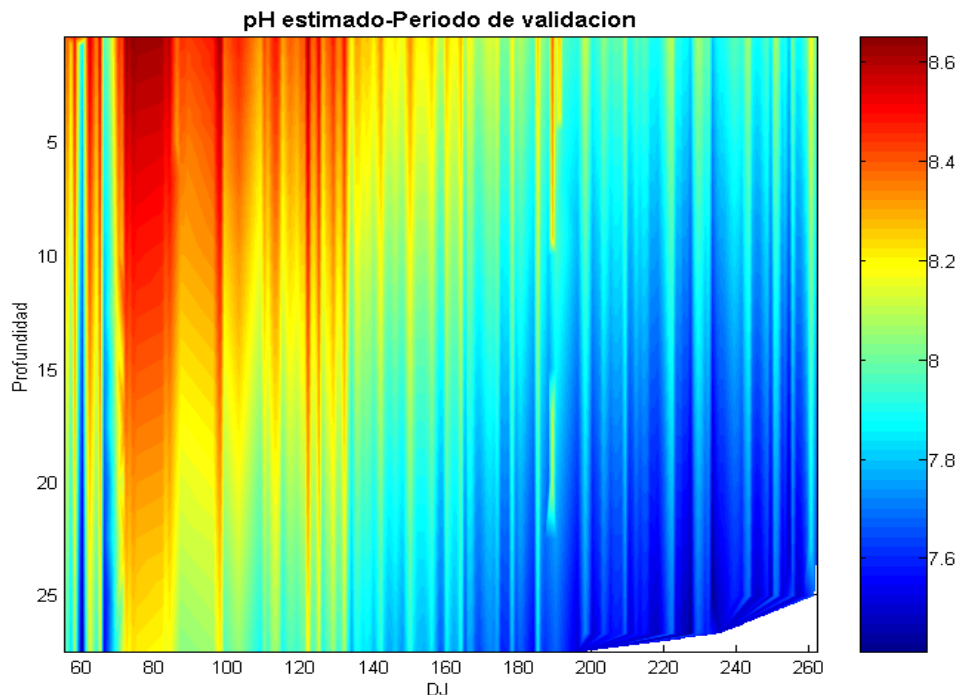
Los valores observados de pH se muestran en la Figura (116).



**FIGURA 116** Gráfico de pH observado durante el periodo de validación

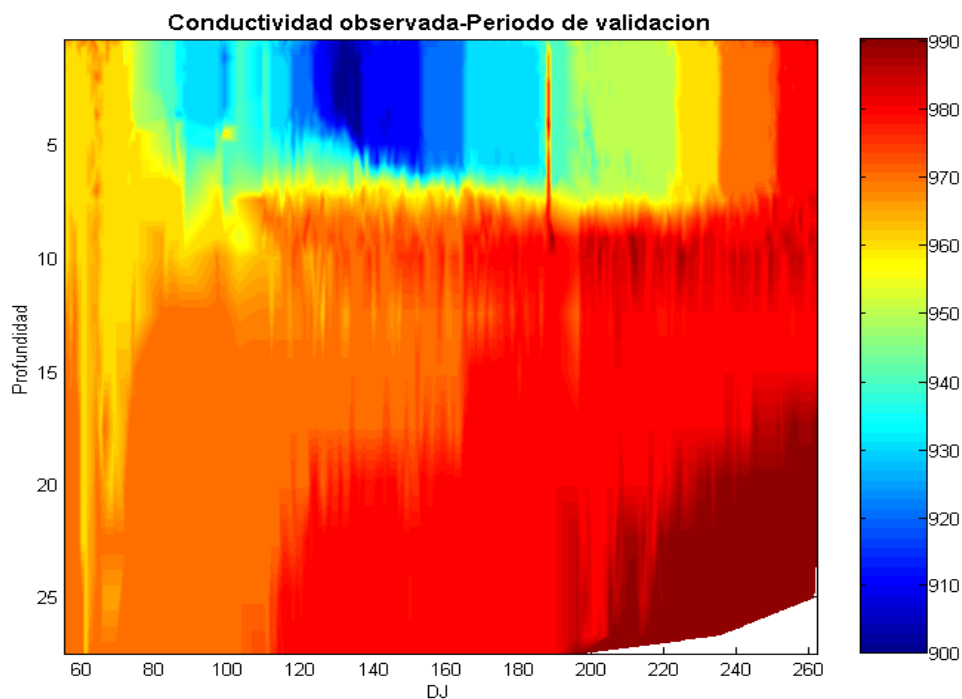
El valor del pH es más o menos uniforme en tanto no ocurra estratificación térmica o que ésta sea débil. Al establecerse la estratificación térmica intensa se forman dos zonas bien diferenciadas (ver Figura 116): la primera desde la superficie del almacenamiento hasta una profundidad de siete metros con un valor de pH aproximado de 8,4; la segunda abarca desde el fin de la primera zona hasta el fondo tomando aproximadamente un valor de 7,8.

De las cuatro variables de calidad del agua analizadas el pH fue la que mostró más problemas para su estimación (Figura 117) con los valores más bajos del coeficiente de determinación. El rango de los valores estimados es más reducido que el de los valores observados. Además, en la parte del periodo en que todavía no se presenta la estratificación térmica intensa (hasta el día 160 aproximadamente) no se representa adecuadamente la situación real de mezclado casi uniforme. Por otra parte, cuando ya está ocurriendo la estratificación térmica la delimitación de las dos zonas no se hace de una forma muy precisa, estando los valores estimados del pH en las dos zonas por debajo de los valores reales.



**FIGURA 117** Gráfico de pH estimado durante el periodo de validación

El comportamiento de los valores observados de la conductividad se puede apreciar en la Figura (118):



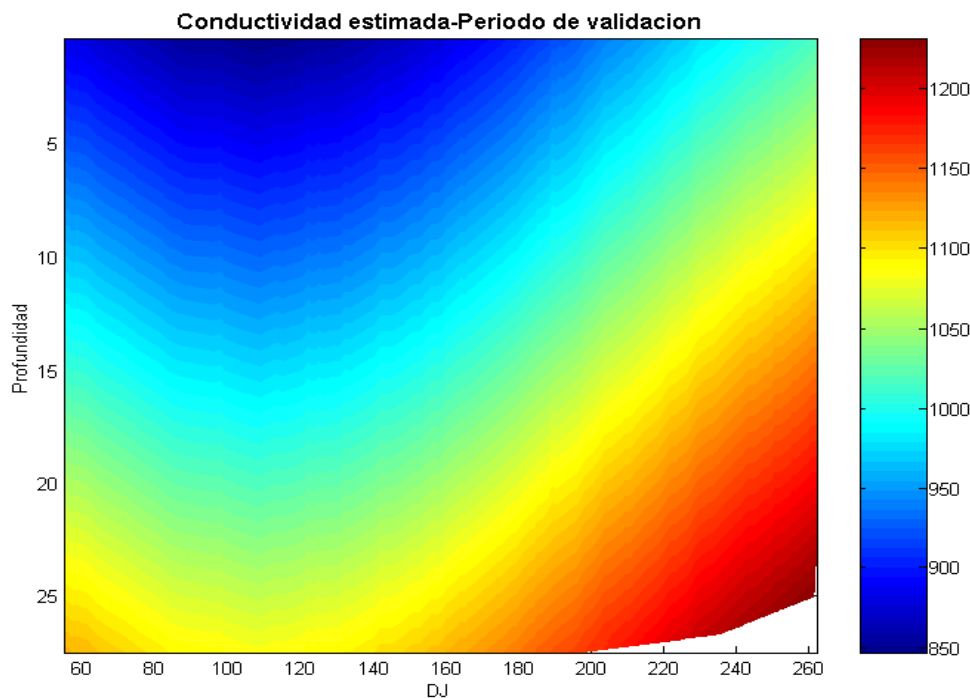
**FIGURA 118** Gráfico de conductividad observada para el periodo de validación

En la Figura (118) se puede observar que la concentración de la conductividad se estratifica siguiendo un patrón similar al de la temperatura. En todos los casos para un

---

día específico los menores valores de la conductividad se encuentran en la parte superior del almacenamiento y los valores más altos se ubican en el fondo del mismo. En el mes de mayo (aproximadamente) es cuando se observan los menores valores de conductividad en la capa más superficial del embalse.

Se deben hacer varias consideraciones acerca de la conductividad estimada (Figura 119). Para empezar la distribución espacio-temporal de la conductividad es compleja (como se puede apreciar al revisar el gráfico de conductividad observada), lo que hace previsible que su estimación precisa sea difícil de lograr. Además, como en la determinación de la ecuación para estimación se tiene que es la variable de calidad del agua que depende de menos variables predictivas su gráfico no mostrará gran diversidad en su comportamiento.



**FIGURA 119** Gráfico de conductividad estimada para el periodo de validación

Al comparar los gráficos de conductividad observada y estimada se tiene que en general los valores correspondientes a la misma ubicación espacio-temporal coinciden aproximadamente. El gráfico de estimados no es capaz de representar la forma de estratificación real de la conductividad que constituye una capa uniforme de aproximadamente siete metros de espesor, casi horizontal y adyacente a la superficie. Asimismo, el rango de valores de la conductividad no es reproducido adecuadamente, pues en el gráfico de valores estimados el límite inferior es de 850, mientras que en el gráfico de observados es de 900; el límite superior se encuentra en peor situación aún

---

---

pues mientras que el de valores reales toma un valor de 990, el de valores estimados llega hasta 1200, lo que es claramente una sobrevaloración excesiva en algunas zonas del gráfico.

#### 9.11 RESULTADOS DEL MODELO MECANICISTA

La calidad de los resultados obtenidos con la aplicación del modelo mecanicista, mediante un programa desarrollado especialmente en el lenguaje Visual Basic 2005, se puede evaluar con la tabla siguiente:

Periodo	Variable	R	R <sup>2</sup>	Error estándar de estimación
Calibración	Ta	0,9703	0,9414	1,56
Validación	Ta	0,9567	0,9152	2,01

**TABLA 118** Resumen de resultados del modelo mecanicista aplicado a temperatura del agua en el embalse

La variabilidad explicada es muy alta (más del 91% para el periodo de validación) y el error estándar de estimación se puede considerar relativamente reducido.

La validación efectuada con la información disponible, en términos generales, fue satisfactoria para la variable de salida temperatura del agua.

Las gráficas mensuales mostradas en el capítulo anterior muestran el perfil de temperatura observado en el embalse y tienen una concordancia entre los datos observados y los simulados que se puede evaluar como excelente para los meses de enero, febrero, marzo, abril y mayo. La simulación de junio y julio se puede considerar como buena y en los meses de agosto y septiembre la relación entre temperaturas observadas y simuladas es regular.

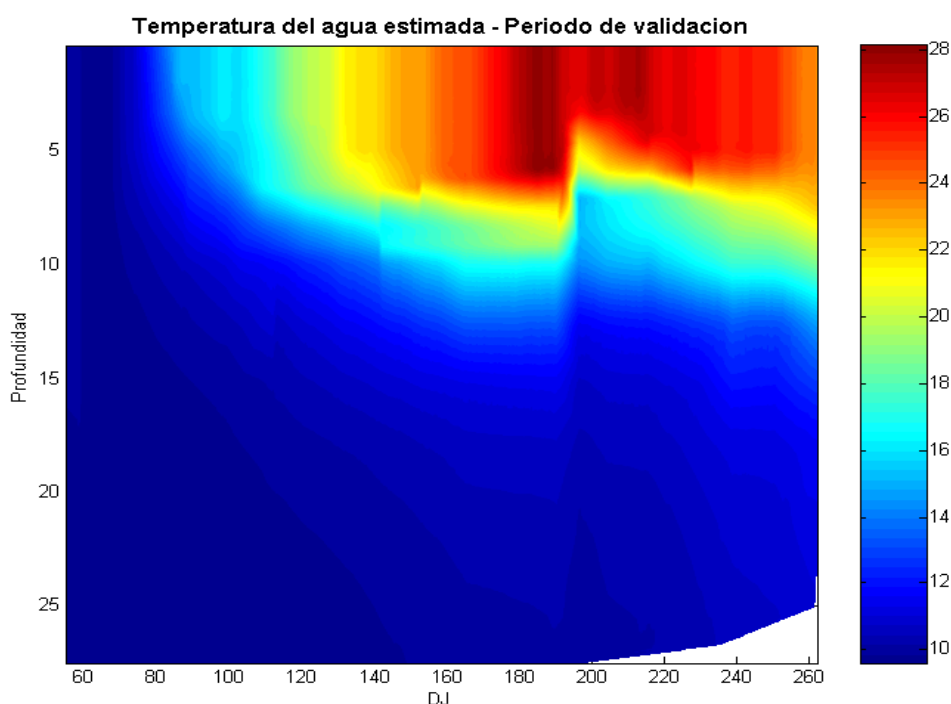
Cuando el almacenamiento está completamente mezclado o con una estratificación moderada el modelo con fundamento físico representa adecuadamente las condiciones de temperatura en su interior. En el caso de una estratificación fuerte la simulación alcanzada con el modelo va de regular a buena.

De los perfiles de temperatura obtenidos en la modelación se observa que para condiciones de mezcla completa los resultados son excelentes; cuando se tiene estratificación moderada la simulación del perfil de la temperatura del agua es muy buena. En el caso de estratificación fuerte el modelo hace una simulación adecuada del perfil de temperatura en los primeros ocho metros de profundidad, en donde los

---

datos observados y valores simulados están sensiblemente cercanos.

El gráfico (interpolado) de la variable temperatura del agua (Figura 120) en el embalse para el periodo de validación, mostrando los valores estimados por el método mecanicista, exhibe buena concordancia con el gráfico de valores observados para el mismo intervalo de tiempo. Se tiene que las características y espesores del epilimnion, metalimnion e hipolimnion son a grandes rasgos bien representados, observándose más deficiencias para simular el espesor del metalimnion.



**FIGURA 120** Gráfico de temperatura del agua estimada con el método mecanicista durante el periodo de validación

## 9.12 COMPARATIVO DE MODELACIÓN CON PLSR Y MÉTODO MECANICISTA

Son varios los aspectos a comparar en la aplicación de los dos métodos al cálculo de la temperatura del agua en el embalse: cantidad de información requerida, flexibilidad del modelo, evaluación de los valores estimados contra los observados y fidelidad en la representación de los perfiles.

Respecto a la información requerida, en el PLSR se requiere aportar 11 variables meteorológicas, 4 variables de estado del embalse y los datos observados de temperatura del agua en perfil. Por otra parte el método mecanicista requiere 6 variables meteorológicas y la cobertura de nubes (calculada), 4 variables de estado del embalse, los datos observados de temperatura del agua en perfil, la temperatura del

---

---

caudal de entrada, 4 parámetros, la tabla de calado-volumen-superficie, espesor mínimo y máximo de capa, ubicación geográfica y altitud del embalse y temperatura del agua superficial. Al comparar se aprecia claramente que el modelo mecanicista requiere más información de entrada que el método PLSR lo que implica un mayor tiempo dedicado a modelar. Además, los parámetros del modelo mecanicista deben ser calibrados lo que supone simular en muchas ocasiones, mientras que en el PLSR sólo es necesario ajustar el número de componentes a extraer, acción que es generalmente más rápida de realizar.

Acerca de la flexibilidad del modelo hay dos aspectos a considerar: adaptación a la unidad temporal en que se expresan los datos y cantidad requerida de los mismos. Respecto a la capacidad de procesamiento para el intervalo temporal, el PLSR tiene gran adaptabilidad y si los datos son proporcionados en otra unidad de tiempo (semanalmente, cada 12 horas, cada 3 horas, etc.), este método tiene la capacidad de obtener la simulación sin hacer modificaciones en él. En cambio el modelo mecanicista comprende expresiones matemáticas adaptadas a una escala temporal diaria, por lo que para trabajar a otra escala diferente se tendrían que efectuar cambios en parte de su estructura, lo que requiere tiempo y conocimientos de programación. La cantidad de datos requerida por el PLSR es para el caso de muestras naturales de hasta centenares de muestras; el motivo es que las muestras de calibración deben contener todas las variaciones (Ferré, 2006) que luego se encontrarán al validar o al predecir. En cambio el modelo con fundamento físico podrá usarse para hacer predicciones, en condiciones similares, con una cantidad de muestras mucho menor debido a que las ecuaciones que lo sustentan almacenan intrínsecamente las relaciones clave entre las variables y requieren por tanto de menos datos observados.

La evaluación de valores estimados contra observados para temperatura del agua con el PLSR para el periodo de validación proporciona un coeficiente de determinación de  $R^2 = 0,7426$  y un error de estimación de  $3,54^{\circ}\text{C}$ ; mientras que con el modelo mecanicista se obtiene para la validación un  $R^2 = 0,9152$  y un error de estimación de  $2,01^{\circ}\text{C}$ . Es evidente que el método mecanicista explica un mayor porcentaje de variabilidad y que el error cometido representa un 57% del correspondiente a la aplicación del método PLSR.

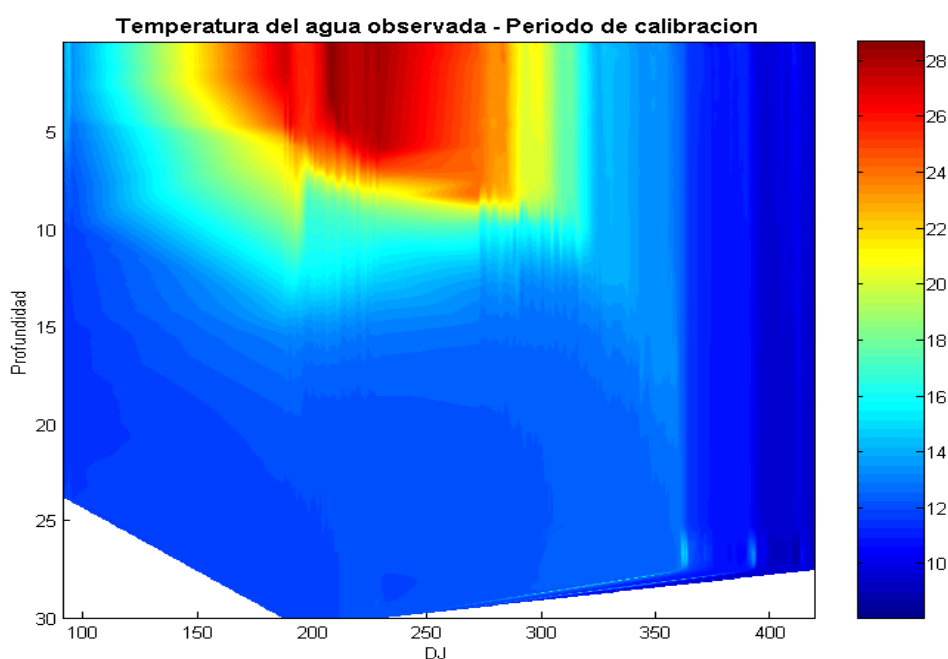
La calidad en la representación de los perfiles de temperatura y la correcta configuración de las capas del embalse (epilimnion, metalimnion e hipolimnion) son factores importantes en la modelación de la temperatura del agua en almacenamientos. Se observa que el PLSR no define muy adecuadamente los perfiles

---

puntuales de las capas, aunque las envolventes generales sí se corresponden con lo observado. En el caso del modelo mecanicista los perfiles de las capas se observan bien definidos y aunque el espesor varía con respecto a los valores observados, principalmente para el metalimnion, la representación en general es mejor que la conseguida por medio del PLSR.

La representación de los perfiles de temperatura es un punto que puede permitir adentrarse en el conocimiento de las simulaciones efectuadas. Debido a ello se procede a revisar con más detalle este aspecto.

Primero se presentarán las representaciones de valores observados de temperatura para el periodo de calibración (Figura 121), así como las correspondientes simulaciones: a) la efectuada con el mejor ajuste hecho con el PLSR (los resultados del modelo reducido en términos lineales ajustados mediante un polinomio cúbico a las observaciones) y b) la realizada por medio del modelo mecanicista.



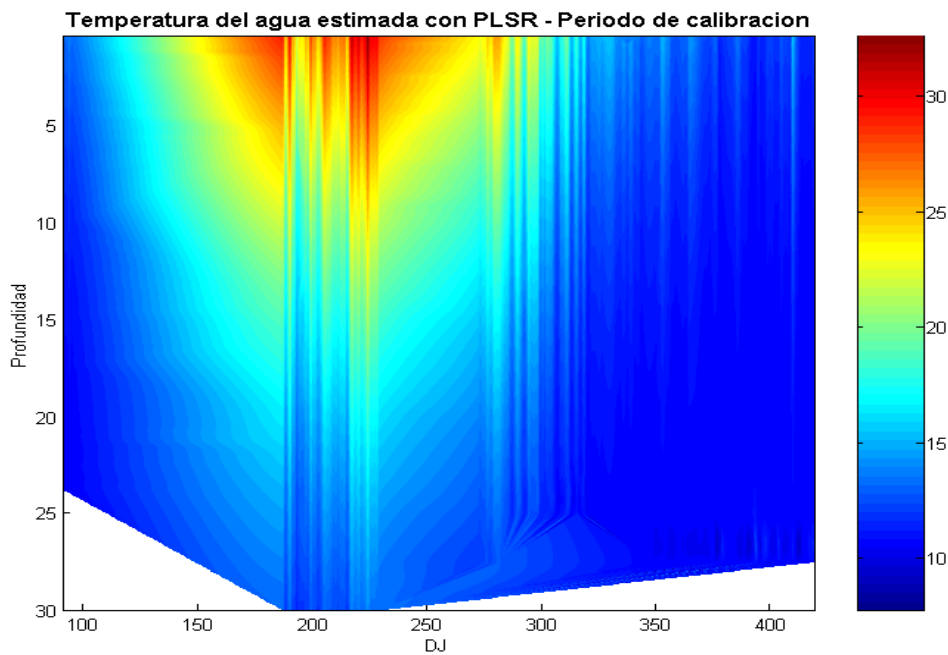
**FIGURA 121** Gráfico de temperatura del agua observada durante el periodo de calibración

Se puede observar que el principio de la primavera, la segunda mitad del otoño y el invierno representan periodos de mezcla térmica homogénea para el Embalse Amadorio. En los periodos restantes del año se presenta estratificación que puede ir de moderada para los meses de abril (la segunda parte de este mes), mayo y octubre; aparte de presentar intensidad fuerte desde el mes de junio hasta septiembre.



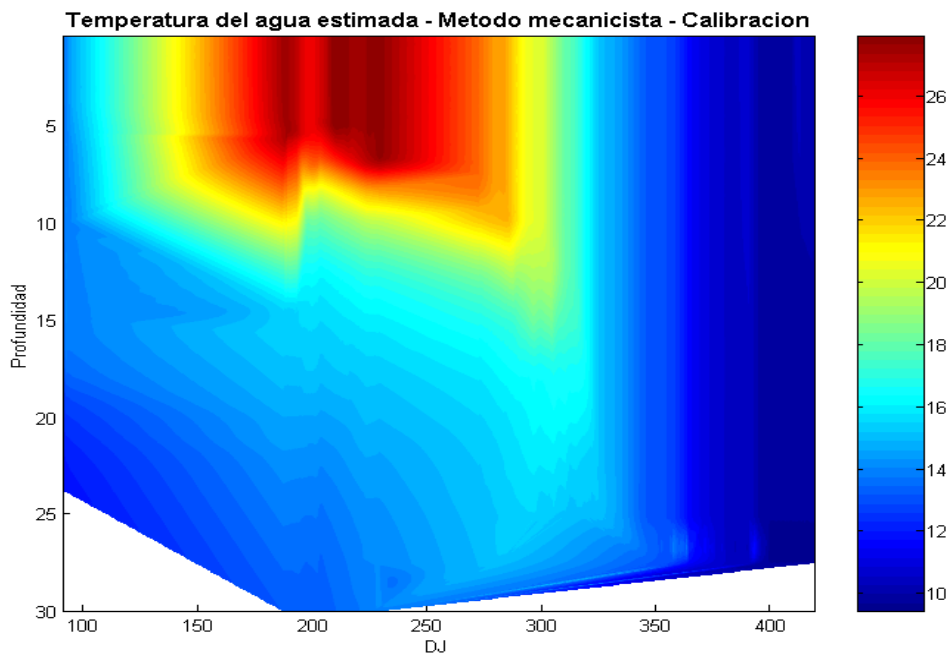
---

Asimismo, la mejor regresión realizada usando métodos multivariantes (ajuste con PLSR empleando términos lineales reducidos y ajuste cúbico a las observaciones) proporciona los valores estimados para calibración que se ven en la Figura (122):



**FIGURA 122** Gráfico de temperatura del agua estimada durante el periodo de calibración para el mejor ajuste efectuado usando el método PLSR

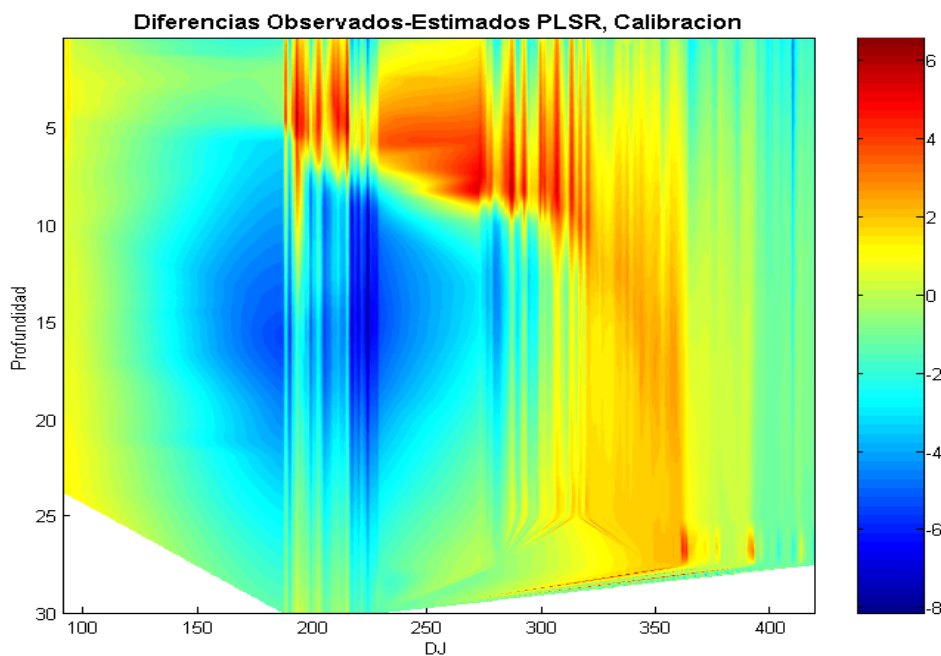
Y los resultados del modelo mecanicista propuesto se muestran en la Figura (123):



**FIGURA 123** Gráfico de temperatura del agua estimada durante el periodo de calibración usando el modelo mecanicista

---

Con el objetivo de evaluar el desempeño de cada una de las metodologías utilizadas se obtendrán gráficos de errores (valor observado – valor estimado), para saber su magnitud y la ubicación espacio temporal de los mismos.



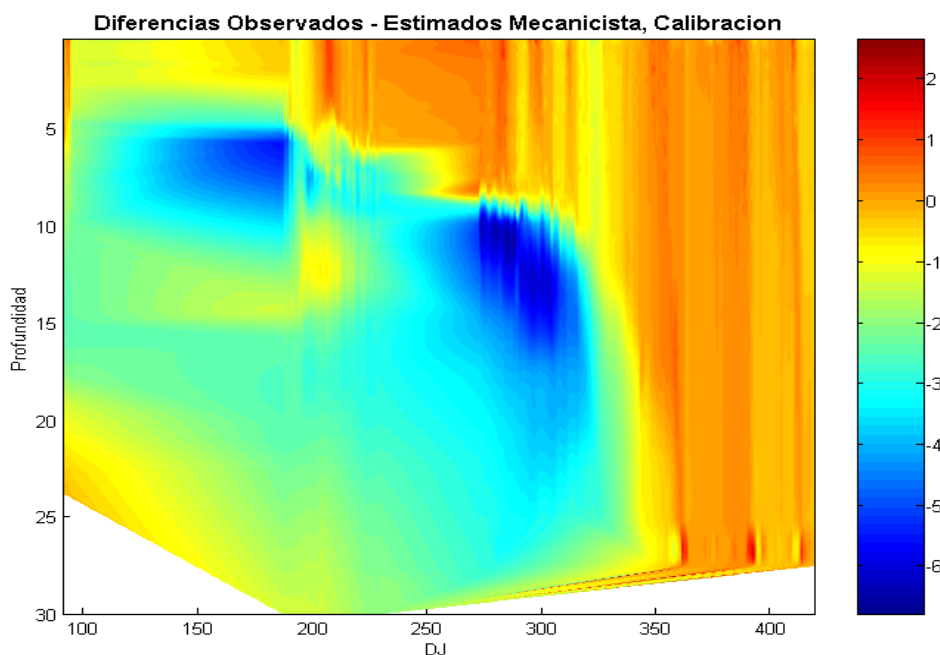
**FIGURA 124** Gráfico de diferencias entre valores observados de temperatura del agua y valores estimados usando el método PLSR durante el periodo de calibración

Se puede apreciar en la Figura (124) que la mayor parte del espacio de estimaciones presenta errores de predicción de valores muy bajos. Se pueden apreciar algunas subestimaciones en el valor predicho de magnitud apreciable (aproximadamente 4°C) a una profundidad de entre 0,5 y 8 metros para el periodo de tiempo en que se presenta la estratificación térmica intensa (desde junio hasta el mes de octubre, ambos inclusive). También se tienen varias sobre-estimaciones (de entre 3 y 5°C) a profundidades de entre 6 y 24 metros para el periodo de tiempo que va del mes de junio al mes de septiembre (coincidiendo con la estratificación térmica intensa).

Por lo tanto, se puede concluir que las estimaciones efectuadas con el modelo desarrollado de ajuste reducido con términos lineales usando el método PLSR y ajuste de los valores estimados a los observados usando un polinomio cúbico son mejores (el valor del error cometido es reducido) en los periodos de mezcla completa del embalse o cuando ocurre estratificación moderada; por otra parte, cuando la estratificación térmica es intensa el modelo tiende a subestimar o sobre-estimar en algunas regiones bien identificadas.

---

Los errores cometidos con el modelo mecanicista se presentan en la Figura (125):



**FIGURA 125** Gráfico de diferencias entre valores observados de temperatura del agua y valores estimados usando el método mecanicista durante el periodo de calibración

Al analizar el gráfico anterior se observa que la mayor parte de las estimaciones efectuadas poseen errores de predicción muy bajos. Sin embargo, se puede ver regiones en donde se subestima la temperatura del agua (aproximadamente entre 1 y 2°C), por ejemplo entre las profundidades de 0,5 y 8 metros para los meses de junio a octubre, ambos inclusive. Además, existen dos regiones claras en donde se sobreestima la temperatura (entre 3 y 6°C); una está a una profundidad de entre 5 y 10 metros y se presenta en los meses de abril a julio; la segunda se ubica a profundidades entre 9 y 25 metros durante los meses de agosto a noviembre.

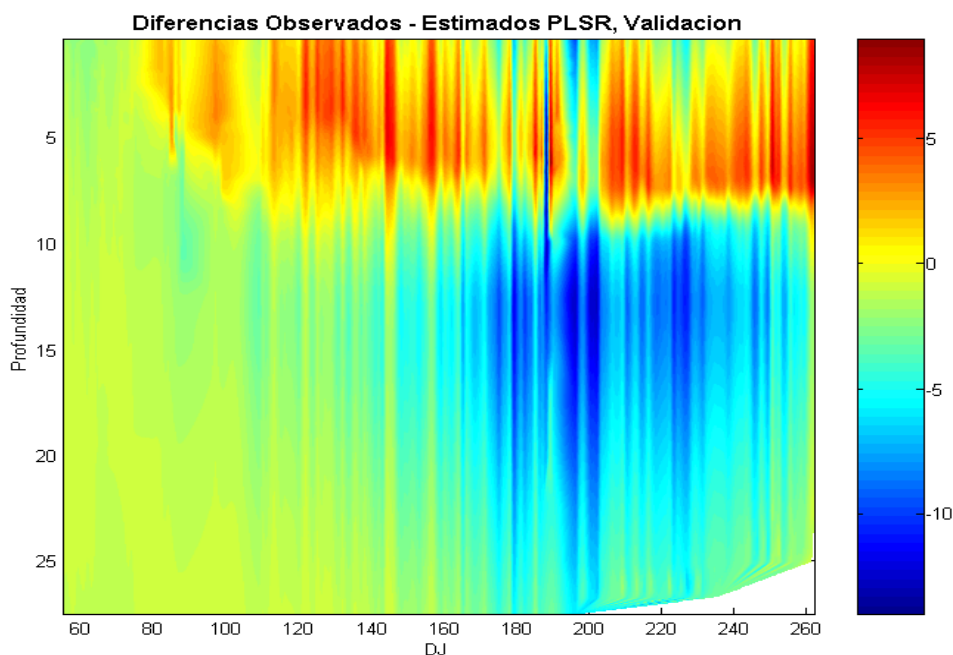
Se puede concluir que la presencia de la estratificación, la moderada y la intensa, afectan a la capacidad de predicción del modelo mecanicista a ciertas profundidades durante el periodo de calibración, pero en comparación con el modelo desarrollado usando la técnica PLSR, la extensión de las regiones de error significativo es menor al aplicar el modelo mecanicista.

En ambos métodos la estratificación genera que las estimaciones, en ciertas regiones espacio-temporales, tengan un error asociado mayor. Cuando el embalse se encuentra completamente mezclado ambos métodos son muy efectivos, representando de manera muy adecuada el perfil de temperaturas observado en el

---

Embalse Amadorio.

Respecto al periodo de validación los valores de temperaturas observadas del agua se muestran en la Figura (112), las temperaturas estimadas con PLSR se representan en la Figura (113) y las temperaturas estimadas con el método mecanicista se muestran en la Figura (120). Se obtendrán las diferencias (errores) entre las observaciones de temperatura y las estimaciones obtenidas con cada uno de los modelos.

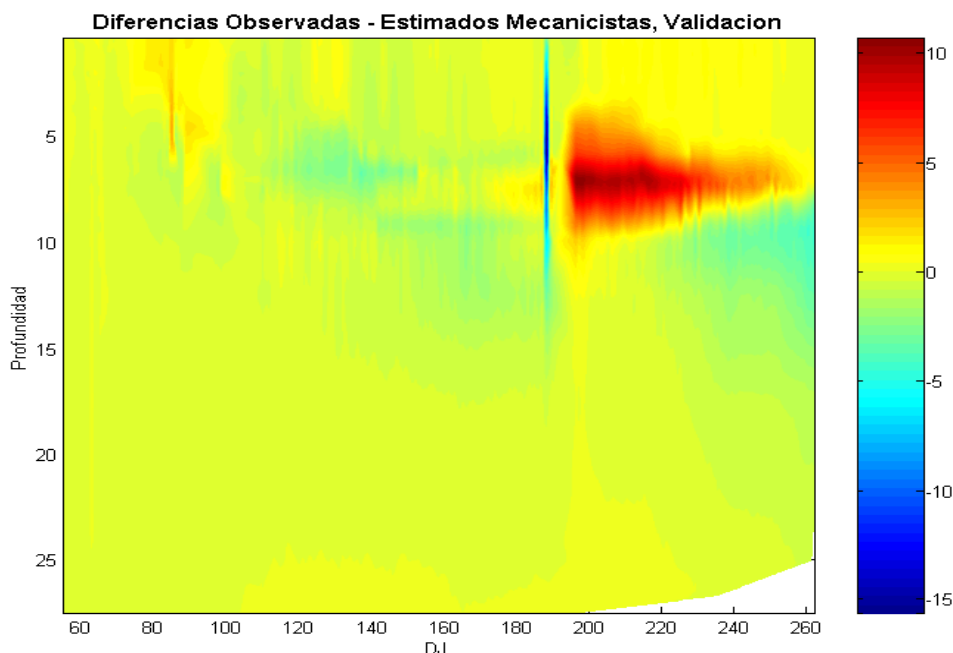


**FIGURA 126** Gráfico de diferencias entre valores observados de temperatura del agua y valores estimados usando el método PLSR durante el periodo de validación

De la Figura (126), que muestra las diferencias entre las observaciones y los valores estimados usando el modelo obtenido por medio del PLSR para el periodo de validación, se puede decir que la mayoría de las estimaciones presenta errores de predicción pequeños. Se observa una subestimación de los valores reales de alrededor de 5°C para la región comprendida entre 0 y 8 metros de profundidad durante los meses de abril a septiembre, ambos inclusive. También se tiene una sobre-estimación de los valores de temperatura del agua (de entre 5 y 10°C) para una región comprendida entre los 9 y los 25 metros de profundidad para la temporada comprendida por los meses de junio hasta septiembre, incluidos ambos. La estratificación térmica también en esta ocasión dificulta la estimación de los valores, siendo los errores de las estimaciones de los perfiles de temperatura mayores al ser más altas las temperaturas del medio ambiente.

---

El gráfico que muestra las diferencias entre los valores observados y los estimados por medio del modelo mecanicista durante el periodo de validación es el de la Figura (127):



**FIGURA 127** Gráfico de diferencias entre valores observados de temperatura del agua y valores estimados usando el modelo mecanicista durante el periodo de validación

De todos los gráficos de diferencias (errores) analizados hasta este punto se tiene que este (Figura 127), en donde se aplica el modelo mecanicista al periodo de validación, es el que tiene la mayor extensión espacio-temporal de estimaciones con errores de predicción muy pequeños. Las únicas excepciones son: una región que comprende de los 5 a los 8 metros de profundidad en el mes de mayo con errores de sobrestimación de entre 2 y 4°C; y una zona de los 4 a los 10 metros de profundidad de los meses de julio a septiembre con errores de subestimación de entre 6 y 10°C. La estratificación térmica también influye en este caso en la obtención de algunas estimaciones al generar diferencias con los valores observados pero afecta menos que en el caso de la aplicación del modelo obtenido aplicando el PLSR. Por tanto, para obtener estimaciones con el menor error de predicción asociado será más recomendable utilizar el modelo mecanicista.

---

---

---

## CAPÍTULO 10. CONCLUSIONES Y LINEAS FUTURAS

### 10.1 CONCLUSIONES

Es importante resaltar que todos los análisis realizados se pudieron efectuar gracias a la existencia de los datos necesarios y oportunos provenientes de la Confederación Hidrográfica del Júcar. Por lo tanto, es necesario mencionar que toda la información que se registraba durante el periodo de estudio, hidrológica, meteorológica y de calidad del agua se debe seguir recabando en el futuro. Aún más, si fuera posible aumentar el tipo de mediciones, por ejemplo sobre los caudales de entrada respecto a parámetros de calidad, como temperatura, oxígeno disuelto, pH, conductividad, etc. Esto con toda seguridad ayudará a generar inferencias de mayor precisión, además de conocer la dinámica de los constituyentes.

En esta tesis doctoral se desarrollaron dos planteamientos: en el primero se utilizaron varios métodos multivariantes para estudiar la evolución de los parámetros de calidad del agua: temperatura, oxígeno disuelto, ph y conductividad en perfil vertical y poder efectuar su estimación. El segundo planteamiento consiste en el desarrollo de un modelo mecanicista para simular la distribución de temperatura también en perfil vertical y tener la opción de realizar predicciones; este modelo fue programado en Visual Basic 2005.

A continuación se llevo a cabo una comparativa de ambos métodos enfocada en conocer aspectos de interés como son: tipo de información de entrada para cada método, ventajas e inconvenientes de aplicación de cada uno de ellos y precisión de las estimaciones efectuadas.

El proceso seguido durante este trabajo para el análisis utilizando métodos multivariantes descrito de forma breve fue el siguiente: se efectuó una exploración inicial de los datos disponibles por medio de estadística descriptiva para clasificar los registros con información válida y completa y se dividió el conjunto total en dos partes, una para calibración de las regresiones y la otra para validación, además se inspeccionó la normalidad de las variables; después se probaron varias transformaciones para las variables eligiéndose como la más prometedora a la transformación probit; luego se aplicó el análisis de componentes principales para detectar datos anómalos y comprobar la homogeneidad de los periodos de calibración y validación. A continuación se usó el análisis factorial para saber como se agrupan las variables (al aplicarles los distintos tratamientos) en factores y se obtuvieron puntuaciones factoriales para probar la normalidad en cada caso; después se empleó

---

---

el análisis cluster no jerárquico para formar conglomerados con base en distintos tratamientos de la información (variables estandarizadas, transformación probit, puntuaciones en componentes principales y puntuaciones factoriales) y a continuación se aplicó el análisis cluster jerárquico a las variables con tratamiento de estandarización y después de transformación probit, con el propósito de saber como se agrupan en los conglomerados. El análisis discriminante sirvió para verificar los resultados del análisis cluster.

El modelo lineal de regresión múltiple se aplicó con fines exploratorios a cada variable respuesta, para saber cuales son las variables con más influencia en la regresión y la posible capacidad predictiva de los modelos formulados para distintos tratamientos, determinándose que la regresión por medio del periodo de calibración completo proporciona mejores resultados que efectuándola por conglomerados. El análisis de correlación canónica permitió establecer relaciones entre el bloque de variables predictivas y el bloque de variables respuesta para todas las opciones de tratamiento planteadas demostrando que el mejor resultado se obtiene usando el periodo de calibración sin ser dividido y con las variables estandarizadas. Por último el PLSR hizo posible evaluar las dependencias de las variables respuesta con respecto a las variables predictivas usándose los distintos tratamientos de las variables, además permitió mejorar por medio de distintos planteamientos la calidad y parsimonia del ajuste de las regresiones para las variables de calidad del agua analizadas. En todas las aplicaciones de regresiones multivariantes se usaron conjuntos de datos independientes para efectuar calibración y validación; así se puede comprobar que las relaciones obtenidas son objetivas y no simple producto de la casualidad. También es interesante mencionar que dentro de las referencias consultadas no se encontraron casos de modelación de la calidad del agua en embalses usando técnicas de regresión multivariantes, por lo que esta aplicación efectuada para los datos disponibles en el Embalse Amadorio podría ser una de las primeras de este tipo.

También fue formulado un modelo mecanicista por capas horizontales el cual fue planteado por el método de diferencias finitas y resolviéndolo por el método de Thomas. Para poder desarrollar este método se efectuaron ajustes en los datos de la tabla calado-volumen-superficie; se propuso una metodología para estimar la cobertura de nubes que es un valor necesario para el desarrollo del modelo; se realizó un análisis de regresión para obtener los datos ausentes de la temperatura del agua en la superficie del embalse; además se procedió a la calibración de cuatro parámetros del modelo usando como valores observados los generados por la sonda



---

---

multiparamétrica.

Como un elemento de atención importante se efectuó un comparativo de modelación con PLSR y el método mecanicista, analizando aspectos como: el volumen de información necesaria para cada uno de ellos, la flexibilidad y adaptabilidad a los cambios en la escala de medida temporal, valoración de los resultados obtenidos en cada caso y efectividad en la obtención de perfiles y representación de capas en el interior del embalse.

Respecto a las técnicas utilizadas se emiten los siguientes comentarios:

El análisis de componentes principales es una herramienta muy útil y se puede aplicar para los siguientes propósitos:

- Para verificar que los datos provenientes de los periodos de calibración y validación son aproximadamente homogéneos y por tanto la regresión efectuada sobre el primero de ellos podrá ser aplicada sobre el segundo (la información pertenece a la misma población). Esta comprobación en conjunto con la revisión variable por variable de la coincidencia aproximada de los rangos, pueden asegurar que los resultados en un periodo son extrapolables al otro.
- La detección de datos anómalos. Al emplear los diagramas de dispersión de las componentes por pares esta localización se facilita enormemente y, en esta aplicación, se puede evaluar como muy efectiva.

Por otra parte el análisis factorial sirvió para detectar, aún antes de realizar cualquier tipo de regresión, que se podrían efectuar inferencias adecuadas para las variables Ta y Odis debido a contar con altas correlaciones con algunas variables meteorológicas e hidrológicas. Asimismo, las correlaciones más bajas tanto en cantidad como en calidad de las variables pH y Con preveían regresiones menos apropiadas.

Asimismo, el FA ayudó a determinar que la variabilidad explicada por los factores es prácticamente la misma cuando a las variables sólo se les aplica una estandarización univariante que al caso de aplicarles una transformación probit. Esto permitió determinar que en este estudio no era una ventaja realizar transformaciones de las variables.

Al aplicar el análisis cluster y el análisis discriminante se pudo apreciar que para una

---

---

misma ubicación en el espacio (el sitio donde se encuentra ubicada la sonda multiparamétrica en el embalse Amadorio) no proporciona buenos resultados al dividir los datos en conglomerados y efectuar regresiones por separado para cada uno de ellos. Estas técnicas han mostrado buenos comportamientos al clasificar distintos sitios (puntos de muestreo en diferentes ubicaciones) y efectuar tratamientos sobre los grupos obtenidos. Por lo tanto, sólo se recomienda cuando en el estudio se vayan a manejar ubicaciones espaciales distintas.

Más sin embargo, en los análisis donde se cuente con una cantidad mayor de datos (cuatro años de información o más), podría resultar interesante formar conglomerados por época del año (primavera, verano, etc.) y probar si una regresión para cada época del año proporciona mejores resultados que una efectuada sobre los datos sin clasificar.

Una manera rápida de saber si los datos elegidos como variables predictivas podrán, posiblemente, servir para efectuar predicciones sobre las variables respuesta es el empleo de los gráficos de caja y bigotes, pues permiten saber de manera sencilla si las variables utilizadas se encuentran aproximadamente dentro de los mismos rangos, pues esta es propiamente la condición de homogeneidad, una de las necesarias para poder efectuar inferencias aceptables.

El uso de la técnica de regresión lineal múltiple sirvió para varios propósitos. El primero de ellos fue conocer de manera preliminar valores iniciales de los parámetros de ajuste, como son el coeficiente de determinación y el error típico de predicción. También permitió tener información anticipada acerca de si el pronóstico de cada una de las variables respuesta iba a ser aceptable.

Además, permitió saber a detalle las consecuencias de dividir la información de calibración en conglomerados constituyendo una herramienta de análisis de regresión fundamental. Debido a todas las suposiciones iniciales que impone (Draper et al., 1998; Etxeberria, 1999; Gil Pascual, 2003; Pérez, 2005; Hultquist, 2006; Cuadras, 2007; Miket, 2007) es difícil que en estudios de este tipo la regresión lineal múltiple sirva para obtener las inferencias definitivas.

Respecto al análisis de correlación canónica sólo en pocos artículos de los consultados se hablaba acerca de la condición que impone este método de normalidad multivariante (Ouarda et al., 2001; Ouarda et al., 2008) (que no se cumple para la información que se tiene en este trabajo). A este respecto se considera que si los

---

---

resultados de validación son congruentes con los obtenidos durante la calibración el método, con sus reservas, sí es posible aplicarlo. Una forma de verificar que su uso es adecuado es por medio de las correlaciones canónicas, tanto en sus  $p$ -valores como en la magnitud de cada una de ellas (Gil Pascual, 2003). En este caso los valores obtenidos son satisfactorios.

Además, el CCA es un método que a pesar de exigir mucha preparación en su aplicación y una revisión extensa de aspectos teóricos, es muy utilizado tanto en hidrología como en meteorología (Ouarda et al., 2001; Gutiérrez et al., 2004; Ouarda et al., 2008). Por tanto, su uso es adecuado para investigaciones de estas características. Además, permite obtener ajustes aceptables utilizando menos variables predictivas.

De la misma forma, el CCA permite conocer relaciones y dependencias entre las variables por medio de la interpretación de las correlaciones canónicas (Calvo, 1993). Así, por ejemplo se pone de manifiesto que el aumento de energía térmica en el ambiente genera una mayor temperatura del agua a todos niveles y una disminución del oxígeno disuelto. El aumento de la temperatura del agua se explica fácilmente debido a la conducción del calor (contacto con el medio ambiente a mayor temperatura que el agua del embalse) y al incremento de la energía absorbida por radiación (tanto la radiación solar de onda corta como la radiación atmosférica de onda larga). La disminución del oxígeno disuelto al aumentar la temperatura ambiente se debe a que ésta produce un incremento en la temperatura del agua lo que genera una disminución de la solubilidad del oxígeno en el líquido, combinado con un aumento del metabolismo microbiano y de la demanda biológica de oxígeno.

Además, con la aplicación del CCA se obtuvo que a mayor profundidad los valores de conductividad aumentan. Esto pudiera estar relacionado con la distribución de sólidos disueltos debido a fuerzas gravitacionales, pues a mayor conductividad se tiene una mayor concentración de sólidos totales disueltos (TDS), lo que hace que la densidad del agua sea mayor. Por lo tanto, parece lógico que el agua con mayor conductividad este en el fondo y la que posee menor conductividad en la superficie del agua. En el caso de la variable pH no se obtuvieron relaciones claras con otras variables.

El método PLSR fue el que mejores resultados permitió obtener dentro de los métodos multivariantes de regresión. Además, esta técnica no requiere que las variables sigan una distribución normal (Zarzo, 2004). Para obtener los mejores resultados en cada una de las variables de calidad del agua se recomienda:

- 
- Primero obtener los coeficientes estandarizados utilizando todas las variables predictivas. Efectuar validación cruzada para obtener el número de componentes óptimo.
  - Usando el número de componentes obtenido en el paso anterior obtener nuevamente los coeficientes estandarizados.
  - Suprimir del proceso las variables que tengan coeficientes estandarizados muy bajos. Esto se puede efectuar de manera eficiente en el gráfico de coeficientes estandarizados.
  - Utilizar nuevamente el PLSR con las variables predictivas que hayan quedado, probando varias alternativas. Estas alternativas (grupos de variables) se probarán en el conjunto de datos de validación.
  - Probar el ajuste obtenido en el paso anterior con el conjunto de datos de validación para el grupo reducido de variables que se hubiera seleccionado, variando el número de componentes para obtener el mejor coeficiente de determinación. La regresión que tenga el  $R^2$  más alto para validación será la seleccionada. En este trabajo prácticamente en todos los casos el número de componentes seleccionado en este paso coincidió con el conseguido por validación cruzada.
  - Luego revisar el gráfico de predicciones-observaciones, para evaluar la calidad del ajuste.
  - A continuación tomar las predicciones y las observaciones de cada una de las variables y efectuar una regresión no lineal sobre ellas, probando varios modelos (polinomial, exponencial, logarítmico, logístico, etc.), hasta encontrar el mejor en cada caso. La única variable que mejoró sus predicciones con este planteamiento fue la temperatura del agua en el embalse ( $T_a$ ). Las otras tres variables tienen mejores ajustes usando solamente expresiones lineales.

Los procesos que efectúa el modelo mecanicista para determinar la temperatura del agua a distintos niveles de profundidad a una escala temporal diaria son los siguientes: caracterización del embalse como un conjunto de capas horizontales de número y espesor variable según las condiciones existentes; mezcla entre capas a consecuencia de la flotación; unión de capas que posean un espesor por debajo de un espesor mínimo especificado; división de capas que tengan un espesor mayor al especificado; cálculo del coeficiente de difusión turbulenta entre las capas del embalse; ubicación en la capa que corresponda de los caudales que ingresen al embalse; determinación de los caudales a extraer de cada una de las capas debido a la extracción por tomas y desagüe del vertedor; calcular los flujos de calor en las

capas del almacenamiento, estos flujos son debidos a la radiación solar neta de onda corta, radiación atmosférica neta de onda larga, radiación de onda larga que sale del agua, conducción y evaporación; modificaciones en las capas debido a precipitación, evaporación e infiltración; cálculo de la temperatura media en cada una de las capas por medio del método de diferencias finitas y el algoritmo de Thomas. Todos estos procesos están desarrollados en un código escrito especialmente para el desarrollo de esta tesis, en Visual Basic 2005.

El modelo mecanicista construido exhibe buenos valores estimados, tanto en la calibración como en la validación, así como perfiles de temperatura muy similares a los reales, lo que hace confiar que los fenómenos simulados incluidos configuran en gran medida una representación fiel de las relaciones físicas que se presentan al interior del almacenamiento.

Además, de la comparación de los gráficos de valores observados y estimaciones efectuadas con el modelo mecanicista se tiene que la mayoría de las ocasiones el error cometido al estimar es de valor reducido y las regiones y periodos de tiempo donde el error se vuelve significativo están bien identificadas.

Los coeficientes de determinación y los gráficos correspondientes al periodo de validación obtenidos con la mejor simulación realizada con PLSR hacen concluir que sólo son aceptables las regresiones obtenidas para las variables temperatura del agua y oxígeno disuelto:

Periodo	Variable respuesta	Coefficiente de correlación, R	Coefficiente de determinación, R <sup>2</sup>	Error estándar de estimación
Validación	Ta	0,8617	0,7426	3,54
	Odis	0,8655	0,7491	2,32

**TABLA 119** Evaluación de los mejores ajustes obtenidos con PLSR para las variables Ta y Odis

Como una aclaración, los coeficientes de determinación fueron calculados en base a valores diarios (cada 24 horas) para cada profundidad correspondiente, por lo que se evaluó la calidad de las estimaciones hechas a escala temporal diaria y a escala espacial unidimensional (en la dirección vertical). Estas escalas utilizadas para el cálculo del coeficiente de determinación fueron empleadas tanto en las regresiones multivariantes como en el modelo mecanicista.

Se observa que son mejores los resultados de temperatura del agua obtenidos

---

---

mediante el modelo mecanicista (Tabla 118). Debido a esto, este modelo en particular puede servir para simular propuestas de gestión, como serían caudales de extracción, caudales que ingresan al embalse, nivel de la toma o de las tomas y modificaciones en el volumen objetivo a mantener en el embalse.

Una de las posibles causas que provocan las diferencias entre los valores observados y simulados es que no se conocen de manera constante las concentraciones de las variables de calidad del agua en los caudales de entrada al Embalse Amadorio. Este hecho afecta tanto la aplicación de los métodos de regresión multivariantes como al modelo mecanicista. A pesar que puede existir la imposibilidad práctica de medir las concentraciones en cada uno de los volúmenes de agua que ingresen al embalse, se debe promover que al menos sean monitoreados de manera constante los caudales más importantes. Esto podría permitir mejorar los resultados de los modelos.

En el comparativo que se efectúa en este trabajo se aprecia que los parámetros de ajuste son mejores para el modelo mecanicista comparándolos con el mejor ajuste efectuado con el PLSR y se resaltan las ventajas y desventajas de cada uno de estos métodos para modelar la temperatura del agua en el embalse, que de manera breve son:

- La metodología basada en PLSR presenta una mayor rapidez de cálculo en comparación con el modelo mecanicista.
- Se requiere tener una mayor cantidad de variables medidas para aplicar el modelo mecanicista en comparación con el basado en PLSR, que requiere menos variables.
- Para poder desarrollar la técnica del PLSR se requiere contar con una gran cantidad de registros (de varios cientos a miles como mínimo), mientras que el modelo mecanicista puede proporcionar resultados aceptables si la cantidad de registros es menor.
- En general, será más complicado operar con el modelo mecanicista en comparación con el formulado con el PLSR.
- El modelo mecanicista tiene una gran capacidad de generalización a otros sitios (embalses) en comparación con las expresiones específicas obtenidas mediante el PLSR.
- Es posible que el modelo mecanicista puede usarse a un plazo de tiempo más amplio para efectuar estimaciones en comparación con las expresiones obtenidas con el PLSR. En ambos modelos es necesario realizar más

---

investigación para determinar los horizontes de predicción y sus niveles de confiabilidad.

- Los resultados obtenidos con el modelo mecanicista fueron, en este caso, mejores que los obtenidos con el método basado en PLSR.
- Es importante resaltar que el uso conjunto de ambas técnicas puede ayudar a comprender la evolución de la variable analizada, así como a tener una mayor seguridad en las predicciones realizadas.

Por otra parte, las aportaciones que se hacen en este trabajo son:

- Se ensayaron y aplicaron técnicas, como el análisis de componentes principales y el análisis factorial que sirvieron para verificar la hipótesis de homogeneidad entre los grupos de datos propuestos para calibración y validación.
- Se mostró que ciertos parámetros meteorológicos y de estado de un embalse tienen correlación e influencia directa sobre algunas variables de calidad del agua, además se usaron estas correlaciones con fines de predicción e interpretación.
- Como una alternativa a los métodos convencionales de modelación en una dimensión de parámetros de calidad del agua en un embalse se probó el análisis de correlación canónica, mostrando que su uso es factible.
- Se usó el PLSR para modelar en una dimensión variables de calidad del agua en un almacenamiento, demostrando que al ser usado para pronosticar, los resultados que proporciona son mejores a los obtenidos con algunos métodos de regresión.
- Se propuso un modelo en diferencias finitas para determinar la temperatura del agua en perfil vertical, muy flexible, que permite determinar la exactitud de las estimaciones efectuadas. Este modelo efectúa numerosas operaciones utilizando información básica y toma en cuenta prácticamente todos los factores que inciden sobre la temperatura del agua al interior del almacenamiento.
- Se propone un método para calcular la cobertura de nubes para cada día de simulación. Esto se efectuó así debido a que la estación de Villajoyosa, al ser una estación automática, no registra esta información. Según información de los investigadores Deas y Lowney (2000) en las estaciones completamente automatizadas la cobertura de nubes es un dato que no se registra. Por lo tanto, al mostrar el método empleado para obtener la nubosidad un buen

---

comportamiento global, se sugiere utilizarlo en otros modelos (calidad del agua, meteorológicos, del medio físico en general) que requieran como entrada esta información.

- La comparación del desempeño de métodos multivariantes y el modelo mecanicista. La aplicación del modelo mecanicista para determinar la temperatura del agua en el embalse permitió evaluar por comparación el desempeño de las técnicas multivariantes en relación con éste, mostrando su grado de efectividad y señalando las ventajas y desventajas existentes al aplicar cada uno de los planteamientos.

## 10.2 APLICACIONES Y LÍNEAS FUTURAS

Antes de precisar algunas de las posibles aplicaciones de los ajustes conseguidos en donde se tiene una mejor predicción, caso de la temperatura del agua y el oxígeno disuelto, se mencionarán los motivos debido a los cuales se considera que estas dos cantidades son las más importantes en lo que se refiere a las variables de calidad del agua.

Para la temperatura del agua, además de los numerosos aspectos ya mencionados en el Capítulo 1, se tiene que su distribución en perfil interviene directamente en el fenómeno de convección (debido a la flotación), así como en la estratificación, determinando así la distribución de los otros parámetros de calidad del agua (Environmental Laboratory, 1995; Deas et al., 2000).

Para el oxígeno disuelto, además de todos los puntos incluidos en el Capítulo 1, este parámetro es fundamental en el logro de los objetivos medioambientales (Environmental Laboratory, 1995) y para suministro de agua para consumo humano.

Los ajustes obtenidos para la temperatura del agua y el oxígeno disuelto pueden tener, entre otras, las siguientes aplicaciones:

- Predicción a distintos niveles de los valores futuros de  $T_a$  y  $O_{dis}$ .
- Simulación de los valores de estos dos parámetros al operar las tomas existentes en la presa.
- Simulación de la operación de propuesta de tomas multinivel.
- Predicción de la calidad de los desfuegos que pasen por el vertedor de las demasías hacia aguas abajo.
- Suministro de datos de entrada para modelos ecológicos de tipo diverso



---

(conservación de especies de interés en el propio embalse o río abajo, etc.).

- Suministro de datos de entrada para modelos económicos que tomen como parámetros para asignar valor económico al agua la temperatura y el oxígeno disuelto, por ser las medidas de calidad más importantes.
- Análisis de los factores que más influyen en los valores y distribución de estos dos parámetros de calidad del agua.
- Establecimiento de metodologías a seguir para estimar otros parámetros de interés en el almacenamiento.
- Reposición de valores perdidos al no efectuarse el muestreo por causas diversas.
- Análisis estacional del comportamiento de estos parámetros de calidad del agua.

Las posibles líneas futuras que se pueden seguir al revisar la información empleada y el seguimiento del trabajo desarrollado hasta el momento son:

- Actualmente existen en el mercado sondas multiparamétricas capaces de medir otros parámetros, además de los ya mencionados, como son: salinidad, turbiedad, clorofila, amonio, nitratos, cloruros, etc. Por lo tanto, sería interesante proponer la ampliación de la sonda actual para que pueda medir mayor cantidad de parámetros a distintas profundidades. Además, sería muy adecuado que se instalarán sondas de este tipo en otros embalses. Para la inferencia a futuro de la calidad una de las propuestas a seguir sería la aplicación del CCA y el PLSR.
- Una posible mejora a los planteamientos presentados aquí sería esta: para que el método sea más efectivo se requiere que la predicción de un dato futuro esté basada principalmente en los últimos valores del parámetro de calidad y que tome menos en cuenta a los datos más “antiguos”. Esto se puede hacer ponderando los datos (Dayal et al., 1997). Por ejemplo, si se van a efectuar pronósticos del mes de diciembre de un año, los datos de noviembre se repiten 11 veces en la base de datos, los de octubre 10 veces, para septiembre sólo 9 veces y se sigue así hasta terminar en el mes de enero (sólo aparecen una vez). Es necesario hacer muchas pruebas con el número de meses previos a incluir y los coeficientes de ponderación más adecuados. Además, se considera después de trabajar con los modelos, que es necesario contar con información continua y que abarque un periodo de al menos cuatro años.
- Con el objetivo de lograr una mejor gestión del embalse es necesario seguir

---

registrando información de la sonda y efectuar predicciones de calidad tanto con las técnicas multivariantes utilizadas aquí (CCA y PLSR) y sus posibles modificaciones, como con otras técnicas alternativas: series temporales univariantes, series temporales multivariantes, redes neuronales artificiales, lógica difusa, modelos de regresión jerárquicos multinivel, modelación gráfica, algoritmo genético, redes bayesianas, etc. Así, se podrán evaluar los distintos desempeños de los métodos y emplear el que mejores resultados reporte para un uso en concreto.

Para poder aplicar series temporales es necesario que las lecturas de las variables de calidad del agua se efectúen siempre a los mismos niveles y a las mismas horas cada día. Para que el análisis de una serie temporal lleve a buenos resultados no basta con utilizar las técnicas adecuadas, sino que es imprescindible que esos datos sean confrontables y por tanto deben ser homogéneos. Si se cambia continuamente la metodología de observación el resultado será una serie temporal conformada por un grupo de registros no comparables porque son muy heterogéneos ([www.eumed.net/cursecon/libreria/2004/jsf/4.pdf](http://www.eumed.net/cursecon/libreria/2004/jsf/4.pdf)). Así, al contar con un conjunto de observaciones registradas de manera uniforme, es posible que la aplicación de las series temporales sea efectiva.

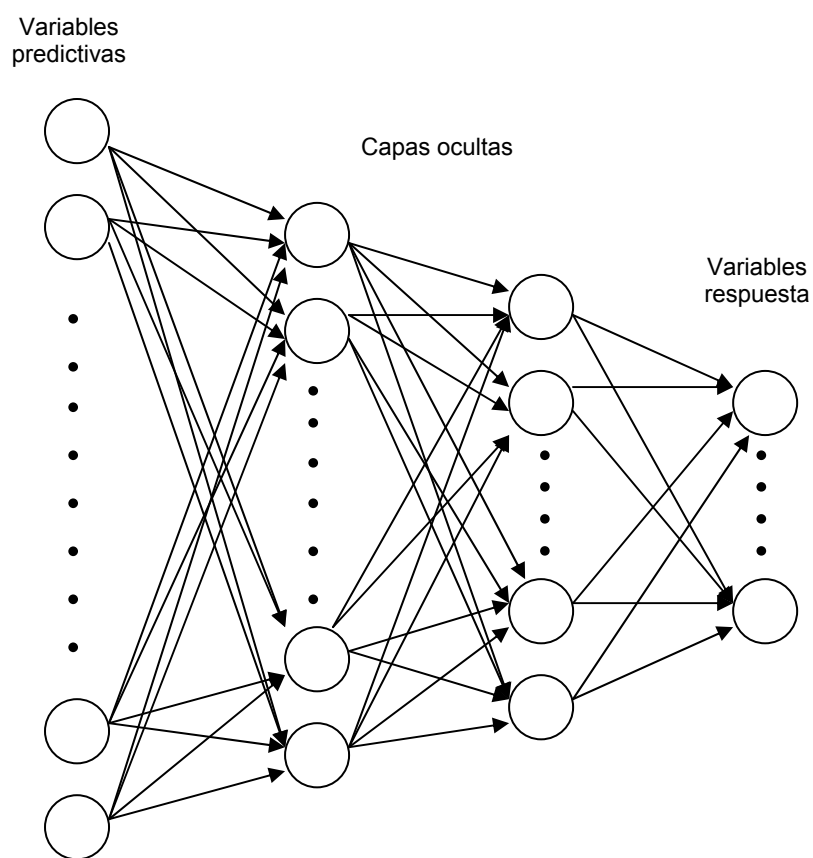
Si se cuenta con datos en intervalos regulares de tiempo y se desea emplear la posible "inercia" en el comportamiento de la serie para pronosticar sus valores futuros, esta clase de análisis se llama univariante debido a que usa únicamente la información de la propia historia de la serie, con fundamento en la hipótesis de que las condiciones en el futuro serán análogas a las que se observaron en el pasado. Los modelos univariantes están especialmente indicados para la estimación a corto plazo (Peña, 1989).

En la estimación a mediano y largo plazo se debe tener en cuenta la evolución de otras variables (en este caso las meteorológicas y las hidrológicas), relacionándolas con la que se busca estimar mediante un modelo de regresión dinámica (serie temporal multivariante) o modelo de función de transferencia. Este tipo de modelos permiten medir cómo se traspasan los efectos entre variables (Peña, 1989).

Respecto al uso de una red neuronal, puede utilizarse una red multicapa (por ejemplo el perceptron multicapa) ([es.wikipedia.org/wiki/](http://es.wikipedia.org/wiki/)). El perceptron

---

multicapa es un tipo de red neuronal con conexiones hacia delante (Figura 128). Distintos autores han demostrado independientemente que el perceptron multicapa es un aproximador universal, lo que quiere decir que cualquier función continua puede aproximarse por medio de él. La capacidad del perceptron multicapa para aprender en base a un conjunto de ejemplos, aproximar funciones no lineales, filtrar ruido de los datos, etc., hace que sea un modelo adecuado para tratar con problemas reales. Dentro del campo de las redes neuronales el perceptron multicapa es actualmente una de las configuraciones más empleadas para resolver problemas (Isasi et al., 2004).



**FIGURA 128** Esquema de una red neuronal artificial para estimar las variables de calidad del agua en un embalse

Las neuronas de la capa de entrada se encargan exclusivamente de recibir los patrones que llegan del exterior y propagan dichos patrones a todas las neuronas de la siguiente capa. La última capa sirve como salida de la red, entregando al exterior la respuesta de la red para cada uno de los patrones de entrada. Las neuronas de las capas ocultas realizan un procesamiento no lineal de los patrones recibidos (Figura 128) (Isasi et al., 2004). En el perceptron multicapa a proponer la capa de entrada contendrá un número de neuronas

---

que represente a las variables predictivas, un cierto número de capas ocultas a determinar y una capa de salida que contenga tantas neuronas como variables de calidad del agua (variables respuesta) se vayan a estimar.

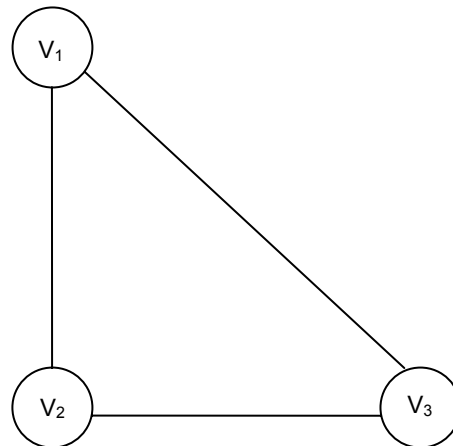
La lógica difusa fue primeramente desarrollada por Zadeh a mediados de la década de 1960 para representar incertidumbre y conocimiento impreciso; esta provee medios aproximados pero efectivos para describir el comportamiento de sistemas que son demasiado complejos, mal definidos o no son fáciles de analizar matemáticamente. Este tipo de lógica emplea enunciados de clasificación en lugar de opciones de verdadero o falso (Jain et al., 1999). La lógica difusa (también llamada lógica borrosa) se adapta especialmente bien a valores poco precisos, como por ejemplo un “gran” caudal de entrada o “poco” volumen almacenado, estratificación “fuerte” o “débil”. Por medio de sus algoritmos podría ser factible encontrar como modificar un modelo para que se fuera adaptando a condiciones cambiantes durante el periodo de estudio, como por ejemplo la estratificación o el cambio de condiciones climatológicas ([es.wikipedia.org/wiki/](http://es.wikipedia.org/wiki/)). Los valores de las variables predictivas se pueden clasificar en intervalos por medio de funciones definidas especialmente, lo que le puede proporcionar flexibilidad al modelo. Se puede conjuntar su uso con el de redes neuronales.

Un modelo de regresión jerárquica multinivel puede ser desarrollado para determinar algunas variables de calidad del agua. Para cada profundidad seleccionada en el embalse se plantea un modelo de regresión y luego se busca la inter-relación entre ellos (por medio de la variación de sus coeficientes). La característica que distingue a los modelos multinivel de los de regresión clásica es que modelan la variación entre grupos (para este caso los conjuntos de datos a distintas profundidades) (Gelman et al., 2007).

La modelación gráfica proporciona una descripción condensada de las interacciones dominantes entre las variables observadas originales, dado un conjunto de datos provenientes de una muestra. La técnica está estrechamente relacionada con el análisis de regresión múltiple. El objetivo de la modelación gráfica es alcanzar una descripción simplificada de las interrelaciones entre variables por medio de pequeñas modificaciones de los elementos de la matriz de covarianza muestral **S**. En un modelo gráfico cada variable es representada mediante un vértice y cualquier restricción de independencia condicional para dos variables es representada por la ausencia de una línea que directamente

---

las conecte (Figura 129). Se deben probar varias configuraciones de las variables respuesta con el conjunto de las variables predictivas. El ajuste del modelo a los datos puede ser formulado como una estimación de máxima verosimilitud de los elementos de la matriz de correlación inversa que se correspondan con las líneas existentes en el modelo gráfico (Callies, 2005).



**FIGURA 129** Modelo gráfico para tres variables (Fuente: Callies, 2005)

Los algoritmos genéticos pueden ser clasificados como algoritmos evolutivos de búsqueda aleatoria guiada que usan la probabilidad para conducir la exploración. Los algoritmos genéticos son creados por analogía con los procesos en la reproducción de los organismos biológicos. Algunos de los procesos más importantes son los de cruce y mutación en cromosomas y genes. Los “mejores” organismos son aquellos que tienen mayor oportunidad que los “peores” de sobrevivir y producir una nueva generación (Jain et al., 1999). Los algoritmos genéticos emplean procesos estocásticos, pero la solución obtenida no es en definitiva aleatoria. La capacidad de los algoritmos genéticos para dirigirse hacia las porciones más prometedoras de un espacio de soluciones proviene de su habilidad para intercambiar cadenas portadoras de soluciones parciales (Pazos Sierra, 1996). El problema de calcular una variable de calidad del agua debe ser parametrizado en un conjunto de variables codificadas en un cromosoma y buscar por medio del algoritmo genético la mejor solución.

Una red bayesiana consiste de un conjunto de variables y un conjunto de arcos dirigidos entre variables. Matemáticamente la estructura es llamada un gráfico dirigido (Jensen, 2002), donde los nodos representan variables que pueden ser de cualquier tipo y los arcos que los unen codifican dependencias

---

condicionales entre las variables ([es.wikipedia.org/wiki/](http://es.wikipedia.org/wiki/)). Si hay un arco que une A con B, se dice que B es un hijo de A y A es un padre de B (Jensen, 2002). El cálculo de las probabilidades se realiza con el Teorema de Bayes ([es.wikipedia.org/wiki/](http://es.wikipedia.org/wiki/)). En este caso, se deben definir las variables predictivas para cada variable respuesta, evaluando la posibilidad de tener variables instrumentales y del conjunto de variables respuesta observadas obtener las probabilidades de ocurrencia.

- El modelo en diferencias finitas para múltiples capas propuesto aquí puede servir de base para el desarrollo de nuevos modelos que determinen otros componentes de calidad del agua, operación del embalse, etc. Debido a que el modelo mecanicista propuesto simula los procesos de entradas y salidas de caudales del embalse y el movimiento del agua en su interior generado por causas relacionadas con la temperatura, puede extenderse a otros componentes de calidad del agua.
- Los más modernos desarrollos actuales se están efectuando en torno al “acoplamiento de modelos”. Así, las técnicas propuestas aquí pudieran integrarse con un modelo climático de alta resolución (por ejemplo el PROMES) y con una técnica de estimación de la demanda hídrica futura en el entorno del Embalse Amadorio. Las salidas del modelo climático y del modelo de balance hídrico se constituirían en las entradas de datos del modelo de calidad del agua. De esta forma se podría contar con estimaciones de la evolución del nivel en el almacenamiento, demanda a suministrar y perfil de calidad. El acoplamiento también pudiera efectuarse con modelos ecológicos, de gestión, económicos, etc.

---

---

## CAPÍTULO 11. REFERENCIAS

Abdi, H. 2003a. "Partial least squares regression (PLS-regression)". In M. Lewis-Beck, A. Bryman, T. Futing (Eds): Encyclopedia for research methods for the social sciences. Thousand Oaks.

Abdi, H. 2003b. "Multivariate Analysis". In M. Lewis-Beck, A. Bryman, T. Futing (Eds): Encyclopedia for research methods for the social sciences. Thousand Oaks.

Abdi, H. 2007. "Partial least square regression (PLS-regression)". In Neil Salkind (Ed.): Encyclopedia of Measurement and Statistics. Thousand Oaks.

Adler, F.J. 2005. "Los embalses y los recursos hídricos superficiales". Cet.

AENOR. 2005. "Calidad del agua". Asociación Española de Normalización y Certificación.

Aluja Banet, T.; Morineau, A. 1999. "Aprender de los datos: el análisis de componentes principales". Ediciones Universitarias de Barcelona.

Angelini, R.; Petreire Jr, M. 2000. "A model for the plankton system of the Broa reservoir, São Carlos, Brazil". Ecological Modelling, Volume 126, Issues 2-3. Elsevier.

Antonopoulos, V. Z.; Gianniou, S. K. 2003. "Simulation of water temperature and dissolved oxygen distribution in Lake Vegoritis, Greece". Ecological Modelling, Volume 160, Issues 1-2. Elsevier.

Arhonditsis, G. B.; Stow, C. A.; Steinberg, L. J.; Kenney, M. A.; Lathrop, R. C.; McBride, S. J.; Reckhow, K. H. 2006. "Exploring ecological patterns with structural equation modelling and Bayesian analysis". Ecological Modelling, Volume 192, Issues 3-4. Elsevier.

Armengol, J. 2000. "Análisis y valoración de los embalses como ecosistemas". Conferencia Iberdrola 2000. Universidad de Barcelona.

Atanasova, N.; Todorovski, L.; Dzeroski, S.; Kompare, B. 2006. "Constructing a library of domain knowledge for automated modelling of aquatic ecosystems". Ecological Modelling, Volume 194, Elsevier.

---

Azevedo Neto, J. M.; Acosta Álvarez, G. 1992. "Manual de hidráulica". Editorial Harla.

Baffi, G.; Martin, E. B.; Morris, A. J. 1999. "Non-linear projection to latent structures revisited (the neural network PLS algorithm)". *Computers & Chemical Engineering* Volume 23, Issue 9. Elsevier.

Balabin, R. M.; Safieva, R. Z.; Lomakina, E. I. 2007. "Comparison of linear and nonlinear calibration models base on near infrared (NIR) spectroscopy data for gasoline properties prediction". *Chemometrics and intelligent laboratory systems* Volume 88, Issue 2. Elsevier.

Bartels, J. H. M.; Janse, T. A. H. M.; Pijpers, F. W. 1985. "Classification of the quality of surface waters by means of pattern recognition". *Analytica Chimica Acta*, Volume 177. Elsevier Science Publishers B. V., Amsterdam.

Bartholomew, D. J. 2006. "Latent-Variable modeling". *Encyclopedia of Statistical Sciences*. Standard article. John Wiley & Sons, Inc.

Baumann, K.; Albert, H.; Von Korff, M. 2002. "A systematic evaluation of the benefits and hazards of variable selection in latent variable regression. Part I. Search algorithm, theory and simulations". *Journal of Chemometrics*, Volume 16, Issue 7. Wiley InterScience.

Berglund, A.; Kettaneh, N.; Uppgård, L. L.; Wold, S.; Bendwell, N.; Cameron, D. R. 2001. "The GIFI approach to non-linear PLS modelling". *Journal of Chemometrics*, Volume 15, Issue 4. John Wiley & Sons, Ltd.

Berglund, A.; Wold, S. 1997. "INLR, Implicit non-linear latent variable regression". *Journal of Chemometrics*, Volume 11, Issue 2. John Wiley & Sons, Ltd.

Blanco, M.; Coello, J.; Iturriaga, H.; Maspoch, S.; Redón, M. 1995. "Partial least-squares regression for multicomponent kinetic determinations in linear and non-linear systems". *Analytica Chimica Acta*, Volume 303, Issues 2-3. Elsevier.

Bowie, G.L., et al. 1985. "Rates, constants, and kinetics formulations in surface water quality modeling". Environmental Research Laboratory. U. S. Environmental Protection Agency.

Bro, R.; Smilde, A. K. 2003. "Centering and scaling in component analysis". *Journal of Chemometrics*, Volume 17, Issue 1. Wiley InterScience.

Brodnjak-Vončina, D.; Dobčnik, D.; Novič, M.; Zupan, J. 2002. "Chemometrics characterisation



---

---

of the quality of river water". *Analytica Chimica Acta*, Volume 462, Issue 1. Elsevier.

Browne, M. W. 2006. "Rotation techniques". *Encyclopedia of Statistical Sciences*. Standard article. John Wiley & Sons, Inc.

Burnham, A. J.; MacGregor, J. F.; Viveros, R. 1996. "Frameworks for latent variable multivariate regression". *Journal of Chemometrics*, Volume 10, Issue 1. John Wiley & Sons, Ltd.

Burnham, A. J.; MacGregor, J. F.; Viveros, R. 2001. "Interpretation of regression coefficients under a latent variable regression model". *Journal of Chemometrics*, Volume 15, Issue 4. John Wiley & Sons, Ltd.

Callies, U. 2005. "Interaction structures analysed from water-quality data". *Ecological Modelling*, Volume 187, Issue 4. Elsevier.

Calvo Gómez, F. 1993. "Técnicas estadísticas multivariantes". Universidad de Deusto, Bilbao.

Çamdevýren, H.; Demýr, N.; Kanik, A.; Keskýn, S. 2005. "Use of principal component scores in multiple linear regression models for prediction of *Chlorophyll-a* in reservoirs". *Ecological Modelling*, Volume 181, Issue 4. Elsevier.

Carbonell, E.; Denis, J. B.; Calvo, R.; González, F.; Pruñonosa, V. 1983. "Regresión lineal. Un enfoque conceptual y práctico". Ministerio de Agricultura, Pesca y Alimentación. Instituto Nacional de Investigaciones Agrarias.

Castañé, P. M.; Rovedatti, M. G.; Topalián, M. L.; Salibián, A. 2006. "Spatial and temporal trends of physicochemical parameters in the water of the Reconquista River (Buenos Aires, Argentina)". *Environmental Monitoring and Assessment*, Volume 117, Numbers 1-3. Springer.

Centre for Water Research. 2007. "Dynamic Reservoir Simulation Model. Science Manual". University of Western Australia.

Chang, E. K. M. 2007. "Assessing the increasing trend in Northern Hemisphere winter storm track activity using surface ship observations and a statistical storm track model". *Journal of Climate*. American Meteorological Society.

Chapra, S. C. 1997. "Surface water-quality modeling". WCB McGraw-Hill.

Chapra, S.;Pelletier, G. 2003. "QUAL2K: A modeling framework for simulating river and stream

---

---

water quality: Documentation and users manual". Civil and Environmental Engineering Dept., Tufts University, Medford, MA.

Chase, W.; Bown, F. 2000. "General statistics". Fourth edition. John Wiley & Sons, Inc.

Chen, T.; Morris, J.; Martin, E. 2007. "Gaussian process regression for multivariate spectroscopic calibration". *Chemometrics and Intelligent Laboratory Systems*, Volume 87, Issue 1. Elsevier.

Cheng, B.; Wu, X. 2006. "An modified PLSR method in prediction". *Journal of Data Science*, Volume 4, Number 3.

Confederación Hidrográfica del Júcar. 1999. "Plan Hidrológico del Júcar". Página electrónica de la CHJ.

Cuadras, C. M. 2007. "Nuevos métodos de análisis multivariante". CMC Editions. Propiedad del autor.

Daszykowski, M.; Serneels, S.; Kaczmarek, K.; Van Espen, P.; Croux, Ch.; Walczak, B. 2007. "TOMCAT: A MATLAB toolbox for multivariate calibration techniques". *Chemometrics and Intelligent Laboratory Systems*, Volume 85, Issue 2. Elsevier.

Davies, A. M. C. 2001. "Uncertainty testing in PLS regression". *Spectroscopy Europe* 13/2.

Dayal, B. S.; MacGregor, J. F. 1997. "Improved PLS algorithms". *Journal of Chemometrics*, Volume 11, Issue 1. John Wiley & Sons, Ltd.

De Carlo, E. H.; Beltran, V. L.; Tomlinson, M. S. 2004. "Composition of water and suspended sediment in streams of urbanized subtropical watersheds in Hawaii". *Applied Geochemistry*, Volume 19, Issue 7. Elsevier.

De Jong, S. 1993. "SIMPLS: an alternative approach to partial least squares regression". *Chemometrics and Intelligent Laboratory Systems*, Volume 18, Issue 3. Elsevier Science Publishers B. V., Amsterdam.

Deas, M.L.; Lowney, C. L. 2000. "Water temperature modelling review". Central Valley. The Bay Delta Modeling Forum.

Denham, M. C. 2000. "Choosing the number of factors in partial least squares regression: estimating and minimizing the mean squared error of prediction". *Journal of Chemometrics*,

Devore, J. L. 1998. "Probabilidad y estadística para ingeniería y ciencias". International Thomson Editores.

Dolz, J.; Puertas, J.; Aguado, A.; Agulló, L. 1995. "Efectos térmicos en presas y embalses". E.T.S.I. Caminos, Canales y Puertos. Universidad Politécnica de Cataluña.

Draper, N. R.; Smith, H. 1998. "Applied regression analysis". John Wiley & Sons, Inc.

Eastment, H. T.; Krzanowski, W. J. 1982. "Cross-Validatory choice of the number of components from a principal component analysis". *Technometrics*, Volume 24, No. 1. American Statistical Association and American Society for Quality.

Edwards, L. J.; Hamilton, S. A. 1995. "Errors-in-variables and the Box-Cox transformation". *Computational Statistics & Data Analysis*, Volume 20, Issue 2. Elsevier.

Einax, J. W.; Aulinger, A.; Tümpling, W. V.; Prange, A. 1999. "Quantitative description of element concentrations in longitudinal river profiles by multiway PLS models". *Fresenius' Journal of Analytical Chemistry*, Volume 363, Number 7. Springer-Verlag.

Engelen, S.; Hubert, M.; Vanden Branden, K.; Verboven, S. 2003. "Robust PCR and robust PLSR: a comparative study". En: *Theory and Applications of Recent Robust Methods* (edited by M. Hubert, G. Pison, A. Struyf and S. Van Aelst). *Statistics for Industry and Technology*. Birkhauser, Basel.

Environmental Laboratory. 1995. "CE-QUAL-R1: A numerical one-dimensional model of reservoir water quality; user's manual". Instruction Report E-82-1 (Revised Edition), U. S. Army Engineer Waterways Experiment Station, Vicksburg, MS.

Ergon, R. 2006. "Reduced PCR/PLSR models by subspace projections". *Chemometrics and Intelligent Laboratory Systems*, Volume 81, Issue 1. Elsevier.

Etxeberria, J. 1999. "Regresión múltiple". Editorial La Muralla, S. A./Editorial Hespérides.

Felipe-Sotelo, M.; Andrade, J. M.; Carlosena, A.; Tauler, R. 2007. "Temporal characterisation of river waters in urban and semi-urban areas using physico-chemical parameters and chemometric methods". *Analytica Chimica Acta*, Volume 583, Issue 1. Elsevier.

Ferrán Aranaz, M. 1996. "SPSS para Windows. Programación y análisis estadístico". McGraw-

---

Hill/Interamericana de España.

Ferrán Aranaz, M. 2002. "Curso de SPSS para Windows". McGraw-Hill/Interamericana de España.

Ferré, J. 2006. "Calibración multivariante en análisis cuantitativo. El modelo inverso". Técnicas de laboratorio, No. 310. Publica, S. A.

Flury, B. D. 2006. "Principal components". Encyclopedia of Statistical Sciences. Standard article. John Wiley & Sons, Inc.

Garreaud S., R.; Meruane N., C. 2005. "Módulo 2 Flujos radiativos". Departamento de Geofísica. Facultad de Ciencias Físicas y Matemáticas. Universidad de Chile.

Geladi, P. 2002. "Some recent trends in the calibration literature". Chemometrics and intelligent laboratory systems, Volume 60, Issues 1-2. Elsevier.

Geladi, P.; Kowalski, B. R. 1986. "Partial Least-Squares Regression: A Tutorial". Analytica Chimica Acta, Volume 185. Elsevier Science Publishers B. V.

Gelda, R.K.; Effler, S.W. 2007. "Simulation of operations and water quality performance of reservoir multilevel intake configurations". Journal of Water Resources Planning and Management, Vol. 133, No. 1. ASCE.

Gelman, A.; Hill, J. 2007. "Data analysis using regression and Multilevel/Hierarchical Models". Cambridge University Press.

Gil Pascual, J. A. 2003. "Métodos de investigación en educación. Volumen III: Análisis multivariante". Universidad Nacional de Educación a Distancia.

Giussani, B.; Dossi, C.; Monticelli, D.; Pozzi, A.; Recchia, S. 2006. "A chemometric approach to the investigation of major and minor ion chemistry in Lake Como (Lombardia, Northern Italy)". Annali di Chimica, Volume 96, Issue 5-6. Società Chimica Italiana.

Gutiérrez, J. M.; Cano, R.; Cofiño, A. S.; Sordo, C. M. 2004. "Redes probabilísticas y neuronales en las ciencias atmosféricas". Centro de Publicaciones. Secretaría General Técnica. Ministerio de Medio Ambiente, España.

Hadjiiski, L.; Geladi, P.; Hopke, P. 1999. "A comparison of modelling nonlinear systems with artificial neural networks and partial least squares". Chemometrics and Intelligent Laboratory

---

Systems, Volume 49, Issue 1. Elsevier.

Han, B. P.; Armengol, J.; Garcia, J. C.; Comerma, M.; Roura, M.; Dolz, J.; Straskraba, M. 2000. "The thermal structure of Sau Reservoir (NE: Spain): a simulation approach". *Ecological Modelling*, Volume 125, Issues 2-3. Elsevier.

Helland, I. 2006. "Partial least squares regression". *Encyclopedia of Statistical Sciences*. Standard article. John Wiley & Sons, Inc.

Henderson-Sellers, B.; Markland, H. R. 1987. "Decaying lakes. The origins and control of cultural eutrophication". John Wiley & Sons.

Hornung, R. 2002. "Numerical modelling of stratification in Lake Constance with the 1-D hydrodynamic model DYRESM". Universität Stuttgart.

[http://ec.europa.eu/environment/water/index\\_en.htm](http://ec.europa.eu/environment/water/index_en.htm)

<http://es.wikipedia.org/wiki/>

<http://folk.uio.no/ohammer/past/multivar.html>

<http://functions.wolfram.com/GammaBetaErf/InverseErf/06/01/>

<http://hispagua.cedex.es/documentacion/suplementos/presas.htm>

<http://momac.uclm.es/>

<http://statisticssolutions.com/Partial-Least-Squares-Regression.htm>

[http://www.cfd-online.com/Wiki/Tridiagonal\\_matrix\\_algorithm\\_-\\_TDMA\\_\(Thomas\\_algorithm\)](http://www.cfd-online.com/Wiki/Tridiagonal_matrix_algorithm_-_TDMA_(Thomas_algorithm))

<http://www.chj.es/>

<http://www.dip-alicante.es/documentación/censo.asp>

<http://www.epa.gov/owow/monitoring/volunteer/stream/vms54.html>

<http://www.epa.gov/owow/monitoring/volunteer/stream/vms59.html> (2007)

---

---

<http://www.eumed.net/cursecon/libreria/2004/jsf/4.pdf>

<http://www.fao.org/docrep/009/y4666s/y4666s0f.htm>

<http://www.geoiberica.com>

<http://www.ign.es/ign/home/calculadora/UTM2LL.jsp>

<http://www.itchihuahua.edu.mx/academic/industrial/estadistica1/cap01.html>

<http://www.spss.com/spss/images/pdfs/S16CMPPhr.pdf>

<http://www.statsoft.com/textbook/stathome.html>

[http://www.umetrics.com/default.asp/pagename/methods\\_over\\_hist/c/2](http://www.umetrics.com/default.asp/pagename/methods_over_hist/c/2)

Hultquist, R. 2006. "Multiple linear regression". Encyclopedia of Statistical Sciences. Standard article. John Wiley & Sons, Inc.

Hydrologic Engineering Center. 1986. "HEC-5. Simulation of flood control and conservation systems. Appendix on water quality analysis". U. S. Army Corps of Engineers.

Instituto Valenciano de Investigaciones Agrarias (I.V.I.A.). 2000-2005. "Datos meteorológicos". I.V.I.A.

Isasi Viñuela, P.; Galván León, I. M. 2004. "Redes de neuronas artificiales. Un enfoque práctico". Pearson Educación, S. A.

Jain, L. C.; Martin, N. M. 1999. "Fusion of neural networks, fuzzy sets, and genetic algorithms. Industrial applications". CRC Press LLC.

Jensen, F. V. 2002. "An introduction to Bayesian networks". Reprinted by Taylor & Francis.

Jobson, J. D. 1991. "Applied multivariate data analysis. Volume I: Regression and experimental design". Springer-Verlag New York, Inc.

Jobson, J. D. 1992. "Applied multivariate data analysis. Volume II: Categorical and multivariate methods". Springer texts in Statistics. Springer-Verlag New York, Inc.

---

---

Joehnk, K. D.; Umlauf, L. 2001. "Modelling the metalimnetic oxygen minimum in a medium sized alpine lake". *Ecological Modelling*, Volume 136, Issue 1. Elsevier.

Johnson, D. E. 2000. "Métodos multivariados aplicados al análisis de datos". International Thomson Editores.

Johnson, R. A.; Wichern, D. W. 1992. "Applied multivariate statistical analysis". Prentice-Hall International, Inc.

Johnson, R.A.; Wichern, D. 2004. "Multivariate analysis". *Encyclopedia of Statistical Sciences*. Standard Article. John Wiley & Sons, Inc.

Juricek, B. C.; Seborg, D. E.; Larimore, W. E. 2005. "Process control applications of subspace and regression-based identification and monitoring methods". American Control Conference. Portland, OR, USA.

Kabir, A. 2000. "The dam debate yet again". The World Commission on Dams.

Kannel, P. R.; Lee, S.; Kanel, S. R.; Khan, S. P. 2007. "Chemometric application in classification and assessment of monitoring locations of an urban river system". *Analytica Chimica Acta*, Volume 582, Issue 2. Elsevier.

Karul, C.; Soyupak, S.; Çilesiz, A. F.; Akbay, N.; Germen, E. 2000. "Case studies on the use of neural networks in eutrophication modeling". *Ecological Modelling*, Volume 134, Issues 2-3. Elsevier.

Kendall, M.; Stuart, A.; Ord, J. K. 1991. "Advanced Theory of Statistics. Volume 2: Classical inference and relationship". The Ipswich Book Company.

Kettaneh, N.; Berglund, A.; Wold, S. 2005. "PCA and PLS with very large data sets". *Computational Statistics & Data Analysis*, Volume 48, Issue 1. Elsevier.

Kettenring, J. R. 2006. "Canonical analysis". *Encyclopedia of Statistical Sciences*. Standard article. John Wiley & Sons, Inc.

Kotz, S.; Balakrishnan, N.; Johnson, N. L. 2000. "Continuous multivariate distributions". Wiley series in probability and statistics. John Wiley & Sons.

Kowalik, C.; Einax, J. W. 2006. "Modern chemometric data analysis – methods for the objective evaluation of load in river systems". *Acta hydrochim. hydrobiol.* 34. WILEY-VCH Verlag GmbH

---

& Co. KGaA.

Kowalkowski, T.; Zbytniewski, R.; Szpejna, J.; Buszewski, B. 2006. "Application of chemometrics in river water classification". *Water Research*, Volume 40, Issue 4. Elsevier.

Lerman, A.; Imboden, D. M.; Gat, J. R. 1995. "Physics and Chemistry of lakes". Springer-Verlag.

Li, B.; Morris, J.; Martin, E. B. 2002. "Model selection for partial least squares regression". *Chemometrics and Intelligent Laboratory Systems*, Volume 64, Issue 1. Elsevier.

Li, Ch.; Huang, H. P. 2003. "Model building by merging submodels using PLSR". *Journal of Chemical Engineering of Japan*, Vol. 36, No. 9. The Society of Chemical Engineers, Japan.

Li, Ch.; Ye, H.; Wang, G.; Zhang, J. 2005. "A recursive nonlinear PLS algorithm for adaptive nonlinear process modelling". *Chemical Engineering and Technology*, Volume 28, Issue 2. Wiley-VCH.

Limno\_Tech, Inc. 2002. "Descriptive inventory of models with prospective relevance to ecological impacts of water withdrawals". The Great Lakes Commission.

Lix, L. M.; Keselman, H. J.; Hinds, A. M. 2004. "Robust tests for the multivariate Behrens-Fisher problem". *Computer Methods and Programs in Biomedicine*. Elsevier Ireland Ltd.

Llorca Llorca, R.; Bautista Carrascosa, I. 2006. "Prácticas de atmósfera, suelo y agua". Editorial Universidad Politécnica de Valencia.

Logofet, D. O.; Korotkov, V. N. 2002. "'Hybrid' optimisation: a heuristic solution to the Markov-chain calibration problem". *Ecological Modelling*, Volume 151, Issues 1. Elsevier.

Manahan, S. E. 2000. "Environmental chemistry". Lewis Publishers.

Marengo, E.; Gennaro, M. C.; Giacosa, D.; Abrigo, C.; Saini, G.; Avignone, M. T. 1995. "How chemometrics can helpfully assist in evaluating environmental data. Lagoon water". *Analytica Chimica Acta*, Volume 317, Issues 1-3. Elsevier.

Margalef, R. 1983. "Limnología". Ediciones Omega.

Markiegi, X.; Rallo, A.; Andía, A. 1999. "Protección de la calidad del agua en los embalses del



---

Sistema Zadorra". Ararteko.

Martens, H. 2001. "Reliable and relevant modelling of real world data: a personal account of the development of PLS regression". *Chemometrics and Intelligent Laboratory Systems*, Volume 58, Issue 2. Elsevier.

Martens, H.; Næs, T. 1989. "Multivariate calibration". John Wiley & Sons.

Martín Martín, Q.; Cabero Morán, M. T.; De Paz Santana, Y. R. 2008. "Tratamiento estadístico de datos con SPSS". International Thomson Editores Spain.

Martín Moneris, M.; Marzal Doménech, P. 1999. "Modelación de la calidad del agua". Departamento de Ingeniería Hidráulica y Medio Ambiente. Editorial de la Universidad Politécnica de Valencia.

Martin, J.L.; McCutcheon, S. C. 1999. "Hydrodynamics and transport for water quality modelling". Lewis Publishers.

Martínez Arias, R. 1999. "El análisis multivariante en la investigación científica". Editorial La Muralla, S. A./Editorial Hespérides.

Maxwell, A. E. 2006. "Factor analysis". *Encyclopedia of Statistical Sciences*. Standard article. John Wiley & Sons, Inc.

McNeil, V. H.; Cox, M. E.; Preda, M. 2005. "Assessment of chemical water types and their spatial variation using multi-stage cluster analysis, Queensland, Australia". *Journal of Hydrology*, Volume 310, Issues 1-4. Elsevier.

Mendenhall, W.; Scheaffer, R. L.; Wackerly, D. D. 1986. "Estadística matemática con aplicaciones". Grupo Editorial Iberoamérica.

Mendiguchía, C.; Moreno, C.; Galindo-Riaño, M. D.; García-Vargas, M. 2004. "Using chemometric tools to assess anthropogenic effects in river water. A case study: Guadalquivir River (Spain)". *Analytica Chimica Acta*, Volume 515, Issue 1. Elsevier.

Miket, M. J. 2007. "Multiple regression – Selecting the best equation". ([math.usask.ca/~miket/S344D.pdf](http://math.usask.ca/~miket/S344D.pdf)). University of Saskatchewan, Canada.

Ministerio de Obras Públicas, Transportes y Medio Ambiente. 1994. "Real Decreto 1541/1994".

---

Boletín Oficial del Estado n. 179 de 28/7/1994. España.

Miró I Martínez, P.; Debón Aucejo, A.; Crespo Abril, F. 2006. "Métodos estadísticos en economía". Editorial de la UPV. Universidad Politécnica de Valencia.

Momen, B.; Eichler, L. W.; Boylen, C. W.; Zehr, J. P. 1996. "Application of multivariate statistics in detecting temporal and spatial patterns of water chemistry in Lake George, New York". Ecological Modelling, Volume 91, Issues 1-3. Elsevier.

Montero Lorenzo, J. M. 2007. "Estadística descriptiva". Thomson.

Montes-Botella, C.; Tenorio, M. D. 2003. "Water characterization and seasonal heavy metal distribution in the Odiel River (Huleva, Spain) by means of principal component analysis". Archives of Environmental Contamination and Toxicology, Volume 45, Number 4. Springer-Verlag New York Inc.

Navidi, W. 2006. "Estadística para ingenieros". McGraw-Hill Interamericana Editores.

Neudecker, H. 2006. "Varimax method". Encyclopedia of Statistical Sciences. Standard Article. John Wiley & Sons, Inc.

Nolberto Sifuentes, V. 2002. "Parte I. Análisis de regresión lineal múltiple". Curso: Análisis multivariante aplicado a la biología pesquera. IMARPE.

Novotny, V. 2003. "Water quality: Diffuse pollution and watershed management". John Wiley & Sons, Inc.

Orlob, G. T. 1981. "Models for stratified impoundments". Department of Civil Engineering, University of California, Davis, CA.

Ouarda, T. B. M. J.; Bâ, K. M.; Diaz-Delgado, C.; Cârsteanu, A.; Chokmani, K.; Gingras, H.; Quentin, E.; Trujillo, E.; Bobée, B. 2008. "Intercomparison of regional flood frequency estimation methods at ungauged sites for a Mexican case study". Journal of Hydrology, Volume 348, Issues 1-2. Elsevier.

Ouarda, T. B. M. J.; Girard, C.; Cavadias, G. S.; Bobée, B. 2001. "Regional flood frequency estimation with canonical correlation analysis". Journal of Hydrology, Volume 254, Issues 1-4. Elsevier Science B. V.

Paredes Arquiola, J.; Andreu Álvarez, J.; Martín Moneris, M.; Solera Solera, A. 2006a.

---

---

“Modelación de la calidad del agua a escala de cuenca”. Departamento de Ingeniería Hidráulica y Medio Ambiente. Editorial de la Universidad Politécnica de Valencia.

Paredes Arquiola, J.; Martín Monerris, M. 2006b. “Tomo III. Estudio de la calidad ambiental de las aguas en el Embalse de Amadorio. Propuesta de actuaciones”. Universidad Politécnica de Valencia.

Parinet, B.; Lhote, A.; Legube, B. 2004. “Principal component analysis: an appropriate tool for water quality evaluation and management-application to a tropical lake system”. *Ecological Modelling*, Volume 178, Issues 3-4. Elsevier.

Park, R. A.; Clough, J. S. 2004. “AQUATOX (RELEASE 2). Modeling environmental fate and ecological effects in aquatic ecosystems. Volume 2: Technical documentation”. U. S. Environmental Protection Agency.

Park, R. A.; Clough, J. S. 2005. “AQUATOX for Windows. A modular fate and effects model for aquatic ecosystems. Release 2.1. Addendum to release 2. Technical documentation”. U.S. Environmental Protection Agency.

Pazos Sierra, A. (Coordinador). 1996. “Redes de neuronas artificiales y algoritmos genéticos”. Universidade da Coruña.

Peña, D. 1989. “Estadística modelos y métodos. 2. Modelos lineales y series temporales”. Alianza Editorial, S. A.

Peña, D. 2002. “Análisis de datos multivariantes”. McGraw-Hill/Interamericana de España.

Peña, D. 2005a. “Análisis de series temporales”. Alianza Editorial, S. A.

Peña, D. 2005b. “Fundamentos de estadística”. Alianza Editorial, S. A.

Pérez López, C. 2005. “Métodos estadísticos avanzados con SPSS”. International Thomson Editores Spain Paraninfo, S. A.

Pérez López, C.; Santín González, D. 2007. “Minería de datos. Técnicas y herramientas”. International Thomson Editores Spain Paraninfo, S. A.

Pérez López, C. 2008. “Técnicas estadísticas con SPSS 12. Aplicaciones al análisis de datos”. Pearson Educación S. A.

---

---

Petersen, W.; Bertino, L.; Callies, U.; Zorita, E. 2001. "Process identification by principal component analysis of river water-quality data". *Ecological Modelling*, Volume 138, Issues 1-3. Elsevier.

Prats i Vime, R.; Correcher Martínez, E. 2004. "Seguimiento en continuo de la calidad de las aguas embalsadas". Grupo de trabajo 24, Teledetección y sensores ambientales. VII Congreso Nacional del Medio Ambiente, España. ([www.conama.org/documentos/GT24.pdf](http://www.conama.org/documentos/GT24.pdf)).

Rodrigues, J.; Alves, A.; Pereira, H.; Da Silva Perez, D.; Chantre, G.; Schwanninger, M. 2006. "NIR PLSR results obtained by calibration with noisy, low-precision referente values: Are the results acceptable?" *Holzforschung*, Volume 60, Issue 4. Walter de Gruyter.

Romero, J. R.; Antenucci, J. P.; Imberger, J. 2004. "One- and three-dimensional biogeochemical simulations of two differing reservoirs". *Ecological Modelling*, Volume 174. Elsevier.

Romero Rojas, J. A. 1999. "Calidad del agua". Alfaomega.

Rueda, F.; Moreno-Ostos, E.; Armengol, J. 2006. "The residence time of river water in reservoirs". *Ecological Modelling*, Volume 191, Issue 2. Elsevier.

Ruiz-Maya Pérez, L.; Martín Pliego, F. J. 2001. "Estadística II: Inferencia". Editorial AC.

Sánchez Carrión, J.J. 1999. "Manual de análisis estadístico de los datos". Alianza Editorial.

Schladow, S. G.; Hamilton, D. P. 1997. "Prediction of water quality in lakes and reservoirs: Part II-Model calibration, sensitivity analysis and application". *Ecological Modelling* 96, Volume 96. Elsevier.

Serway, R. A.; Faughn, J. S. 2005. "Fundamentos de Física". Internacional Thomson Editores.

Shrestha, S.; Kazama, F. 2007. "Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan". *Environmental Modelling & Software*, Volume 22, Issue 4. Elsevier.

Simeonov, V.; Einax, J. W.; Stanimirova, I.; Kraft, J. 2002. "Environmetric modelling and interpretation of river water monitoring data". *Analytical and Bioanalytical Chemistry*, Volume 374, Number 5. Springer-Verlag.

Singh, K. P.; Malik, A.; Basant, N.; Saxena, P. 2007. "Multi-way partial least squares modelling

---

of water quality data". *Analytica Chimica Acta*, Volume 584, Issue 2. Elsevier.

Singh, K. P.; Malik, A.; Mohan, D.; Sinha, S. 2004. "Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India) – a case study". *Water Research*, Volume 38, Issue 18. Elsevier.

Singh, K. P.; Malik, A.; Sinha, S. 2005. "Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques- a case study". *Analytica Chimica Acta*, Volume 538, Issues 1-2. Elsevier.

Spiegel, M. R.; Schiller, J. J.; Srinivasan, R. A. 2001. "Teoría y problemas de probabilidad y estadística". McGRAW-HILL INTERAMERICANA, S. A.

SPSS, 2006. "SPSS 15.0 Brief Guide". SPSS Inc.

Srivastava, D. K.; Mudholkar, G. S. 2001. "Trimmed  $\tilde{T}^2$ : A robust analog of Hotelling's  $T^2$ ". *Journal of statistical planning and inference*. Elsevier Science B. V.

StatPoint, Inc. 2005. "Statgraphics Centurion XV. User Manual". (<http://www.statgraphics.com/>)

Stiling, Peter. 2002. "Ecology. Theories and Applications". Prentice Hall.

Sundberg, R. 2006. "Small-sample and selection bias effects in multivariate calibration, exemplified for OLS and PLS regressions". *Chemometrics and Intelligent Laboratory Systems*, Volume 84, Issues 1-2. Elsevier.

Tebbutt, T. H. Y. 1990. "Fundamentos de control de la calidad del agua". Editorial Limusa.

Terrado, M.; Barceló, D.; Tauler, R. 2006. "Identification and distribution of contamination sources in the Ebro river basin by chemometrics modelling coupled to geographical information systems". *Talanta*, Volume 70, Issue 4. Elsevier.

Terrado, M.; Kuster, M.; Raldúa, D.; López de Alda, M.; Barceló, D.; Tauler, R. 2007. "Use of chemometric and geostatistical methods to evaluate pesticide pollution in the irrigation and drainage channels of the Ebro river delta during the rice-growing season". *Analytical and Bioanalytical Chemistry*, Volume 387, Number 4. Springer-Verlag.

Thomann, R. V.; Mueller, J. A. 1987. "Principles of surface water quality modelling and control".

---

---

Harper Collins Publishers Inc.

Tufford, D. L.; Mckellar, H. N. 1999. "Spatial and temporal hydrodynamic and water quality modeling analysis of a large reservoir on the South Carolina (USA) coastal plain". *Ecological Modelling*, Volume 114, Issues 2-3. Elsevier.

Tusell, F. 2007. "Análisis multivariante". Edición electrónica en línea.

U. S. Army Corps of Engineers. 1987. "Engineering and design. Reservoir water quality analysis". Department of the Army. Engineer manual 1110-2-1201.

Van Albada, S. J.; Robinson, P. A. 2007. "Transformation of arbitrary distributions to the normal distribution with application to EEG test-retest reliability". *Journal of Neuroscience Methods*, Volume 161, Issue 2. Elsevier.

Vega, M.; Pardo, R.; Barrado, E.; Debán, L. 1998. "Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis". *Water Research*, Volume 32, Issue 12. Elsevier Science Ltd.

Verboven, S.; Hubert, M. 2005. "LIBRA: a MATLAB library for robust analysis". *Chemometrics and Intelligent Laboratory Systems*, Volume 75, Issue 2. Elsevier.

Visauta Vinacua, B. 2002. "Análisis estadístico con SPSS 11.0 para Windows. Volumen I. Estadística básica". McGraw-Hill/Interamericana de España.

Visauta Vinacua, B.; Martori I Cañas, J. C. 2003. "Análisis estadístico con SPSS para Windows". Volumen II, Estadística multivariante. McGraw-Hill/Interamericana de España, S. A. U.

Viscarra Rossel, R. A. 2007. "ParLes: Software for chemometric analysis of spectroscopic data". *Chemometrics and Intelligent Laboratory Systems*, Volume 90, Issue 1. Elsevier.

Wackerly, D. D.; Mendenhall III, W.; Scheaffer, R. L. 2002. "Mathematical statistics with applications". Wadsworth Group. Duxbury and Brooks/Cole are imprints of the Wadsworth Group, a division of Thomson Learning, Inc.

Walpole, R. E.; Myers, R. H.; Myers, S. L. 1999. "Probabilidad y estadística para ingenieros". Prentice-Hall Hispanoamericana, S. A.

---

---

Wetzel, R. G. 2001. "Limnology". Third Edition. Academic Press.

Whigham, P. A.; Recknagel, F. 2001. "Predicting chlorophyll-a in freshwater lakes by hybridising process-based models and genetic algorithms". *Ecological Modelling*, Volume 146, Issues 1-3. Elsevier.

Wold, S. 2001. "Personal memories of the early PLS development". *Chemometrics and Intelligent Laboratory Systems*, Volume 58, Issue 2. Elsevier.

Wold, S.; Sjöström, M.; Eriksson, L. 2001b. "PLS-regression: a basic tool of chemometrics". *Chemometrics and Intelligent Laboratory Systems*, Volume 58, Issue 2. Elsevier.

Wold, S.; Trygg, J.; Berglund, A.; Antti, H. 2001. "Some recent developments in PLS modeling". *Chemometrics and Intelligent Laboratory Systems*, Volume 58, Issue 2. Elsevier.

Wu, W.; Manne, R. 2000. "Fast regression methods in a Lanczos (or PLS-1) basis. Theory and applications". *Chemometrics and Intelligent Laboratory Systems*, Volume 51, Issue 2. Elsevier.

Wüest, A.; Lorke, A. 2003. "Small-scale hydrodynamics in lakes". *Annu. Rev. Fluid Mech.* 35: 373-412.

Wunderlin, D. A.; Díaz, M. P.; Amé, M. V.; Pesce, S. F.; Hued, A. C.; Bistoni, M. A. 2001. "Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquía River Basin (Córdoba-Argentina)". *Water Research*, Volume 35, Issue 12. Elsevier Science Ltd.

Wurbs, R. A. 1994. "Computer models for water resources planning and management". U.S. Army Corps of Engineers. Institute for Water Resources. Water Resources Support Center. IWR Report 94-NDS-7.

Zarzo Castelló, M. 2004. "Aplicación de técnicas estadísticas multivariantes al control de calidad de procesos por lotes". Tesis doctoral. Departamento de Estadística e Investigación Operativa Aplicadas y Calidad. Universidad Politécnica de Valencia.





## **ANEXOS DIGITALES (CD)**

**ANEXO 1. ESTRUCTURA INICIAL DE LOS DATOS DEL  
MODELO**



Fecha	Hidrológicas				Profundidad (m)	Calidad del agua			
	VolExp (m3)	Qent (m3/día)	Qsal (m3/día)	Filt (m3/día)		Ta (°C)	Odis (mg/l)	pH	Con (mS/cm)
31/03/2004	7814000	37171,6	4320,0	3,4E-12	0,42	14,06	13,52	8,51	1050,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	0,83	13,95	14,11	8,51	1050,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	1,25	13,88	14,30	8,49	1051,67
31/03/2004	7814000	37171,6	4320,0	3,4E-12	1,67	13,85	14,37	8,49	1051,67
31/03/2004	7814000	37171,6	4320,0	3,4E-12	2,09	13,76	14,38	8,49	1053,33
31/03/2004	7814000	37171,6	4320,0	3,4E-12	2,92	13,62	13,93	8,46	1056,67
31/03/2004	7814000	37171,6	4320,0	3,4E-12	3,34	13,74	14,59	8,46	1050,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	3,75	13,61	13,96	8,43	1054,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	4,59	13,47	12,59	8,22	1066,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	5,42	12,98	9,95	7,82	1098,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	6,26	12,55	6,97	7,63	1114,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	7,09	12,28	6,28	7,61	1120,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	7,92	12,15	6,15	7,62	1122,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	8,34	12,07	6,41	7,65	1120,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	8,76	12,01	6,12	7,64	1123,33
31/03/2004	7814000	37171,6	4320,0	3,4E-12	9,17	11,93	6,34	7,65	1130,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	9,59	11,91	6,20	7,66	1126,67
31/03/2004	7814000	37171,6	4320,0	3,4E-12	10,01	11,85	6,26	7,66	1128,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	10,84	11,75	6,27	7,67	1130,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	11,68	11,57	6,17	7,67	1130,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	12,51	11,50	6,02	7,66	1135,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	13,34	11,47	5,88	7,67	1140,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	14,18	11,46	5,72	7,66	1145,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	15,01	11,43	5,67	7,66	1145,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	15,85	11,43	5,61	7,66	1155,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	16,68	11,44	5,48	7,65	1155,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	17,51	11,44	5,33	7,64	1160,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	18,35	11,44	5,13	7,63	1160,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	19,18	11,44	5,15	7,63	1160,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	20,02	11,45	5,01	7,63	1170,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	20,43	11,46	4,84	7,62	1170,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	21,27	11,49	4,58	7,61	1170,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	22,1	11,51	4,28	7,60	1180,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	22,94	11,59	3,56	7,55	1200,00
31/03/2004	7814000	37171,6	4320,0	3,4E-12	23,77	11,62	3,27	7,55	1210,00
01/04/2004	7864000	56480,2	4320,0	3,6E-12	0,42	13,90	13,90	8,55	1050,00
01/04/2004	7864000	56480,2	4320,0	3,6E-12	0,83	13,88	14,41	8,52	1050,00
01/04/2004	7864000	56480,2	4320,0	3,6E-12	1,25	13,86	14,55	8,51	1050,00
01/04/2004	7864000	56480,2	4320,0	3,6E-12	1,67	13,83	14,63	8,50	1050,00
01/04/2004	7864000	56480,2	4320,0	3,6E-12	2,09	13,81	14,62	8,49	1050,00
01/04/2004	7864000	56480,2	4320,0	3,6E-12	2,92	13,75	14,61	8,49	1050,00
01/04/2004	7864000	56480,2	4320,0	3,6E-12	3,75	13,71	14,61	8,49	1050,00
01/04/2004	7864000	56480,2	4320,0	3,6E-12	4,59	13,65	14,37	8,47	1053,33
01/04/2004	7864000	56480,2	4320,0	3,6E-12	5,42	13,60	13,90	8,38	1056,67
01/04/2004	7864000	56480,2	4320,0	3,6E-12	6,26	12,70	8,13	7,73	1113,33
01/04/2004	7864000	56480,2	4320,0	3,6E-12	7,09	12,34	6,60	7,70	1120,00
01/04/2004	7864000	56480,2	4320,0	3,6E-12	7,92	12,17	6,30	7,69	1120,00
01/04/2004	7864000	56480,2	4320,0	3,6E-12	8,76	12,00	6,24	7,69	1126,67
01/04/2004	7864000	56480,2	4320,0	3,6E-12	9,59	11,92	6,26	7,69	1130,00
01/04/2004	7864000	56480,2	4320,0	3,6E-12	10,01	11,82	6,19	7,70	1130,00
01/04/2004	7864000	56480,2	4320,0	3,6E-12	10,84	11,73	6,20	7,70	1130,00
01/04/2004	7864000	56480,2	4320,0	3,6E-12	11,68	11,60	6,25	7,71	1130,00
01/04/2004	7864000	56480,2	4320,0	3,6E-12	12,51	11,54	6,04	7,69	1130,00
01/04/2004	7864000	56480,2	4320,0	3,6E-12	13,34	11,47	5,87	7,68	1140,00
01/04/2004	7864000	56480,2	4320,0	3,6E-12	14,18	11,44	5,80	7,68	1143,33
01/04/2004	7864000	56480,2	4320,0	3,6E-12	15,01	11,42	5,75	7,68	1150,00
01/04/2004	7864000	56480,2	4320,0	3,6E-12	15,85	11,42	5,61	7,66	1153,33
01/04/2004	7864000	56480,2	4320,0	3,6E-12	16,68	11,42	5,48	7,66	1156,67
01/04/2004	7864000	56480,2	4320,0	3,6E-12	17,51	11,42	5,33	7,65	1160,00
01/04/2004	7864000	56480,2	4320,0	3,6E-12	18,35	11,44	5,19	7,65	1166,67
01/04/2004	7864000	56480,2	4320,0	3,6E-12	19,18	11,44	5,08	7,64	1166,67
01/04/2004	7864000	56480,2	4320,0	3,6E-12	20,02	11,46	4,80	7,63	1170,00
01/04/2004	7864000	56480,2	4320,0	3,6E-12	20,43	11,47	4,67	7,63	1170,00
01/04/2004	7864000	56480,2	4320,0	3,6E-12	21,27	11,48	4,47	7,62	1173,33
01/04/2004	7864000	56480,2	4320,0	3,6E-12	22,1	11,51	4,20	7,61	1180,00
01/04/2004	7864000	56480,2	4320,0	3,6E-12	22,94	11,55	3,92	7,59	1190,00
01/04/2004	7864000	56480,2	4320,0	3,6E-12	23,77	11,60	3,60	7,58	1200,00
02/04/2004	7921000	63233,4	4320,0	1,8E-12	0,42	14,68	14,03	8,61	1040,00
02/04/2004	7921000	63233,4	4320,0	1,8E-12	0,83	14,43	14,79	8,52	1040,00
02/04/2004	7921000	63233,4	4320,0	1,8E-12	1,25	14,36	14,99	8,48	1040,00
02/04/2004	7921000	63233,4	4320,0	1,8E-12	1,67	14,31	15,10	8,47	1045,00
02/04/2004	7921000	63233,4	4320,0	1,8E-12	2,09	14,09	15,04	8,44	1045,00
02/04/2004	7921000	63233,4	4320,0	1,8E-12	2,92	13,90	14,33	8,37	1055,00
02/04/2004	7921000	63233,4	4320,0	1,8E-12	3,75	13,76	13,56	8,30	1060,00

Nomenclatura y unidades empleadas para las variables:

Clasificación de la variable	Variable y sus unidades de medida
Temporal	DJ, Día juliano
Meteorológicas	V, Velocidad media del viento (Km/h)
	DV, Dirección media del viento (°)
	Vx, Velocidad máxima del viento (Km/h)
	T, Temperatura media (°C)
	Tn, Temperatura mínima (°C)
	Tx, Temperatura máxima del aire (°C)
	Horas frío, Hf, Número de horas por debajo de 7 °C
	HR, Humedad relativa media del aire (%)
	Rad, Radiación solar global (W/m <sup>2</sup> )
	P, Precipitación total (mm)
Hidrológicas	ETo, Evapotranspiración de referencia (mm)
	VolExp, Volumen de explotación (m <sup>3</sup> )
	Qent, Caudal de entrada al embalse (m <sup>3</sup> /día)
	Qsal, Caudal de salida del embalse (m <sup>3</sup> /día)
	Filt, Filtración (m <sup>3</sup> /día)
	Profundidad (m)
Calidad del agua	Ta, Temperatura del agua (°C)
	Odis, Oxígeno disuelto (mg/l)
	pH
	Con, Conductividad (μS/cm)

## **ANEXO 2. ESTADÍSTICA DESCRIPTIVA DE PERIODOS DE CALIBRACIÓN Y VALIDACIÓN**

Parámetro	Media	Error típico	Mediana	Moda	Desviación estándar	Varianza de la muestra	Curiosis	Coefficiente de asimetría	Rango	Mínimo	Máximo
Día juliano	207,7	1,848	224,0	91,0	122,6	15030	-1,309	-0,4807	359,0	5,0	364,0
V (Km/h)	1,312	0,0049	1,3	1,2	0,3278	0,1074	2,448	1,023	2,2	0,5	2,7
DV (°)	265,3	1,112	289,0	292,0	73,76	5441	2,789	-1,734	357,0	0,0	357,0
Vx (Km/h)	5,200	0,0229	4,8	3,8	1,520	2,311	0,2013	0,8673	7,6	2,5	10,1
T (°C)	16,85	0,0983	15,0	10,7	6,522	42,53	-1,215	0,2077	25,2	3,9	29,1
Tn (°C)	12,82	0,0952	11,1	7,3	6,316	39,89	-1,184	0,1566	24,2	-0,1	24,1
Tx (°C)	21,45	0,0985	20,1	27,0	6,532	42,67	-1,179	0,1599	24,9	8,1	33,0
Hf (horas)	1,506	0,0571	0,0	0,0	3,788	14,35	7,721	2,840	20,0	0,0	20,0
HR (%)	57,60	0,2030	60,0	69,0	13,47	181,3	-0,2626	-0,7357	61,0	20,0	81,0
Rad (W/m <sup>2</sup> )	158,5	1,111	141,0	115,0	73,69	5431	-0,6900	0,4711	298,0	21,0	319,0
P (mm)	0,5692	0,0363	0,0	0,0	2,405	5,786	39,01	5,810	22,4	0,0	22,4
ETo (mm)	2,691	0,0231	2,0	1,9	1,531	2,343	-0,7486	0,7685	6,1	0,8	6,9
VolExp (m <sup>3</sup> )	8477988	19800	8575000	6720000	1313501	1,725E+12	-1,504	-0,0284	4032000	6709000	10741000
Qent (m <sup>3</sup> /día)	32560	529,7	32694	0,0	35137	1234609169	12,36	2,961	232298	0,0	232298
Qsal (m <sup>3</sup> /día)	25523	169,6	28080	20304	11251	126577117,8	-1,122	-0,2989	42422	777,6	43200
Filt (m <sup>3</sup> /día)	-613,4	84,19	0,0	0,0	5585	31195578	155,8	-12,41	72000	-72000	1,330E-11
Profundidad (m)	9,141	0,1171	6,67	1,67	7,766	60,31	-0,1881	0,9642	29,58	0,42	30,0
Ta (°C)	15,36	0,0828	13,26	13,30	5,491	30,16	-0,2318	1,020	19,24	9,367	28,61
Odis (mg/l)	8,231	0,0651	9,483	0,200	4,319	18,65	-0,7031	-0,8093	15,42	0,0367	15,46
pH	8,108	0,0061	8,248	8,390	0,4028	0,1623	1,879	-1,012	3,1675	5,885	9,053
Con (µS/cm)	1013	1,513	980	970	100,4	10077	-0,8492	0,4707	370,0	845,0	1215

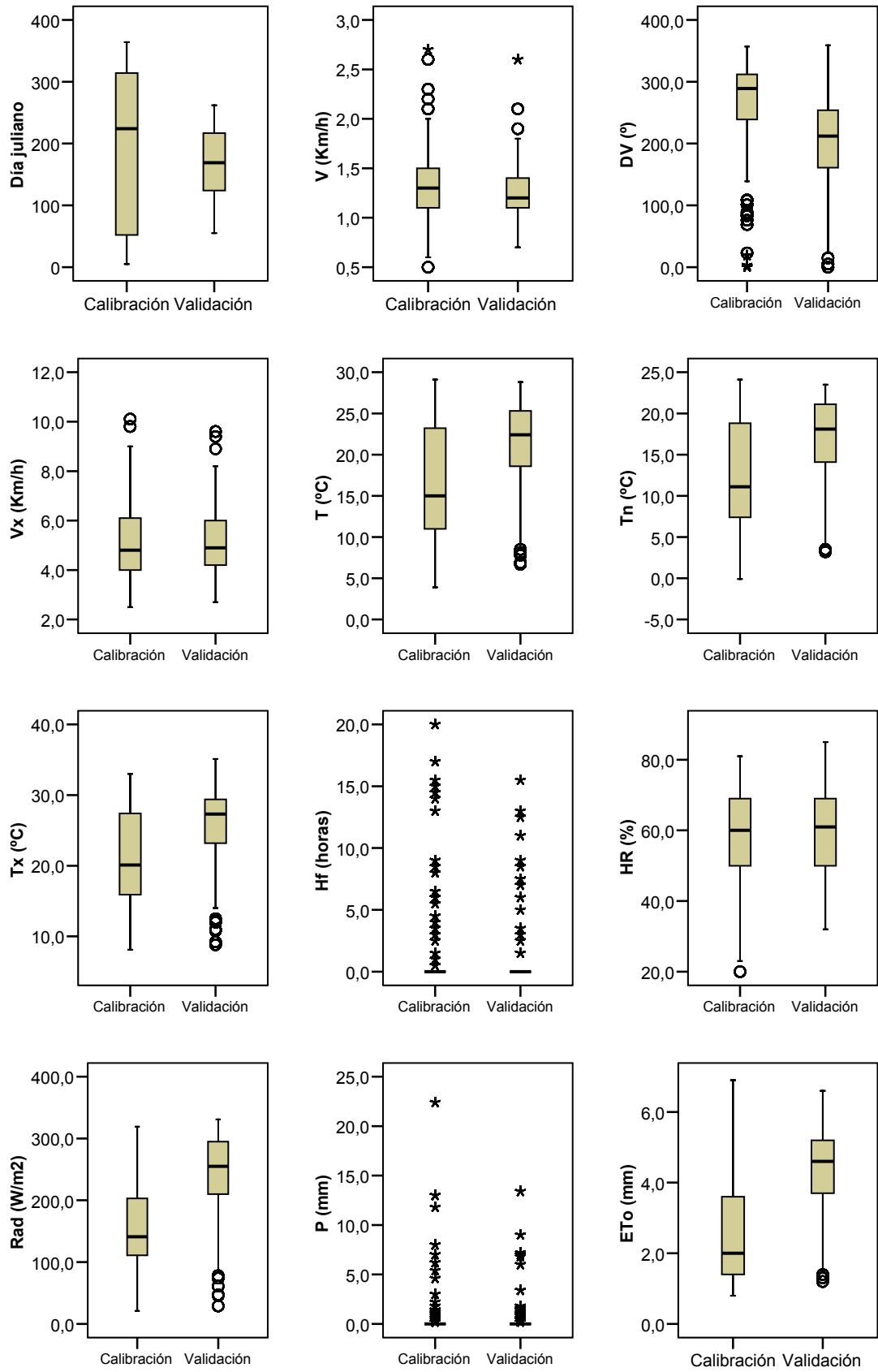
Estadísticas descriptivas del periodo de calibración (31/03/2004 a 23/02/2005)

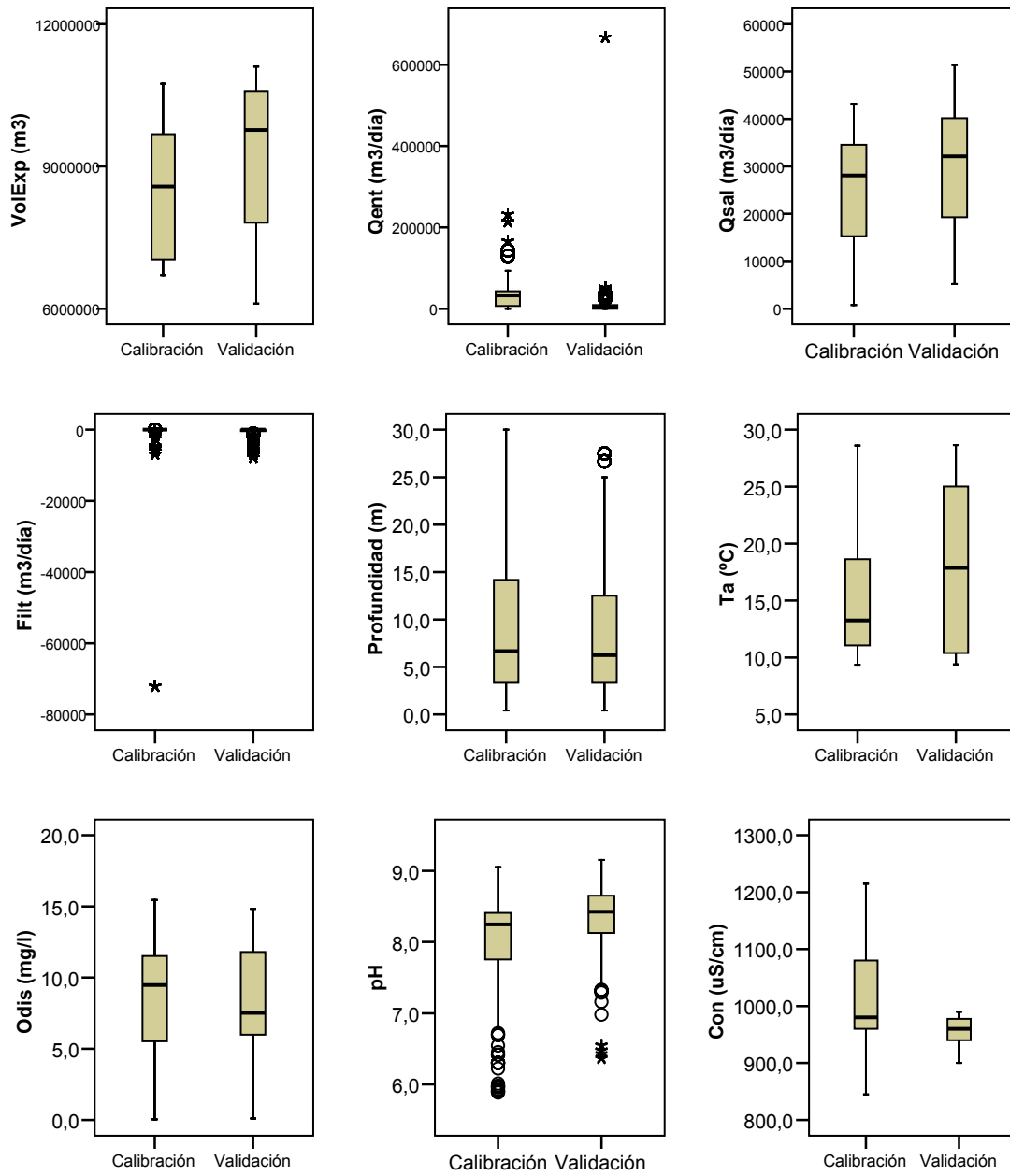
Parámetro	Media	Error típico	Mediana	Moda	Desviación estándar	Varianza de la muestra	Curtosis	Coefficiente de asimetría	Rango	Mínimo	Máximo
Día juliano	167,0	0,862	169,0	196,0	58,31	3400	-0,992	-0,2131	207,0	55,0	262,0
V (Km/h)	1,255	0,0041	1,2	1,1	0,2742	0,0752	2,402	1,060	1,9	0,7	2,6
DV (°)	205,9	1,103	212,0	237,0	74,57	5561	-0,1906	-0,403	359,0	0,0	359,0
Vx (Km/h)	5,136	0,0178	4,9	4,1	1,206	1,456	0,9403	0,9805	6,9	2,7	9,6
T (°C)	21,20	0,0763	22,4	25,6	5,162	26,65	0,3668	-1,068	22,1	6,7	28,8
Tn (°C)	16,82	0,0749	18,1	17,5	5,067	25,67	0,1693	-0,9768	20,3	3,2	23,5
Tx (°C)	25,73	0,0784	27,3	29,2	5,302	28,11	0,8118	-1,112	26,3	8,8	35,1
Hf (horas)	0,537	0,0324	0,0	0,0	2,194	4,81	21,55	4,577	15,5	0,0	15,5
HR (%)	58,71	0,1796	61,0	70,0	12,15	147,5	-0,8289	-0,2943	53,0	32,0	85,0
Rad (W/m <sup>2</sup> )	242,4	0,974	255,0	263,0	65,86	4338	0,5292	-0,9882	302,0	29,0	331,0
P (mm)	0,3550	0,0232	0,0	0,0	1,569	2,461	34,53	5,616	13,4	0,0	13,4
ETo (mm)	4,297	0,0179	4,6	4,5	1,208	1,460	-0,1259	-0,7820	5,4	1,2	6,6
VolExp (m <sup>3</sup> )	9215464	23543	9769000	10128000	1592040	2,535E+12	-1,158	-0,5323	4992000	6110000	11102000
Qent (m <sup>3</sup> /día)	12717	770,3	3351	0,0	52090	2713362517	144,5	11,74	667336	0,0	667336
Qsal (m <sup>3</sup> /día)	28606	177,5	32141	40176	12005	144120851	-0,970	-0,3675	46224	5184	51408
Filt (m <sup>3</sup> /día)	-803,0	24,81	0,0	0,0	1678	2814038	4,2	-2,22	7809	-7809	3,274E-11
Profundidad (m)	8,639	0,1107	6,250	1,25	7,488	56,08	0,0804	1,087	27,08	0,42	27,5
Ta (°C)	18,05	0,1012	17,86	9,72	6,844	46,83	-1,618	0,055	19,26	9,390	28,65
Odis (mg/l)	7,857	0,0607	7,523	0,125	4,104	16,85	-0,8161	-0,4168	14,73	0,1000	14,83
pH	8,360	0,0056	8,425	8,73	0,3753	0,1409	0,942	-0,742	2,793	6,36	9,153
Con (µS/cm)	957	0,3359	960	980	22,7	515,9	-0,5290	-0,6820	90,0	900,0	990

Estadísticas descriptivas del periodo de validación (24/02/2005 a 19/09/2005)



**ANEXO 3. DIAGRAMAS DE CAJA Y BIGOTES AGRUPADOS EN PERIODOS DE CALIBRACIÓN Y VALIDACIÓN PARA CADA UNA DE LAS VARIABLES**

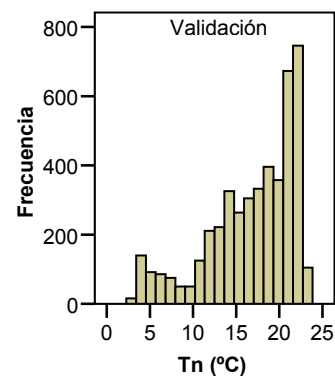
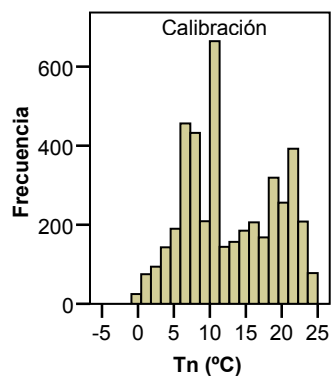
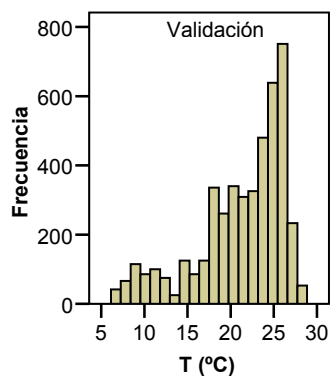
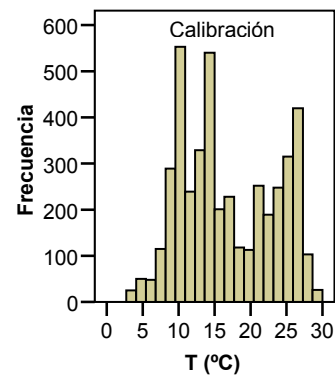
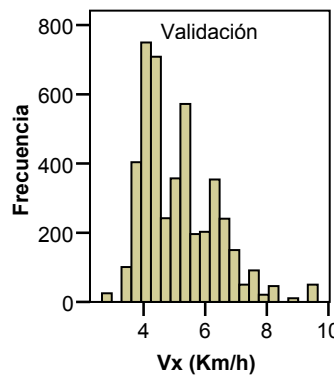
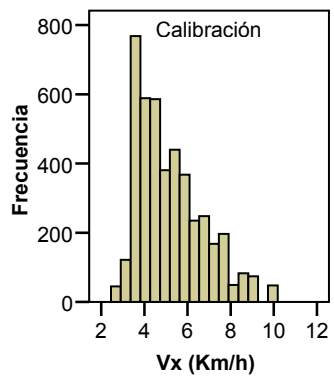
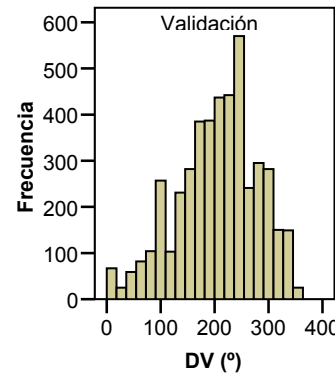
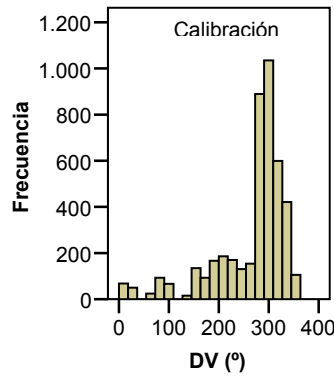
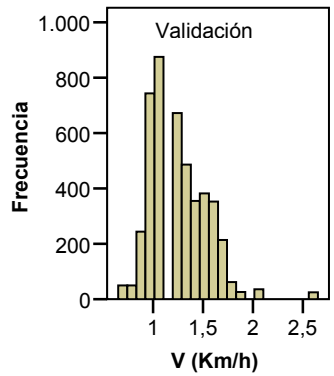
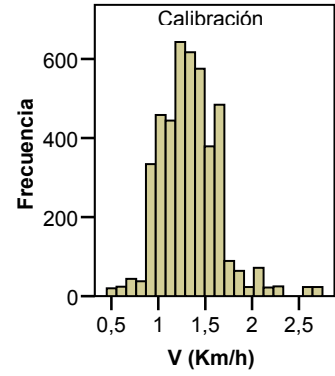
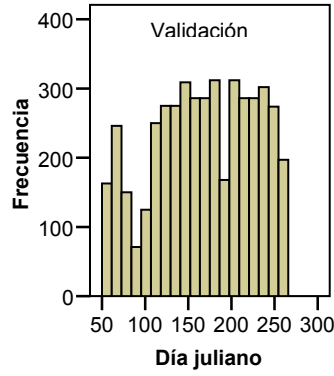
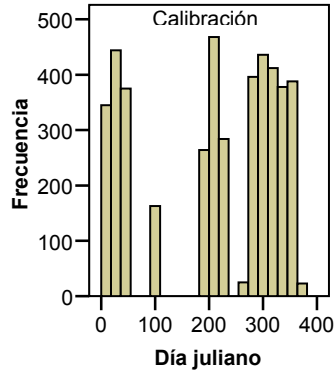


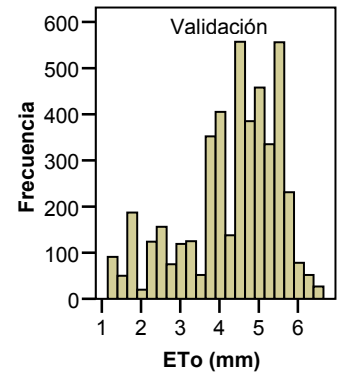
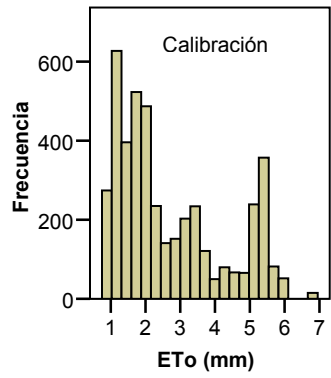
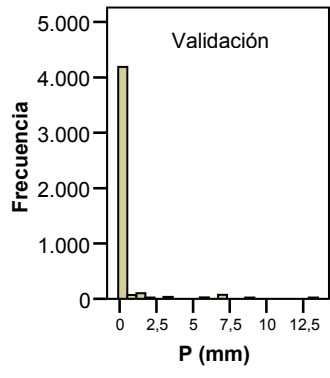
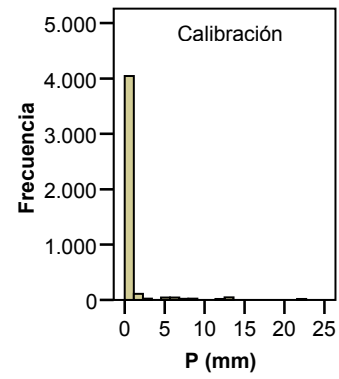
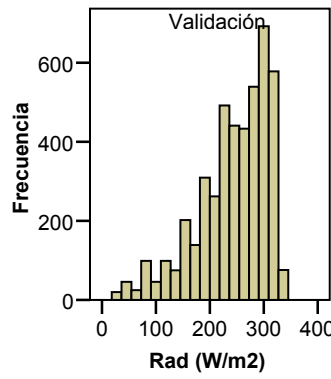
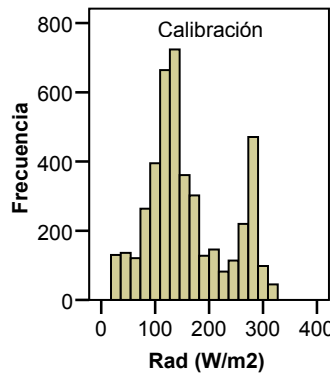
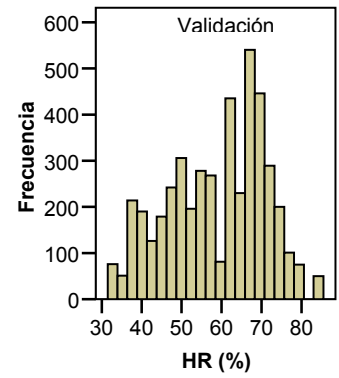
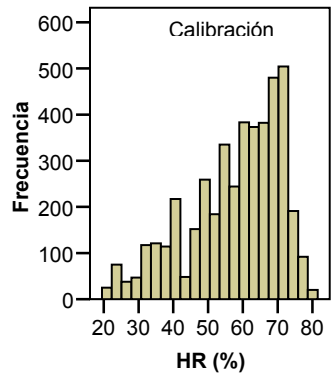
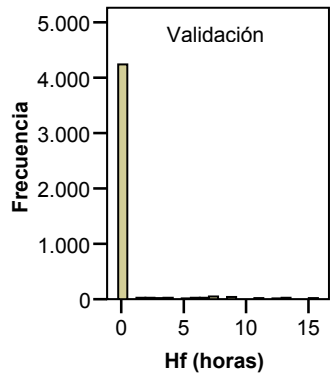
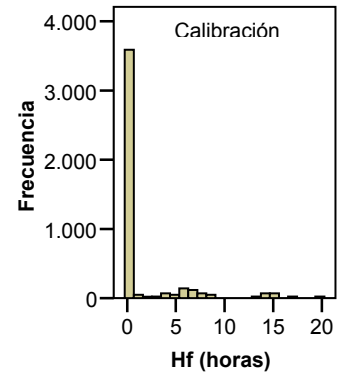
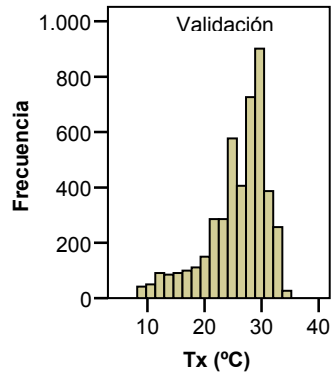
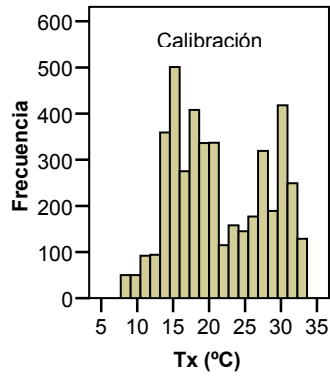


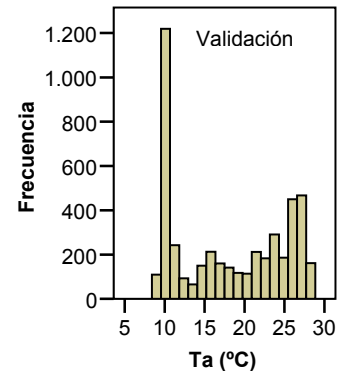
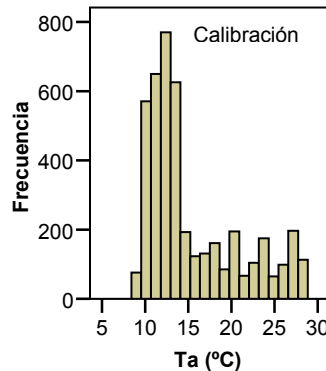
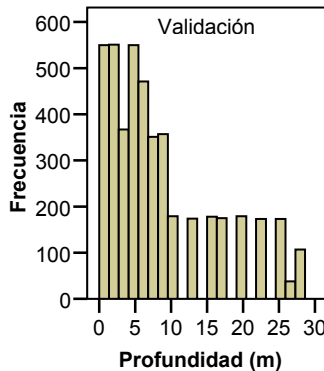
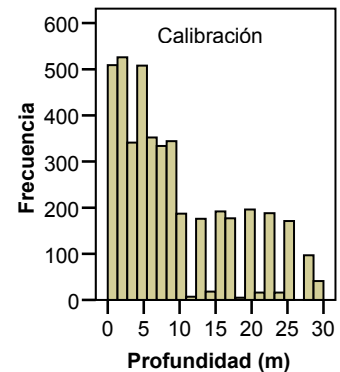
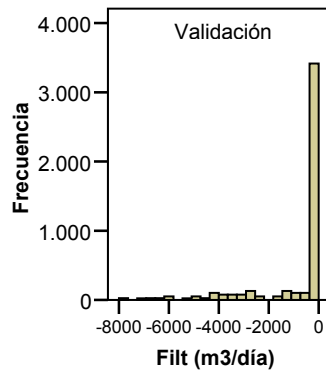
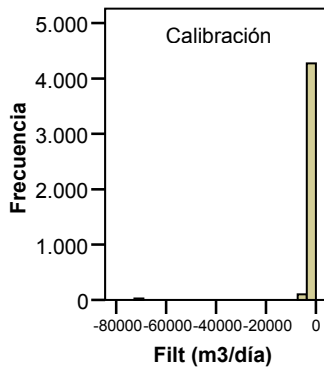
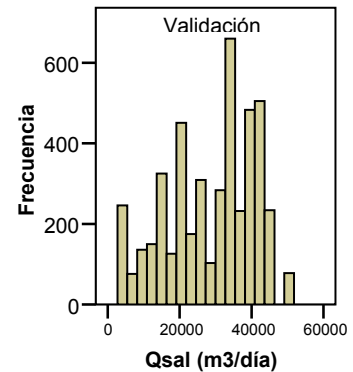
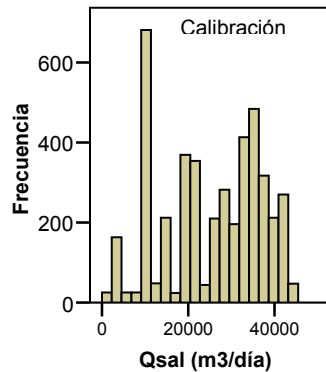
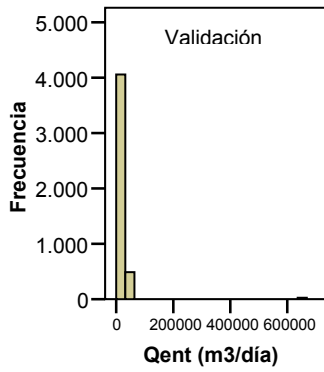
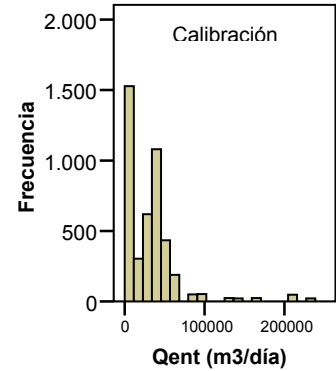
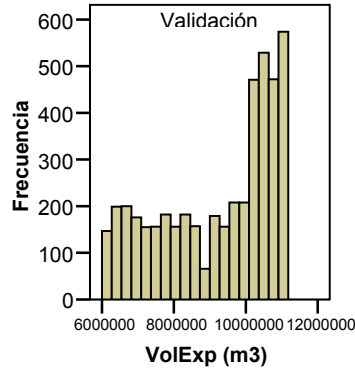
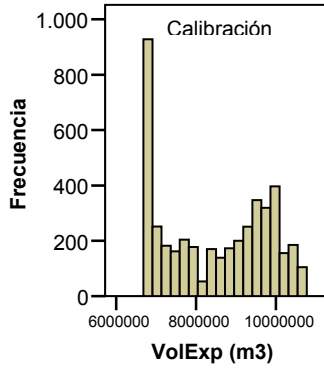
Periodo de calibración (31/03/2004 a 23/02/2005)

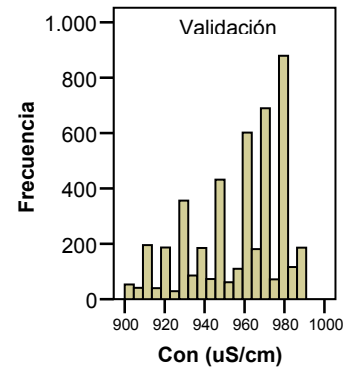
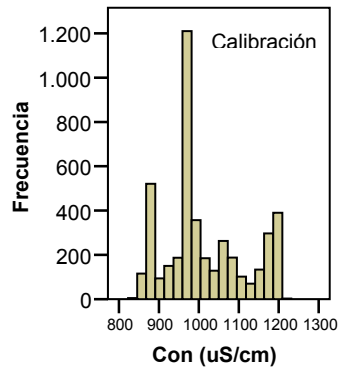
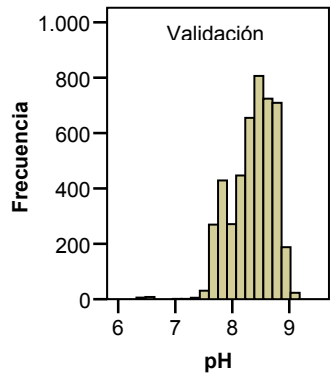
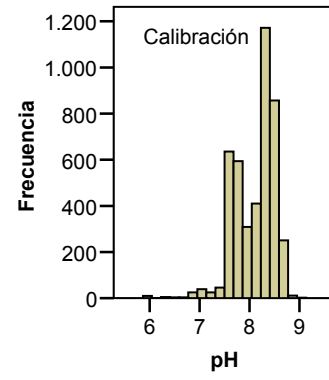
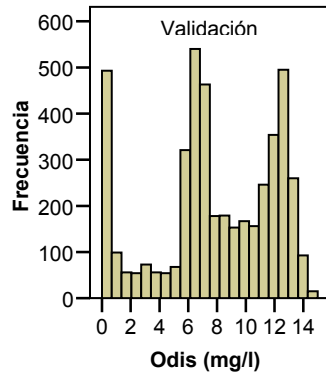
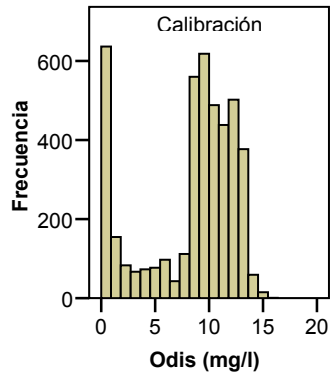
Periodo de validación (24/02/2005 a 19/09/2005)

**ANEXO 4. HISTOGRAMAS DE LOS DATOS INICIALES EN  
PERIODOS DE CALIBRACIÓN Y VALIDACIÓN POR  
CADA VARIABLE**









Periodo de calibración (31/03/2004 a 23/02/2005)

Periodo de validación (24/02/2005 a 19/09/2005)



**ANEXO 5. DIAGRAMAS Q-Q DE DATOS INICIALES PARA EL  
PERIODO COMPLETO DE ESTUDIO**

Gráfico Q-Q normal de Día juliano

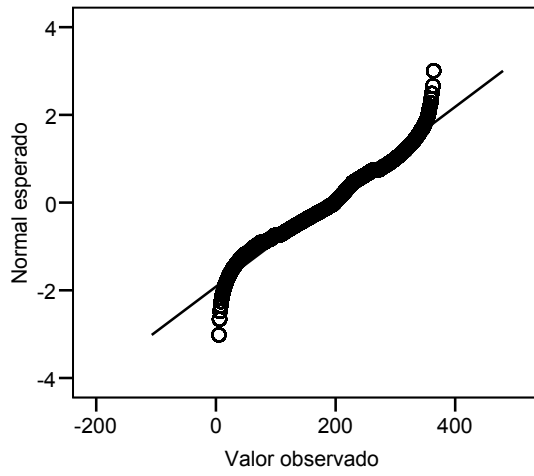


Gráfico Q-Q normal de V (Km/h)

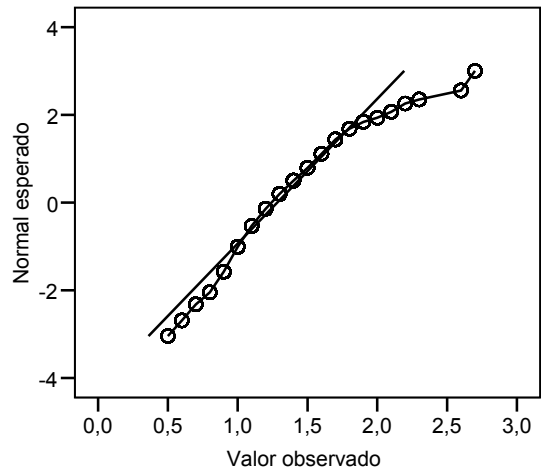


Gráfico Q-Q normal de DV (°)

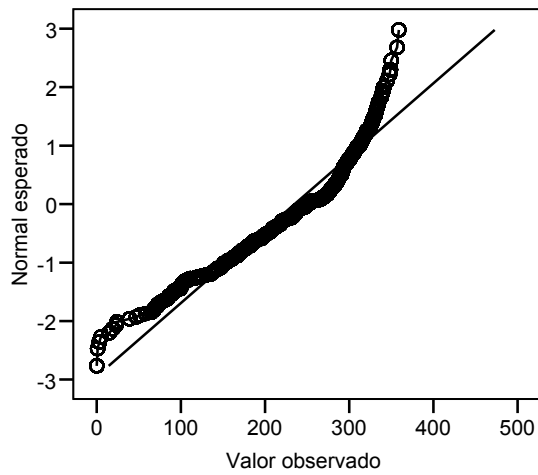


Gráfico Q-Q normal de Vx (Km/h)

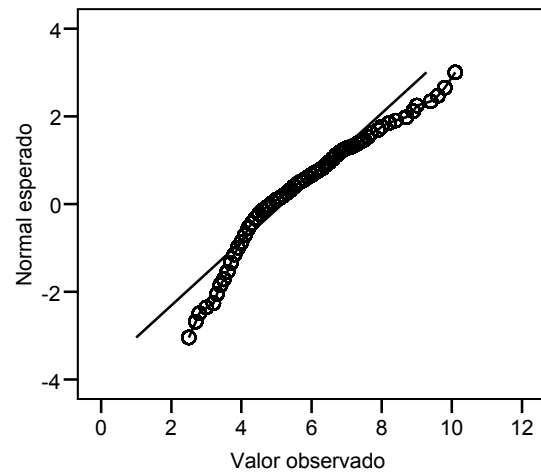


Gráfico Q-Q normal de T (°C)

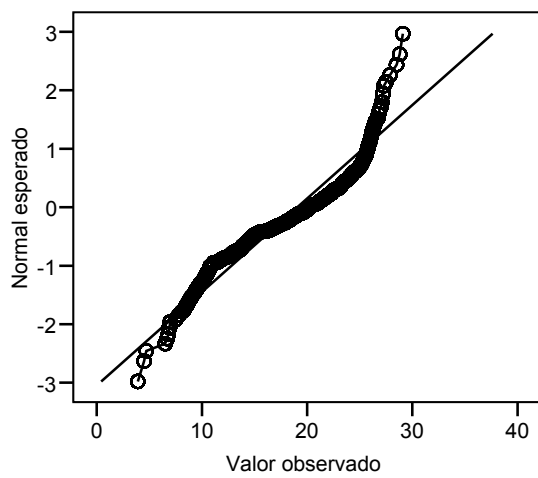


Gráfico Q-Q normal de Tn (°C)

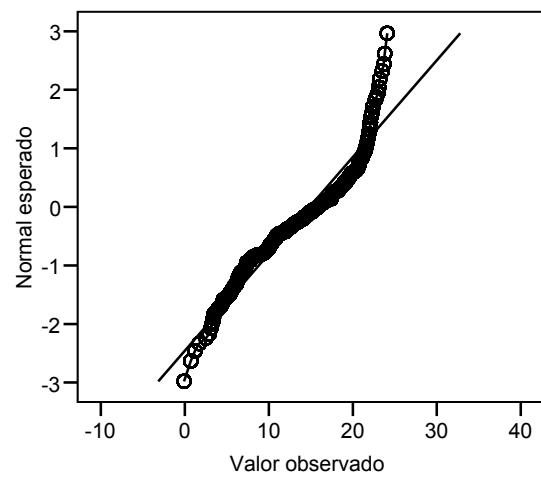


Gráfico Q-Q normal de Tx (°C)

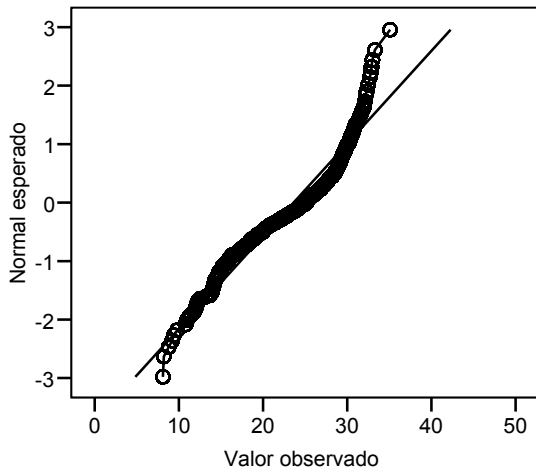


Gráfico Q-Q normal de Hf (horas)

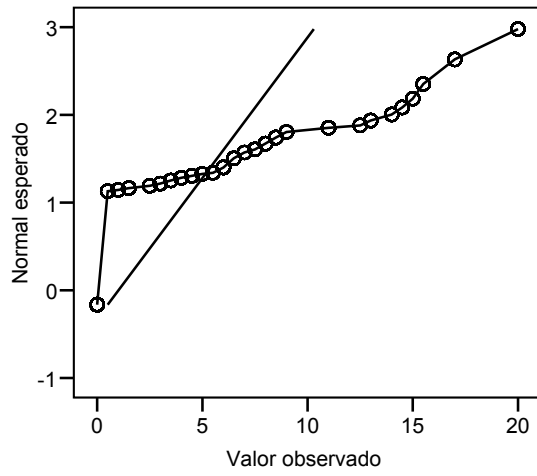


Gráfico Q-Q normal de HR (%)

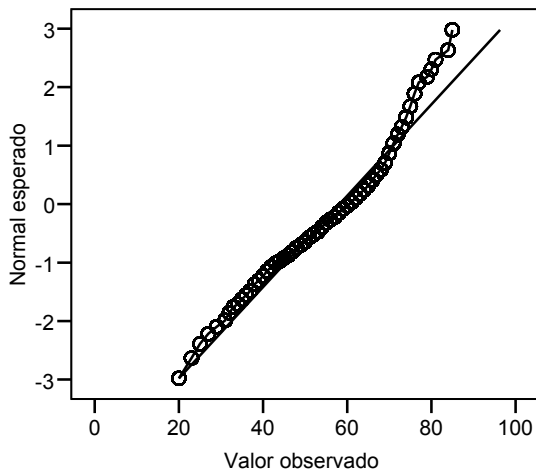


Gráfico Q-Q normal de Rad (W/m2)

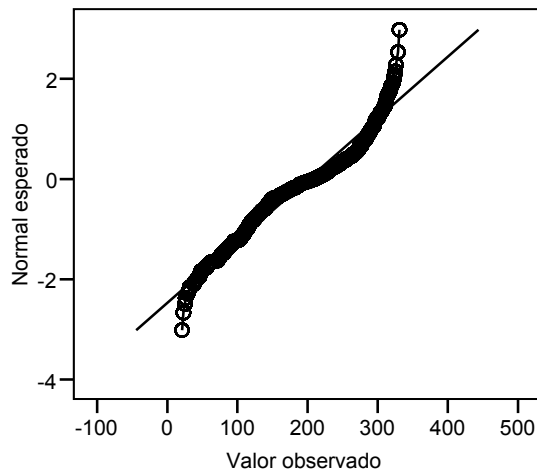


Gráfico Q-Q normal de P (mm)

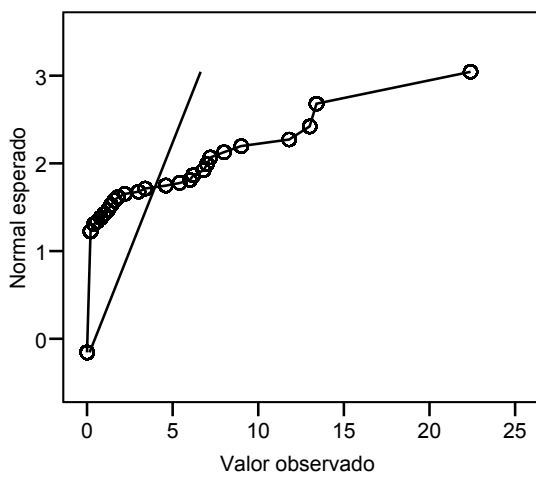


Gráfico Q-Q normal de ETo (mm)

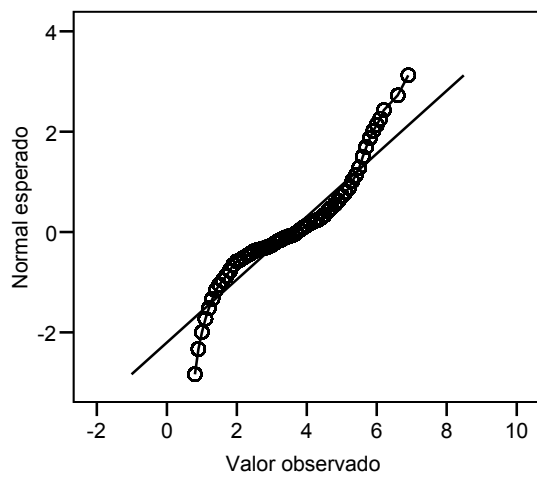


Gráfico Q-Q normal de VolExp (m3)

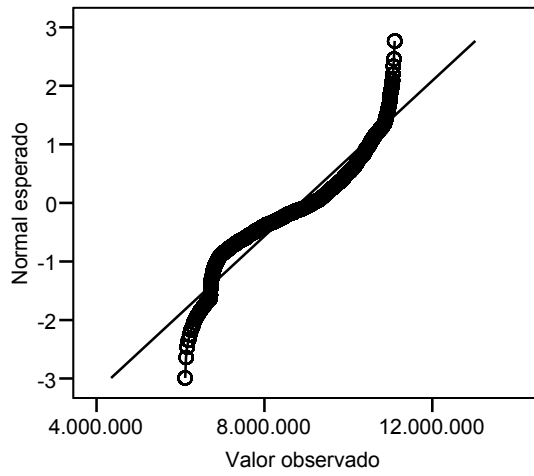


Gráfico Q-Q normal de Qent (m3/día)

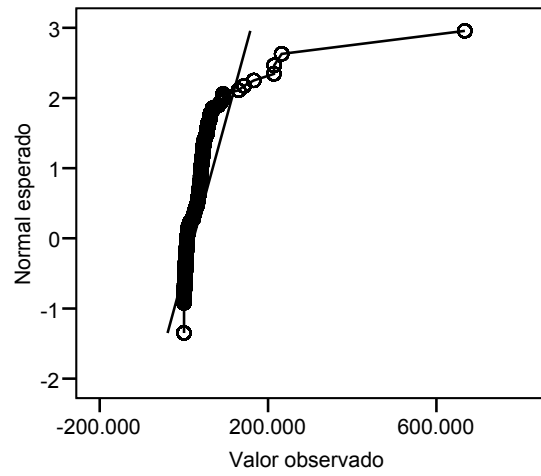


Gráfico Q-Q normal de Qsal (m3/día)

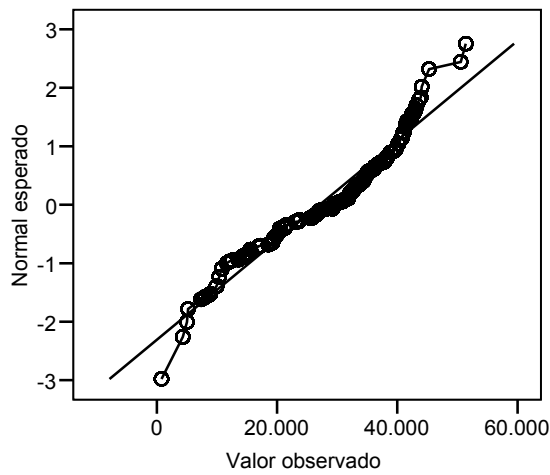


Gráfico Q-Q normal de Filt (m3/día)

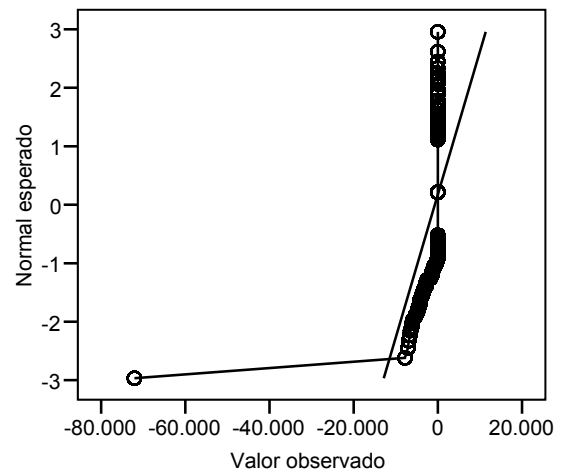


Gráfico Q-Q normal de Profundidad (m)

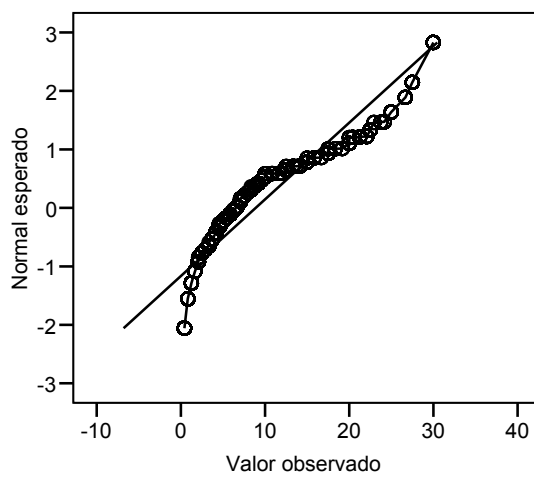


Gráfico Q-Q normal de Ta (°C)

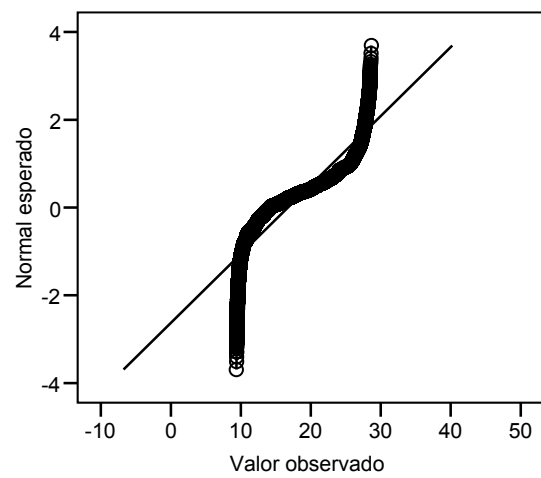


Gráfico Q-Q normal de Odis (mg/l)

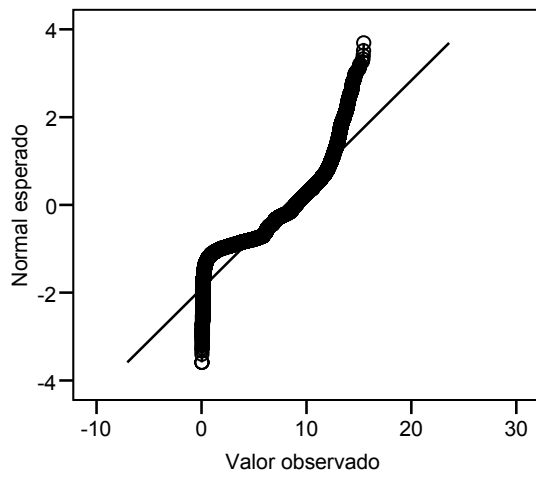


Gráfico Q-Q normal de pH

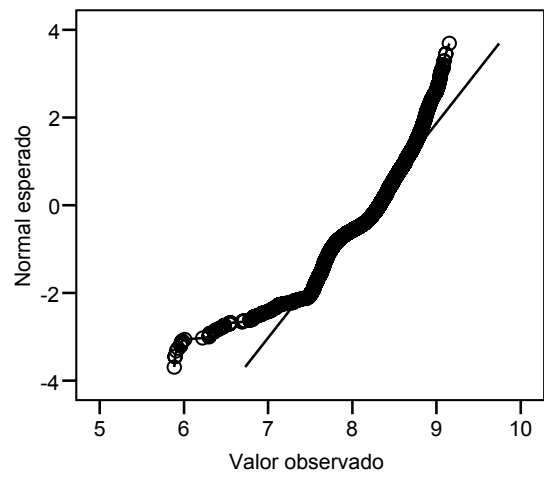
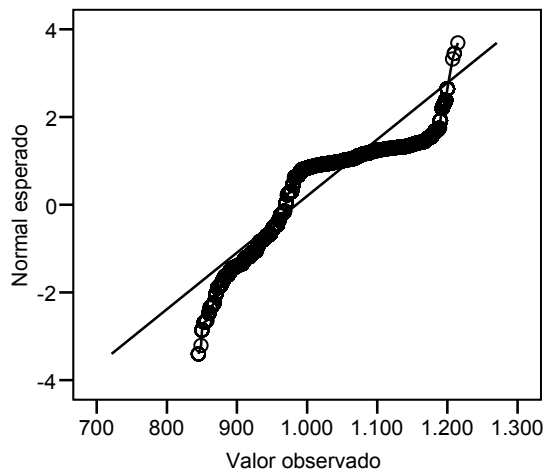
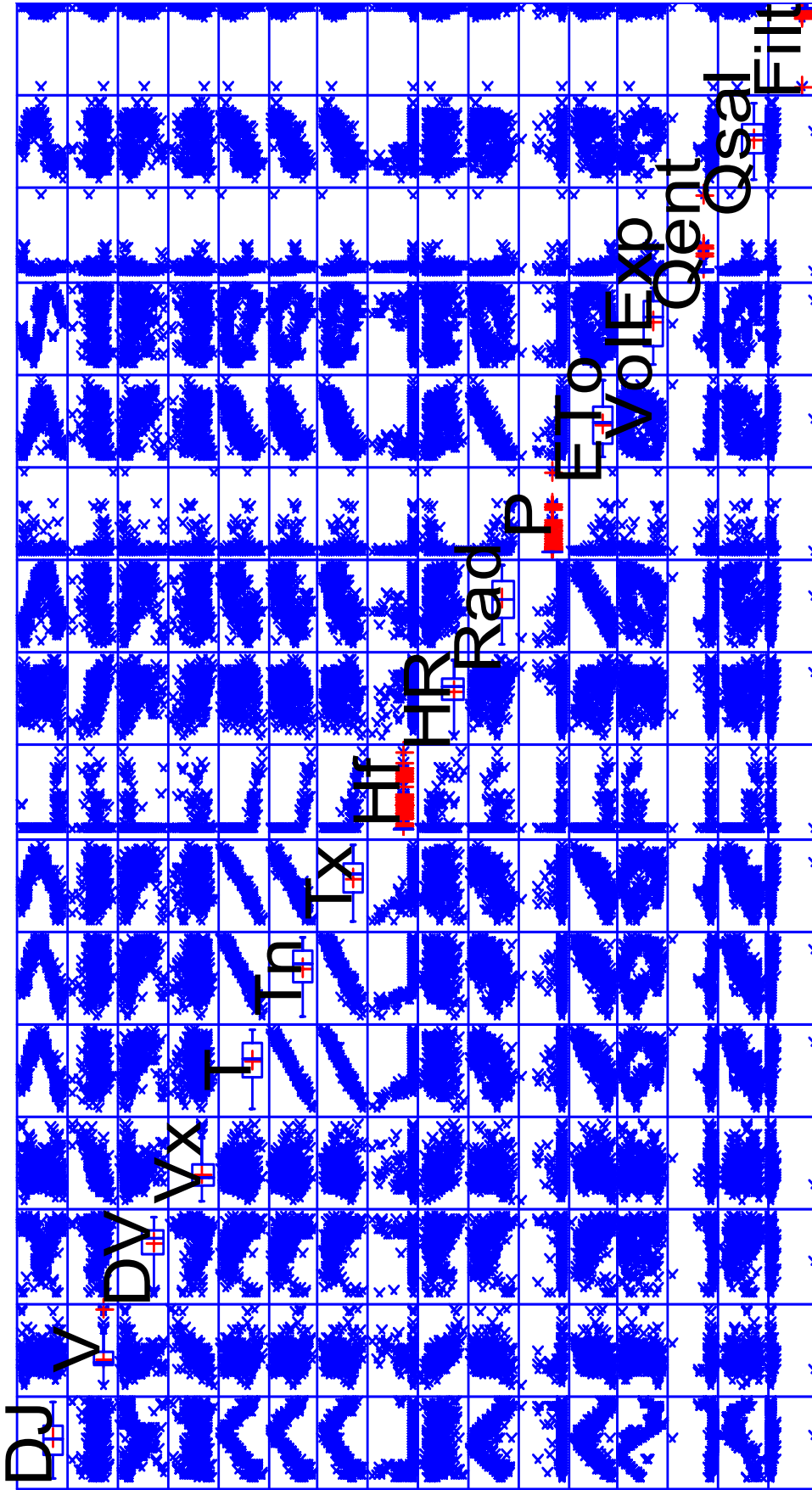


Gráfico Q-Q normal de Con (uS/cm)

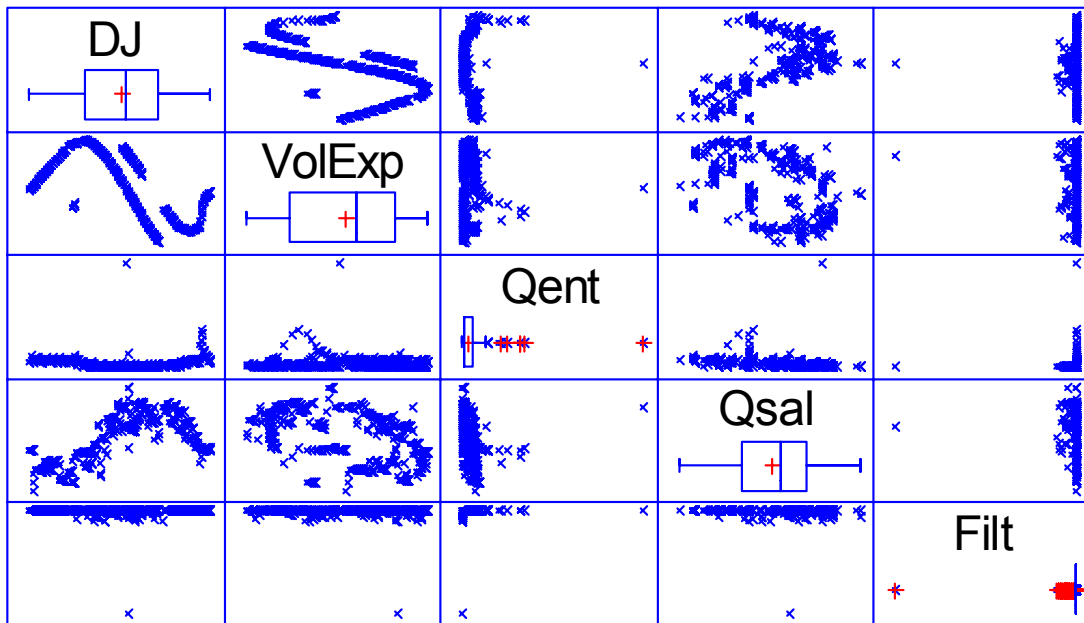


Periodo de estudio completo (31/03/2004 a 19/09/2005)

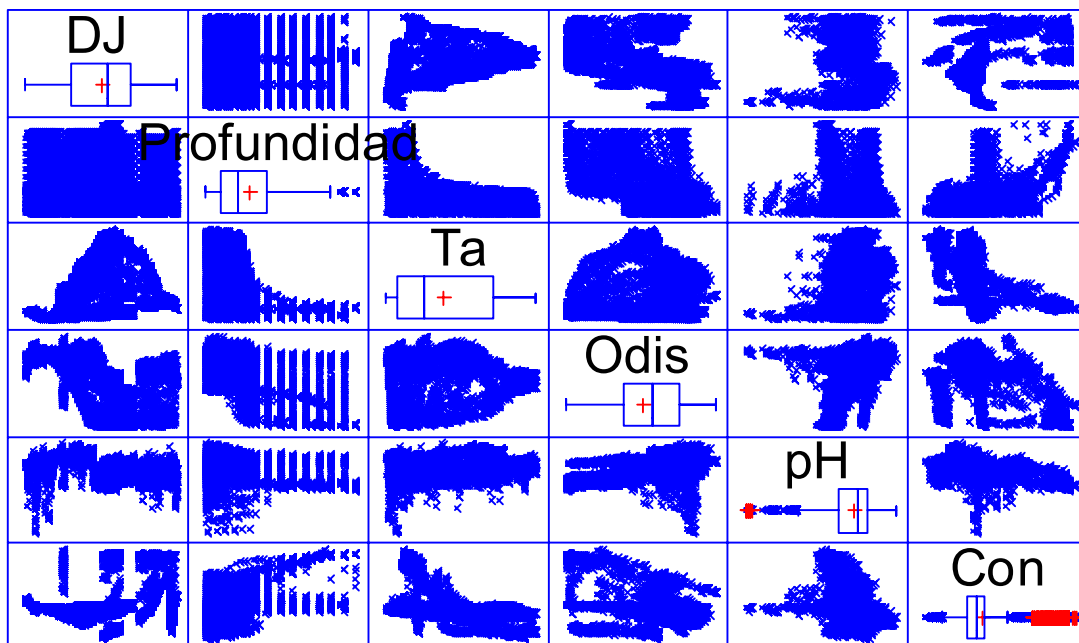
**ANEXO 6. GRÁFICOS DE DISPERSIÓN MATRICIAL PARA LOS  
DATOS METEOROLÓGICOS, HIDROLÓGICOS Y DE  
CALIDAD DEL AGUA EN EL PERIODO COMPLETO  
DE ESTUDIO**



DATOS METEOROLÓGICOS



DATOS HIDROLÓGICOS



DATOS DE CALIDAD DEL AGUA



**ANEXO 7. PCA DE LOS DATOS METEOROLÓGICOS,  
HIDROLÓGICOS Y DE CALIDAD DEL AGUA EN  
EL PERIODO DE CALIBRACIÓN, DATOS  
ORIGINALES**

Variable	Componentes						
	1	2	3	4	5	6	7
DJ	0,134344	0,371215	-0,069300	-0,332976	-0,001983	-0,077019	0,078531
V	-0,054194	-0,258480	0,274217	-0,404803	-0,135416	-0,081706	0,058346
DV	-0,206907	0,065572	0,028698	-0,260413	0,331011	0,079670	0,045238
Vx	-0,024976	-0,162526	0,206584	-0,408870	-0,485762	-0,134045	-0,112845
T	0,375898	0,018683	-0,023962	-0,010627	-0,062084	0,037994	0,002141
Tn	0,372289	0,048966	-0,039415	-0,003146	-0,053546	0,023360	-0,027970
Tx	0,371293	0,003470	-0,024354	-0,003343	-0,031251	0,069052	0,042544
Hf	-0,198504	-0,251637	0,136428	0,106961	0,369485	-0,111238	-0,266187
HR	0,018820	0,247806	-0,270606	0,426459	-0,119554	0,011712	-0,017541
Rad	0,290548	-0,261377	0,135362	0,108942	0,010109	0,103157	0,038903
P	-0,061232	0,229358	-0,117962	0,003346	-0,293545	-0,243116	-0,786717
ETo	0,336220	-0,205651	0,127691	0,004569	-0,077207	0,037001	-0,070170
VolExp	0,082030	-0,404050	0,086351	0,353197	-0,175572	-0,002178	-0,092796
Qent	-0,216803	0,116647	-0,064483	0,034106	-0,401812	0,099364	0,324327
Qsal	0,285674	0,200727	-0,066627	-0,217541	0,231297	-0,041039	-0,040710
Filt	-0,063614	0,018968	0,022872	-0,136527	-0,013260	0,912368	-0,333095
Profundidad	0,013624	0,175023	0,432751	0,216761	-0,143239	0,079461	0,072626
Ta	0,282310	-0,080991	-0,299427	-0,172867	0,090590	-0,031996	-0,052403
Odis	-0,227576	-0,254965	-0,358857	-0,098802	0,004540	-0,002016	-0,005006
pH	-0,060446	-0,098104	-0,407869	-0,054493	-0,327100	0,142024	0,198437
Con	-0,046205	0,380305	0,380468	0,058782	-0,060759	0,034680	0,051922

Variable	Componentes						
	8	9	10	11	12	13	14
DJ	-0,293678	0,090536	0,260681	-0,157436	-0,095358	-0,067833	0,046454
V	0,043012	0,183877	-0,122506	-0,519624	-0,427793	-0,086769	0,064140
DV	0,464746	-0,645168	0,062962	0,101059	-0,303389	-0,075204	0,045167
Vx	-0,031457	0,032041	-0,107330	0,615732	-0,148544	0,165686	-0,021042
T	0,041643	-0,119827	0,022591	0,139539	0,031754	0,010538	0,113979
Tn	0,043839	-0,095084	-0,002888	0,083291	-0,040305	-0,042518	0,095300
Tx	0,069888	-0,133497	0,024468	0,188531	0,079414	0,056787	0,144930
Hf	-0,236517	0,104339	0,497707	0,165630	-0,186908	0,403792	0,165671
HR	-0,006320	0,084101	-0,258214	0,075829	-0,733706	0,174242	0,004711
Rad	-0,008098	-0,223508	0,000558	-0,278499	-0,028837	0,417076	-0,072646
P	0,124791	-0,206310	0,110440	-0,253398	0,092409	-0,017187	-0,000608
ETo	-0,003531	-0,166346	0,021497	-0,169899	0,001493	0,153511	-0,014988
VolExp	-0,022236	-0,087474	0,130887	-0,003042	-0,123173	-0,438504	0,312940
Qent	-0,460393	-0,453663	0,340809	-0,069205	-0,050992	0,006480	0,057539
Qsal	0,004741	0,208297	0,210501	0,068588	-0,142872	-0,115519	0,426799
Filt	-0,132280	0,095802	-0,030938	0,015501	-0,046684	-0,044975	-0,001337
Profundidad	0,346183	0,195275	0,458103	0,093855	-0,085971	-0,140529	-0,410947
Ta	-0,088770	-0,025300	0,162017	-0,023564	-0,140438	0,105264	-0,601896
Odis	0,043828	0,017605	-0,079119	0,048528	0,111133	0,158835	0,013218
pH	0,498674	0,236688	0,353621	-0,127735	0,061703	0,249076	0,220675
Con	0,080355	-0,014925	-0,165606	-0,110967	0,136319	0,482291	0,227522

Variable	Componentes						
	15	16	17	18	19	20	21
DJ	-0,625818	-0,006590	-0,060728	0,340967	0,028351	-0,025519	0,015575
V	0,201171	0,279438	0,097795	0,004513	0,123681	-0,062045	-0,012638
DV	-0,131951	-0,021695	-0,040586	-0,011926	-0,013454	0,015827	-0,006227
Vx	-0,087916	-0,198517	-0,041896	-0,024929	-0,049228	-0,039546	0,017872
T	0,046784	0,306190	0,122081	0,078308	0,020136	-0,073746	-0,824230
Tn	0,060432	0,428077	0,080407	0,021992	-0,641607	-0,238728	0,400401
Tx	0,092214	0,208158	0,189849	0,156246	0,703254	0,009205	0,397046
Hf	-0,034104	0,216474	0,175526	-0,033158	-0,009426	0,006954	0,000683
HR	-0,017861	-0,023229	0,017143	0,097651	0,037099	0,089308	-0,011015
Rad	-0,008369	-0,474884	-0,002590	0,273004	-0,049035	-0,441101	-0,008602
P	0,102864	-0,026868	0,042909	0,048303	0,060678	-0,054570	0,002369
ETo	-0,081856	-0,059905	-0,057920	0,037824	-0,167234	0,831155	0,037685
VolExp	-0,322193	0,020035	-0,410922	-0,155588	0,105985	-0,138141	0,010118
Qent	0,327291	0,027132	-0,006862	-0,062192	-0,014769	0,029431	0,006022
Qsal	0,469575	-0,374440	-0,307893	-0,050935	-0,039986	0,037407	-0,007679
Filt	-0,018851	0,013333	-0,006416	-0,008513	0,005281	-0,011096	-0,003443
Profundidad	0,147011	0,048170	-0,133539	0,296566	-0,005660	0,020107	-0,002245
Ta	0,008323	0,118702	-0,333035	-0,448564	0,137750	-0,092340	0,006012
Odis	0,153653	0,262314	-0,522078	0,573777	-0,012174	0,040440	-0,005030
pH	-0,154178	-0,051814	0,147814	-0,200886	-0,038980	0,007847	0,002554
Con	-0,093373	0,234539	-0,450250	-0,271146	0,073434	-0,051773	0,003629

**Tabla de pesos de los componentes**

Esta tabla muestra las ecuaciones de los componentes principales. Por ejemplo, el primer componente principal tiene la ecuación:

$$\begin{aligned} \text{Componente 1} = & 0,134344(\text{DJ}) - 0,054194(\text{V}) - 0,206907(\text{DV}) - 0,024976(\text{Vx}) + \\ & 0,375898(\text{T}) + 0,372289(\text{Tn}) + 0,371293(\text{Tx}) - 0,198504(\text{Hf}) + 0,018820(\text{HR}) + \\ & 0,290548(\text{Rad}) - 0,061232(\text{P}) + 0,336220(\text{ETo}) + 0,082030(\text{VolExp}) - 0,216803(\text{Qent}) \\ & + 0,285674(\text{Qsal}) - 0,063614(\text{Filt}) + 0,013624(\text{Profundidad}) + 0,282310(\text{Ta}) - \\ & 0,227576(\text{Odis}) - 0,060446(\text{pH}) - 0,046205(\text{Con}) \end{aligned}$$

Donde los valores de las variables en la ecuación están estandarizados (substrayendo sus medias y dividiéndolos por sus desviaciones típicas).

Diagrama de dispersión

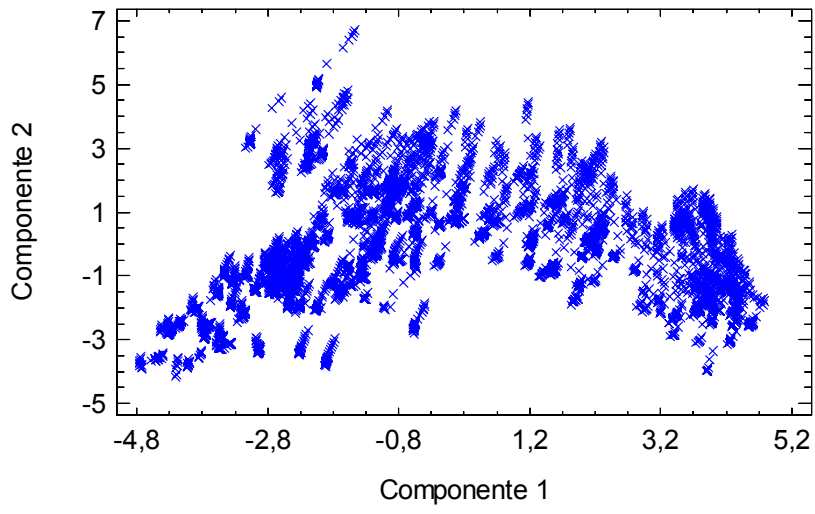


Diagrama de dispersión

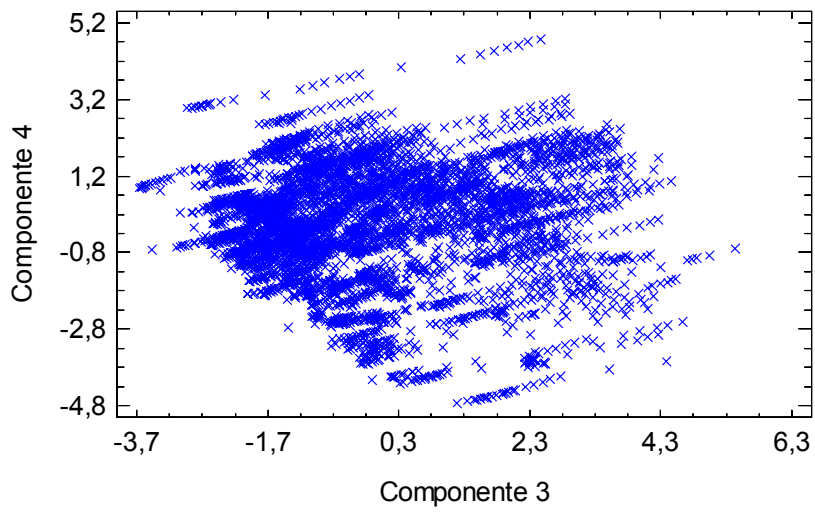


Diagrama de dispersión

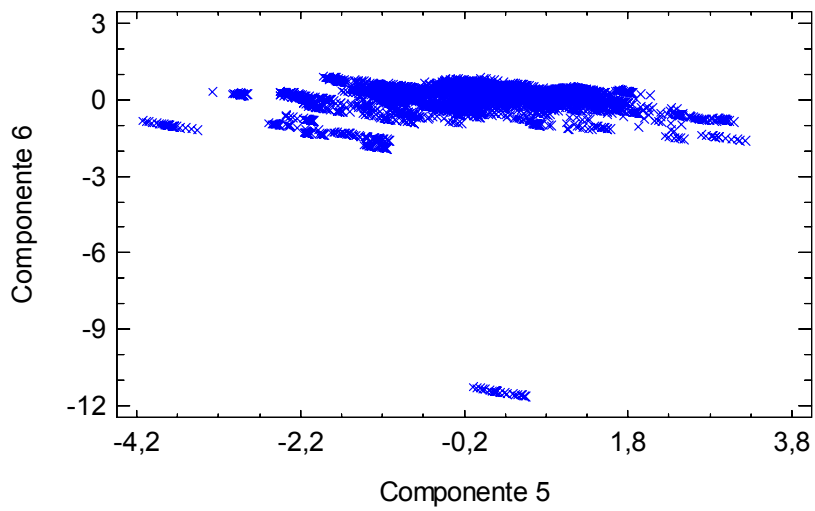


Diagrama de dispersión

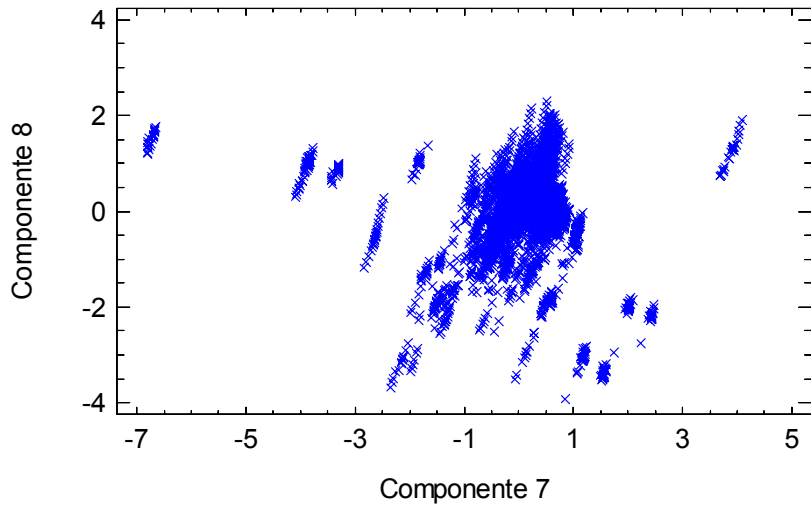


Diagrama de dispersión

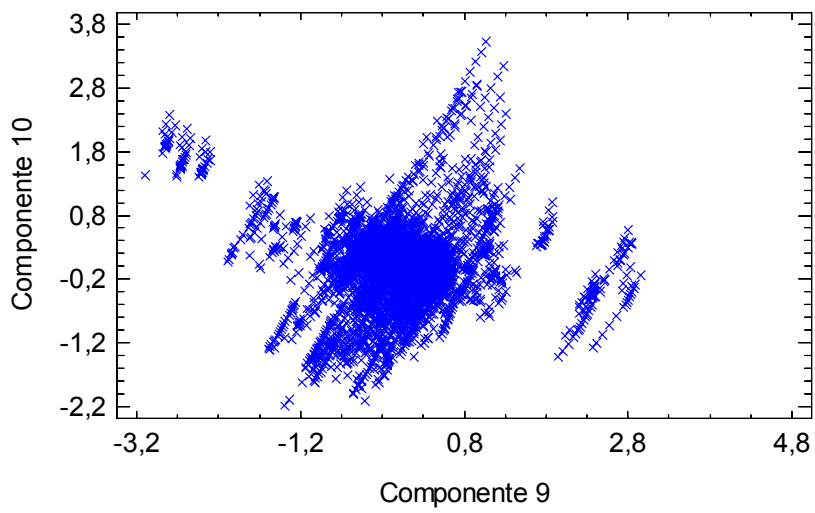


Diagrama de dispersión

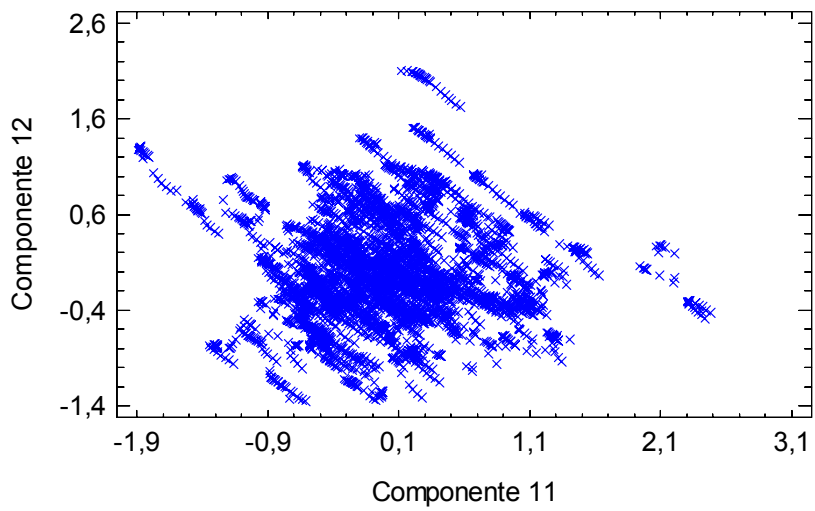


Diagrama de dispersión

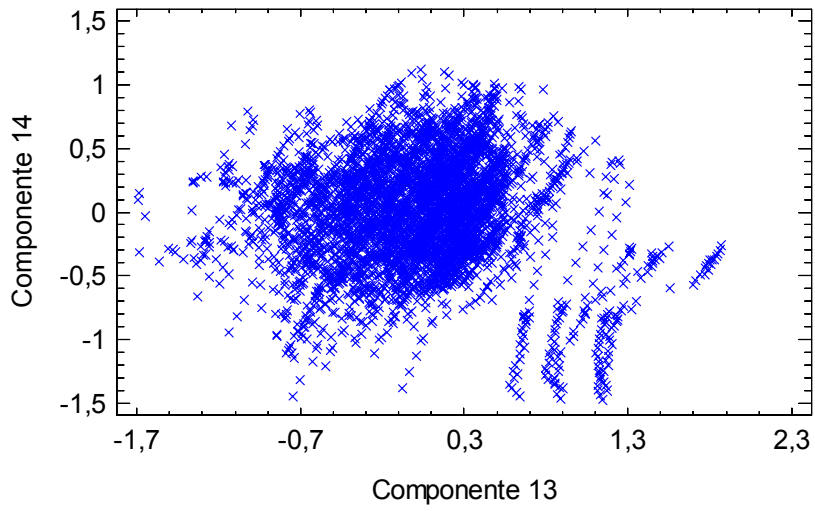


Diagrama de dispersión

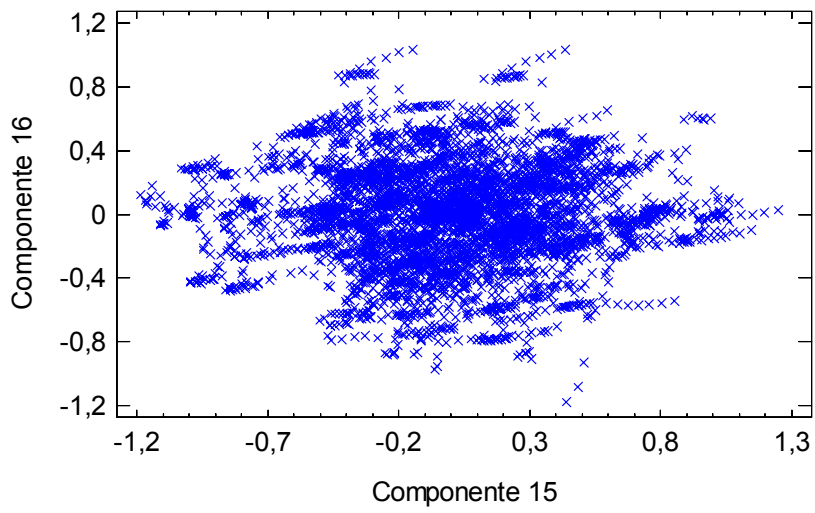


Diagrama de dispersión

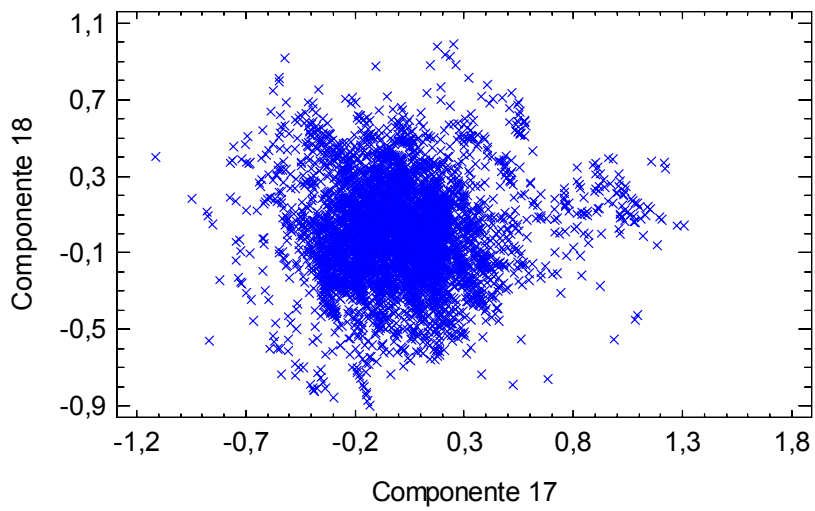


Diagrama de dispersión

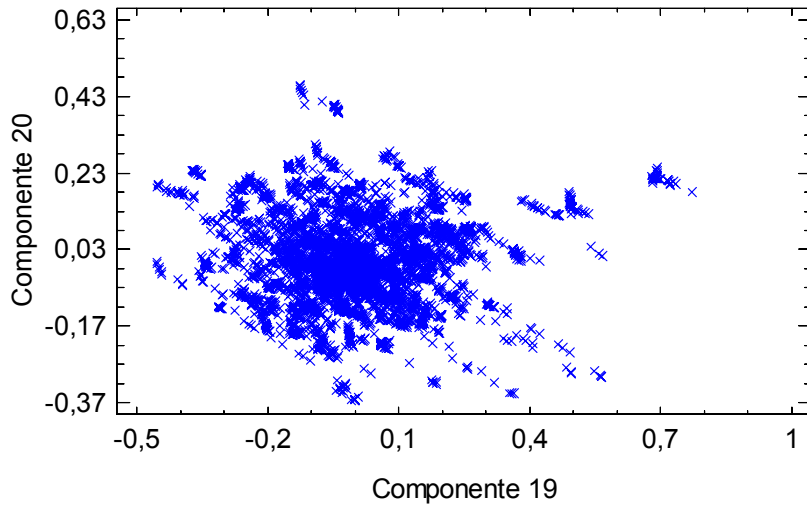
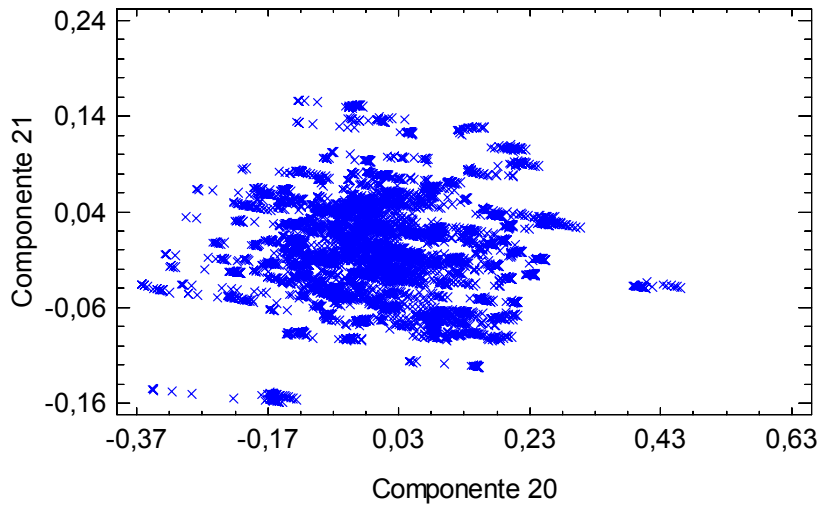


Diagrama de dispersión



**Diagramas de dispersión de los componentes por pares**

Gráfico de Pesos del Componente

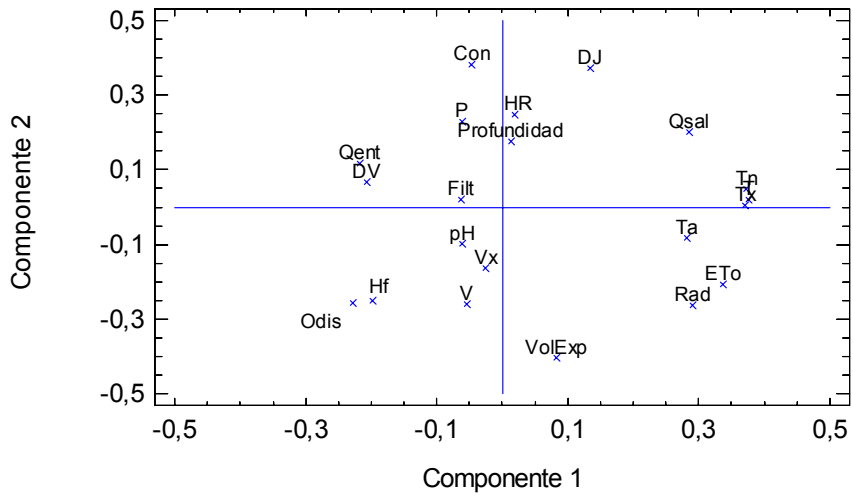


Gráfico de Pesos del Componente

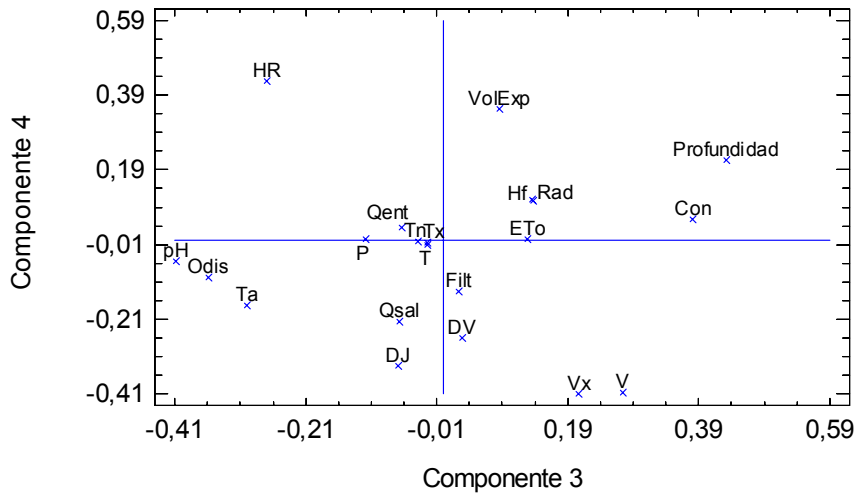


Gráfico de Pesos del Componente

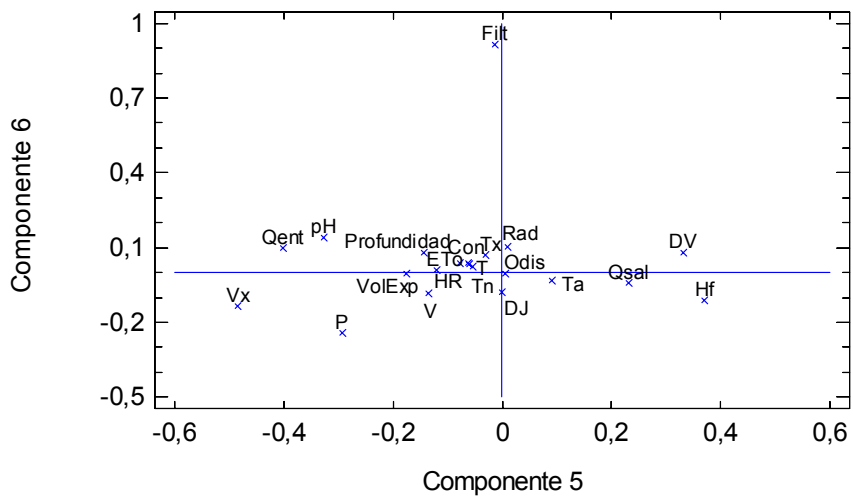




Gráfico de Pesos del Componente

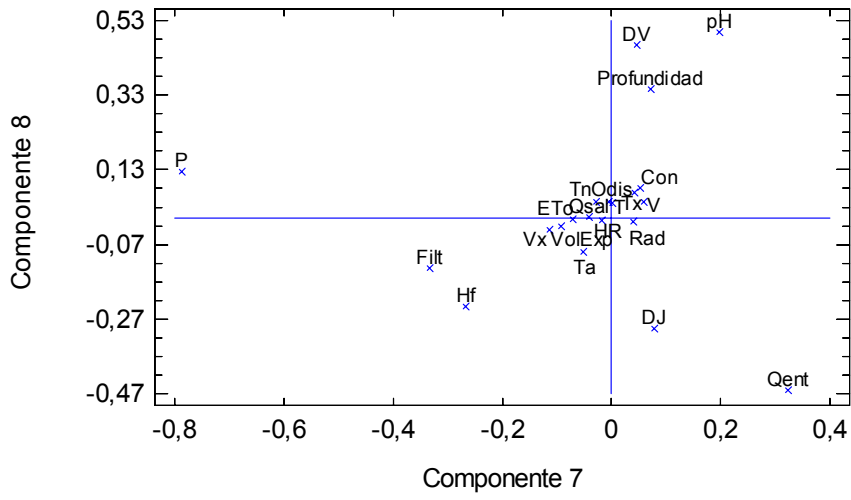


Gráfico de Pesos del Componente

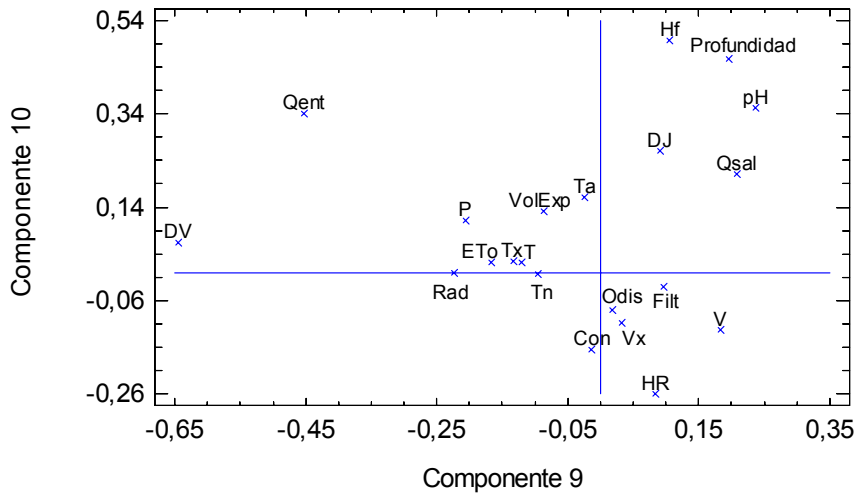


Gráfico de Pesos del Componente

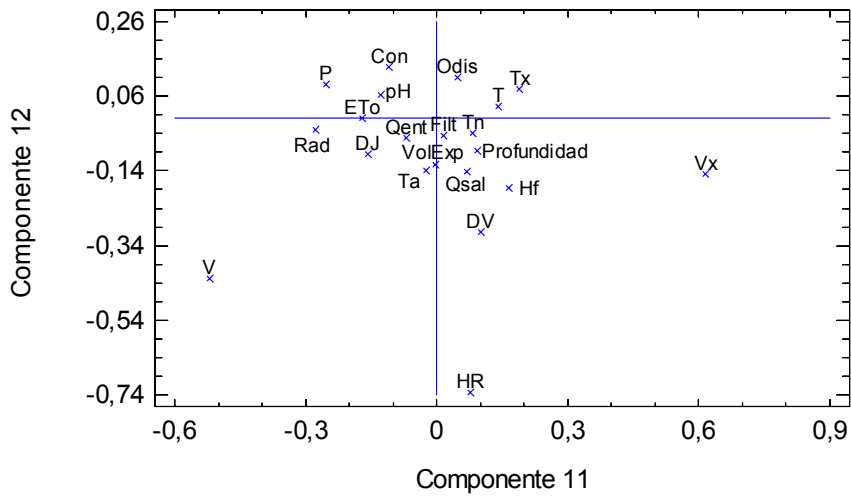


Gráfico de Pesos del Componente

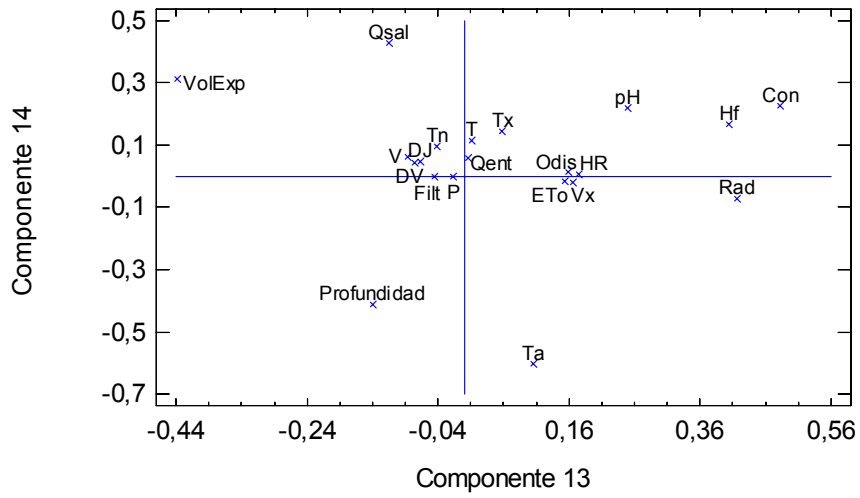


Gráfico de Pesos del Componente

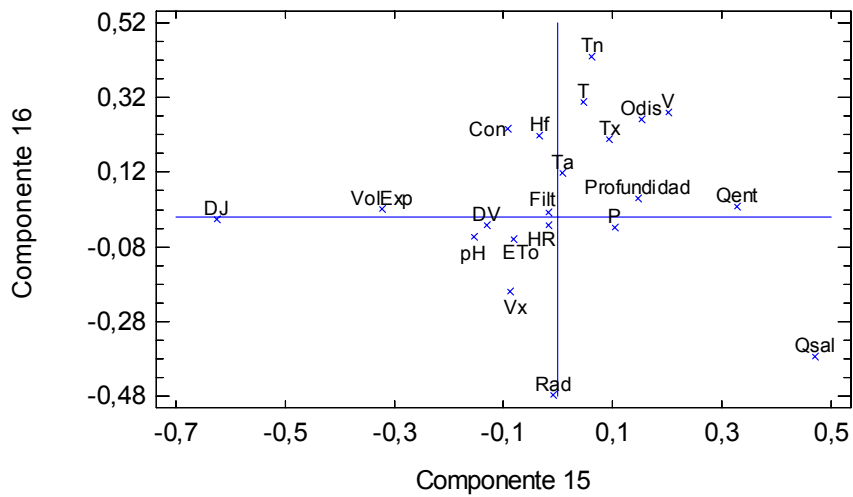


Gráfico de Pesos del Componente

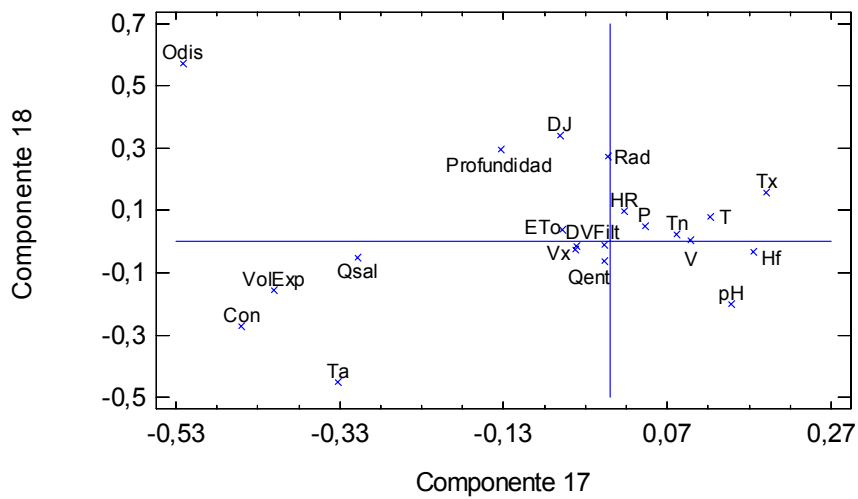


Gráfico de Pesos del Componente

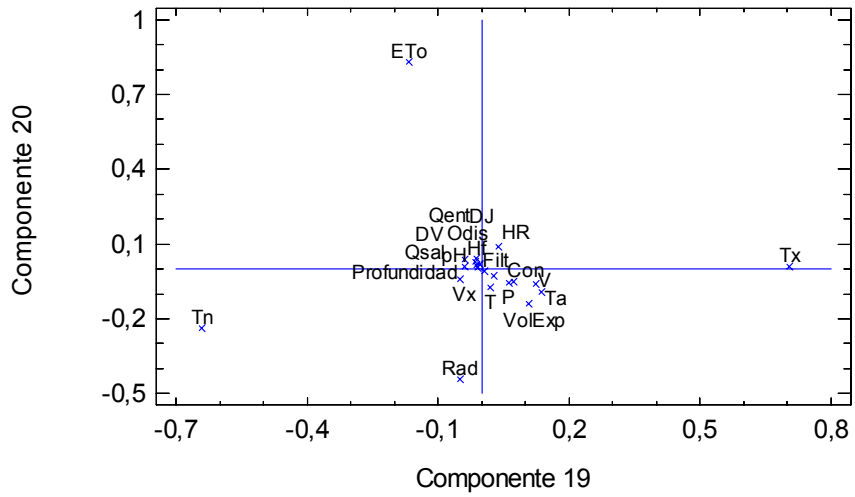
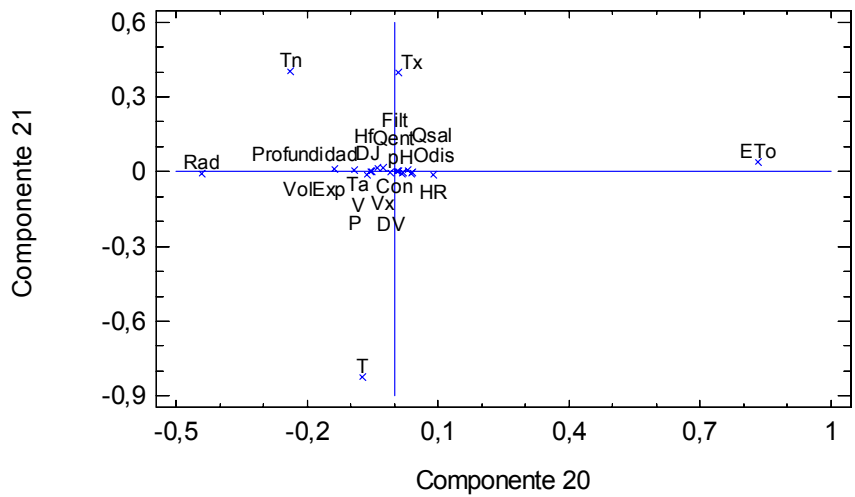


Gráfico de Pesos del Componente



Gráficos de pesos de los componentes

**ANEXO 8. PCA DE LOS DATOS METEOROLÓGICOS,  
HIDROLÓGICOS Y DE CALIDAD DEL AGUA EN  
EL PERIODO DE CALIBRACIÓN, CORREGIDOS  
(SIN DATOS ANÓMALOS)**

## Resumen del Análisis

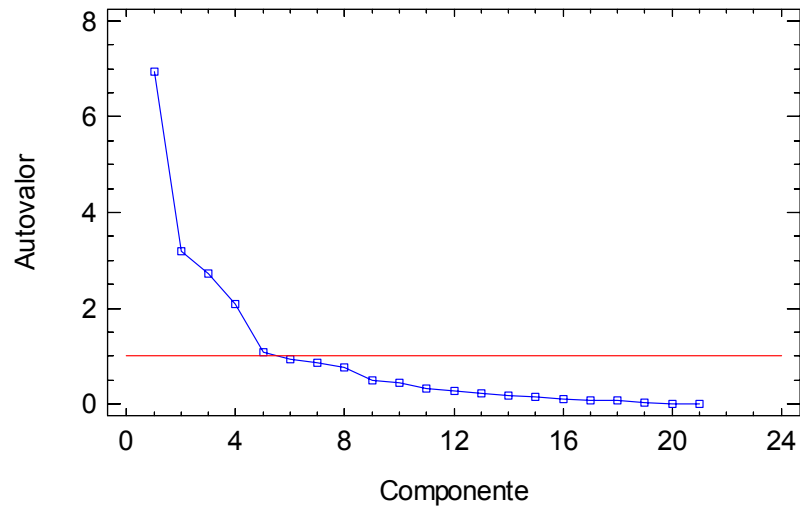
Datos/VARIABLES: DJ, V, DV, Vx, T, Tn, Tx, Hf, HR, Rad, P, ETo, VolExp, Qent, Qsal, Filt, Profundidad, Ta, Odis, pH, y Con.

Entrada de datos: observaciones  
Número de casos completos: 4401  
Tratamiento de valor perdido: lista considerada  
Estandarizado: si  
Número de componentes extraídos: 21

## Análisis de Componentes Principales

Componente Número	Autovalor	% de Varianza	Porcentaje acumulado
1	6,93217	33,010	33,010
2	3,18022	15,144	48,154
3	2,7298	12,999	61,153
4	2,10131	10,006	71,160
5	1,08017	5,144	76,303
6	0,935969	4,457	80,760
7	0,874795	4,166	84,926
8	0,759152	3,615	88,541
9	0,50628	2,411	90,952
10	0,453618	2,160	93,112
11	0,327645	1,560	94,672
12	0,270731	1,289	95,961
13	0,22685	1,080	97,042
14	0,180031	0,857	97,899
15	0,151283	0,720	98,619
16	0,100599	0,479	99,098
17	0,0775182	0,369	99,467
18	0,0711552	0,339	99,806
19	0,0252075	0,120	99,926
20	0,0128686	0,061	99,988
21	0,00261867	0,012	100

Gráfico de Sedimentación



Variable	Componentes						
	1	2	3	4	5	6	7
DJ	0,130690	0,373395	-0,071215	0,332869	-0,015873	0,024458	0,114902
V	-0,049956	-0,261919	0,265863	0,415980	0,135491	-0,064511	0,105873
DV	-0,221010	0,064182	0,013028	0,246552	-0,243002	-0,085617	-0,043906
Vx	-0,025040	-0,162144	0,199901	0,431981	0,457208	-0,059158	-0,169403
T	0,371367	0,025341	-0,026758	0,011611	0,065776	0,049275	-0,053364
Tn	0,367750	0,055389	-0,041658	0,003571	0,056594	0,015957	-0,065838
Tx	0,366912	0,010119	-0,027458	0,002774	0,038960	0,082839	-0,023322
Hf	-0,193170	-0,256818	0,136668	-0,109028	-0,396219	-0,208387	-0,112598
HR	0,017971	0,247446	-0,259979	-0,435212	0,139887	-0,086975	0,027623
Rad	0,290678	-0,258365	0,131831	-0,110060	0,001361	0,058402	0,015891
P	-0,063658	0,229459	-0,113815	0,005845	0,257427	-0,533049	-0,666909
ETo	0,334624	-0,201042	0,123091	-0,000780	0,075981	0,001421	-0,096417
VolExp	0,083609	-0,402597	0,088046	-0,346370	0,177608	-0,001544	-0,153782
Qent	-0,221688	0,117910	-0,061233	-0,030587	0,406951	0,247446	0,239749
Qsal	0,280792	0,205845	-0,070341	0,216695	-0,250860	-0,042103	0,003185
Filt	-0,125411	0,058749	-0,004538	0,051680	-0,119483	0,746732	-0,610375
Profundidad	0,013410	0,170572	0,438148	-0,208188	0,168440	0,039670	0,019026
Ta	0,279156	-0,073362	-0,305300	0,167287	-0,100758	-0,023386	-0,036103
Odis	-0,224411	-0,254642	-0,361786	0,091250	-0,004146	-0,012250	0,004268
pH	-0,060245	-0,094988	-0,409588	0,046302	0,371206	0,114467	0,108664
Con	-0,047029	0,374958	0,385819	-0,053199	0,070398	0,018309	0,047863

Variable	Componentes						
	8	9	10	11	12	13	14
DJ	-0,297121	0,136405	-0,224550	-0,096017	0,166157	-0,053959	0,062297
V	0,012227	0,199930	0,187420	-0,379842	0,520822	-0,016625	0,062435
DV	0,531838	-0,596790	-0,178465	0,152060	0,317362	-0,021015	0,043789
Vx	-0,092202	0,026368	0,129320	0,628054	-0,058393	0,175955	-0,022375
T	0,045925	-0,111345	-0,049682	0,115940	-0,065672	0,004338	0,109636
Tn	0,040851	-0,083315	-0,014093	0,081239	0,017575	-0,035637	0,089828
Tx	0,088111	-0,135456	-0,061194	0,146893	-0,128417	0,042918	0,139765
Hf	-0,255167	0,201571	-0,458304	0,213555	0,071233	0,419991	0,171067
HR	-0,015903	0,062907	0,312919	0,265829	0,61517	0,285312	0,007695
Rad	0,022970	-0,222728	-0,051637	-0,274880	0,068606	0,429787	-0,060601
P	-0,049144	-0,063408	-0,126844	-0,294551	-0,038107	-0,021573	-0,004830
ETo	-0,007955	-0,143579	-0,053974	-0,174539	0,036641	0,161881	-0,008604
VolExp	-0,048396	-0,029574	-0,126158	0,045482	0,212347	-0,411917	0,314941
Qent	-0,413116	-0,402060	-0,429496	-0,049734	0,108765	0,017832	0,055032
Qsal	0,026034	0,228239	-0,159242	0,119793	0,140720	-0,097794	0,418451
Filt	-0,017373	0,086956	0,038414	-0,043194	0,132895	0,062425	0,000783
Profundidad	0,339486	0,302629	-0,401423	0,116835	0,081467	-0,120532	-0,417374
Ta	-0,091975	0,014774	-0,156677	0,010211	0,142845	0,137407	-0,598289
Odis	0,038975	0,003810	0,077360	0,008557	-0,160747	0,135961	0,008664
pH	0,484046	0,334577	-0,296850	-0,139967	-0,066391	0,234080	0,227687
Con	0,082399	-0,045916	0,151086	-0,155168	-0,182238	0,450505	0,236010

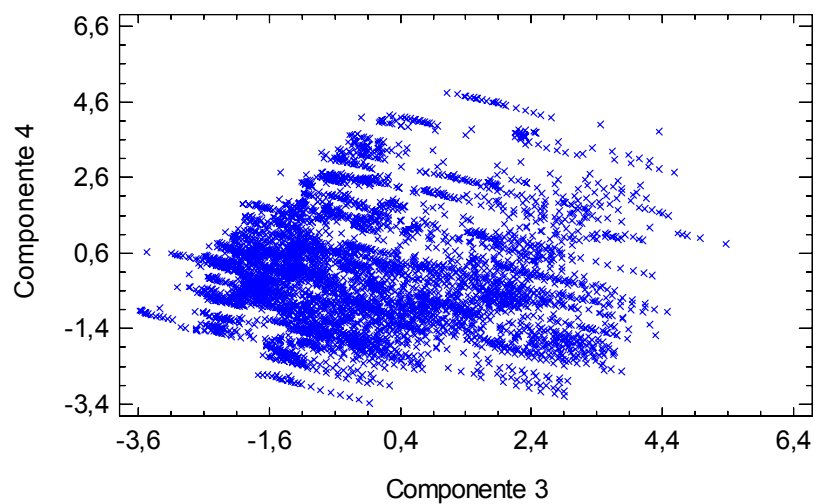
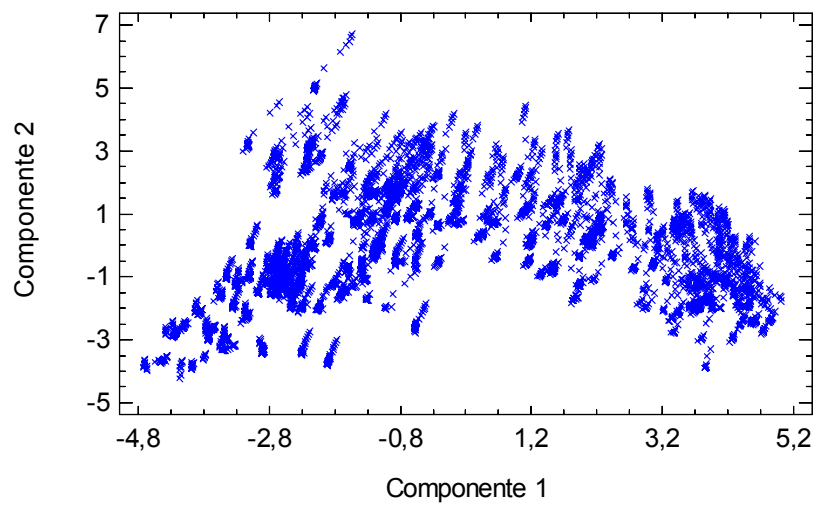
Variable	Componentes						
	15	16	17	18	19	20	21
DJ	-0,627591	0,020141	0,060903	0,336122	0,029035	-0,024332	0,015583
V	0,227100	0,277400	-0,095083	0,014647	0,122037	-0,060466	-0,012818
DV	-0,137866	0,014962	0,030682	-0,010261	-0,011888	0,015878	-0,006219
Vx	-0,084447	-0,191356	0,034149	-0,029058	-0,047398	-0,041014	0,018635
T	0,067465	0,305755	-0,114852	0,086092	0,019756	-0,075657	-0,824298
Tn	0,087734	0,427753	-0,074997	0,031176	-0,642095	-0,238626	0,400749
Tx	0,108754	0,207449	-0,181555	0,165096	0,702580	0,012899	0,396761
Hf	-0,013153	0,222763	-0,174314	-0,024645	-0,010434	0,007387	0,000161
HR	0,005448	-0,007370	-0,025230	0,102312	0,036448	0,090729	-0,010693
Rad	-0,041272	-0,472051	0,005509	0,267895	-0,046845	-0,442216	-0,007490
P	0,108153	-0,032538	-0,041464	0,050485	0,061456	-0,055640	0,002739
ETo	-0,086629	-0,056772	0,054794	0,036269	-0,170243	0,830934	0,035060
VolExp	-0,309409	0,040512	0,397511	-0,165823	0,106867	-0,135659	0,011703
Qent	0,329461	0,011197	0,016808	-0,057632	-0,016963	0,030368	0,005545
Qsal	0,454155	-0,397153	0,307703	-0,059312	-0,040178	0,036997	-0,007146
Filt	0,013232	0,012898	-0,019379	0,014149	0,000188	-0,002503	-0,000181
Profundidad	0,137840	0,035323	0,144175	0,294311	-0,006013	0,020376	-0,002244
Ta	0,001481	0,121417	0,321011	-0,456198	0,137185	-0,090320	0,006565
Odis	0,150397	0,238071	0,545334	0,561115	-0,012761	0,040978	-0,005197
pH	-0,151424	-0,040612	-0,156470	-0,197917	-0,038246	0,006377	0,002270
Con	-0,089291	0,235663	0,445567	-0,283226	0,073387	-0,051019	0,003706

### Tabla de pesos de los componentes

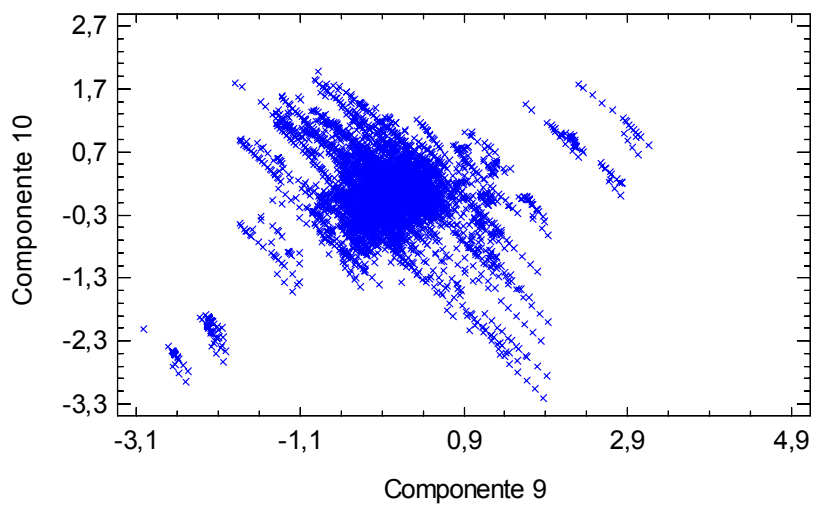
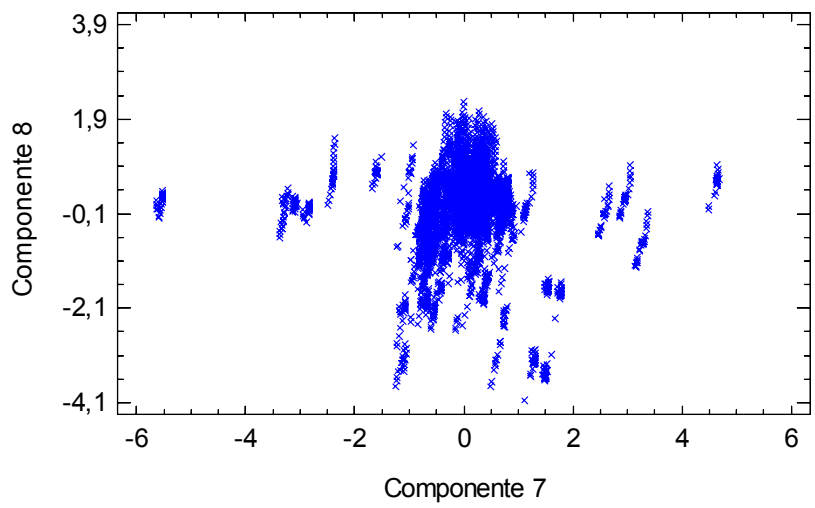
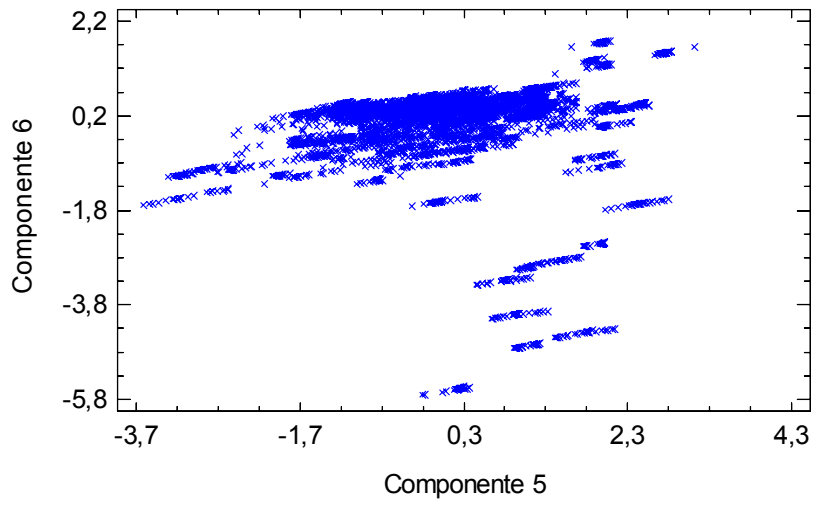
Esta tabla muestra las ecuaciones de los componentes principales. Por ejemplo, el primer componente principal tiene la ecuación:

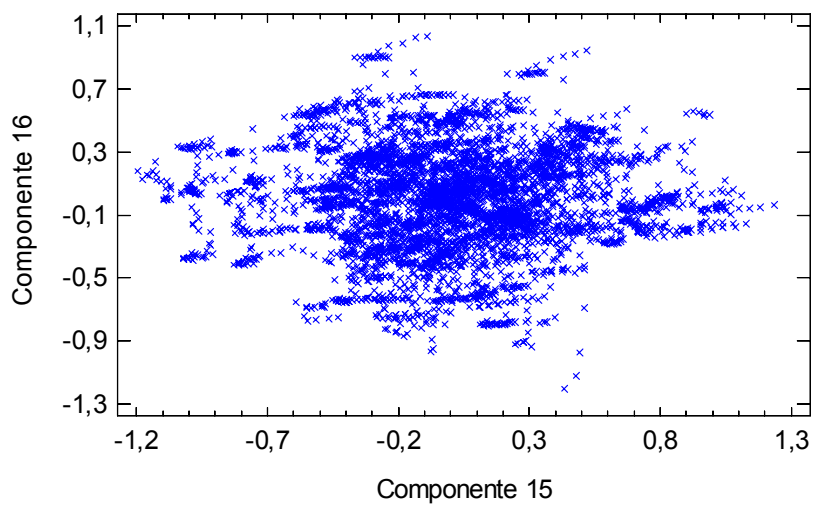
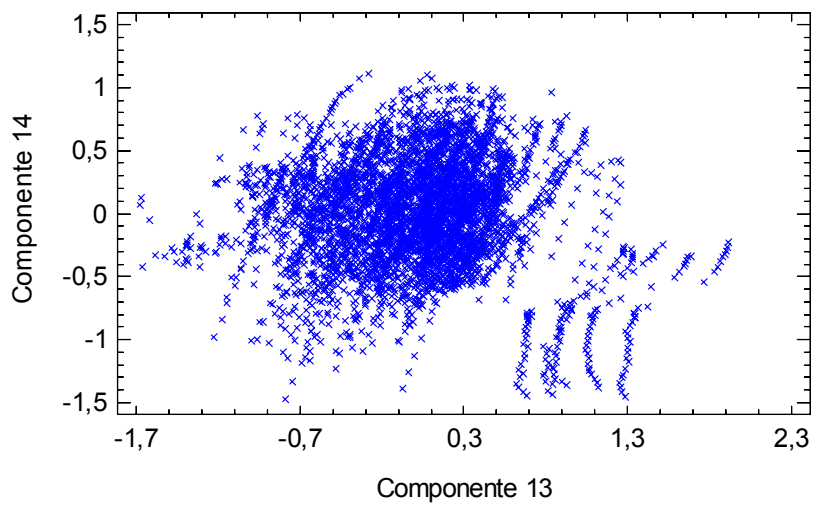
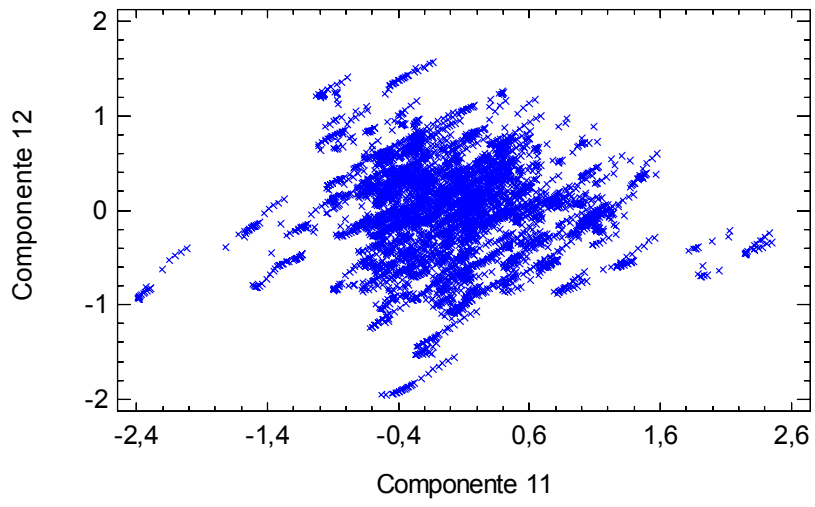
$$\begin{aligned} \text{Componente 1} = & 0,131148(\text{DJ}) - 0,050672(\text{V}) - 0,205235(\text{DV}) - 0,023723(\text{Vx}) + \\ & 0,372801(\text{T}) + 0,369223(\text{Tn}) + 0,368254(\text{Tx}) - 0,194988(\text{Hf}) + 0,018449(\text{HR}) + \\ & 0,290790(\text{Rad}) - 0,061391(\text{P}) + 0,335430(\text{ETo}) + 0,083428(\text{VolExp}) - 0,223087(\text{Qent}) \\ & + 0,282427(\text{Qsal}) - 0,125857(\text{Filt}) + 0,013309(\text{Profundidad}) + 0,280270(\text{Ta}) - \\ & 0,224890(\text{Odis}) - 0,059600(\text{pH}) - 0,047345(\text{Con}) \end{aligned}$$

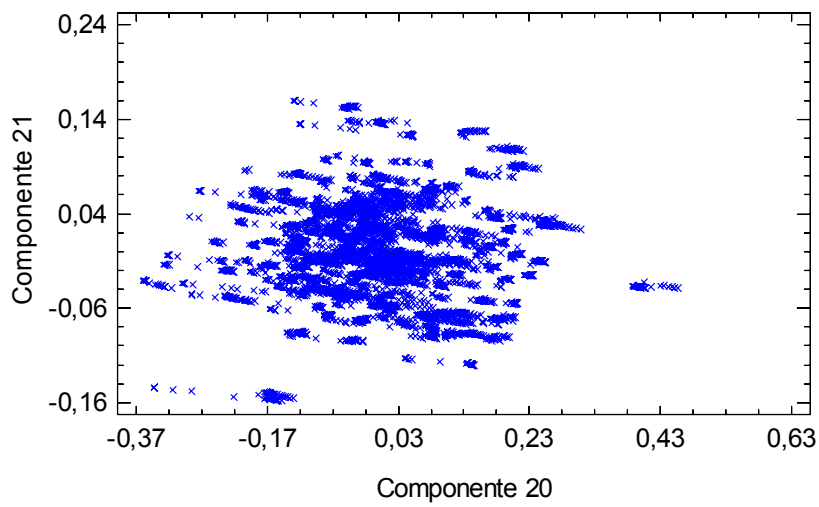
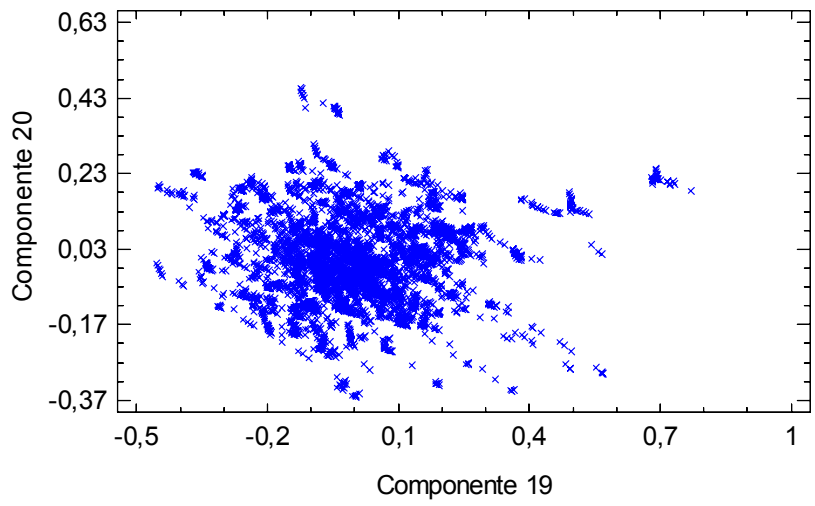
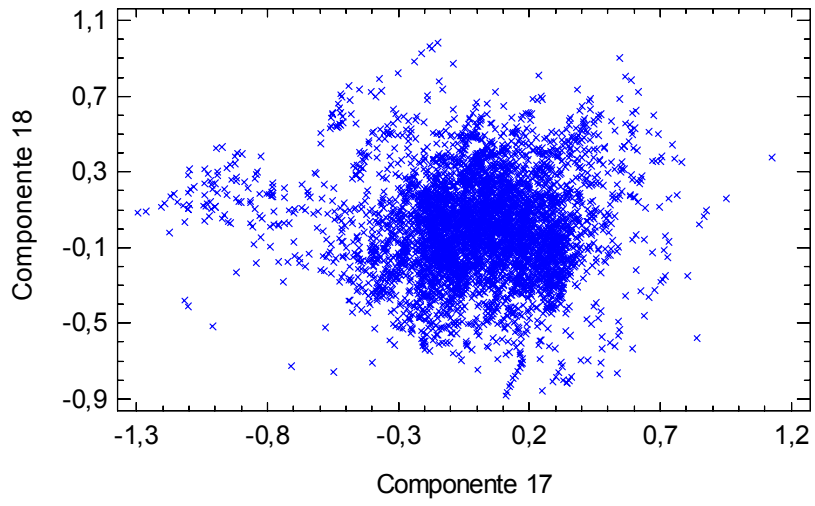
Donde los valores de las variables en la ecuación se encuentran estandarizados (se calculan substrayendo sus medias y dividiéndolos por sus desviaciones típicas).











**Diagramas de dispersión de los componentes por pares**

Gráfico de Pesos del Componente

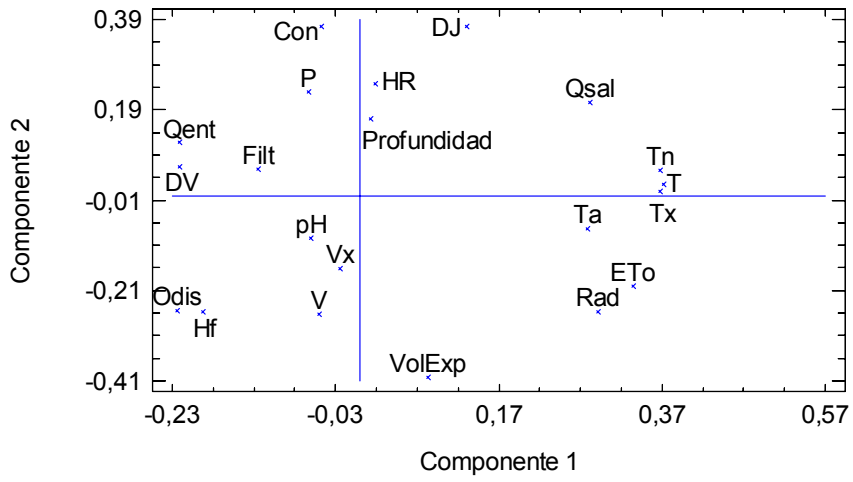


Gráfico de Pesos del Componente

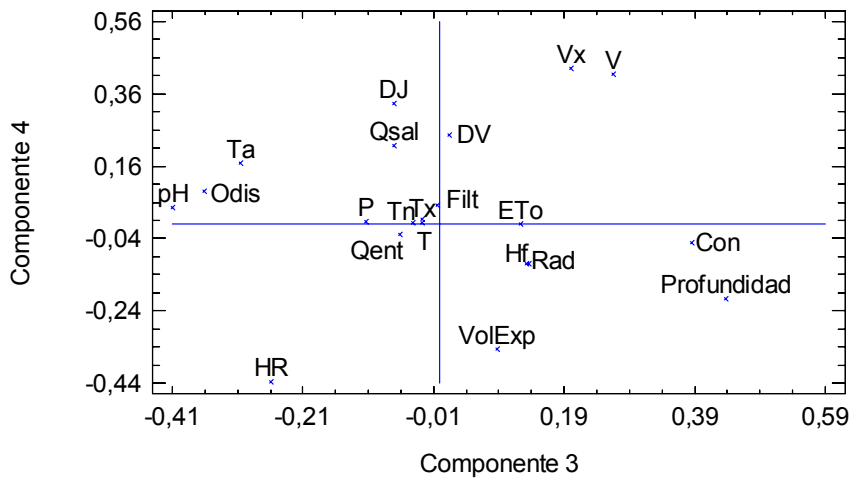


Gráfico de Pesos del Componente

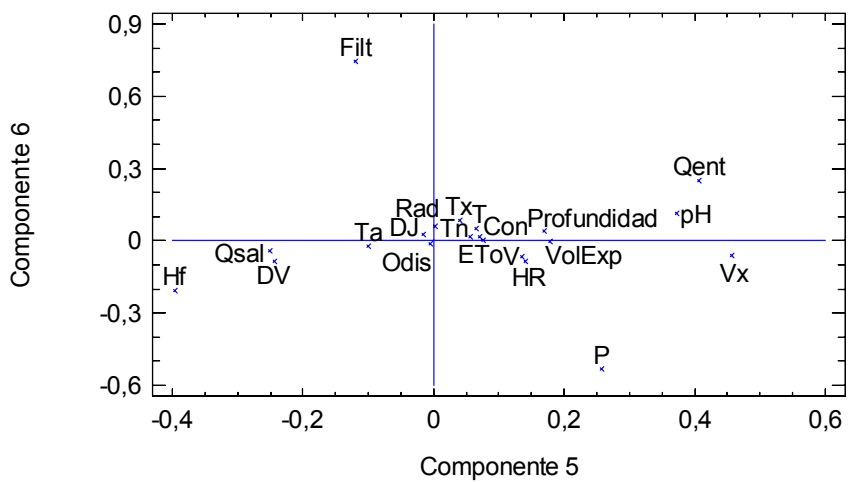


Gráfico de Pesos del Componente

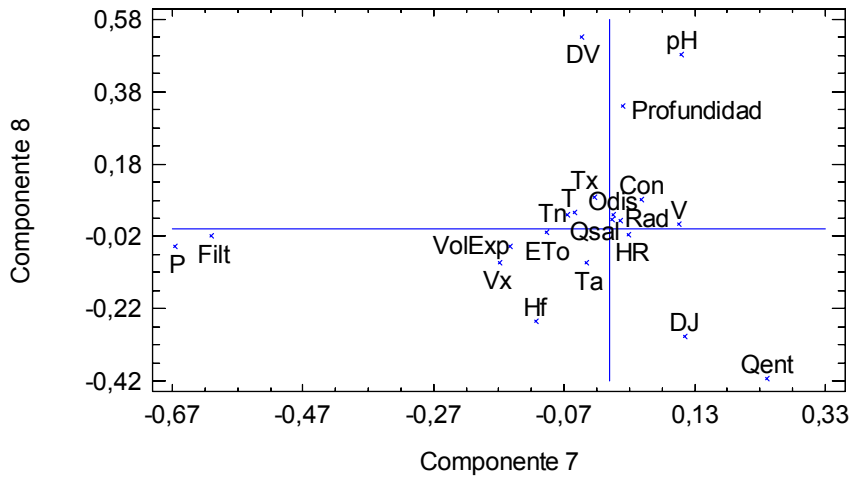


Gráfico de Pesos del Componente

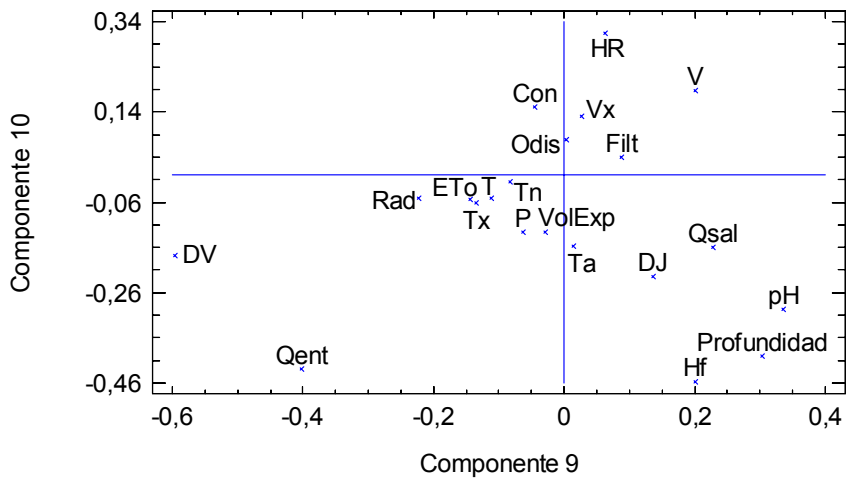


Gráfico de Pesos del Componente

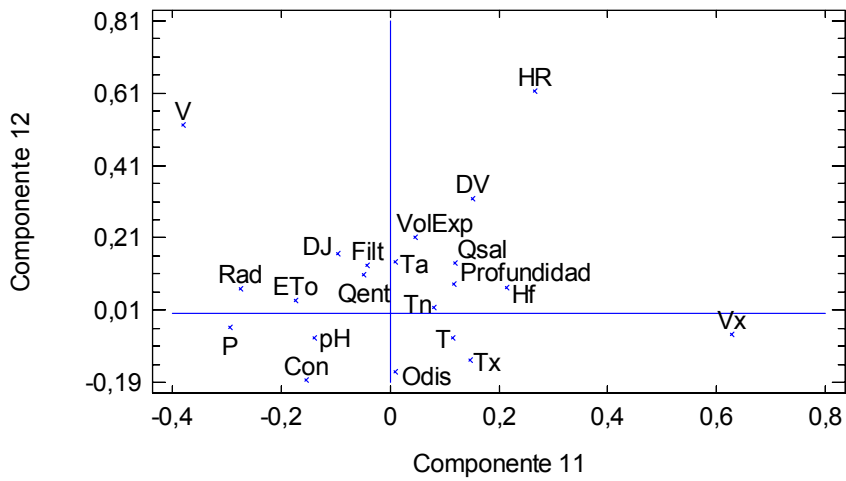


Gráfico de Pesos del Componente

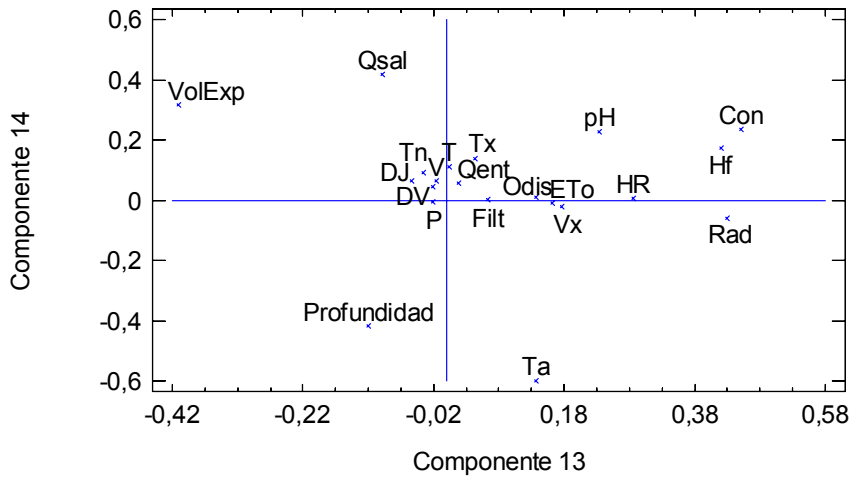


Gráfico de Pesos del Componente

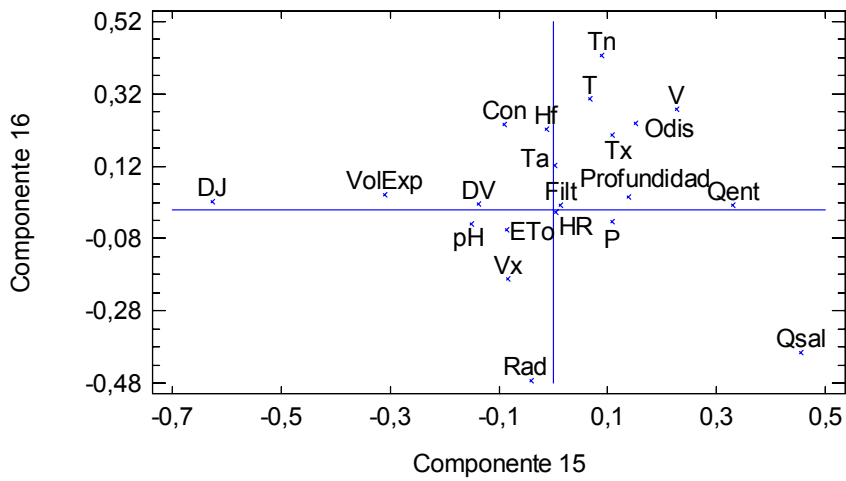


Gráfico de Pesos del Componente

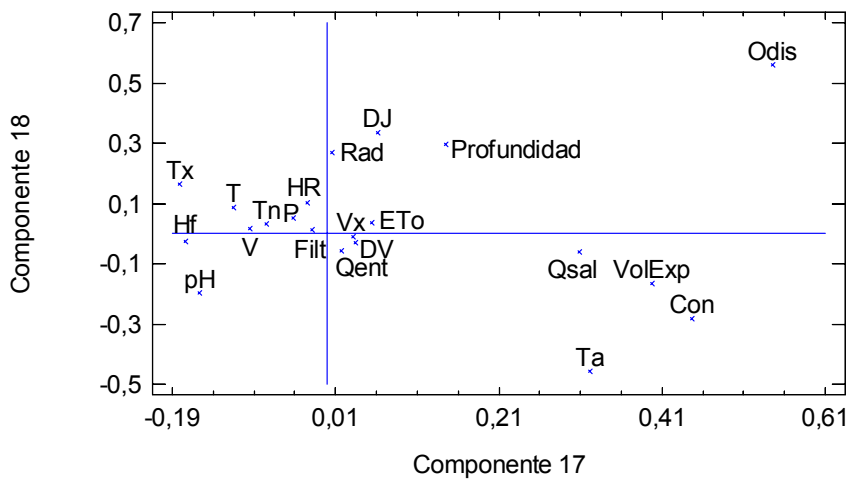


Gráfico de Pesos del Componente

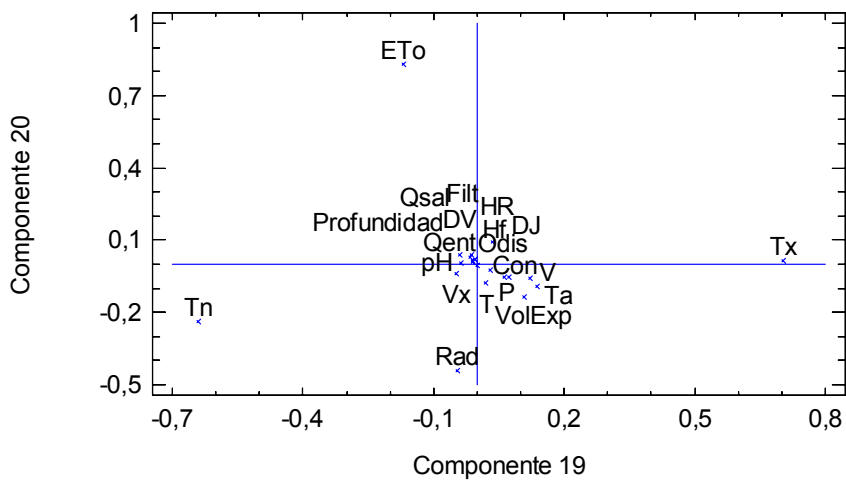
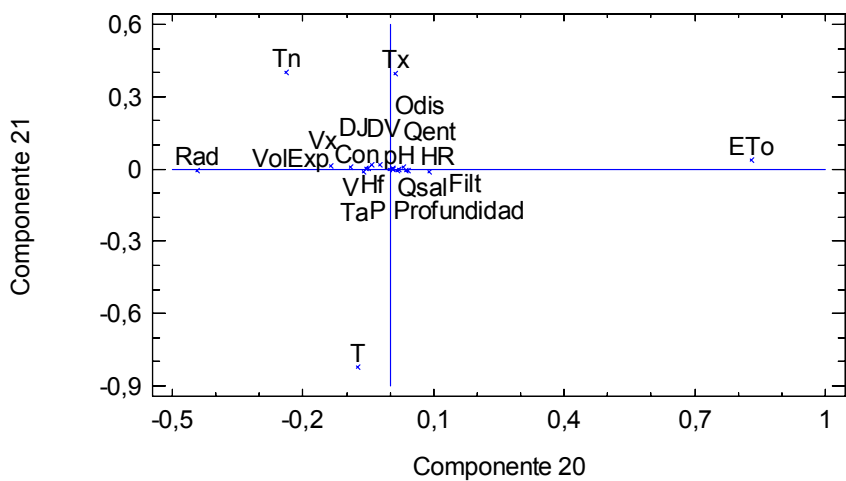


Gráfico de Pesos del Componente



**Gráficos de pesos de los componentes**

**ANEXO 9. PCA DE LOS DATOS METEOROLÓGICOS,  
HIDROLÓGICOS Y DE CALIDAD DEL AGUA EN  
EL PERIODO DE VALIDACIÓN, DATOS  
ORIGINALES**



Variable	Componentes						
	1	2	3	4	5	6	7
DJ	0,342634	-0,087420	0,065500	-0,191805	0,131710	0,063165	0,024325
V	-0,013914	0,141149	-0,487776	-0,322884	-0,078609	0,125258	-0,074290
DV	-0,085970	-0,068613	-0,031913	-0,023379	0,658694	0,438669	0,158853
Vx	-0,048987	0,077253	-0,366329	-0,421119	-0,329498	0,156712	0,042756
T	0,364366	0,074193	0,022624	0,040972	-0,057890	-0,029852	0,058652
Tn	0,359860	0,041692	0,067337	0,010686	-0,101520	-0,049536	0,036350
Tx	0,358238	0,089585	-0,012378	0,045112	-0,027084	0,001286	0,102834
Hf	-0,208161	-0,123894	-0,132215	-0,210167	0,089931	-0,264312	-0,476478
HR	-0,011224	-0,197900	0,459412	0,065368	-0,203225	-0,144658	-0,061228
Rad	0,162198	0,293830	-0,217866	0,384799	0,086751	0,082813	-0,068866
P	-0,042893	-0,133760	0,200685	-0,245166	-0,279417	0,433183	0,378001
ETo	0,264012	0,268988	-0,242433	0,222964	-0,014124	0,050803	0,005764
VolExp	-0,264685	0,213301	-0,069064	0,352160	-0,244178	0,011624	0,036399
Qent	-0,044284	0,012315	-0,146867	-0,100870	-0,097897	-0,500131	0,602320
Qsal	0,337133	-0,039342	0,034036	-0,133681	-0,020690	-0,069276	-0,085658
Filt	-0,067009	-0,002078	-0,132473	-0,042596	0,403384	-0,396226	0,339395
Profundidad	0,019076	-0,377614	-0,251048	0,356109	-0,176601	0,102083	0,095191
Ta	0,188323	0,334978	0,248511	-0,291007	0,113689	-0,061010	-0,067369
Odis	-0,266784	0,321483	0,117606	-0,055085	0,015311	-0,044192	-0,017696
pH	-0,191454	0,302671	0,182417	0,045209	0,002678	0,205151	0,261684
Con	0,082996	-0,467366	-0,159305	0,021331	0,087980	0,043225	0,063384

Variable	Componentes						
	8	9	10	11	12	13	14
DJ	0,013422	0,055895	0,014333	0,206008	-0,044651	-0,177181	-0,153026
V	0,035394	0,108748	0,248448	0,028143	-0,006984	-0,260646	0,297146
DV	-0,279034	0,062489	0,311276	-0,273696	-0,060519	0,267650	0,035900
Vx	0,210107	0,207096	0,230894	-0,110547	0,052032	0,252074	-0,252728
T	0,031811	0,023475	-0,011796	-0,059974	0,010596	0,292924	-0,125988
Tn	0,046464	0,057146	0,019332	-0,101030	0,000781	0,257128	0,003204
Tx	0,008333	-0,013899	-0,031539	-0,012787	0,039968	0,328675	-0,257421
Hf	-0,208761	-0,383052	0,120854	0,345125	-0,192814	0,383682	-0,148490
HR	0,089528	0,116263	0,781337	-0,001871	0,112865	-0,028851	0,014990
Rad	-0,082184	-0,262043	0,336923	0,274082	0,159188	-0,242936	-0,080929
P	0,005633	-0,671034	-0,006893	0,084629	0,045859	-0,022393	0,012152
ETo	-0,052997	-0,187215	0,094983	0,075529	0,100378	0,014228	-0,017816
VolExp	0,038212	-0,099588	0,047789	-0,231221	0,023120	0,251822	0,165279
Qent	-0,570626	0,042194	0,099029	0,029938	-0,027797	-0,033433	0,000396
Qsal	-0,028736	-0,096360	0,004736	-0,037023	-0,076993	0,090821	0,761257
Filt	0,681212	-0,249988	0,103146	-0,010947	-0,033211	0,036069	0,033339
Profundidad	0,068016	0,052880	0,044644	0,149412	-0,411303	0,269243	0,182212
Ta	-0,042954	-0,007229	0,030729	0,130635	-0,128712	0,109860	0,103580
Odis	-0,016805	-0,001890	-0,077449	0,040672	0,507837	0,330327	0,210060
Ph	0,115623	0,330912	-0,005914	0,636298	-0,259485	0,114252	0,073830
Con	0,007686	0,155454	-0,083383	0,383144	0,618980	0,130263	0,101204

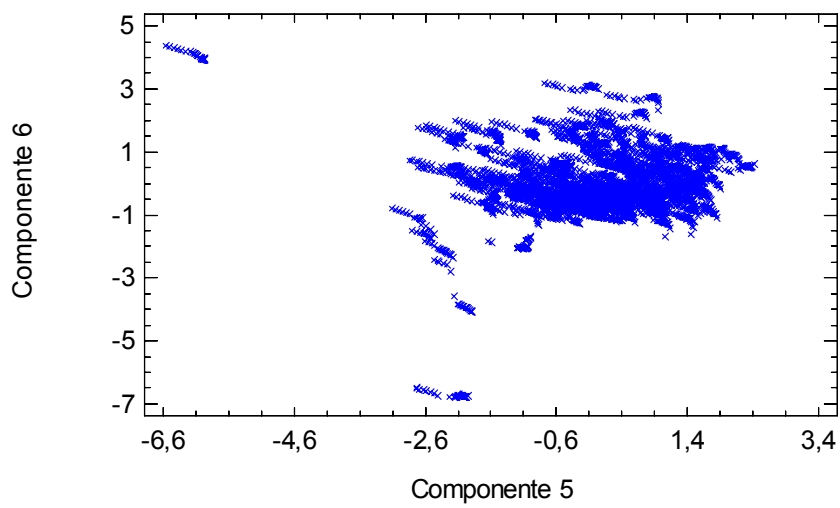
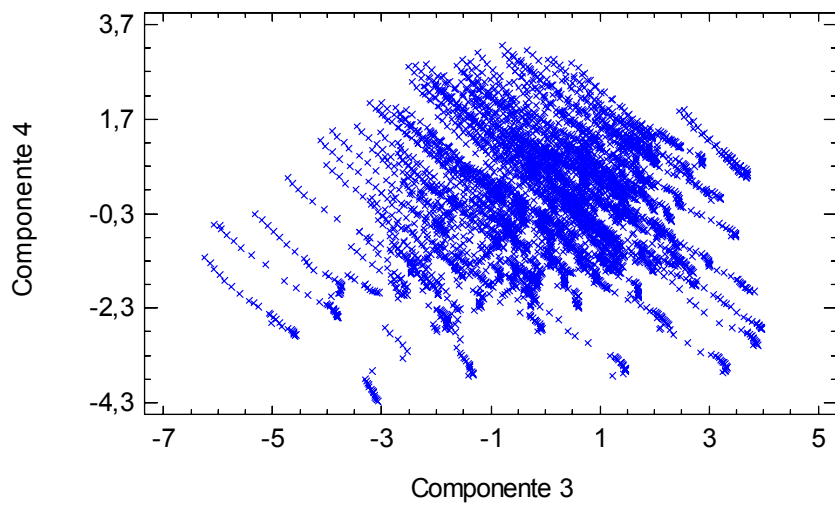
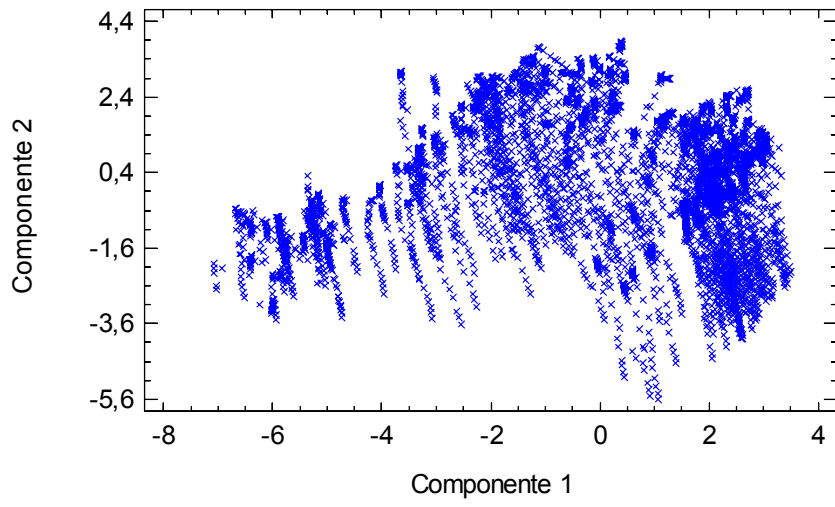
Variable	Componentes						
	15	16	17	18	19	20	21
DJ	0,032100	0,084749	-0,310335	-0,324376	-0,611505	0,348819	0,062018
V	-0,575217	-0,054455	-0,076609	0,171506	-0,079624	-0,054001	-0,007687
DV	0,010331	-0,044803	0,017442	-0,042844	0,009591	0,017121	-0,000033
Vx	0,486389	0,052533	0,066957	-0,130160	0,034609	-0,022245	0,000251
T	-0,201828	-0,120746	-0,026750	-0,034830	-0,059001	-0,138813	-0,815339
Tn	-0,351023	-0,110865	0,027501	-0,584782	0,260267	-0,204293	0,424491
Tx	-0,048451	-0,149136	-0,149338	0,660269	-0,186179	-0,101835	0,380726
Hf	-0,067954	-0,183329	-0,015292	-0,063827	-0,021579	0,034560	0,005338
HR	-0,061901	-0,011732	-0,044106	0,111526	0,039069	0,115960	-0,016038
Rad	0,238005	0,080857	-0,012143	-0,122728	-0,094560	-0,479811	0,013159
P	-0,092614	0,000567	0,006904	-0,001219	0,032210	-0,033126	-0,015497
ETo	0,021424	0,005753	0,039814	0,027746	0,396120	0,723420	-0,019198
VolExp	-0,094118	-0,145932	0,350617	-0,127626	-0,575352	0,176898	0,066983
Qent	0,036053	0,015016	-0,005476	-0,039474	-0,021426	-0,008332	-0,002653
Qsal	0,414653	-0,259472	-0,066934	0,042571	-0,032353	-0,037347	-0,003232
Filt	-0,012661	-0,004446	0,007740	-0,012013	0,000182	0,004930	0,003598
Profundidad	-0,042199	0,514757	-0,204762	0,025731	0,014862	-0,019087	-0,001483
Ta	-0,051686	0,588560	0,511042	0,104442	-0,059588	-0,019882	0,018184
Odis	-0,045994	0,298254	-0,540959	-0,056223	0,000660	0,000392	-0,012730
pH	0,010876	-0,326070	0,004288	-0,017638	0,071872	0,015271	-0,007566
Con	-0,045188	-0,011546	0,376902	0,016120	-0,046843	0,013364	0,005903

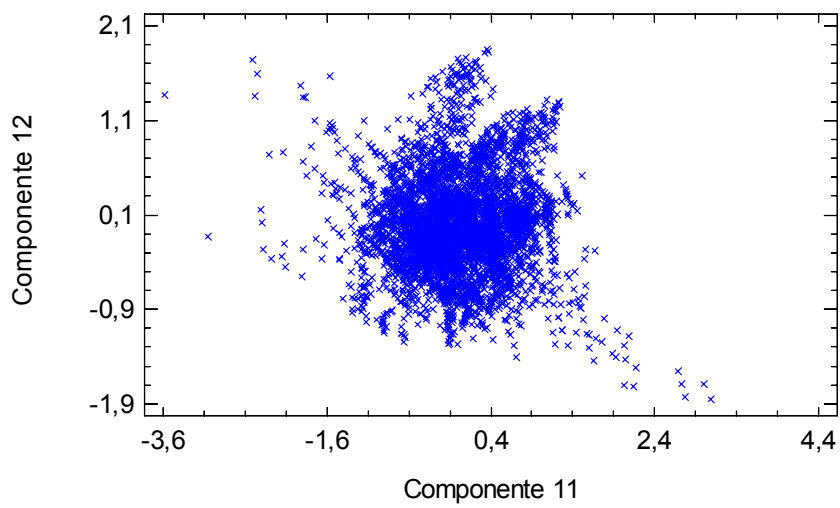
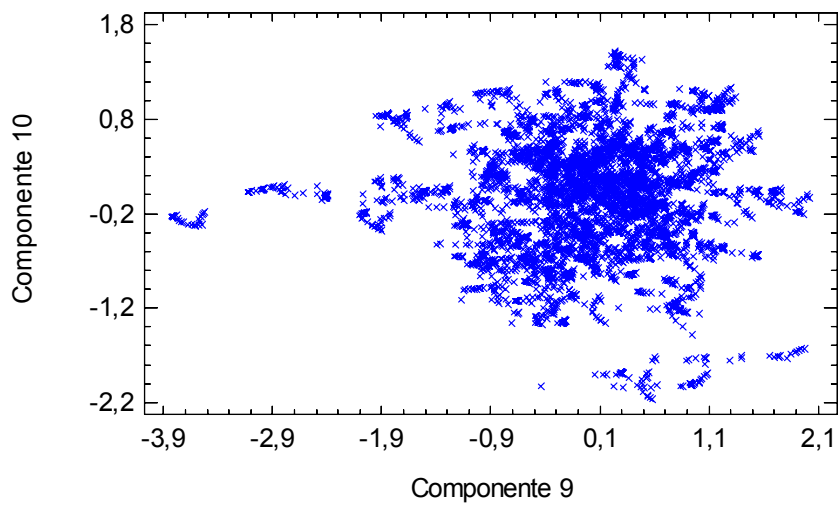
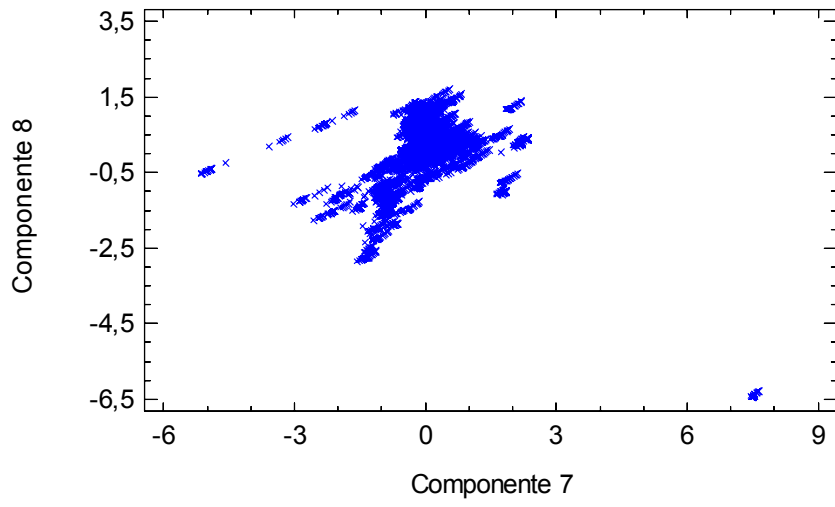
**Tabla de pesos de los componentes**

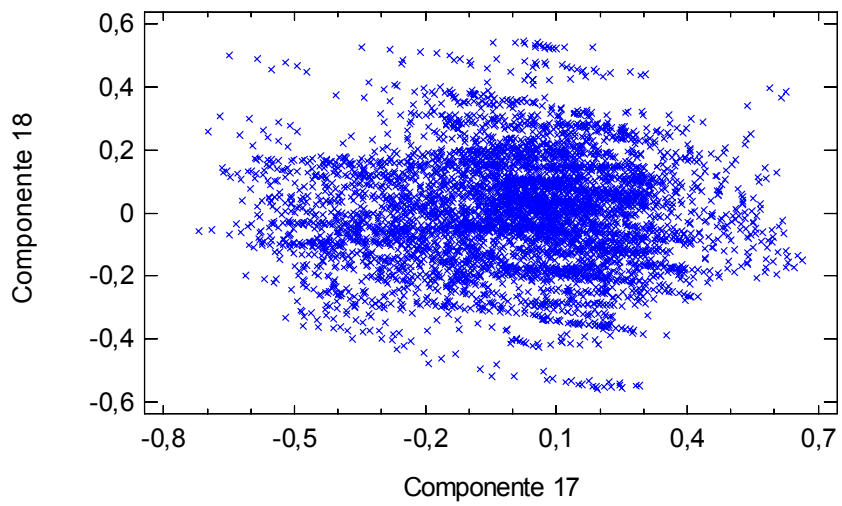
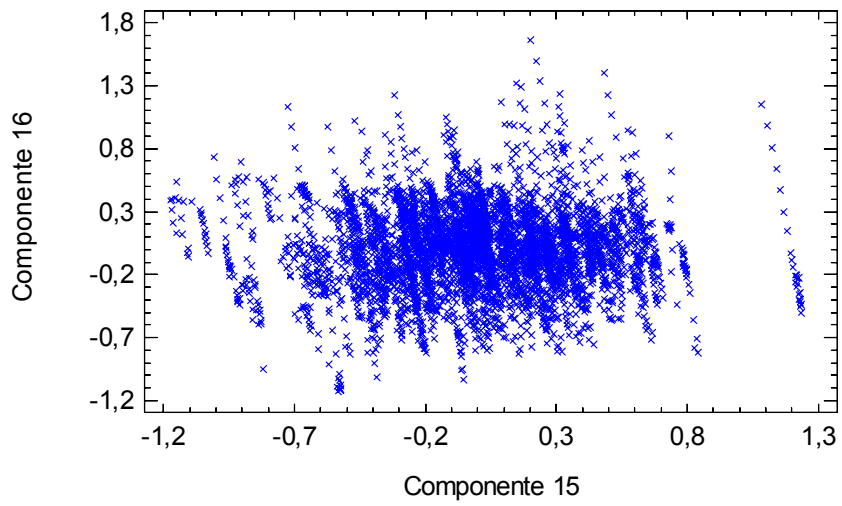
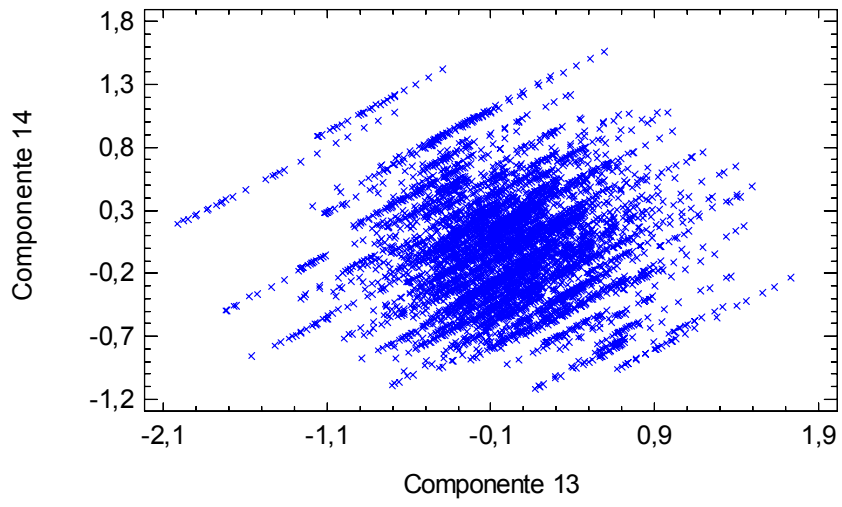
Esta tabla muestra las ecuaciones de los componentes principales. Por ejemplo, el primer componente principal tiene la ecuación:

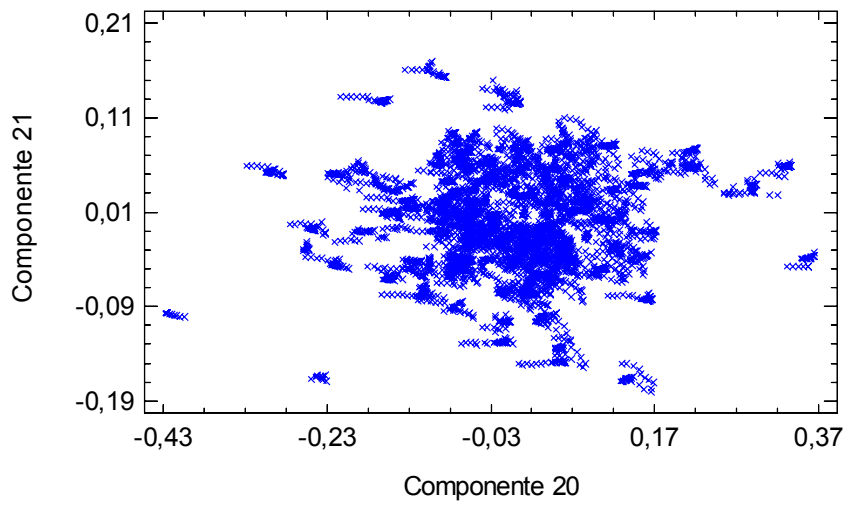
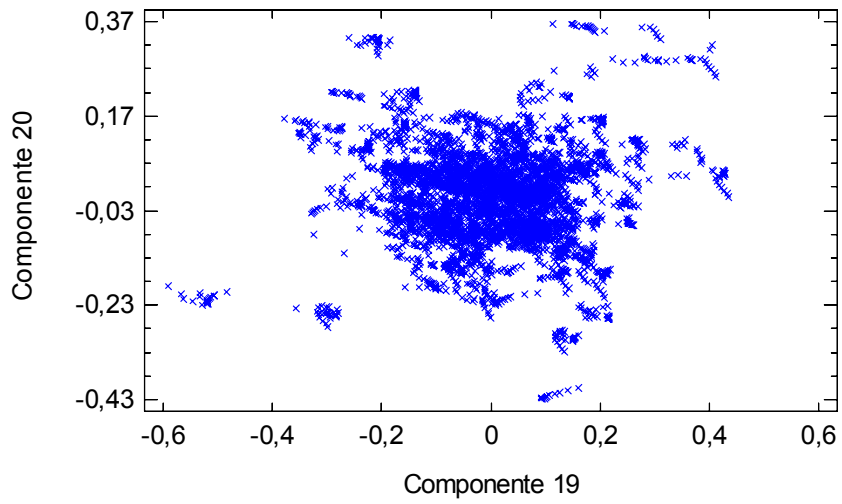
$$\begin{aligned} \text{Componente 1} = & 0,342634(\text{DJ}) - 0,013914(\text{V}) - 0,085970(\text{DV}) - 0,048987(\text{Vx}) + \\ & 0,364366(\text{T}) + 0,359860(\text{Tn}) + 0,358238(\text{Tx}) - 0,208161(\text{Hf}) - 0,011224(\text{HR}) + \\ & 0,162198(\text{Rad}) - 0,042893(\text{P}) + 0,264012(\text{ETo}) - 0,264685(\text{VolExp}) - 0,044284(\text{Qent}) \\ & + 0,337133(\text{Qsal}) - 0,067009(\text{Filt}) + 0,019076(\text{Profundidad}) + 0,188323(\text{Ta}) - \\ & 0,266784(\text{Odis}) - 0,191454(\text{pH}) + 0,082996(\text{Con}) \end{aligned}$$

Donde los valores de las variables en la ecuación están estandarizados (substrayendo sus medias y dividiéndolos por sus desviaciones típicas).

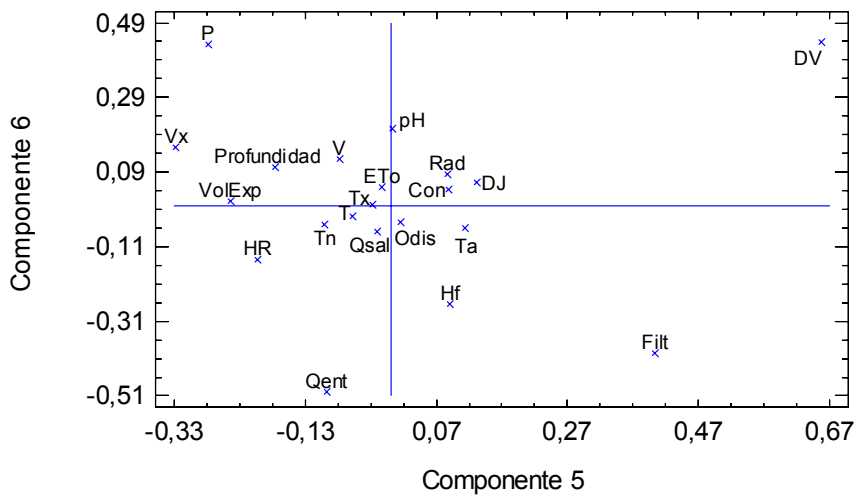
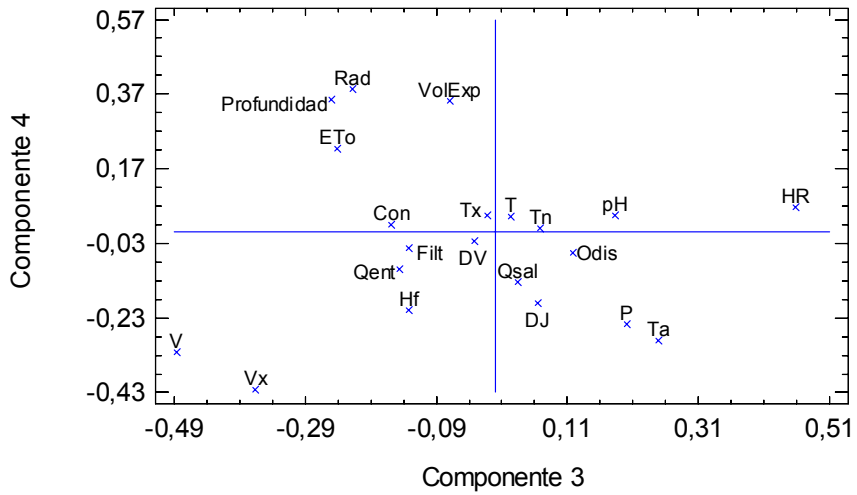
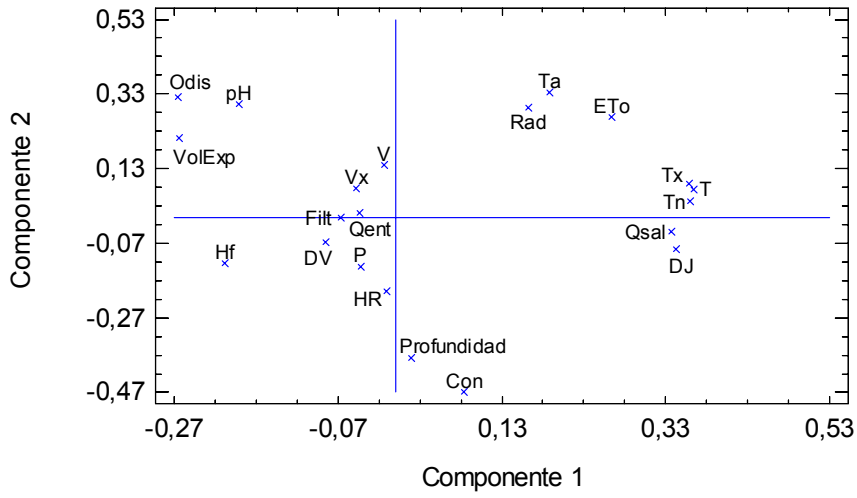


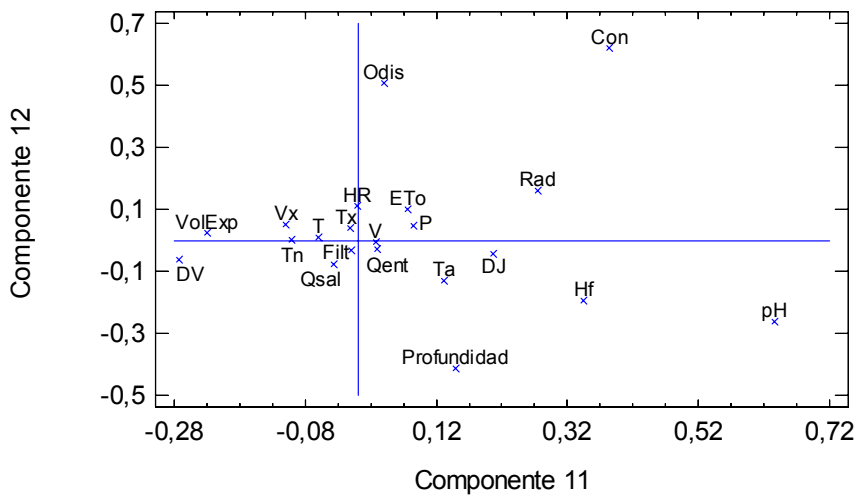
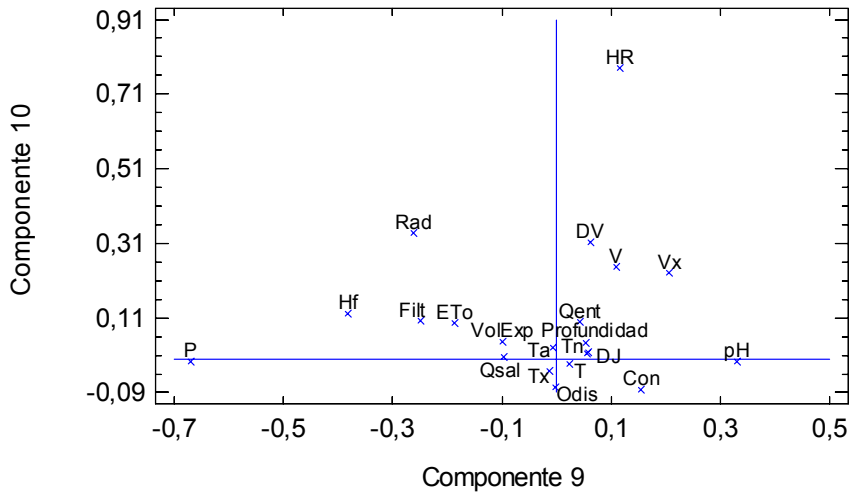
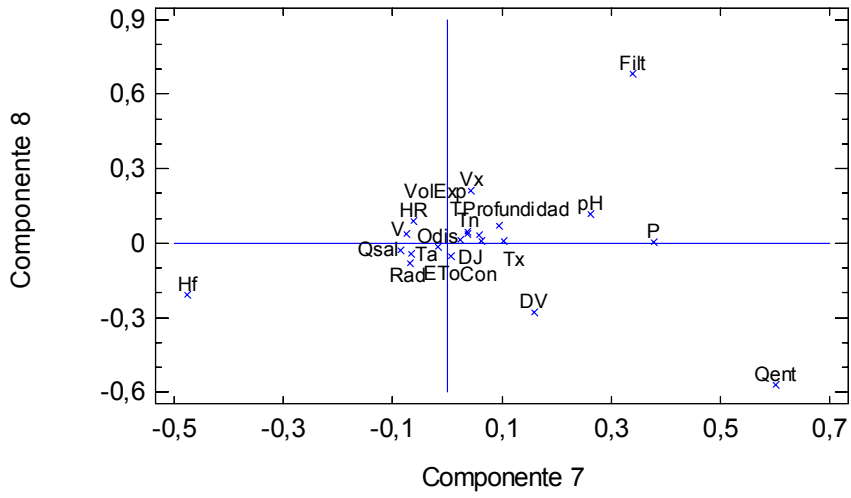




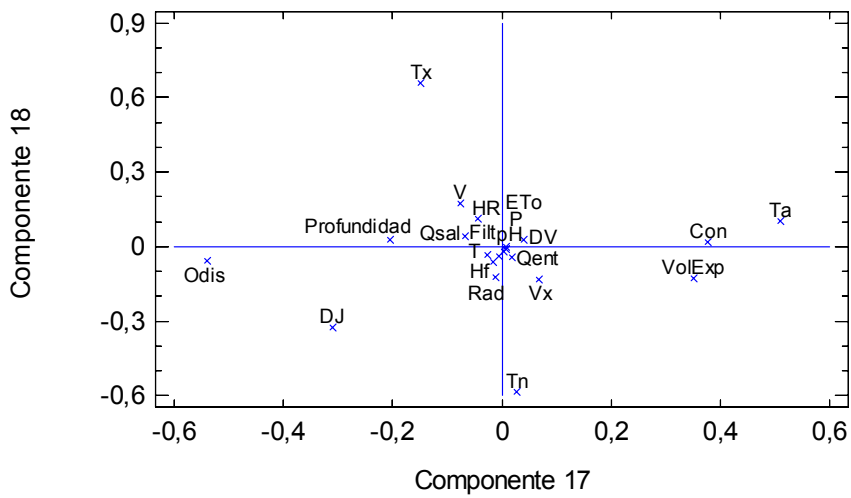
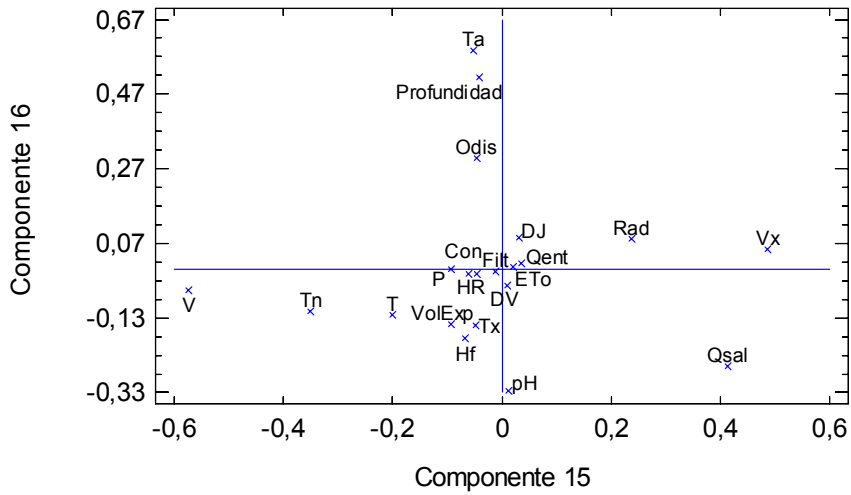
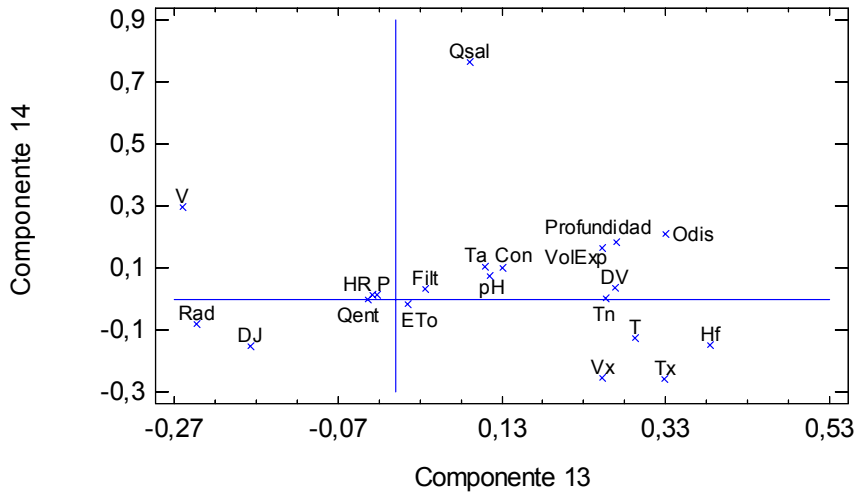


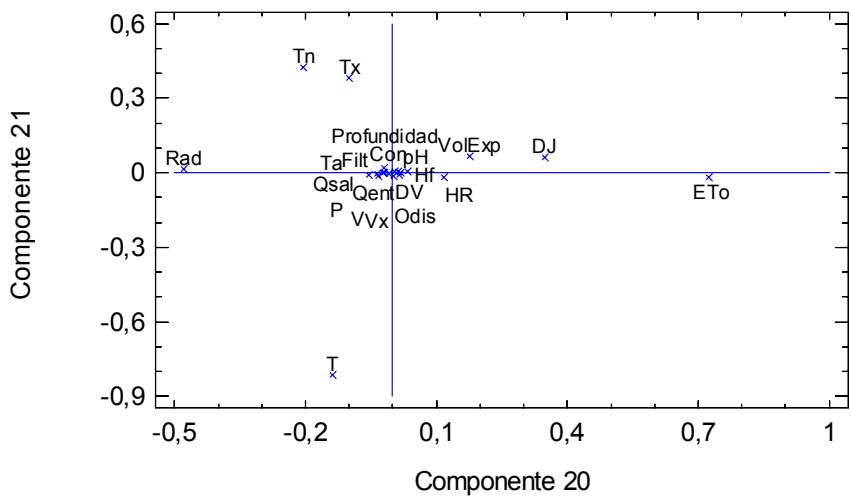
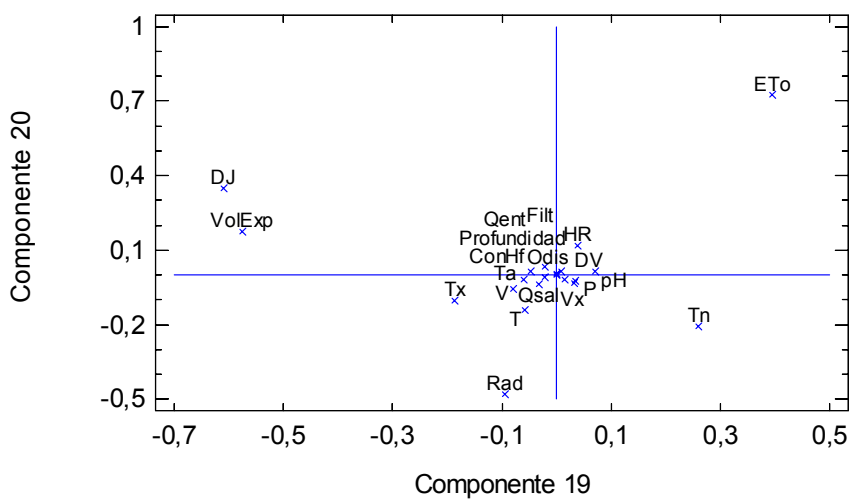
**Diagramas de dispersión de los componentes por pares**











**Gráficos de pesos de los componentes**

**ANEXO 10. PCA DE LOS DATOS METEOROLÓGICOS,  
HIDROLÓGICOS Y DE CALIDAD DEL AGUA EN  
EL PERIODO DE VALIDACIÓN, CORREGIDOS  
(SIN DATOS ANÓMALOS)**

## Resumen del Análisis

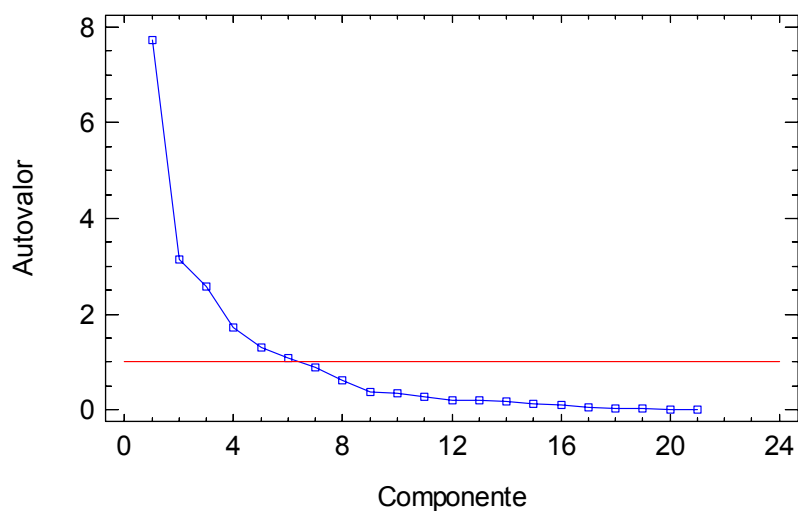
Datos/Variables: DJ, V, DV, Vx, T, Tn, Tx, Hf, HR, Rad, P, ETo, VolExp, Qent, Qsal, Filt, Profundidad, Ta, Odis, pH, y Con.

Entrada de datos: observaciones  
 Número de casos completos: 4573  
 Tratamiento de valor perdido: lista considerada  
 Estandarizado: si  
 Número de componentes extraídos: 21

### Análisis de Componentes Principales

Componente Número	Autovalor	% de Varianza	Porcentaje acumulado
1	7,72942	36,807	36,807
2	3,13491	14,928	51,735
3	2,57818	12,277	64,012
4	1,71952	8,188	72,200
5	1,29475	6,165	78,366
6	1,08345	5,159	83,525
7	0,883742	4,208	87,733
8	0,631541	3,007	90,740
9	0,370144	1,763	92,503
10	0,359236	1,711	94,214
11	0,279491	1,331	95,545
12	0,204631	0,974	96,519
13	0,192406	0,916	97,435
14	0,179038	0,853	98,288
15	0,12756	0,607	98,895
16	0,103649	0,494	99,389
17	0,0578408	0,275	99,664
18	0,0346013	0,165	99,829
19	0,0236953	0,113	99,942
20	0,00910451	0,043	99,985
21	0,00310098	0,015	100

Gráfico de Sedimentación



Variable	Componentes						
	1	2	3	4	5	6	7
DJ	0,325228	-0,102042	0,061353	-0,206803	-0,115388	0,079695	0,013647
V	-0,015512	0,117058	-0,501357	-0,321982	0,127675	0,042792	-0,030111
DV	-0,086790	-0,070931	-0,032149	-0,058796	-0,514069	0,654122	-0,197458
Vx	-0,048630	0,061443	-0,373882	-0,408392	0,372610	0,016341	0,190711
T	0,349791	0,057932	0,012340	0,033083	0,016323	-0,036904	0,076722
Tn	0,345340	0,028408	0,058565	0,006383	0,054375	-0,076418	0,077124
Tx	0,343494	0,070927	-0,022217	0,034826	-0,005821	0,017851	0,085268
Hf	-0,202913	-0,120804	-0,122688	-0,201984	-0,088351	-0,337557	-0,451034
HR	-0,010706	-0,174595	0,475270	0,074436	0,147248	-0,218051	0,055372
Rad	0,166808	0,280675	-0,227920	0,370305	-0,088567	0,056722	-0,114725
P	-0,034412	-0,117109	0,218471	-0,225148	0,402298	0,460511	0,213320
ETo	0,257336	0,244801	-0,261082	0,208101	-0,002517	0,047671	-0,036519
VolExp	-0,243843	0,228348	-0,072857	0,370246	0,227202	-0,020494	0,055460
Qent	-0,299866	-0,084069	-0,060084	-0,079131	-0,190697	-0,151369	0,069078
Qsal	0,320329	-0,054268	0,028098	-0,139494	-0,008835	-0,109229	-0,062059
Filt	-0,071701	-0,010503	-0,135401	-0,051069	-0,469049	-0,226781	0,746881
Profundidad	0,014088	-0,381439	-0,230186	0,371847	0,183592	0,085985	0,101907
Ta	0,184873	0,331480	0,225352	-0,307490	-0,130008	-0,061702	-0,068580
Odis	-0,248403	0,336524	0,106719	-0,053922	-0,014622	-0,032991	-0,020000
pH	-0,173440	0,319637	0,171831	0,043035	0,039104	0,267595	0,226688
Con	0,068383	-0,476791	-0,135942	0,022771	-0,066476	0,080718	0,039201

Variable	Componentes						
	8	9	10	11	12	13	14
DJ	0,049544	0,213481	0,003097	-0,022807	0,141834	-0,144047	-0,136231
V	0,103683	0,073766	0,256497	0,023402	0,267622	-0,051110	0,259123
DV	0,054958	-0,197798	0,342740	-0,074209	-0,276416	0,015898	0,078508
Vx	0,220344	-0,061960	0,241890	0,025256	-0,280130	-0,032233	-0,191145
T	0,049012	-0,088919	-0,014014	0,016151	-0,284028	0,093361	-0,121170
Tn	0,080120	-0,108237	0,027940	-0,000100	-0,243226	0,065841	0,020500
Tx	0,018673	-0,079913	-0,051905	0,061331	-0,311205	0,171767	-0,288563
Hf	-0,430383	0,368453	0,100798	-0,117752	-0,361626	0,147448	-0,151227
HR	0,134673	0,054052	0,759220	0,164488	0,009529	-0,021641	-0,004985
Rad	-0,233628	0,201429	0,309063	0,253753	0,307490	0,068874	-0,167975
P	-0,617884	-0,049839	-0,014355	0,117178	0,100137	0,194693	-0,084242
ETo	-0,147650	-0,006212	0,071256	0,149551	0,004696	0,127416	-0,092620
VolExp	-0,088601	-0,202958	0,081862	-0,002947	-0,245641	0,012982	0,211723
Qent	0,197584	-0,302254	-0,010545	0,149524	0,250311	0,690073	-0,297799
Qsal	-0,078359	-0,059488	0,026443	-0,038399	0,083357	0,389432	0,683403
Filt	-0,314898	0,066088	0,118213	-0,058807	-0,074482	-0,100065	0,085107
Profundidad	0,050198	0,187602	0,060184	-0,348894	-0,165930	0,312034	0,129029
Ta	-0,004212	0,113965	0,025776	-0,075014	-0,023849	0,233789	0,032025
Odis	-0,011267	-0,002527	-0,110476	0,488872	-0,323065	0,015588	0,267942
pH	0,301704	0,646971	-0,060212	-0,140368	-0,031785	0,247412	-0,001549
Con	0,136537	0,301699	-0,159890	0,658790	-0,110755	0,013483	0,095798

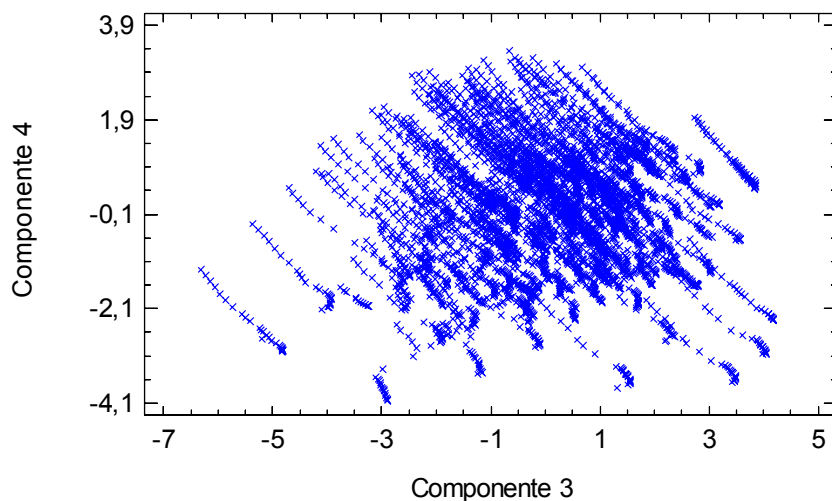
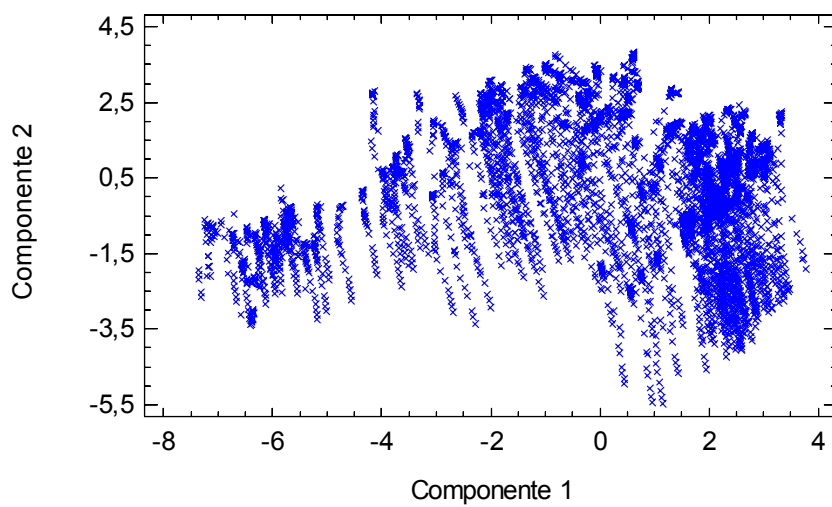
Variable	Componentes						
	15	16	17	18	19	20	21
DJ	0,066105	0,098967	-0,279310	-0,148943	0,026338	-0,769382	0,049426
V	-0,585979	-0,040425	-0,077183	0,036756	-0,188748	0,008090	-0,008494
DV	0,021648	-0,038421	0,025798	-0,020698	0,028476	0,004325	0,001226
Vx	0,512488	0,047868	0,074206	-0,069707	0,090889	0,025768	0,000073
T	-0,194419	-0,108030	-0,015685	-0,153142	-0,070314	0,006573	-0,820084
Tn	-0,328070	-0,080904	0,073774	-0,537989	0,422455	0,125074	0,419198
Tx	-0,072156	-0,162589	-0,191617	0,289395	-0,587028	0,058233	0,380445
Hf	-0,062511	-0,177069	-0,010448	-0,049164	0,041589	-0,057171	0,004323
HR	-0,077904	-0,021805	-0,062098	0,172962	0,010457	0,000521	-0,012058
Rad	0,255082	0,080498	0,002909	-0,410769	-0,190251	0,145059	0,000563
P	-0,107479	-0,005362	-0,004486	-0,042235	0,015819	0,009138	-0,017668
ETo	-0,030087	-0,030599	-0,020711	0,580218	0,577724	-0,109366	-0,001151
VolExp	-0,067273	-0,124673	0,381905	-0,076131	-0,176854	-0,568824	0,057010
Qent	-0,068304	-0,034754	-0,047104	-0,102060	0,058964	-0,163063	-0,010517
Qsal	0,367393	-0,279138	-0,071060	0,009047	-0,051869	-0,009982	-0,004717
Filt	-0,001604	0,002168	0,015976	0,014244	0,004076	0,016368	0,004898
Profundidad	-0,048495	0,510999	-0,207899	-0,002020	-0,010053	0,012399	-0,001897
Ta	-0,053304	0,580238	0,497839	0,106718	-0,102554	-0,007143	0,018134
Odis	-0,020871	0,311210	-0,526626	-0,073447	0,031651	-0,010624	-0,013292
pH	-0,017323	-0,333928	-0,002927	-0,009344	0,064771	0,023450	-0,007051
Con	-0,030181	-0,006983	0,375824	0,040871	-0,035602	-0,022031	0,006096

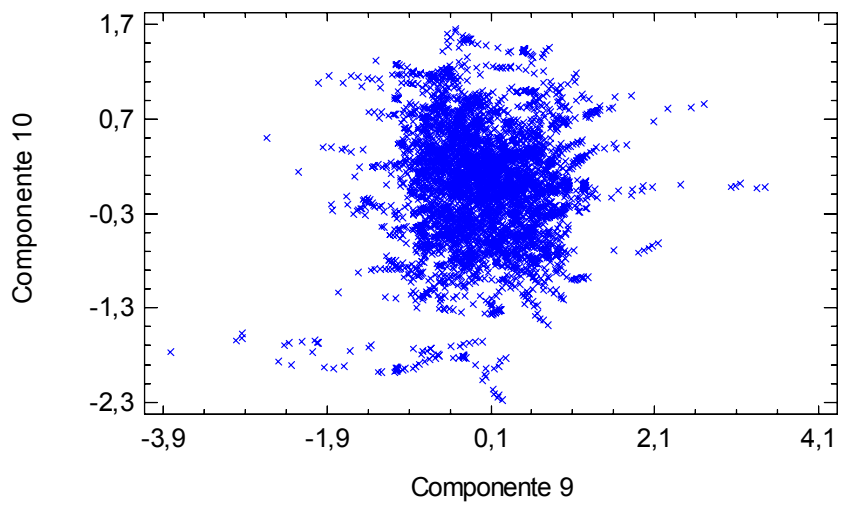
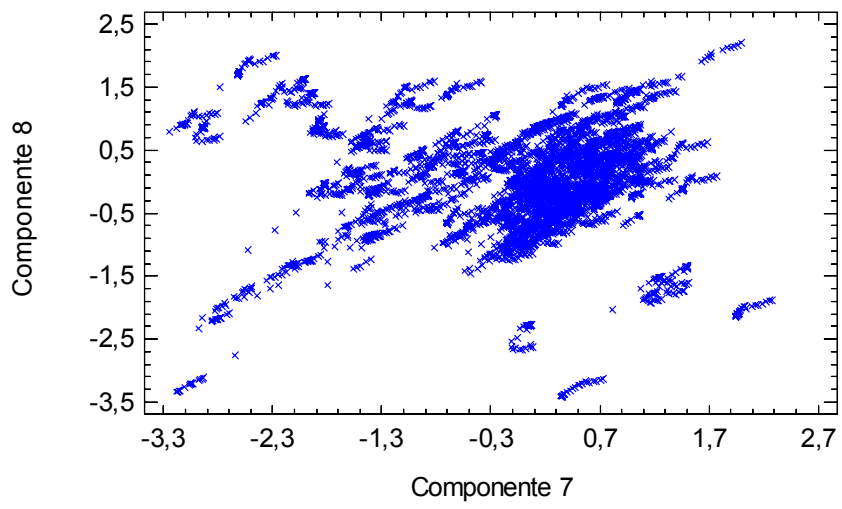
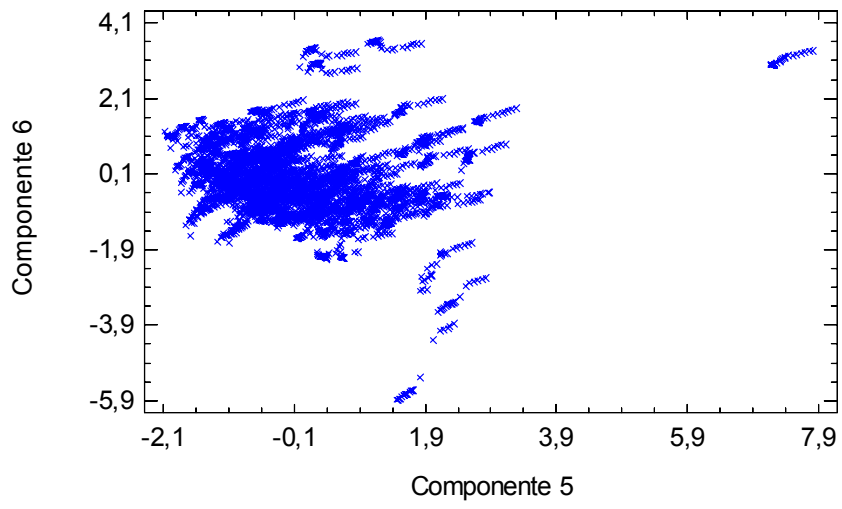
### Tabla de pesos de los componentes

Esta tabla muestra las ecuaciones de los componentes principales. Por ejemplo, el primer componente principal tiene la ecuación:

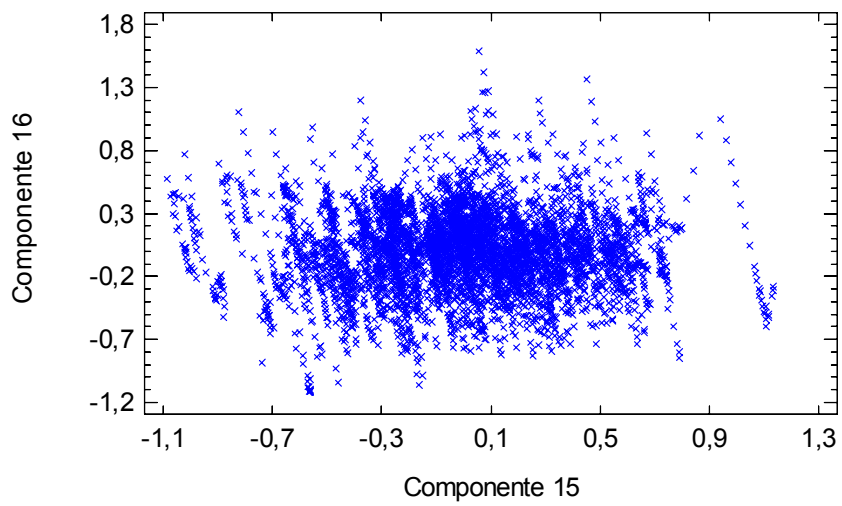
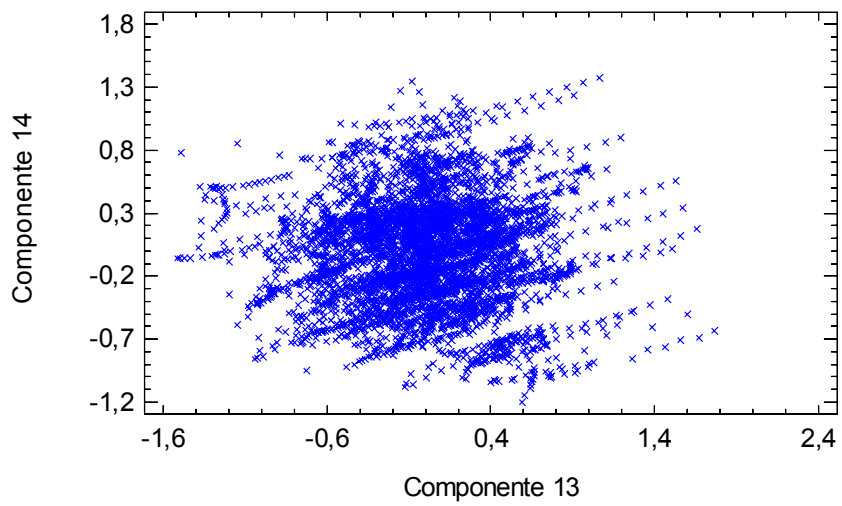
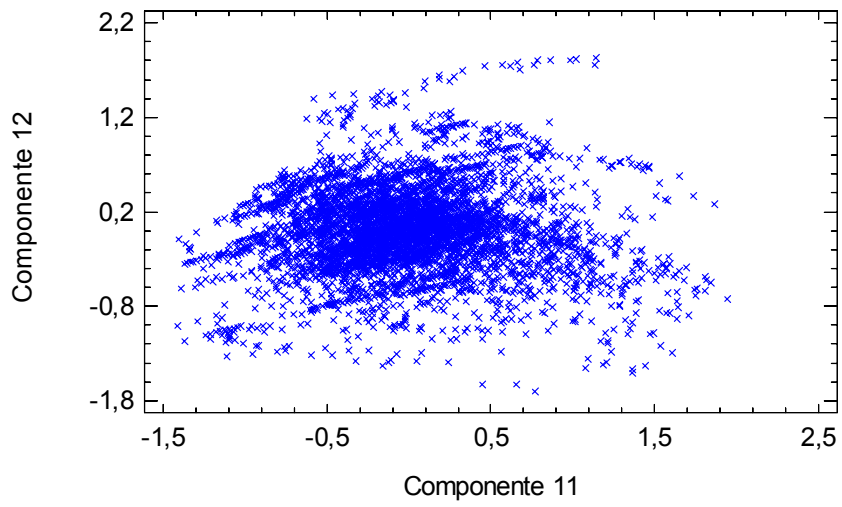
$$\begin{aligned} \text{Componente 1} = & 0,325228(\text{DJ}) - 0,015512(\text{V}) - 0,086790(\text{DV}) - 0,048630(\text{Vx}) + \\ & 0,349791(\text{T}) + 0,345340(\text{Tn}) + 0,343494(\text{Tx}) - 0,202913(\text{Hf}) - 0,010706(\text{HR}) + \\ & 0,166808(\text{Rad}) - 0,034412(\text{P}) + 0,257336(\text{ETo}) - 0,243843(\text{VolExp}) - 0,299866(\text{Qent}) \\ & + 0,320329(\text{Qsal}) - 0,071701(\text{Filt}) + 0,014088(\text{Profundidad}) + 0,184873(\text{Ta}) - \\ & 0,248403(\text{Odis}) - 0,173440(\text{pH}) + 0,068383(\text{Con}) \end{aligned}$$

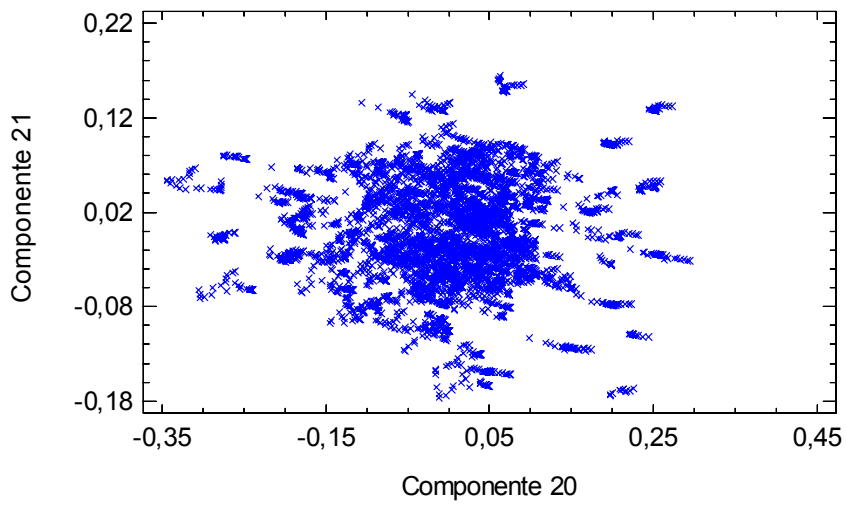
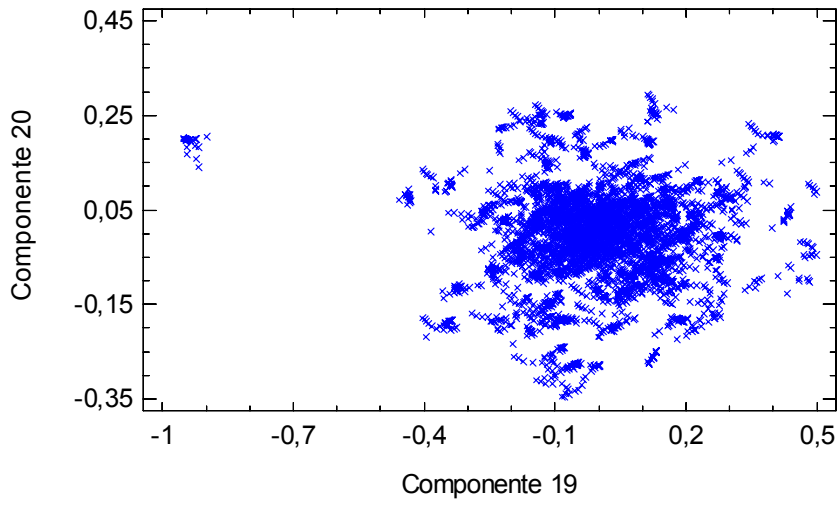
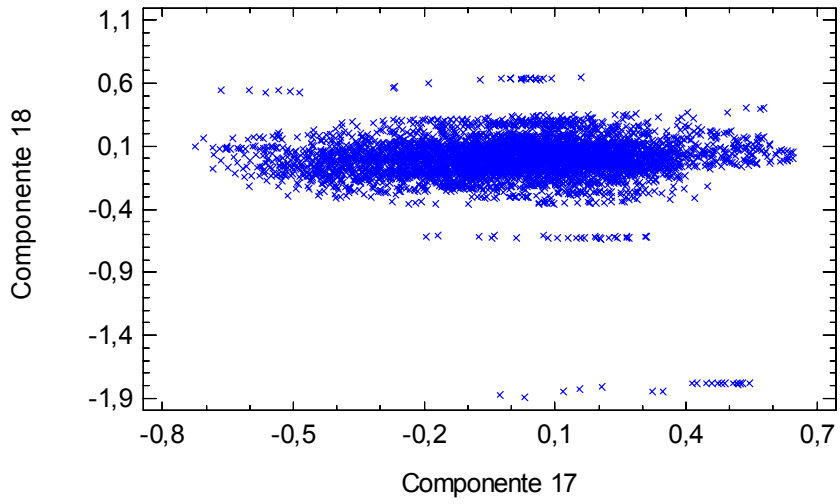
Donde los valores de las variables en la ecuación se encuentran estandarizados (se calculan substrayendo sus medias y dividiéndolos por sus desviaciones típicas).



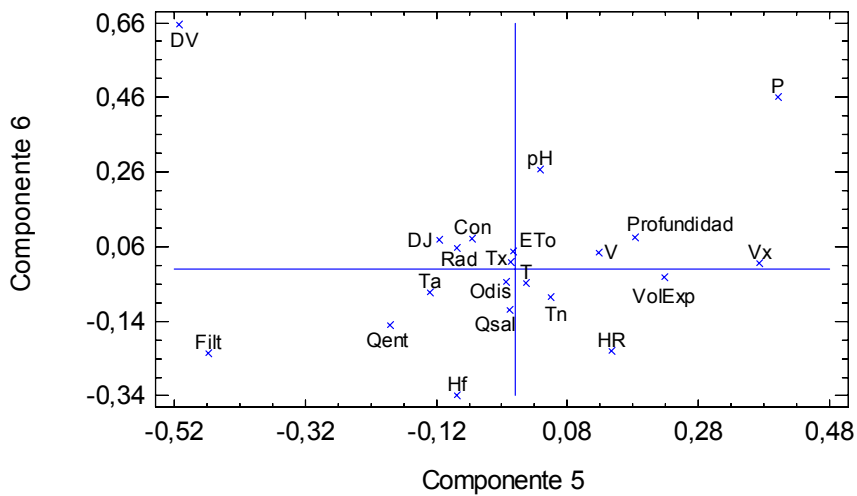
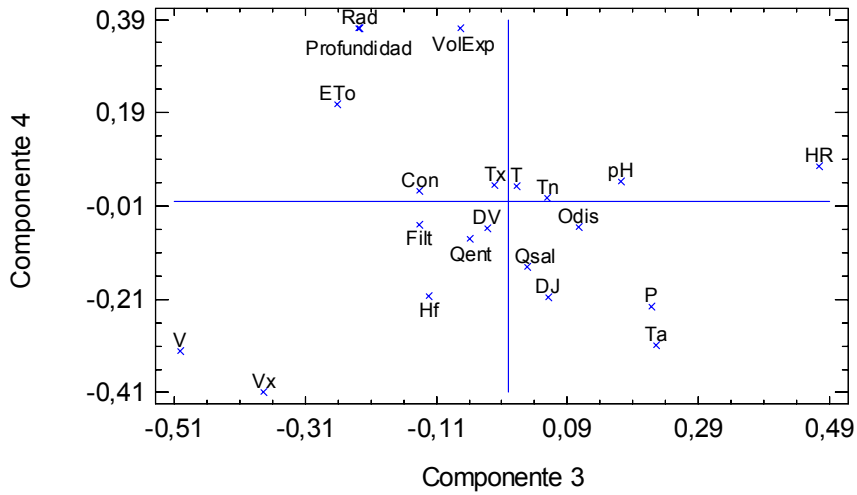
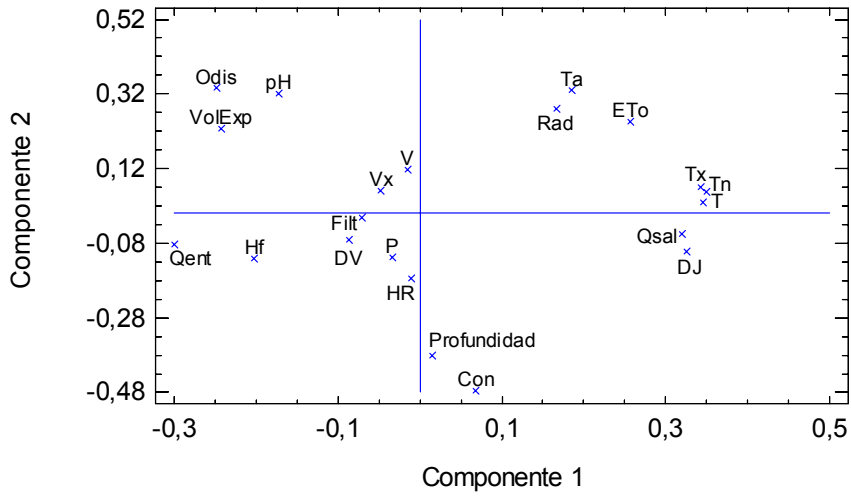


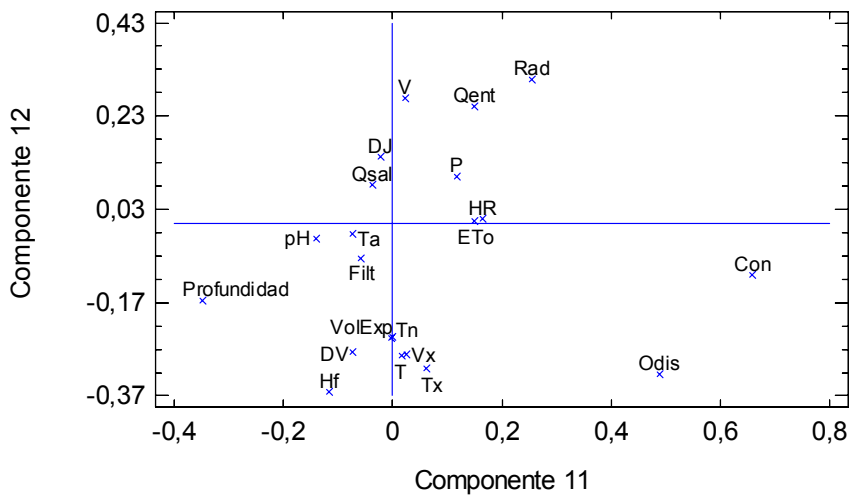
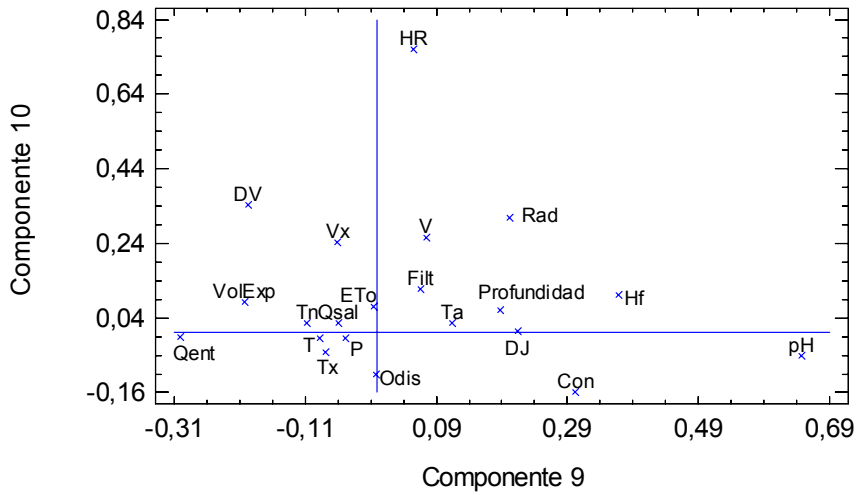
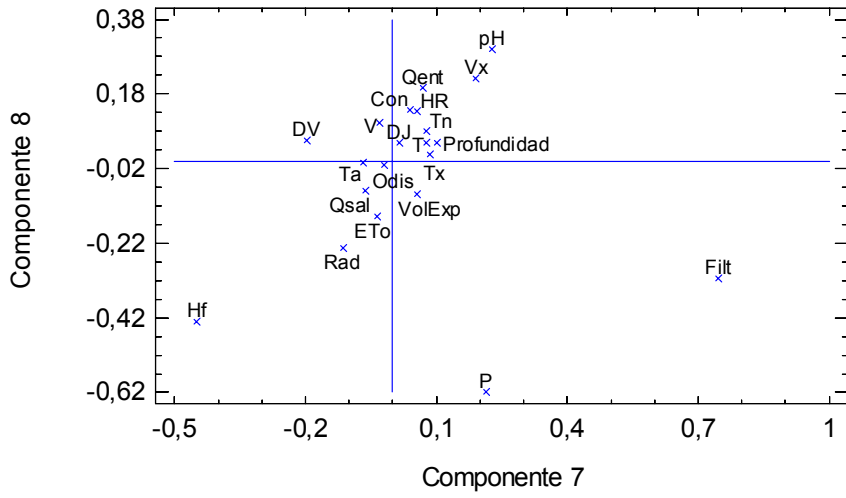


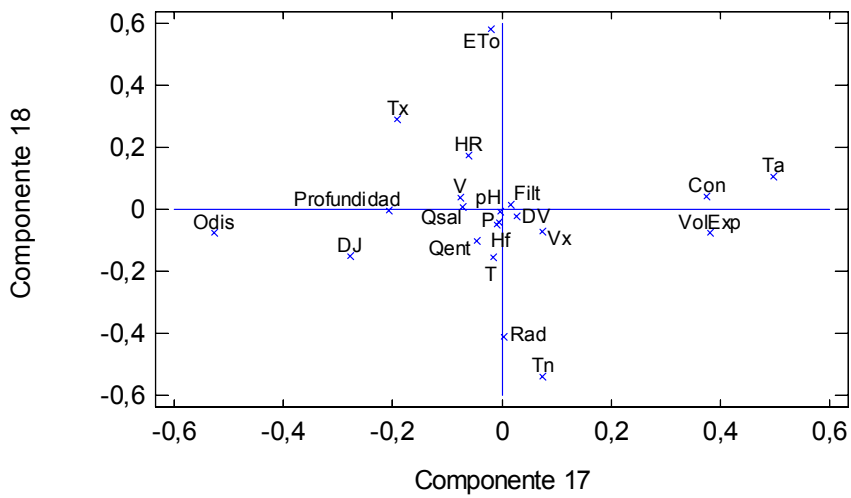
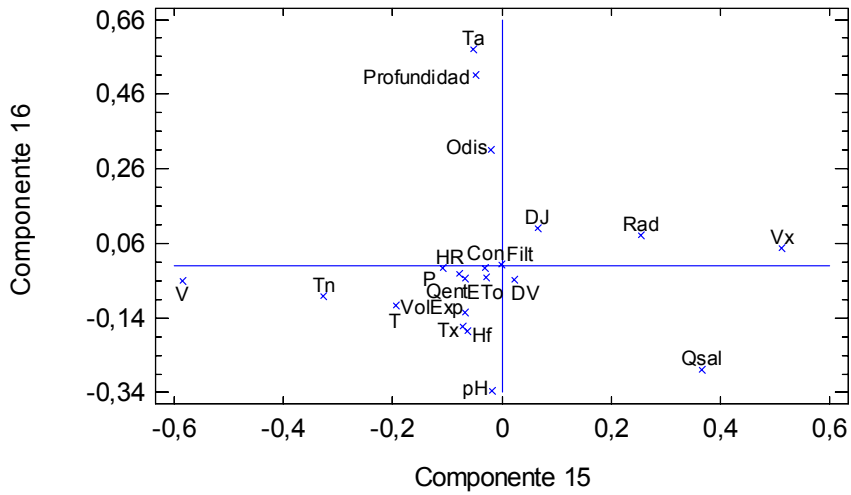
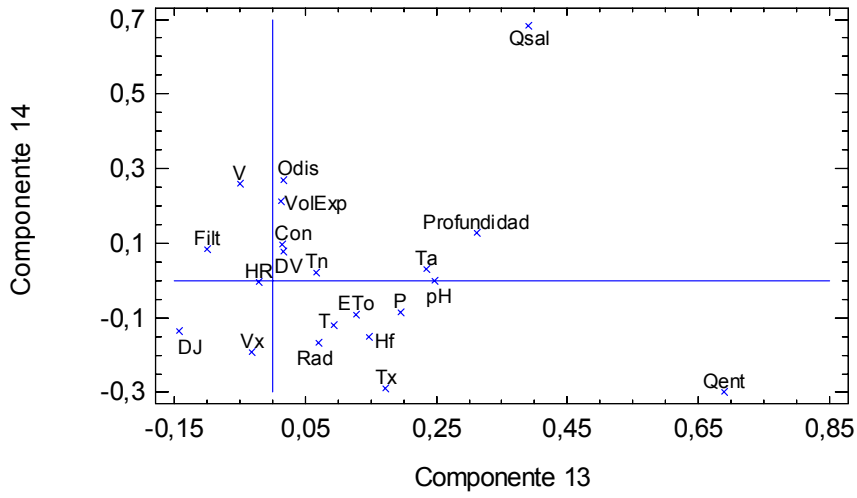


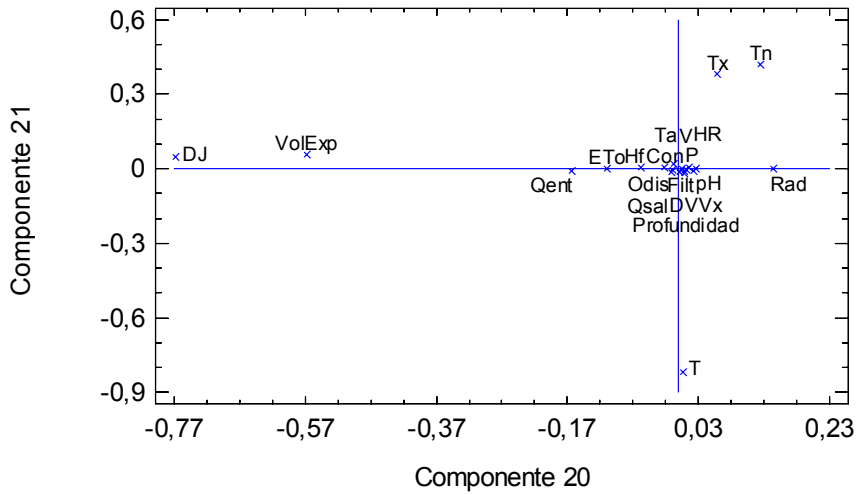
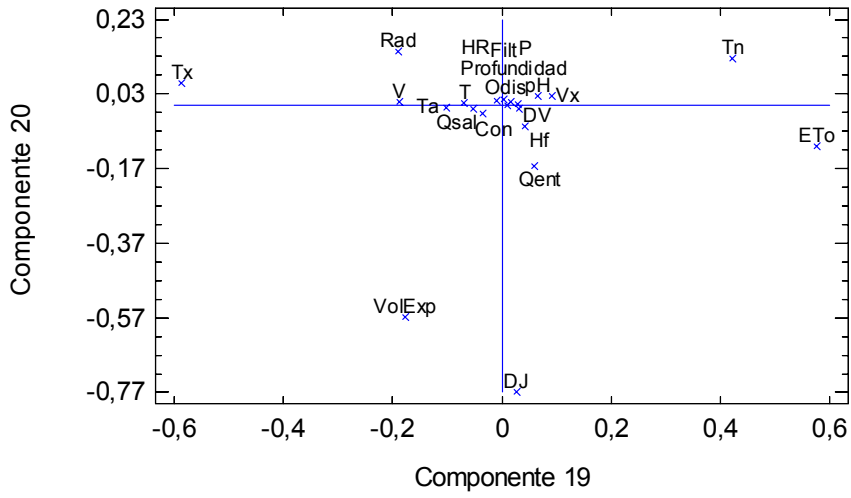


**Diagramas de dispersión de los componentes por pares**









**Gráficos de pesos de los componentes**

**ANEXO 11. PUNTUACIONES DE LOS COMPONENTES EN LOS  
PERIODOS DE CALIBRACIÓN Y VALIDACIÓN (SIN  
DATOS ANÓMALOS)**

## CALIBRACIÓN

Variable	Componentes					
	1	2	3	4	5	6
DJ	-5,73398E-5	-3,75816E-4	7,61838E+3	-3,11047E-3	0,863246	0,486670
V	2,05661E-8	-4,69268E-7	-7,1363E-6	-2,09352E-5	2,19174E-5	8,24045E-4
DV	-2,14556E-5	2,95748E-4	-2,89508E-3	5,92619E-3	-0,496175	0,773584
Vx	2,97511E-8	2,64799E-6	-1,4613E-5	4,00684E-6	1,87413E-3	3,64537E-3
T	9,3327E-7	-9,59452E-5	4,65719E-4	-5,19276E-4	1,70669E-2	-2,06092E-2
Tn	7,41448E-7	-9,43308E-5	4,52298E-4	-5,4212E-4	1,74064E-2	-1,81438E-2
Tx	9,58448E-7	-9,62738E-5	4,54476E-4	-4,76493E-4	1,27124E-2	-2,30657E-2
Hf	7,93742E-7	1,52555E-5	-1,64246E-4	1,77348E-4	-1,12854E-2	8,16419E-3
HR	-3,73036E-7	4,63447E-5	1,68823E-4	-1,37802E-3	1,14816E-2	-4,15525E-2
Rad	3,18944E-5	-8,51589E-4	2,4421E-3	-8,26485E-3	8,63814E-2	-0,401850
P	-4,1029E-7	5,15793E-6	0	4,50352E-6	2,24591E-3	8,05544E-3
ETo	5,84669E-7	-2,1403E-5	7,18036E-5	-1,42521E-4	3,83321E-3	-5,34059E-3
VolExp	0,999984	4,46519E-3	3,35641E-3	1,19965E-4	1,1593E-5	5,33231E-5
Qent	-5,00693E-3	0,982864	0,184254	-1,75546E-3	-9,91943E-4	-4,57382E-4
Qsal	-2,47764E-3	-0,184221	0,982725	1,53707E-2	-8,11113E-3	-4,68935E-4
Filt	-9,04726E-5	4,54721E-3	-1,47227E-2	0,999822	6,50631E-3	-6,4777E-3
Profundidad	4,16889E-7	-7,93989E-7	-1,16016E-5	-2,07218E-5	2,82636E-3	-2,62194E-3

Variable	Componentes					
	7	8	9	10	11	12
DJ	0,130435	2,14758E-2	-2,32096E-3	-1,92282E-2	-8,67791E-4	5,09001E-3
V	1,84789E-3	-1,58885E-2	5,31357E-4	-7,06134E-3	-3,01735E-2	-1,38139E-2
DV	0,385466	8,16061E-2	0	8,37029E-3	9,51266E-4	3,89481E-3
Vx	5,11413E-3	-6,0397E-2	3,64147E-3	7,06451E-2	-1,68736E-2	-0,152525
T	3,33468E-2	7,40618E-2	-2,14918E-4	0,516821	0,243004	8,35378E-2
Tn	2,56222E-2	9,402E-2	7,35522E-4	0,490523	0,239962	-2,46536E-3
Tx	3,72456E-2	7,09598E-2	-2,02229E-3	0,514673	0,219256	0,239157
Hf	-7,60801E-3	-8,95865E-2	-1,10545E-3	-0,432291	0,856291	0,252508
HR	-0,105855	0,977112	-1,6713E-2	-0,155345	6,81762E-3	3,06783E-2
Rad	0,905260	6,83749E-2	-7,46833E-3	-7,73224E-2	-1,26738E-2	-2,14633E-2
P	-1,34203E-2	4,26031E-2	2,75911E-3	4,26459E-2	0,315983	-0,918987
ETo	1,46462E-2	-7,97258E-3	1,09869E-3	5,56639E-2	2,84807E-2	-5,10409E-2
VolExp	-1,64876E-5	4,99607E-8	-1,7899E-7	-4,5139E-7	0	-3,93046E-8
Qent	3,08491E-4	-5,29937E-5	3,80223E-6	1,23367E-5	1,44044E-5	0
Qsal	-2,12888E-3	-3,55349E-4	4,99336E-5	-3,94803E-4	-1,44607E-4	-9,36877E-5
Filt	5,54417E-3	1,6385E-3	-7,23408E-5	-9,3025E-5	1,09546E-4	-3,24468E-5
Profundidad	5,35434E-3	1,70055E-2	0,999816	-3,33804E-3	4,57456E-4	4,30203E-3



Variable	Componentes				
	13	14	15	16	17
DJ	-2,21174E-4	3,53374E-3	-1,04571E-3	-8,17139E-4	1,06276E-4
V	8,81392E-2	-6,91007E-2	-7,55524E-2	0,394125	0,908298
DV	9,00435E-4	-1,87796E-3	7,25827E-4	4,63622E-4	-2,6158E-4
Vx	0,958988	-0,163077	-6,64416E-2	-0,113287	-6,52374E-2
T	3,48511E-2	-3,52217E-2	0,803766	-4,51237E-2	9,49829E-2
Tn	-0,182583	-0,690057	-0,420884	-6,49022E-2	-2,82716E-2
Tx	9,96148E-2	0,668984	-0,404088	4,41682E-2	4,55995E-3
Hf	7,60067E-2	-4,24958E-2	-4,11028E-3	1,05487E-2	1,18397E-2
HR	8,05509E-2	-3,38334E-3	4,69379E-3	1,61746E-2	2,1383E-3
Rad	1,29428E-3	-2,14736E-3	-1,3847E-3	-1,16411E-2	3,0196E-3
P	-0,11069	0,192906	-2,27109E-2	-2,62181E-2	3,25198E-2
ETo	5,98017E-2	-6,79509E-2	5,33754E-2	0,906882	-0,399606
VolExp	-9,71301E-8	2,4074E-7	-7,2583E-8	-1,48209E-7	4,07122E-8
Qent	0	0	-8,53229E-7	2,68462E-6	-6,27473E-8
Qsal	3,26936E-5	-5,3352E-6	6,24428E-6	8,9022E-7	-6,056E-6
Filt	7,52728E-5	-7,76194E-5	0	1,94049E-5	2,61332E-5
Profundidad	-1,51687E-3	1,91481E-3	1,23789E-5	-4,00494E-4	2,26343E-4

**Tabla de pesos de los Componentes**

Esta tabla muestra las ecuaciones de los componentes principales. Por ejemplo, el primer componente principal tiene la ecuación:

$$\begin{aligned} \text{Componente 1} = & - 5,73398E-5(\text{DJ}) + 2,05661E-8(\text{V}) - 2,14556E-5(\text{DV}) + \\ & 2,97511E-8(\text{Vx}) + 9,3327E-7(\text{T}) + 7,41448E-7(\text{Tn}) + 9,58448E-7(\text{Tx}) + \\ & 7,93742E-7(\text{Hf}) - 3,73036E-7(\text{HR}) + 3,18944E-5(\text{Rad}) - 4,1029E-7(\text{P}) + \\ & 5,84669E-7(\text{ETo}) + 0,999984(\text{VolExp}) - 5,00693E-3(\text{Qent}) - 2,47764E-3(\text{Qsal}) - \\ & 9,04726E-5(\text{Filt}) + 4,16889E-7(\text{Profundidad}) \end{aligned}$$

En la Tabla de pesos de los componentes y en la ecuación anterior, las variables se expresan con sus valores y unidades originales, es decir, sin estandarizar.

En la tabla siguiente se muestran las puntuaciones de los componentes para los primeros registros en el periodo de Calibración:

**Componente**

No de registro	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	7813678	70630	37322	938,437	-21,8978	583,173	131,778	91,5008	-3,014	-3,4481	8,2002	3,3524	12,4832	4,507	-1,4357	0,0849	1,72596
2	7813678	70630	37322	938,437	-21,8967	583,172	131,781	91,5078	-2,604	-3,4495	8,2004	3,3542	12,4826	4,5078	-1,4357	0,0847	1,72605
3	7813678	70630	37322	938,437	-21,8955	583,171	131,783	91,515	-2,184	-3,4509	8,2006	3,356	12,4819	4,5086	-1,4357	0,0846	1,72615
4	7813678	70630	37322	938,437	-21,8943	583,17	131,785	91,5221	-1,764	-3,4523	8,2008	3,3578	12,4813	4,5094	-1,4357	0,0844	1,72624
5	7813678	70630	37322	938,437	-21,8931	583,168	131,787	91,5292	-1,344	-3,4537	8,201	3,3596	12,4807	4,5102	-1,4357	0,0842	1,72634
6	7813678	70630	37322	938,437	-21,8907	583,166	131,792	91,5434	-0,514	-3,4565	8,2014	3,3632	12,4794	4,5118	-1,4357	0,0839	1,72653
7	7813678	70630	37322	938,437	-21,8896	583,165	131,794	91,5505	-0,094	-3,4579	8,2016	3,365	12,4788	4,5126	-1,4357	0,0837	1,72662
8	7813678	70630	37322	938,437	-21,8884	583,164	131,796	91,5575	0,316	-3,4592	8,2018	3,3668	12,4781	4,5134	-1,4357	0,0836	1,72672
9	7813678	70630	37322	938,437	-21,886	583,162	131,801	91,5718	1,156	-3,462	8,2021	3,3704	12,4769	4,515	-1,4357	0,0832	1,72691
10	7813678	70630	37322	938,437	-21,8837	583,16	131,805	91,5859	1,985	-3,4648	8,2025	3,3739	12,4756	4,5166	-1,4357	0,0829	1,72709
11	7813678	70630	37322	938,437	-21,8813	583,158	131,81	91,6002	2,825	-3,4676	8,2029	3,3776	12,4743	4,5182	-1,4356	0,0825	1,72728
12	7813678	70630	37322	938,437	-21,879	583,155	131,814	91,6143	3,655	-3,4704	8,2033	3,3811	12,4731	4,5198	-1,4356	0,0822	1,72747
13	7813678	70630	37322	938,437	-21,8766	583,153	131,819	91,6284	4,485	-3,4732	8,2037	3,3847	12,4718	4,5214	-1,4356	0,0819	1,72766
14	7813678	70630	37322	938,437	-21,8754	583,152	131,821	91,6355	4,905	-3,4746	8,2039	3,3865	12,4712	4,5222	-1,4356	0,0817	1,72775
15	7813678	70630	37322	938,437	-21,8742	583,151	131,823	91,6427	5,325	-3,476	8,2041	3,3883	12,4705	4,523	-1,4356	0,0815	1,72785
16	7813678	70630	37322	938,437	-21,8731	583,15	131,825	91,6496	5,735	-3,4773	8,2042	3,3901	12,4699	4,5237	-1,4356	0,0814	1,72794
17	7813678	70630	37322	938,437	-21,8719	583,149	131,828	91,6568	6,155	-3,4787	8,2044	3,3919	12,4693	4,5245	-1,4356	0,0812	1,72804
18	7813678	70630	37322	938,437	-21,8707	583,148	131,83	91,6639	6,575	-3,4801	8,2046	3,3937	12,4686	4,5254	-1,4356	0,081	1,72813
19	7813678	70630	37322	938,437	-21,8684	583,146	131,834	91,678	7,404	-3,4829	8,205	3,3973	12,4674	4,5269	-1,4356	0,0807	1,72832
20	7813678	70630	37322	938,437	-21,866	583,143	131,839	91,6923	8,244	-3,4857	8,2054	3,4009	12,4661	4,5286	-1,4356	0,0804	1,72851
21	7813678	70630	37322	938,437	-21,8636	583,141	131,843	91,7064	9,074	-3,4885	8,2058	3,4044	12,4649	4,5301	-1,4356	0,08	1,7287
22	7813678	70630	37322	938,437	-21,8613	583,139	131,848	91,7206	9,904	-3,4912	8,2061	3,408	12,4636	4,5317	-1,4356	0,0797	1,72889
23	7813678	70630	37322	938,437	-21,8589	583,137	131,852	91,7348	10,74	-3,494	8,2065	3,4116	12,4623	4,5333	-1,4355	0,0794	1,72908
24	7813678	70630	37322	938,437	-21,8566	583,135	131,857	91,749	11,57	-3,4968	8,2069	3,4152	12,4611	4,5349	-1,4355	0,079	1,72926
25	7813678	70630	37322	938,437	-21,8542	583,132	131,861	91,7632	12,41	-3,4996	8,2073	3,4188	12,4598	4,5365	-1,4355	0,0787	1,72945
26	7813678	70630	37322	938,437	-21,8519	583,13	131,866	91,7774	13,24	-3,5024	8,2077	3,4224	12,4585	4,5381	-1,4355	0,0784	1,72964
27	7813678	70630	37322	938,436	-21,8495	583,128	131,87	91,7915	14,07	-3,5052	8,2081	3,426	12,4573	4,5397	-1,4355	0,078	1,72983
28	7813678	70630	37322	938,436	-21,8471	583,126	131,874	91,8058	14,91	-3,508	8,2084	3,4296	12,456	4,5413	-1,4355	0,0777	1,73002
29	7813678	70630	37322	938,436	-21,8448	583,124	131,879	91,8199	15,74	-3,5107	8,2088	3,4331	12,4547	4,5429	-1,4355	0,0774	1,73021
30	7813678	70630	37322	938,436	-21,8424	583,121	131,883	91,8342	16,58	-3,5135	8,2092	3,4368	12,4535	4,5445	-1,4355	0,077	1,7304
31	7813678	70630	37322	938,436	-21,8413	583,12	131,886	91,8411	16,99	-3,5149	8,2094	3,4385	12,4528	4,5453	-1,4355	0,0769	1,73049
32	7813678	70630	37322	938,436	-21,8389	583,118	131,89	91,8554	17,83	-3,5177	8,2098	3,4421	12,4516	4,5469	-1,4355	0,0765	1,73068
33	7813678	70630	37322	938,436	-21,8365	583,116	131,895	91,8695	18,66	-3,5205	8,2102	3,4457	12,4503	4,5485	-1,4355	0,0762	1,73087
34	7813678	70630	37322	938,436	-21,8342	583,114	131,899	91,8838	19,5	-3,5233	8,2105	3,4493	12,449	4,5501	-1,4354	0,0759	1,73106
35	7813678	70630	37322	938,436	-21,8318	583,112	131,904	91,8979	20,33	-3,5261	8,2109	3,4529	12,4478	4,5517	-1,4354	0,0755	1,73125
36	7863581	89831	41047	909,777	-32,998	542,195	221,745	102,987	-3,748	-11,531	7,2995	1,6028	12,3544	4,3217	-1,3601	-0,275	1,96489
37	7863581	89831	41047	909,777	-32,9968	542,194	221,747	102,994	-3,338	-11,532	7,2997	1,6045	12,3537	4,3225	-1,3601	-0,275	1,96498
38	7863581	89831	41047	909,777	-32,9956	542,192	221,749	103,001	-2,918	-11,534	7,2998	1,6063	12,3531	4,3233	-1,3601	-0,275	1,96508
39	7863581	89831	41047	909,777	-32,9944	542,191	221,752	103,008	-2,498	-11,535	7,3	1,6081	12,3525	4,3241	-1,3601	-0,275	1,96517
40	7863581	89831	41047	909,777	-32,9933	542,19	221,754	103,015	-2,079	-11,537	7,3002	1,6099	12,3518	4,3249	-1,3601	-0,275	1,96527

## VALIDACIÓN

Variable	Componentes					
	1	2	3	4	5	6
DJ	-3,29408E-5	-1,68437E-3	-3,55091E-4	1,71024E-3	2,16727E-2	0,105182
V	5,90335E-9	1,2705E-6	4,17567E-6	1,1237E-5	3,02279E-4	1,41413E-3
DV	-2,54663E-6	1,98672E-3	-2,2716E-3	1,58111E-4	0,958824	-0,282381
Vx	5,07218E-8	1,11373E-5	1,12728E-5	3,45476E-6	-2,85769E-3	3,12602E-4
T	-1,85812E-6	-2,60928E-4	9,43213E-5	2,2586E-4	-1,80388E-3	1,97605E-2
Tn	-1,88982E-6	-2,50184E-4	1,10773E-4	1,10494E-4	-6,12466E-3	1,14234E-2
Tx	-1,86411E-6	-2,56791E-4	8,00301E-5	2,80062E-4	1,94079E-3	2,60059E-2
Hf	2,27322E-7	8,05363E-5	1,57146E-5	-7,77307E-5	-2,61168E-3	-9,71658E-4
HR	-6,62691E-7	4,0177E-5	-1,60707E-5	-1,35121E-3	-4,93745E-2	-7,7975E-2
Rad	4,65728E-6	-2,91262E-3	2,82585E-4	6,33443E-3	0,278686	0,949458
P	-5,3043E-8	0	-5,18557E-5	-1,1202E-4	-1,65805E-3	-1,44884E-2
ETo	-9,69943E-8	-6,36744E-5	2,04581E-5	1,08738E-4	3,35706E-3	1,35716E-2
VolExp	0,99998	-5,55921E-3	2,9295E-3	7,89852E-5	1,36963E-5	-2,42754E-5
Qent	3,30964E-3	0,861112	0,505764	-5,1649E-2	2,60826E-4	2,91214E-3
Qsal	-5,34202E-3	-0,50675	0,862034	-7,88351E-3	2,36709E-3	-2,4941E-3
Filt	4,99392E-5	4,05586E-2	3,29627E-2	0,998612	-1,99155E-3	-6,14888E-3
Profundidad	1,21484E-7	-1,03258E-5	5,5306E-6	-3,75758E-5	-2,27211E-4	-6,67571E-4

Variable	Componentes					
	7	8	9	10	11	12
DJ	0,677423	0,684336	2,18032E-2	-0,232037	-7,53918E-2	1,21236E-2
V	-1,07066E-2	9,76078E-3	-2,74004E-4	-1,49998E-2	-2,19668E-2	2,86092E-2
DV	2,57074E-2	-9,36037E-3	1,27251E-4	1,02235E-2	-5,46705E-3	1,84583E-3
Vx	-3,09364E-2	4,91355E-2	2,8271E-4	-5,01333E-2	-1,30573E-2	0,280973
T	7,93786E-2	9,89542E-2	2,05841E-3	0,546703	-8,52051E-2	3,56971E-2
Tn	9,57362E-2	7,73419E-2	1,25799E-3	0,504105	-0,123395	-0,135467
Tx	6,52381E-2	0,118764	3,30759E-3	0,595032	6,0699E-2	0,353170
Hf	-8,0361E-2	-7,25628E-2	-1,42672E-3	-0,131814	-0,847643	0,471721
HR	0,716062	-0,687598	-5,61476E-3	2,17097E-2	-8,44038E-3	2,96345E-2
Rad	-1,30106E-2	-0,141856	-2,32995E-3	-6,08754E-3	1,43876E-2	3,96763E-3
P	6,68789E-3	-4,23753E-3	-1,82232E-3	-0,114948	0,498679	0,742151
ETo	-9,77162E-3	2,47014E-2	1,15963E-3	7,17606E-2	1,48141E-3	3,90755E-2
VolExp	1,81672E-5	1,90211E-5	3,90148E-7	0	0	5,36922E-7
Qent	1,09611E-3	9,26297E-4	3,71889E-5	-1,07874E-4	-3,11355E-5	5,6859E-5
Qsal	-2,98971E-4	-2,37494E-4	-1,9343E-5	-1,91315E-4	3,19438E-5	-1,29655E-5
Filt	-1,09966E-4	-1,22913E-3	7,2382E-6	7,97758E-5	3,19873E-5	1,90305E-5
Profundidad	-1,13711E-2	-1,99611E-2	0,999732	9,67507E-4	1,45621E-3	7,486E-4

Variable	Componentes				
	13	14	15	16	17
DJ	-3,22049E-2	-6,05327E-4	3,80821E-3	6,47717E-3	-1,25504E-3
V	0,159159	2,59479E-2	4,34262E-2	0,174946	0,969415
DV	4,68884E-3	-2,03708E-4	3,84586E-5	-9,20592E-5	-2,78485E-4
Vx	0,860473	0,374546	-4,14673E-2	-0,115324	-0,138828
T	2,18802E-2	-8,01E-2	-0,816556	3,56997E-2	3,40154E-2
Tn	0,346186	-0,621608	0,428943	-7,34027E-2	-3,69012E-2
Tx	-0,297211	0,515576	0,367877	-8,84363E-2	3,41233E-2
Hf	-0,121424	-0,121544	1,24853E-2	4,23893E-3	-1,34589E-2
HR	3,62839E-2	5,29537E-2	-2,04871E-3	1,88444E-2	3,54587E-3
Rad	8,92754E-3	-8,05426E-3	-9,08379E-4	-1,05775E-2	6,26126E-4
P	-5,59552E-2	-0,425001	-5,57874E-2	9,31602E-3	9,13734E-3
ETo	7,2564E-2	4,62618E-2	8,37704E-2	0,97005	-0,192352
VolExp	-9,2531E-7	6,47871E-8	1,20976E-7	7,43331E-8	1,0731E-8
Qent	-3,52641E-5	-4,79389E-5	-4,7237E-6	7,99339E-6	0
Qsal	-1,88237E-5	4,2277E-5	0	0	3,72791E-7
Filt	5,66405E-5	1,46744E-7	1,54949E-5	8,74718E-7	2,25874E-7
Profundidad	8,6984E-4	-1,56741E-3	-3,29696E-4	-7,70227E-4	4,37801E-4

**Tabla de pesos de los Componentes**

Esta tabla muestra las ecuaciones de los componentes principales. Por ejemplo, el primer componente principal tiene la ecuación

$$\text{Componente 1} = -3,29408E-5(DJ) + 5,90335E-9(V) - 2,54663E-6(DV) + 5,07218E-8(Vx) - 1,85812E-6(T) - 1,88982E-6(Tn) - 1,86411E-6(Tx) + 2,27322E-7(Hf) - 6,62691E-7(HR) + 4,65728E-6(Rad) - 5,3043E-8(P) - 9,69943E-8(ETo) + 0,99998(VolExp) + 3,30964E-3(Qent) - 5,34202E-3(Qsal) + 4,99392E-5(Filt) + 1,21484E-7(Profundidad)$$

En la Tabla de pesos de los componentes y en la ecuación anterior, las variables se expresan con sus valores y unidades originales, es decir, sin estandarizar.

En la tabla siguiente se muestran las puntuaciones de los componentes para los primeros registros en el periodo de Validación:

No de registro	Componente																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	10104844.9	-29747	63159,773	-1381,2791	507,682674	-98,43071	315,41423	193,7561	6,09286	-1,29601	-11,7995	20,7784	-6,952823	5,703919	0,5925899	1,24092	1,091861
2	10104844.9	-29747	63159,773	-1381,2791	507,682581	-98,43098	315,40956	193,7479	6,50275	-1,29561	-11,7989	20,7787	-6,952466	5,703276	0,5924547	1,2406	1,092041
3	10104844.9	-29747	63159,773	-1381,2791	507,682485	-98,43126	315,40479	193,7395	6,92263	-1,29582	-11,7983	20,7779	-6,952101	5,702618	0,5923163	1,24028	1,092225
4	10104844.9	-29747	63159,773	-1381,2792	507,68239	-98,43154	315,40001	193,7311	7,34252	-1,2948	-11,7977	20,7793	-6,951736	5,701959	0,5921778	1,23995	1,092409
5	10104844.9	-29747	63159,773	-1381,2792	507,682297	-98,43182	315,39535	193,7229	7,75241	-1,2944	-11,7971	20,7796	-6,951379	5,701317	0,5920426	1,23964	1,092588
6	10104844.9	-29747	63159,773	-1381,2792	507,682201	-98,4321	315,39057	193,7145	8,1723	-1,29399	-11,7965	20,7799	-6,951014	5,700658	0,5919042	1,23931	1,092772
7	10104844.9	-29747	63159,773	-1381,2792	507,682013	-98,43265	315,38114	193,698	9,00208	-1,29319	-11,7953	20,7806	-6,950292	5,699357	0,5916305	1,23868	1,093135
8	10104844.9	-29747	63159,773	-1381,2792	507,681917	-98,43293	315,37636	193,6896	9,42196	-1,29278	-11,7947	20,7809	-6,949926	5,698699	0,591492	1,23835	1,093319
9	10104844.9	-29747	63159,773	-1381,2793	507,681822	-98,43321	315,37158	193,6812	9,84185	-1,29238	-11,7941	20,7812	-6,949561	5,698041	0,5913536	1,23803	1,093503
10	10104844.9	-29747	63159,773	-1381,2793	507,681729	-98,43349	315,36692	193,6731	10,2517	-1,29198	-11,7935	20,7815	-6,949204	5,697398	0,5912184	1,23771	1,093683
11	10104844.9	-29747	63159,773	-1381,2793	507,681633	-98,43377	315,36215	193,6646	10,6716	-1,29157	-11,7928	20,7818	-6,948839	5,69674	0,5910799	1,23739	1,093867
12	10104844.9	-29747	63159,773	-1381,2793	507,681445	-98,43432	315,35271	193,6481	11,5014	-1,29077	-11,7916	20,7824	-6,948117	5,695439	0,5908063	1,23675	1,09423
13	10104844.9	-29747	63159,773	-1381,2794	507,681254	-98,43488	315,34316	193,6313	12,3412	-1,28996	-11,7904	20,7831	-6,947386	5,694122	0,5905293	1,2361	1,094598
14	10104844.9	-29747	63159,773	-1381,2794	507,681161	-98,43516	315,33849	193,6231	12,7511	-1,28956	-11,7898	20,7834	-6,94703	5,69348	0,5903941	1,23579	1,094777
15	10104844.9	-29747	63159,773	-1381,2794	507,681065	-98,43544	315,33372	193,6147	13,171	-1,28916	-11,7892	20,7837	-6,946664	5,692821	0,5902557	1,23546	1,094961
16	10104844.9	-29747	63159,773	-1381,2794	507,680877	-98,43599	315,32428	193,5982	14,0007	-1,28835	-11,788	20,7843	-6,945942	5,69152	0,589982	1,23482	1,095324
17	10104844.9	-29747	63159,773	-1381,2794	507,680686	-98,43655	315,31473	193,5814	14,8405	-1,28754	-11,7868	20,7849	-6,945212	5,690204	0,5897051	1,23418	1,095692
18	10104844.9	-29747	63159,773	-1381,2795	507,680497	-98,43711	315,30529	193,5648	15,6703	-1,28674	-11,7856	20,7856	-6,944472	5,688903	0,5894314	1,23354	1,096056
19	10104844.9	-29747	63159,773	-1381,2796	507,679929	-98,43877	315,27686	193,5149	18,1696	-1,28432	-11,7819	20,7874	-6,942315	5,684984	0,5886072	1,23161	1,09715
20	10104844.9	-29747	63159,773	-1381,2797	507,679361	-98,44044	315,24844	193,465	20,6689	-1,2819	-11,7783	20,7893	-6,940141	5,681066	0,587783	1,22969	1,098245
21	10104844.9	-29747	63159,773	-1381,2798	507,678793	-98,44211	315,22001	193,4151	23,1683	-1,27948	-11,7746	20,7912	-6,937966	5,677147	0,5869587	1,22776	1,099339
22	10104844.9	-29747	63159,773	-1381,2799	507,678225	-98,44378	315,19158	193,3652	25,6676	-1,27706	-11,771	20,793	-6,935791	5,673229	0,5861345	1,22584	1,100434
23	10104844.9	-29747	63159,773	-1381,2799	507,677657	-98,44545	315,16315	193,3153	28,1669	-1,27464	-11,7674	20,7949	-6,933617	5,66931	0,5853102	1,22391	1,101528
24	10104844.9	-29747	63159,773	-1381,28	507,677089	-98,44712	315,13472	193,2654	30,6663	-1,27222	-11,7637	20,7968	-6,931442	5,665392	0,584486	1,22198	1,102623
25	10104844.9	-29747	63159,773	-1381,2801	507,676521	-98,44879	315,1063	193,2155	33,1656	-1,26981	-11,7601	20,7987	-6,929268	5,661473	0,5836618	1,22006	1,103717
26	10127840	-31110,2	62501,222	-1306,0432	508,116997	-218,7664	316,66087	209,6545	6,33138	-2,12557	-9,75644	18,0599	-6,072591	4,087485	1,2625135	1,9376	0,974749
27	10127840	-31110,2	62501,222	-1306,0432	508,116904	-218,7666	316,65621	209,6464	6,74127	-2,12517	-9,75584	18,0602	-6,072235	4,086842	1,2623783	1,93728	0,974929
28	10127840	-31110,2	62501,222	-1306,0432	508,116808	-218,7669	316,65143	209,638	7,16115	-2,12477	-9,75523	18,0605	-6,07187	4,086184	1,2622398	1,93696	0,975113
29	10127840	-31110,2	62501,222	-1306,0432	508,116713	-218,7672	316,64666	209,6296	7,58104	-2,12436	-9,75462	18,0609	-6,071504	4,085525	1,2621013	1,93664	0,975297
30	10127840	-31110,2	62501,222	-1306,0432	508,11662	-218,7675	316,64199	209,6214	7,99093	-2,12396	-9,75402	18,0612	-6,071148	4,084883	1,2619662	1,93632	0,975476
31	10127840	-31110,2	62501,222	-1306,0432	508,116524	-218,7677	316,63722	209,613	8,41082	-2,12356	-9,75341	18,0615	-6,070782	4,084224	1,2618277	1,936	0,975566
32	10127840	-31110,2	62501,222	-1306,0433	508,116336	-218,7683	316,62778	209,5984	9,2406	-2,12275	-9,7522	18,0621	-6,07006	4,082925	1,261554	1,93536	0,976023
33	10127840	-31110,2	62501,222	-1306,0433	508,11624	-218,7686	316,623	209,5864	9,66048	-2,12235	-9,75159	18,0624	-6,069695	4,082265	1,2614156	1,93503	0,976207
34	10127840	-31110,2	62501,222	-1306,0433	508,116145	-218,7689	316,61823	209,5797	10,0804	-2,12194	-9,75098	18,0627	-6,06933	4,081607	1,2612771	1,93471	0,976391
35	10127840	-31110,2	62501,222	-1306,0433	508,116052	-218,7691	316,61357	209,5715	10,4903	-2,12154	-9,75038	18,063	-6,068973	4,080964	1,2611419	1,9344	0,976571
36	10127840	-31110,2	62501,222	-1306,0433	508,115956	-218,7694	316,60879	209,5631	10,9101	-2,12114	-9,74977	18,0634	-6,068608	4,080306	1,2610034	1,93407	0,976755
37	10127840	-31110,2	62501,222	-1306,0434	508,115768	-218,77	316,59935	209,5465	11,7399	-2,12034	-9,74856	18,064	-6,067886	4,079005	1,2607298	1,93343	0,977118
38	10127840	-31110,2	62501,222	-1306,0434	508,115577	-218,7705	316,5898	209,5298	12,5797	-2,11952	-9,74734	18,0646	-6,067155	4,077688	1,2604529	1,93279	0,977486
39	10127840	-31110,2	62501,222	-1306,0434	508,115484	-218,7708	316,58514	209,5216	12,9896	-2,11913	-9,74674	18,0649	-6,066798	4,077046	1,2603177	1,93247	0,977665
40	10127840	-31110,2	62501,222	-1306,0434	508,115388	-218,7711	316,58036	209,5132	13,4095	-2,11872	-9,74613	18,0652	-6,066433	4,076387	1,2601792	1,93215	0,977849

Nomenclatura y unidades empleadas para las variables independientes o predictivas:

Clasificación de la variable	Variable y sus unidades de medida
Temporal	DJ, Día juliano
Meteorológicas	V, Velocidad media del viento (Km/h)
	DV, Dirección media del viento (°)
	Vx, Velocidad máxima del viento (Km/h)
	T, Temperatura media (°C)
	Tn, Temperatura mínima (°C)
	Tx, Temperatura máxima del aire (°C)
	Horas frío, Hf, Número de horas por debajo de 7 °C
	HR, Humedad relativa media del aire (%)
	Rad, Radiación solar global (W/m <sup>2</sup> )
	P, Precipitación total (mm)
ETo, Evapotranspiración de referencia (mm)	
Hidrológicas	VolExp, Volumen de explotación (m <sup>3</sup> )
	Qent, Caudal de entrada al embalse (m <sup>3</sup> /día)
	Qsal, Caudal de salida del embalse (m <sup>3</sup> /día)
	Filt, Filtración (m <sup>3</sup> /día)
	Profundidad (m)

**ANEXO 12. ANÁLISIS FACTORIAL PARA EL PERIODO  
COMPLETO DE ESTUDIO, SIN DATOS  
ANÓMALOS Y LAS VARIABLES SIN  
TRANSFORMAR**

## PERIODO COMPLETO DE ESTUDIO

### Estadísticos descriptivos

	Media	Desviación típica	N del análisis
DJ	186,95	97,561	8974
V	1,283	,303	8974
DV	235,85	79,106	8974
Vx	5,168	1,3697	8974
T	19,066	6,260	8974
Tn	14,859	6,052	8974
Tx	23,630	6,312	8974
Hf	1,01	3,119	8974
HR	58,16	12,821	8974
Rad	201,76	81,054	8974
P	,46	2,025	8974
ETo	3,51	1,593	8974
VolExp	8853793,2	1507776,4	8974
Qent	20238,4	28660,2	8974
Qsal	27093,9	11742,2	8974
Filt	-501,2	1390,7	8974
Profundidad	8,885	7,629	8974
Ta	16,73	6,361	8974
Odis	8,04	4,215	8974
pH	8,24	,409	8974
Con	984,63	77,421	8974

a) Extracción de los factores por el Método de Componentes principales, sin rotación

### KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,798
	Chi-cuadrado aproximado	212407,580
Prueba de esfericidad de Bartlett	gl	210
	Sig.	,000



**Matriz de correlaciones (a)**

		DJ	V	DV	Vx	T	Tn	Tx	Hf	HR	Rad	P
Correlación	DJ	1,000	-0,062	0,017	-0,014	0,362	0,389	0,336	-0,421	0,036	-0,141	0,145
	V	-0,062	1,000	0,126	0,689	-0,158	-0,179	-0,153	0,208	-0,646	0,068	-0,188
	DV	0,017	0,126	1,000	0,006	-0,486	-0,494	-0,454	0,199	-0,202	-0,428	0,108
	Vx	-0,014	0,689	0,006	1,000	-0,072	-0,101	-0,068	0,065	-0,463	-0,047	0,045
	T	0,362	-0,158	-0,486	-0,072	1,000	0,986	0,987	-0,576	0,041	0,668	-0,132
	Tn	0,389	-0,179	-0,494	-0,101	0,986	1,000	0,956	-0,585	0,110	0,610	-0,090
	Tx	0,336	-0,153	-0,454	-0,068	0,987	0,956	1,000	-0,562	0,002	0,684	-0,159
	Hf	-0,421	0,208	0,199	0,065	-0,576	-0,585	-0,562	1,000	-0,194	-0,238	-0,059
	HR	0,036	-0,646	-0,202	-0,463	0,041	0,110	0,002	-0,194	1,000	-0,191	0,245
	Rad	-0,141	0,068	-0,428	-0,047	0,668	0,610	0,684	-0,238	-0,191	1,000	-0,331
	P	0,145	-0,188	0,108	0,045	-0,132	-0,090	-0,159	-0,059	0,245	-0,331	1,000
	ETo	0,022	0,122	-0,485	0,069	0,825	0,777	0,830	-0,349	-0,242	0,930	-0,248
	VolExp	-0,670	0,032	-0,274	0,038	-0,085	-0,118	-0,085	0,167	-0,050	0,398	-0,149
	Qent	0,014	0,010	0,319	0,073	-0,612	-0,606	-0,606	0,196	0,064	-0,567	0,087
	Qsal	0,622	-0,094	-0,270	-0,101	0,755	0,773	0,737	-0,415	0,035	0,303	-0,023
	Filt	-0,015	0,057	0,196	0,042	-0,251	-0,263	-0,237	0,117	-0,126	-0,216	-0,035
	Profundidad	-0,024	0,004	-0,010	0,006	-0,002	-0,002	-0,002	0,010	0,000	0,015	-0,008
	Ta	0,319	-0,099	-0,287	-0,076	0,633	0,636	0,617	-0,336	0,040	0,395	-0,059
Odis	-0,449	0,072	0,197	0,070	-0,553	-0,562	-0,536	0,307	-0,053	-0,248	0,020	
pH	-0,249	-0,100	-0,033	-0,016	-0,099	-0,115	-0,086	-0,179	0,111	0,064	0,051	
Con	0,309	0,000	0,230	0,000	-0,188	-0,167	-0,193	-0,048	0,044	-0,336	0,142	
Sig. (Unilateral)	DJ		0,000	0,050	0,095	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	V	0,000		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	DV	0,050	0,000		0,298	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	Vx	0,095	0,000	0,298		0,000	0,000	0,000	0,000	0,000	0,000	0,000
	T	0,000	0,000	0,000	0,000		0,000	0,000	0,000	0,000	0,000	0,000
	Tn	0,000	0,000	0,000	0,000	0,000		0,000	0,000	0,000	0,000	0,000
	Tx	0,000	0,000	0,000	0,000	0,000	0,000		0,000	0,439	0,000	0,000
	Hf	0,000	0,000	0,000	0,000	0,000	0,000	0,000		0,000	0,000	0,000
	HR	0,000	0,000	0,000	0,000	0,000	0,000	0,439	0,000		0,000	0,000
	Rad	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000		0,000
	P	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
	ETo	0,018	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	VolExp	0,000	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	Qent	0,096	0,176	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	Qsal	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,013
	Filt	0,074	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	Profundidad	0,012	0,348	0,166	0,297	0,439	0,416	0,424	0,166	0,485	0,073	0,217
	Ta	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Odis	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,029	
pH	0,000	0,000	0,001	0,068	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
Con	0,000	0,485	0,000	0,485	0,000	0,000	0,000	0,000	0,000	0,000	0,000	

a Determinante = 5,13E-011

Matriz de correlaciones (continuación)

		ETo	VolExp	Qent	Qsal	Filt	Profundidad	Ta	Odis	pH	Con
Correlación	DJ	0,022	-0,670	0,014	0,622	-0,015	-0,024	0,319	-0,449	-0,249	0,309
	V	0,122	0,032	0,010	-0,094	0,057	0,004	-0,099	0,072	-0,100	0,000
	DV	-0,485	-0,274	0,319	-0,270	0,196	-0,010	-0,287	0,197	-0,033	0,230
	Vx	0,069	0,038	0,073	-0,101	0,042	0,006	-0,076	0,070	-0,016	0,000
	T	0,825	-0,085	-0,612	0,755	-0,251	-0,002	0,633	-0,553	-0,099	-0,188
	Tn	0,777	-0,118	-0,606	0,773	-0,263	-0,002	0,636	-0,562	-0,115	-0,167
	Tx	0,830	-0,085	-0,606	0,737	-0,237	-0,002	0,617	-0,536	-0,086	-0,193
	Hf	-0,349	0,167	0,196	-0,415	0,117	0,010	-0,336	0,307	-0,179	-0,048
	HR	-0,242	-0,050	0,064	0,035	-0,126	0,000	0,040	-0,053	0,111	0,044
	Rad	0,930	0,398	-0,567	0,303	-0,216	0,015	0,395	-0,248	0,064	-0,336
	P	-0,248	-0,149	0,087	-0,023	-0,035	-0,008	-0,059	0,020	0,051	0,142
	ETo	1,000	0,277	-0,637	0,478	-0,240	0,015	0,499	-0,372	-0,016	-0,315
	VolExp	0,277	1,000	-0,099	-0,476	-0,060	0,036	-0,161	0,342	0,323	-0,383
	Qent	-0,637	-0,099	1,000	-0,485	0,255	0,002	-0,416	0,293	0,025	0,264
	Qsal	0,478	-0,476	-0,485	1,000	-0,185	-0,021	0,566	-0,563	-0,269	-0,008
	Filt	-0,240	-0,060	0,255	-0,185	1,000	-0,003	-0,181	0,116	-0,042	0,120
	Profundidad	0,015	0,036	0,002	-0,021	-0,003	1,000	-0,563	-0,582	-0,322	0,520
	Ta	0,499	-0,161	-0,416	0,566	-0,181	-0,563	1,000	0,014	0,216	-0,450
	Odis	-0,372	0,342	0,293	-0,563	0,116	-0,582	0,014	1,000	0,578	-0,447
pH	-0,016	0,323	0,025	-0,269	-0,042	-0,322	0,216	0,578	1,000	-0,465	
Con	-0,315	-0,383	0,264	-0,008	0,120	0,520	-0,450	-0,447	-0,465	1,000	
Sig. (Unilateral)	DJ	0,018	0,000	0,096	0,000	0,074	0,012	0,000	0,000	0,000	0,000
	V	0,000	0,001	0,176	0,000	0,000	0,348	0,000	0,000	0,000	0,485
	DV	0,000	0,000	0,000	0,000	0,000	0,166	0,000	0,000	0,001	0,000
	Vx	0,000	0,000	0,000	0,000	0,000	0,297	0,000	0,000	0,068	0,485
	T	0,000	0,000	0,000	0,000	0,000	0,439	0,000	0,000	0,000	0,000
	Tn	0,000	0,000	0,000	0,000	0,000	0,416	0,000	0,000	0,000	0,000
	Tx	0,000	0,000	0,000	0,000	0,000	0,424	0,000	0,000	0,000	0,000
	Hf	0,000	0,000	0,000	0,000	0,000	0,166	0,000	0,000	0,000	0,000
	HR	0,000	0,000	0,000	0,000	0,000	0,485	0,000	0,000	0,000	0,000
	Rad	0,000	0,000	0,000	0,000	0,000	0,073	0,000	0,000	0,000	0,000
	P	0,000	0,000	0,000	0,013	0,000	0,217	0,000	0,029	0,000	0,000
	ETo		0,000	0,000	0,000	0,000	0,082	0,000	0,000	0,063	0,000
	VolExp	0,000		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	Qent	0,000	0,000		0,000	0,000	0,433	0,000	0,000	0,009	0,000
	Qsal	0,000	0,000	0,000		0,000	0,023	0,000	0,000	0,000	0,223
	Filt	0,000	0,000	0,000	0,000		0,387	0,000	0,000	0,000	0,000
	Profundidad	0,082	0,000	0,433	0,023	0,387		0,000	0,000	0,000	0,000
	Ta	0,000	0,000	0,000	0,000	0,000	0,000		0,095	0,000	0,000
	Odis	0,000	0,000	0,000	0,000	0,000	0,000	0,095		0,000	0,000
pH	0,063	0,000	0,009	0,000	0,000	0,000	0,000	0,000		0,000	
Con	0,000	0,000	0,000	0,223	0,000	0,000	0,000	0,000	0,000		

**Inversa de la matriz de correlaciones**

	DJ	V	DV	Vx	T	Tn	Tx	Hf	HR	Rad	P	ETo	VolExp	Qent	Qsal	Filt	Profundidad	Ta	Odis	pH	Con
DJ	3,330	-0,101	0,042	0,069	-3,832	1,394	2,187	0,610	0,475	0,223	-0,212	0,712	0,637	-0,714	-1,328	-0,001	0,388	-0,561	0,836	-0,156	-0,497
V	-0,101	3,706	-0,209	-1,894	6,142	-3,645	0,708	0,027	0,837	0,710	0,673	-3,349	0,539	0,106	-0,392	0,052	0,116	0,020	0,175	-0,012	0,004
DV	0,042	-0,209	1,775	0,320	0,638	0,532	-0,966	0,213	0,569	-0,520	-0,210	1,185	0,490	0,105	0,236	-0,020	-0,034	-0,049	-0,045	0,049	-0,021
Vx	0,069	-1,894	0,320	2,578	-4,809	3,153	0,626	0,046	0,111	1,416	-0,441	-0,617	-0,071	-0,194	0,398	0,024	-0,085	0,111	-0,230	-0,053	-0,020
T	-3,832	6,142	0,638	-4,809	251,16	-127,7	-110,4	0,248	2,418	0,431	1,504	-13,56	-2,178	-0,963	1,138	-0,448	0,187	-2,438	3,488	-0,414	0,872
Tn	1,394	-3,645	0,532	3,153	-127,69	87,45	43,23	1,541	-4,034	5,551	-0,797	-1,430	0,345	1,290	-2,447	0,417	-0,506	-1,108	-0,347	1,138	0,022
Tx	2,187	0,708	0,638	0,320	-110,39	43,23	69,14	0,560	0,248	0,373	0,530	-1,471	2,779	0,203	-1,602	0,014	0,122	1,286	-0,855	-0,455	-0,130
Hf	0,610	0,027	0,213	0,046	0,248	1,541	0,560	2,398	0,240	-0,210	0,008	0,164	-0,011	0,199	-0,371	0,039	-0,333	1,286	0,025	1,007	0,544
HR	0,475	0,837	0,569	0,111	2,418	-4,034	0,028	0,240	2,859	-1,914	-0,366	4,032	-0,269	-0,147	0,104	0,199	0,180	-0,164	0,620	-0,206	-0,009
Rad	0,223	0,710	-0,520	1,416	0,431	5,551	0,373	-0,210	-1,914	16,30	1,474	-21,09	0,288	-0,023	0,023	0,041	0,096	0,144	-0,062	-0,670	-0,382
P	-0,212	0,673	-0,210	-0,441	1,504	-0,797	0,530	0,008	-0,366	1,474	1,465	-2,232	0,100	0,177	-0,056	0,064	0,089	0,065	0,059	-0,189	-0,129
ETo	0,712	-3,349	1,185	-0,617	-13,564	-1,430	-1,471	0,164	4,032	-21,09	-2,232	37,93	-3,202	0,617	0,460	0,178	0,025	-1,114	1,311	0,766	0,470
VolExp	0,637	0,539	0,490	-0,071	-2,178	0,345	2,779	-0,011	-0,269	0,288	0,100	-3,202	3,769	0,046	1,237	0,011	-0,571	0,782	-1,183	-0,086	0,480
Qent	-0,714	0,106	0,105	-0,194	-0,963	1,290	0,203	0,199	-0,147	-0,023	0,177	0,617	0,046	2,159	0,766	-0,127	0,009	0,100	-0,199	0,161	-0,057
Qsal	-1,328	-0,392	0,236	0,398	1,138	-2,447	-1,602	-0,371	0,104	0,023	-0,056	0,460	1,237	0,766	5,109	0,061	-0,339	-0,499	0,049	0,408	0,459
Filt	-0,001	0,052	-0,020	0,024	-0,448	0,417	0,014	0,039	0,199	0,041	0,064	0,178	0,011	-0,127	0,061	1,135	0,043	0,027	-0,026	0,057	-0,047
Profundidad	0,388	0,116	-0,034	0,024	0,187	-0,506	0,122	-0,333	0,180	0,096	0,089	0,064	0,025	0,009	0,043	0,043	3,509	1,751	2,412	-0,903	-0,741
Ta	-0,561	0,020	-0,049	0,111	-2,438	3,488	-0,414	0,872	-0,020	0,020	0,020	0,020	0,020	0,020	0,020	0,020	0,020	0,020	0,020	0,020	0,020
Odis	0,836	0,175	-0,045	0,049	-0,021	0,004	0,004	0,004	0,004	0,004	0,004	0,004	0,004	0,004	0,004	0,004	0,004	0,004	0,004	0,004	0,004
pH	-0,156	-0,012	0,004	-0,021	0,872	0,022	0,130	0,544	-0,009	-0,382	-0,129	0,470	0,480	-0,057	0,459	0,047	-0,741	0,741	-0,903	-0,804	0,141
Con	-0,497	0,004	-0,021	-0,020	0,872	0,022	-0,130	0,544	-0,009	-0,382	-0,129	0,470	0,480	-0,057	0,459	0,047	-0,741	0,741	-0,903	-0,804	0,141

## Matriz anti-imagen

Correlación

anti-imagen	DJ	V	DV	Vx	T	Tn	Tx	Hf	HR	Rad	P	ETo	VolExp	Qent	Qsal	Filt	Profundidad	Ta	Odis	pH	Con
DJ	0,815	-0,029	0,017	0,023	-0,132	0,082	0,144	0,216	0,154	0,030	-0,096	0,063	0,180	-0,266	-0,322	-0,001	0,114	-0,143	0,183	-0,054	-0,169
V	-0,029	0,605	-0,082	-0,613	0,201	-0,202	0,044	0,009	0,257	0,091	0,289	-0,283	0,144	0,038	-0,090	0,025	0,032	0,005	0,036	-0,004	0,001
DV	0,017	-0,082	0,890	0,150	0,030	0,043	-0,087	0,103	0,253	-0,097	-0,130	0,144	0,190	0,054	0,078	-0,014	-0,014	-0,017	-0,013	0,023	-0,010
Vx	0,023	-0,613	0,150	0,552	-0,189	0,210	0,047	0,019	0,041	0,218	-0,227	-0,062	-0,023	-0,082	0,110	0,014	-0,028	0,032	-0,057	-0,021	-0,008
T	-0,132	0,201	0,030	-0,189	0,778	-0,862	-0,838	0,010	0,090	0,007	0,078	-0,139	-0,071	-0,041	0,032	-0,027	0,006	-0,072	0,088	-0,016	0,034
Tn	0,082	-0,202	0,043	0,210	-0,862	0,810	0,556	0,106	-0,255	0,147	-0,070	-0,025	0,019	0,094	-0,116	0,042	-0,029	-0,055	-0,015	0,077	0,001
Tx	0,144	0,044	-0,087	0,047	-0,838	0,556	0,832	0,043	0,002	0,011	0,053	-0,029	0,172	0,017	-0,085	0,002	0,008	0,072	-0,041	-0,035	-0,010
Hf	0,216	0,009	0,103	0,019	0,010	0,106	0,043	0,841	0,092	-0,034	0,004	0,017	-0,004	0,088	-0,106	0,024	-0,115	-0,202	0,006	0,410	0,218
HR	0,154	0,257	0,253	0,041	0,090	-0,255	0,002	0,092	0,625	-0,280	-0,179	0,387	-0,082	-0,059	0,027	0,110	0,057	-0,051	0,147	-0,077	-0,003
Rad	0,030	0,091	-0,097	0,218	0,007	0,147	0,011	-0,034	-0,280	0,780	0,302	-0,848	0,037	-0,004	0,002	0,009	0,013	0,017	-0,006	-0,105	-0,059
P	-0,096	0,289	-0,130	-0,227	0,078	-0,070	0,053	0,004	-0,179	0,302	0,497	-0,299	0,042	0,100	-0,021	0,049	0,039	0,025	0,019	-0,098	-0,066
ETo	0,063	-0,283	0,144	-0,062	-0,139	-0,025	-0,029	0,017	0,387	-0,848	-0,299	0,794	-0,268	0,068	0,033	0,027	0,002	-0,084	0,085	0,078	0,047
VolExp	0,180	0,144	0,190	-0,023	-0,071	0,019	0,172	-0,004	-0,082	0,037	0,042	-0,268	0,775	0,016	0,282	0,005	-0,157	0,188	-0,244	-0,028	0,154
Qent	-0,266	0,038	0,054	-0,082	-0,041	0,094	0,017	0,088	-0,059	-0,004	0,100	0,068	0,016	0,933	0,231	-0,081	0,003	0,032	-0,054	0,069	-0,024
Qsal	-0,322	-0,090	0,078	0,110	0,032	-0,116	-0,085	-0,106	0,027	0,002	-0,021	0,033	0,282	0,231	0,919	0,026	-0,080	-0,103	0,009	0,114	0,126
Filt	-0,001	0,025	-0,014	0,014	-0,028	0,006	0,006	0,022	0,948	0,022	0,012	0,010	0,034	-0,081	0,026	0,948	0,022	0,012	-0,010	0,034	-0,027
Profundidad	0,114	0,032	0,032	0,032	0,006	0,006	0,006	0,022	0,605	0,022	0,039	0,002	-0,157	0,003	0,032	0,003	0,605	0,436	0,515	-0,304	-0,246
Ta	-0,143	0,005	-0,017	-0,013	0,032	0,006	0,006	0,022	0,882	0,032	0,032	0,002	-0,157	0,003	-0,103	0,022	0,605	0,436	0,515	-0,304	-0,246
Odis	0,183	0,036	-0,013	-0,057	0,088	-0,015	-0,041	0,006	0,147	-0,006	0,019	0,085	-0,244	-0,054	0,009	-0,010	0,515	0,882	-0,152	-0,398	0,146
pH	-0,054	-0,004	0,023	-0,010	-0,008	0,034	0,001	0,001	-0,003	0,034	-0,066	0,047	0,154	-0,024	0,126	-0,027	-0,246	-0,237	0,041	0,631	0,256
Con	-0,169	0,001	-0,010	-0,008	0,034	0,001	-0,010	0,218	-0,003	-0,059	-0,066	0,047	0,154	-0,024	0,126	-0,027	-0,246	0,041	0,146	0,256	0,853

**Varianza total explicada**

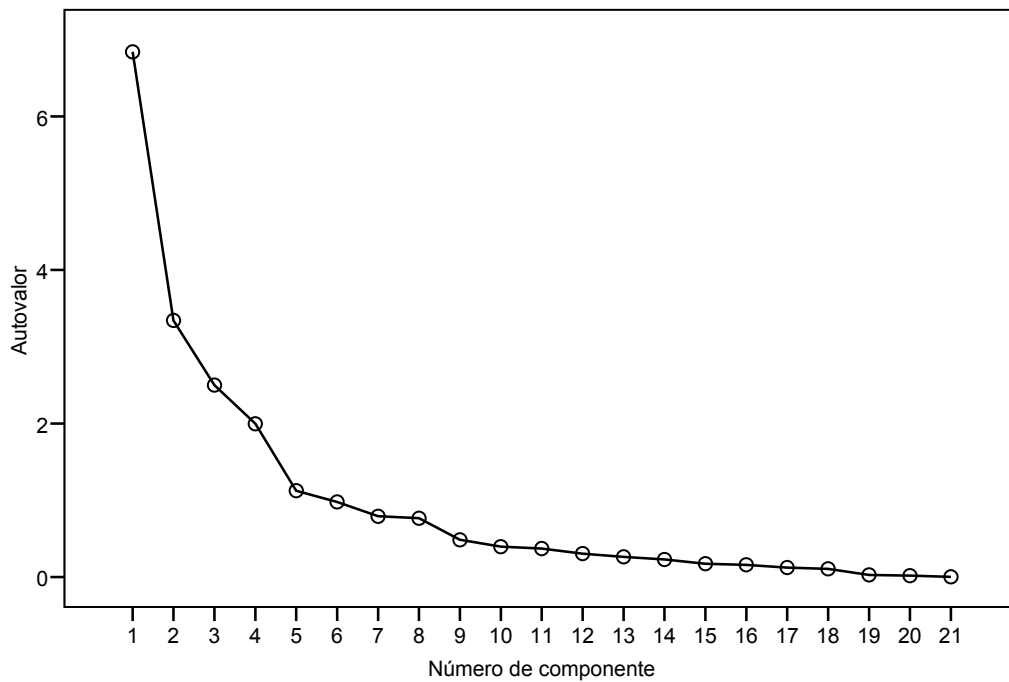
Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	6,841	32,576	32,576	6,841	32,576	32,576
2	3,342	15,915	48,491	3,342	15,915	48,491
3	2,500	11,903	60,394	2,500	11,903	60,394
4	1,996	9,507	69,901	1,996	9,507	69,901
5	1,125	5,357	75,258	1,125	5,357	75,258
6	,977	4,655	79,912			
7	,791	3,767	83,680			
8	,766	3,648	87,327			
9	,485	2,308	89,635			
10	,396	1,886	91,521			
11	,371	1,767	93,287			
12	,306	1,455	94,742			
13	,262	1,250	95,992			
14	,229	1,091	97,083			
15	,174	,828	97,911			
16	,160	,760	98,671			
17	,123	,588	99,258			
18	,106	,506	99,765			
19	,028	,135	99,899			
20	,018	,088	99,987			
21	,003	,013	100,000			

Método de extracción: Análisis de Componentes principales

**Comunalidades**

	Inicial	Extracción
DJ	1,000	,794
V	1,000	,840
DV	1,000	,457
Vx	1,000	,819
T	1,000	,962
Tn	1,000	,950
Tx	1,000	,936
Hf	1,000	,544
HR	1,000	,748
Rad	1,000	,844
P	1,000	,585
ETo	1,000	,936
VolExp	1,000	,831
Qent	1,000	,526
Qsal	1,000	,831
Filt	1,000	,252
Profundidad	1,000	,814
Ta	1,000	,851
Odis	1,000	,898
pH	1,000	,647
Con	1,000	,737

Gráfico de sedimentación



Matriz de componentes(a)

	Componente				
	1	2	3	4	5
DJ	,368	-,667	-,117	,447	,032
V	-,136	,184	,756	,439	,150
DV	-,526	-,236	,049	,292	-,193
Vx	-,103	,130	,594	,435	,499
T	,980	-,031	,001	-,004	,019
Tn	,970	-,078	-,049	-,007	,037
Tx	,967	-,012	,023	,001	-,009
Hf	-,567	,194	,250	-,082	-,342
HR	,034	-,206	-,729	-,398	,120
Rad	,714	,453	,264	-,229	-,085
P	-,154	-,267	-,326	,074	,615
ETo	,850	,337	,302	-,097	,017
VolExp	-,082	,743	,122	-,486	,147
Qent	-,679	-,206	-,106	,100	,046
Qsal	,787	-,370	-,042	,247	-,110
Filt	-,307	-,094	,088	,137	-,349
Profundidad	-,026	-,434	,445	-,630	,173
Ta	,670	,229	-,333	,476	-,112
Odis	-,548	,639	-,317	,297	-,009
pH	-,090	,585	-,435	,135	,298
Con	-,216	-,763	,244	-,200	,088

a 5 componentes extraídos

**Correlaciones reproducidas**

		DJ	V	DV	Vx	T	Tn	Tx	Hf	HR	Rad	P
Correlación reproducida	DJ	0,794 <sup>b</sup>	-0,060	0,082	0,016	0,380	0,412	0,361	-0,414	0,061	-0,175	0,212
	V	-0,060	0,840 <sup>b</sup>	0,164	0,754	-0,137	-0,181	-0,117	0,214	-0,751	0,073	-0,150
	DV	0,082	0,164	0,457 <sup>b</sup>	0,083	-0,513	-0,504	-0,503	0,307	-0,144	-0,520	0,031
	Vx	0,016	0,754	0,083	0,819 <sup>b</sup>	-0,097	-0,124	-0,092	0,026	-0,577	0,000	0,126
	T	0,380	-0,137	-0,513	-0,097	0,962 <sup>b</sup>	0,954	0,948	-0,567	0,042	0,685	-0,132
	Tn	0,412	-0,181	-0,504	-0,124	0,954	0,950 <sup>b</sup>	0,937	-0,589	0,092	0,643	-0,090
	Tx	0,361	-0,117	-0,503	-0,092	0,948	0,937	0,936 <sup>b</sup>	-0,542	0,017	0,692	-0,159
	Hf	-0,414	0,214	0,307	0,026	-0,567	-0,589	-0,542	0,544 <sup>b</sup>	-0,249	-0,203	-0,262
	HR	0,061	-0,751	-0,144	-0,577	0,042	0,092	0,017	-0,249	0,748 <sup>b</sup>	-0,181	0,332
	Rad	-0,175	0,073	-0,520	0,000	0,685	0,643	0,692	-0,203	-0,181	0,844 <sup>b</sup>	-0,386
	P	0,212	-0,150	0,031	0,126	-0,132	-0,090	-0,159	-0,262	0,332	-0,386	0,585 <sup>b</sup>
	ETo	0,010	0,135	-0,543	0,102	0,823	0,784	0,824	-0,338	-0,220	0,860	-0,316
	VolExp	-0,752	0,049	-0,297	0,040	-0,098	-0,134	-0,087	0,210	-0,034	0,409	-0,171
	Qent	-0,054	0,025	0,421	0,047	-0,658	-0,636	-0,657	0,294	0,062	-0,633	0,230
	Qsal	0,648	-0,114	-0,236	-0,101	0,780	0,788	0,766	-0,511	0,021	0,336	-0,058
	Filt	-0,011	0,099	0,296	-0,043	-0,305	-0,309	-0,291	0,286	-0,152	-0,240	-0,161
	Profundidad	-0,049	0,009	-0,079	0,022	-0,006	-0,003	-0,012	0,034	0,036	0,032	0,035
	Ta	0,342	-0,108	-0,262	-0,086	0,645	0,641	0,638	-0,419	0,015	0,395	-0,089
	Odis	-0,458	0,081	0,211	0,076	-0,559	-0,568	-0,545	0,334	-0,038	-0,253	0,034
	pH	-0,302	-0,105	-0,130	0,035	-0,102	-0,101	-0,107	-0,057	0,175	0,029	0,193
Con	0,314	-0,001	0,230	0,024	-0,185	-0,158	-0,195	0,022	0,062	-0,397	0,196	
Residual(a)	DJ		-0,002	-0,065	-0,030	-0,018	-0,024	-0,025	-0,006	-0,025	0,034	-0,067
	V	-0,002		-0,038	-0,064	-0,021	0,002	-0,035	-0,007	0,105	-0,004	-0,037
	DV	-0,065	-0,038		-0,078	0,027	0,010	0,049	-0,107	-0,058	0,092	0,077
	Vx	-0,030	-0,064	-0,078		0,025	0,023	0,024	0,039	0,114	-0,047	-0,081
	T	-0,018	-0,021	0,027	0,025		0,033	0,039	-0,009	-0,001	-0,018	0,000
	Tn	-0,024	0,002	0,010	0,023	0,033		0,019	0,004	0,018	-0,033	0,000
	Tx	-0,025	-0,035	0,049	0,024	0,039	0,019		-0,020	-0,015	-0,008	0,000
	Hf	-0,006	-0,007	-0,107	0,039	-0,009	0,004	-0,020		0,055	-0,035	0,204
	HR	-0,025	0,105	-0,058	0,114	-0,001	0,018	-0,015	0,055		-0,010	-0,087
	Rad	0,034	-0,004	0,092	-0,047	-0,018	-0,033	-0,008	-0,035	-0,010		0,055
	P	-0,067	-0,037	0,077	-0,081	0,000	0,000	0,000	0,204	-0,087	0,055	
	ETo	0,012	-0,013	0,058	-0,033	0,001	-0,007	0,005	-0,011	-0,022	0,071	0,068
	VolExp	0,082	-0,017	0,023	-0,002	0,013	0,016	0,002	-0,043	-0,016	-0,011	0,022
	Qent	0,068	-0,015	-0,102	0,026	0,046	0,030	0,051	-0,099	0,002	0,066	-0,142
	Qsal	-0,026	0,020	-0,034	0,000	-0,025	-0,015	-0,029	0,096	0,014	-0,033	0,035
	Filt	-0,005	-0,042	-0,099	0,084	0,054	0,046	0,054	-0,169	0,026	0,024	0,126
	Profundidad	0,025	-0,005	0,069	-0,017	0,004	0,000	0,009	-0,024	-0,035	-0,016	-0,043
	Ta	-0,023	0,009	-0,025	0,010	-0,011	-0,005	-0,021	0,083	0,024	0,001	0,030
	Odis	0,009	-0,009	-0,013	-0,007	0,006	0,005	0,009	-0,027	-0,016	0,005	-0,014
	pH	0,054	0,005	0,096	-0,050	0,004	-0,013	0,021	-0,122	-0,065	0,035	-0,142
Con	-0,005	0,002	0,000	-0,024	-0,003	-0,010	0,002	-0,070	-0,018	0,061	-0,054	

a Los residuos se calculan entre las correlaciones observadas y reproducidas.

Hay 54 (25,0%) residuales no redundantes con valores absolutos mayores que 0,05.

b Comunalidades reproducidas

**Correlaciones reproducidas (Continuación)**

		ETo	VolExp	Qent	Qsal	Filt	Profundidad	Ta	Odis	pH	Con
Correlación reproducida	DJ	0,010	-0,752	-0,054	0,648	-0,011	-0,049	0,342	-0,458	-0,302	0,314
	V	0,135	0,049	0,025	-0,114	0,099	0,009	-0,108	0,081	-0,105	-0,001
	DV	-0,543	-0,297	0,421	-0,236	0,296	-0,079	-0,262	0,211	-0,130	0,230
	Vx	0,102	0,040	0,047	-0,101	-0,043	0,022	-0,086	0,076	0,035	0,024
	T	0,823	-0,098	-0,658	0,780	-0,305	-0,006	0,645	-0,559	-0,102	-0,185
	Tn	0,784	-0,134	-0,636	0,788	-0,309	-0,003	0,641	-0,568	-0,101	-0,158
	Tx	0,824	-0,087	-0,657	0,766	-0,291	-0,012	0,638	-0,545	-0,107	-0,195
	Hf	-0,338	0,210	0,294	-0,511	0,286	0,034	-0,419	0,334	-0,057	0,022
	HR	-0,220	-0,034	0,062	0,021	-0,152	0,036	0,015	-0,038	0,175	0,062
	Rad	0,860	0,409	-0,633	0,336	-0,240	0,032	0,395	-0,253	0,029	-0,397
	P	-0,316	-0,171	0,230	-0,058	-0,161	0,035	-0,089	0,034	0,193	0,196
	ETo	0,936 <sup>b</sup>	0,268	-0,687	0,506	-0,285	0,030	0,497	-0,375	-0,019	-0,346
	VolExp	0,268	0,831 <sup>b</sup>	-0,152	-0,480	-0,152	0,065	-0,173	0,335	0,368	-0,410
	Qent	-0,687	-0,152	0,526 <sup>b</sup>	-0,434	0,216	0,005	-0,424	0,303	0,014	0,262
	Qsal	0,506	-0,480	-0,434	0,831 <sup>b</sup>	-0,139	-0,053	0,586	-0,580	-0,269	0,043
	Filt	-0,285	-0,152	0,216	-0,139	0,252 <sup>b</sup>	-0,059	-0,152	0,124	-0,151	0,102
	Profundidad	0,030	0,065	0,005	-0,053	-0,059	0,814 <sup>b</sup>	-0,585	-0,593	-0,479	0,587
	Ta	0,497	-0,173	-0,424	0,586	-0,152	-0,585	0,851 <sup>b</sup>	0,028	0,250	-0,506
	Odis	-0,375	0,335	0,303	-0,580	0,124	-0,593	0,028	0,898 <sup>b</sup>	0,599	-0,507
	pH	-0,019	0,368	0,014	-0,269	-0,151	-0,479	0,250	0,599	0,647 <sup>b</sup>	-0,535
Con	-0,346	-0,410	0,262	0,043	0,102	0,587	-0,506	-0,507	-0,535	0,737 <sup>b</sup>	
Residual(a)	DJ	0,012	0,082	0,068	-0,026	-0,005	0,025	-0,023	0,009	0,054	-0,005
	V	-0,013	-0,017	-0,015	0,020	-0,042	-0,005	0,009	-0,009	0,005	0,002
	DV	0,058	0,023	-0,102	-0,034	-0,099	0,069	-0,025	-0,013	0,096	0,000
	Vx	-0,033	-0,002	0,026	0,000	0,084	-0,017	0,010	-0,007	-0,050	-0,024
	T	0,001	0,013	0,046	-0,025	0,054	0,004	-0,011	0,006	0,004	-0,003
	Tn	-0,007	0,016	0,030	-0,015	0,046	0,000	-0,005	0,005	-0,013	-0,010
	Tx	0,005	0,002	0,051	-0,029	0,054	0,009	-0,021	0,009	0,021	0,002
	Hf	-0,011	-0,043	-0,099	0,096	-0,169	-0,024	0,083	-0,027	-0,122	-0,070
	HR	-0,022	-0,016	0,002	0,014	0,026	-0,035	0,024	-0,016	-0,065	-0,018
	Rad	0,071	-0,011	0,066	-0,033	0,024	-0,016	0,001	0,005	0,035	0,061
	P	0,068	0,022	-0,142	0,035	0,126	-0,043	0,030	-0,014	-0,142	-0,054
	ETo		0,009	0,050	-0,028	0,045	-0,016	0,001	0,003	0,003	0,030
	VolExp	0,009		0,054	0,004	0,092	-0,029	0,012	0,007	-0,044	0,026
	Qent	0,050	0,054		-0,051	0,038	-0,003	0,008	-0,010	0,011	0,002
	Qsal	-0,028	0,004	-0,051		-0,046	0,032	-0,021	0,017	0,000	-0,051
	Filt	0,045	0,092	0,038	-0,046		0,056	-0,028	-0,008	0,109	0,018
	Profundidad	-0,016	-0,029	-0,003	0,032	0,056		0,021	0,011	0,156	-0,067
	Ta	0,001	0,012	0,008	-0,021	-0,028	0,021		-0,014	-0,034	0,056
	Odis	0,003	0,007	-0,010	0,017	-0,008	0,011	-0,014		-0,021	0,060
	pH	0,003	-0,044	0,011	0,000	0,109	0,156	-0,034	-0,021		0,069
Con	0,030	0,026	0,002	-0,051	0,018	-0,067	0,056	0,060	0,069		

a Los residuos se calculan entre las correlaciones observadas y reproducidas.

Hay 54 (25,0%) residuales no redundantes con valores absolutos mayores que 0,05.

b Comunalidades reproducidas



b) Extracción de los factores por el Método de Componentes principales, rotación Varimax

**Matriz de componentes rotados**

	Componente				
	1	2	3	4	5
DJ	,134	,861	-,102	-,001	,158
V	-,058	-,038	-,021	,904	-,135
DV	-,599	,222	-,002	,125	-,183
Vx	-,043	-,021	-,003	,868	,252
T	,934	,288	-,002	-,073	,047
Tn	,909	,321	-,010	-,116	,086
Tx	,926	,274	,003	-,057	,011
Hf	-,473	-,331	-,034	,126	-,440
HR	-,036	,005	-,007	-,777	,379
Rad	,841	-,291	,048	,093	-,202
P	-,228	,149	-,017	-,069	,711
ETo	,936	-,112	,026	,190	-,101
VolExp	,170	-,887	,119	,032	,007
Qent	-,719	,029	-,040	-,005	,085
Qsal	,630	,654	-,046	-,063	-,024
Filt	-,338	,105	-,004	,039	-,353
Profundidad	,044	-,182	-,879	-,003	,080
Ta	,570	,384	,613	-,055	-,014
Odis	-,478	-,370	,727	,060	-,005
pH	-,028	-,333	,644	-,066	,341
Con	-,300	,306	-,737	-,010	,097

Método de rotación: Normalización Varimax con Kaiser.

c) Extracción de los factores por el Método de la Máxima verosimilitud, incluyendo una rotación con el Método Varimax

**Comunalidades(a)**

Variable	Inicial
DJ	,700
V	,730
DV	,437
Vx	,612
T	,996
Tn	,989
Tx	,986
Hf	,583
HR	,650
Rad	,939
P	,318
ETo	,974
VolExp	,735
Qent	,537
Qsal	,804
Filt	,119
Profundidad	,715
Ta	,783
Odis	,840
pH	,602
Con	,614

Método de extracción: Máxima verosimilitud.

a Se han encontrado una o más estimaciones de comunalidad mayores que 1 durante las iteraciones. La solución resultante deberá ser interpretada con precaución.

### Varianza total explicada

Factor	Autovalores iniciales		
	Total	% de la varianza	% acumulado
1	6,841	32,576	32,576
2	3,342	15,915	48,491
3	2,500	11,903	60,394
4	1,996	9,507	69,901
5	1,125	5,357	75,258
6	,977	4,655	79,912
7	,791	3,767	83,680
8	,766	3,648	87,327
9	,485	2,308	89,635
10	,396	1,886	91,521
11	,371	1,767	93,287
12	,306	1,455	94,742
13	,262	1,250	95,992
14	,229	1,091	97,083
15	,174	,828	97,911
16	,160	,760	98,671
17	,123	,588	99,258
18	,106	,506	99,765
19	,028	,135	99,899
20	,018	,088	99,987
21	,003	,013	100,000

Método de extracción: Máxima verosimilitud.

### Varianza total explicada (para 5 iteraciones, obtención de la matriz de componentes)

Factor	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	6,841	32,576	32,576	6,589	31,374	31,374	6,093	29,014	29,014
2	3,342	15,915	48,491	2,504	11,922	43,296	2,499	11,902	40,916
3	2,500	11,903	60,394	2,418	11,513	54,809	2,407	11,461	52,376
4	1,996	9,507	69,901	1,802	8,581	63,391	1,975	9,403	61,779
5	1,125	5,357	75,258	,746	3,551	66,941	1,084	5,162	66,941
6	,977	4,655	79,912						
7	,791	3,767	83,680						
8	,766	3,648	87,327						
9	,485	2,308	89,635						
10	,396	1,886	91,521						
11	,371	1,767	93,287						
12	,306	1,455	94,742						
13	,262	1,250	95,992						
14	,229	1,091	97,083						
15	,174	,828	97,911						
16	,160	,760	98,671						
17	,123	,588	99,258						
18	,106	,506	99,765						
19	,028	,135	99,899						
20	,018	,088	99,987						
21	,003	,013	100,000						

Método de extracción: Máxima verosimilitud.

**Matriz factorial(a)**

	Factor				
	1	2	3	4	5
DJ	,361	-,562	,161	,323	-,268
V	-,158	,399	,256	,724	,161
DV	-,486	-,169	,076	,183	-,186
Vx	-,072	,200	,157	,654	,457
T	1,000	-,001	,000	-1,55E-005	,000
Tn	,986	-,069	-,006	-,012	,002
Tx	,987	,030	,003	,004	-,019
Hf	-,575	,222	,057	,036	-,083
HR	,041	-,440	-,209	-,571	-,003
Rad	,668	,688	,024	-,121	-,136
P	-,132	-,290	-,034	-,025	,074
ETo	,825	,548	,058	,034	-,029
VolExp	-,085	,690	-,183	-,376	,390
Qent	-,612	-,234	,012	,019	,070
Qsal	,755	-,305	,092	,212	-,272
Filt	-,251	-,062	,039	,074	-,005
Profundidad	-,002	-,021	,785	-,309	,202
Ta	,633	-,023	-,531	,252	-,250
Odis	-,553	,233	-,721	,088	,096
pH	-,098	,199	-,603	-,090	,154
Con	-,189	-,363	,647	-,037	-,072

Método de extracción: Máxima verosimilitud.

a Se han intentado extraer 5 factores. Se hicieron 5 iteraciones en total.  
(Convergencia=,035). Se ha terminado la extracción.

**Matriz de factores rotados(a)**

	Factor				
	1	2	3	4	5
DJ	,216	,752	-,124	-,004	-,142
V	-,149	,009	-,024	,866	,167
DV	-,542	,202	-,024	,060	-,020
Vx	-,013	-,060	-,003	,818	-,183
T	,971	,216	-,015	-,036	,091
Tn	,953	,248	-,021	-,070	,042
Tx	,957	,207	-,010	-,028	,125
Hf	-,561	-,192	-,001	,113	,165
HR	,051	-,018	,000	-,680	-,314
Rad	,690	-,263	,037	,075	,634
P	-,135	,094	-,016	-,100	-,268
ETo	,839	-,123	,014	,195	,479
VolExp	,093	-,869	,132	,034	,166
Qent	-,603	-,016	-,034	-,017	-,264
Qsal	,632	,618	-,059	-,036	,064
Filt	-,258	,027	-,022	,056	-,058
Profundidad	,027	-,175	-,849	-,010	-,030
Ta	,566	,363	,582	-,034	,133
Odis	-,473	-,346	,739	,044	-,073
pH	-,014	-,336	,565	-,083	-,073
Con	-,255	,271	-,663	-,029	-,114

Método de extracción: Máxima verosimilitud.

Método de rotación: Normalización Varimax con Kaiser.

a La rotación ha convergido en 8 iteraciones.

**ANEXO 13. ANÁLISIS FACTORIAL DEL PERIODO DE CALIBRACIÓN, SIN DATOS ANÓMALOS Y VARIABLES SIN TRANSFORMAR**

## PERIODO DE CALIBRACIÓN

### Estadísticos descriptivos

	Media	Desviación típica	N del análisis
DJ	207,72	122,598	4401
V	1,312	,3278	4401
DV	266,93	71,330	4401
Vx	5,200	1,5203	4401
T	16,845	6,5218	4401
Tn	12,824	6,3155	4401
Tx	21,446	6,5321	4401
Hf	1,51	3,788	4401
HR	57,60	13,465	4401
Rad	158,54	73,692	4401
P	,57	2,405	4401
ETo	2,69	1,531	4401
VolExp	8477988	1313500,6	4401
Qent	32140,2	34964,2	4401
Qsal	25522,57	11250,6	4401
Filt	-187,1	903,7	4401
Profundidad	9,141	7,766	4401
Ta	15,3576	5,49152	4401
Odis	8,2321	4,31904	4401
pH	8,1095	,40284	4401
Con	1013,28	100,382	4401

### KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,763
	Chi-cuadrado aproximado	114258,68
Prueba de esfericidad de Bartlett	gl	210
	Sig.	,000

**Matriz de correlaciones(a)**

		DJ	V	DV	Vx	T	Tn	Tx	Hf	HR	Rad	P
Correlación	DJ	1,000	-0,103	-0,076	0,014	0,352	0,381	0,312	-0,480	0,048	-0,143	0,175
	V	-0,103	1,000	0,165	0,638	-0,183	-0,201	-0,194	0,199	-0,667	0,113	-0,221
	DV	-0,076	0,165	1,000	0,099	-0,522	-0,517	-0,500	0,206	-0,229	-0,486	0,133
	Vx	0,014	0,638	0,099	1,000	-0,034	-0,074	-0,048	0,022	-0,521	0,011	0,025
	T	0,352	-0,183	-0,522	-0,034	1,000	0,987	0,987	-0,554	0,068	0,706	-0,117
	Tn	0,381	-0,201	-0,517	-0,074	0,987	1,000	0,960	-0,572	0,123	0,661	-0,069
	Tx	0,312	-0,194	-0,500	-0,048	0,987	0,960	1,000	-0,537	0,052	0,722	-0,168
	Hf	-0,480	0,199	0,206	0,022	-0,554	-0,572	-0,537	1,000	-0,267	-0,124	-0,081
	HR	0,048	-0,667	-0,229	-0,521	0,068	0,123	0,052	-0,267	1,000	-0,162	0,262
	Rad	-0,143	0,113	-0,486	0,011	0,706	0,661	0,722	-0,124	-0,162	1,000	-0,362
	P	0,175	-0,221	0,133	0,025	-0,117	-0,069	-0,168	-0,081	0,262	-0,362	1,000
	ETo	0,040	0,143	-0,526	0,131	0,844	0,811	0,837	-0,261	-0,204	0,932	-0,234
	VolExp	-0,614	0,082	-0,395	0,026	0,188	0,154	0,193	0,275	-0,036	0,568	-0,224
	Qent	0,036	-0,073	0,200	0,051	-0,510	-0,511	-0,513	0,068	0,130	-0,489	0,114
	Qsal	0,658	-0,142	-0,242	-0,107	0,721	0,737	0,708	-0,453	0,026	0,290	-0,004
	Filt	-0,037	-0,046	0,198	0,022	-0,272	-0,274	-0,264	0,082	-0,076	-0,281	0,049
	Profundidad	-0,039	0,011	-0,041	0,012	0,026	0,024	0,024	0,004	0,004	0,063	-0,012
	Ta	0,366	-0,128	-0,357	-0,064	0,716	0,718	0,700	-0,366	0,035	0,480	-0,069
	Odis	-0,395	0,099	0,321	0,059	-0,560	-0,566	-0,539	0,341	-0,058	-0,389	0,025
	pH	-0,094	-0,082	0,081	-0,025	-0,102	-0,107	-0,086	-0,175	0,169	-0,185	0,066
Con	0,255	-0,045	0,128	-0,008	-0,112	-0,093	-0,125	-0,142	0,076	-0,211	0,156	
Sig. (Unilateral)	DJ		0,000	0,000	0,184	0,000	0,000	0,000	0,000	0,001	0,000	0,000
	V	0,000		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	DV	0,000	0,000		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	Vx	0,184	0,000	0,000		0,012	0,000	0,001	0,071	0,000	0,225	0,051
	T	0,000	0,000	0,000	0,012		0,000	0,000	0,000	0,000	0,000	0,000
	Tn	0,000	0,000	0,000	0,000	0,000		0,000	0,000	0,000	0,000	0,000
	Tx	0,000	0,000	0,000	0,001	0,000	0,000		0,000	0,000	0,000	0,000
	Hf	0,000	0,000	0,000	0,071	0,000	0,000	0,000		0,000	0,000	0,000
	HR	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000		0,000	0,000
	Rad	0,000	0,000	0,000	0,225	0,000	0,000	0,000	0,000	0,000		0,000
	P	0,000	0,000	0,000	0,051	0,000	0,000	0,000	0,000	0,000	0,000	
	ETo	0,004	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	VolExp	0,000	0,000	0,000	0,044	0,000	0,000	0,000	0,000	0,008	0,000	0,000
	Qent	0,009	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	Qsal	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,040	0,000	0,386
	Filt	0,007	0,001	0,000	0,076	0,000	0,000	0,000	0,000	0,000	0,000	0,001
	Profundidad	0,004	0,241	0,004	0,219	0,045	0,057	0,057	0,390	0,398	0,000	0,215
	Ta	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,011	0,000	0,000
	Odis	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,047
	pH	0,000	0,000	0,000	0,052	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Con	0,000	0,001	0,000	0,309	0,000	0,000	0,000	0,000	0,000	0,000	0,000	

a Determinante = 5,04E-012

**Matriz de correlaciones (continuación)**

		ETo	VolExp	Qent	Qsal	Filt	Profundidad	Ta	Odis	pH	Con
Correlación	DJ	0,040	-0,614	0,036	0,658	-0,037	-0,039	0,366	-0,395	-0,094	0,255
	V	0,143	0,082	-0,073	-0,142	-0,046	0,011	-0,128	0,099	-0,082	-0,045
	DV	-0,526	-0,395	0,200	-0,242	0,198	-0,041	-0,357	0,321	0,081	0,128
	Vx	0,131	0,026	0,051	-0,107	0,022	0,012	-0,064	0,059	-0,025	-0,008
	T	0,844	0,188	-0,510	0,721	-0,272	0,026	0,716	-0,560	-0,102	-0,112
	Tn	0,811	0,154	-0,511	0,737	-0,274	0,024	0,718	-0,566	-0,107	-0,093
	Tx	0,837	0,193	-0,513	0,708	-0,264	0,024	0,700	-0,539	-0,086	-0,125
	Hf	-0,261	0,275	0,068	-0,453	0,082	0,004	-0,366	0,341	-0,175	-0,142
	HR	-0,204	-0,036	0,130	0,026	-0,076	0,004	0,035	-0,058	0,169	0,076
	Rad	0,932	0,568	-0,489	0,290	-0,281	0,063	0,480	-0,389	-0,185	-0,211
	P	-0,234	-0,224	0,114	-0,004	0,049	-0,012	-0,069	0,025	0,066	0,156
	ETo	1,000	0,502	-0,554	0,450	-0,289	0,058	0,594	-0,479	-0,198	-0,195
	VolExp	0,502	1,000	-0,188	-0,289	-0,131	0,071	0,036	0,018	-0,005	-0,399
	Qent	-0,554	-0,188	1,000	-0,471	0,190	-0,017	-0,416	0,274	0,169	0,150
	Qsal	0,450	-0,289	-0,471	1,000	-0,171	-0,028	0,632	-0,492	-0,123	0,012
	Filt	-0,289	-0,131	0,190	-0,171	1,000	-0,010	-0,213	0,149	0,025	0,079
	Profundidad	0,058	0,071	-0,017	-0,028	-0,010	1,000	-0,426	-0,622	-0,303	0,643
	Ta	0,594	0,036	-0,416	0,632	-0,213	-0,426	1,000	-0,054	0,188	-0,541
	Odis	-0,479	0,018	0,274	-0,492	0,149	-0,622	-0,054	1,000	0,577	-0,582
	pH	-0,198	-0,005	0,169	-0,123	0,025	-0,303	0,188	0,577	1,000	-0,450
Con	-0,195	-0,399	0,150	0,012	0,079	0,643	-0,541	-0,582	-0,450	1,000	
Sig. (Unilateral)	DJ	0,004	0,000	0,009	0,000	0,007	0,004	0,000	0,000	0,000	0,000
	V	0,000	0,000	0,000	0,000	0,001	0,241	0,000	0,000	0,000	0,001
	DV	0,000	0,000	0,000	0,000	0,000	0,004	0,000	0,000	0,000	0,000
	Vx	0,000	0,044	0,000	0,000	0,076	0,219	0,000	0,000	0,052	0,309
	T	0,000	0,000	0,000	0,000	0,000	0,045	0,000	0,000	0,000	0,000
	Tn	0,000	0,000	0,000	0,000	0,000	0,057	0,000	0,000	0,000	0,000
	Tx	0,000	0,000	0,000	0,000	0,000	0,057	0,000	0,000	0,000	0,000
	Hf	0,000	0,000	0,000	0,000	0,000	0,390	0,000	0,000	0,000	0,000
	HR	0,000	0,008	0,000	0,040	0,000	0,398	0,011	0,000	0,000	0,000
	Rad	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	P	0,000	0,000	0,000	0,386	0,001	0,215	0,000	0,047	0,000	0,000
	ETo		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	VolExp	0,000		0,000	0,000	0,000	0,000	0,008	0,116	0,361	0,000
	Qent	0,000	0,000		0,000	0,000	0,125	0,000	0,000	0,000	0,000
	Qsal	0,000	0,000	0,000		0,000	0,030	0,000	0,000	0,000	0,208
	Filt	0,000	0,000	0,000	0,000		0,249	0,000	0,000	0,051	0,000
	Profundidad	0,000	0,000	0,125	0,030	0,249		0,000	0,000	0,000	0,000
	Ta	0,000	0,008	0,000	0,000	0,000	0,000		0,000	0,000	0,000
	Odis	0,000	0,116	0,000	0,000	0,000	0,000	0,000		0,000	0,000
	pH	0,000	0,361	0,000	0,000	0,051	0,000	0,000	0,000		0,000
Con	0,000	0,000	0,000	0,208	0,000	0,000	0,000	0,000	0,000		



### Comunalidades

	Inicial	Extracción
DJ	1,000	,809
V	1,000	,812
DV	1,000	,544
Vx	1,000	,815
T	1,000	,965
Tn	1,000	,955
Tx	1,000	,937
Hf	1,000	,714
HR	1,000	,801
Rad	1,000	,871
P	1,000	,303
ETo	1,000	,952
VolExp	1,000	,871
Qent	1,000	,576
Qsal	1,000	,861
Filt	1,000	,141
Profundidad	1,000	,740
Ta	1,000	,882
Odis	1,000	,930
pH	1,000	,665
Con	1,000	,880

Método de extracción: Análisis de Componentes principales

### Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	6,932	33,010	33,010	6,932	33,010	33,010
2	3,180	15,144	48,154	3,180	15,144	48,154
3	2,730	12,999	61,153	2,730	12,999	61,153
4	2,101	10,006	71,160	2,101	10,006	71,160
5	1,080	5,144	76,303	1,080	5,144	76,303
6	,936	4,457	80,760			
7	,875	4,166	84,926			
8	,759	3,615	88,541			
9	,506	2,411	90,952			
10	,454	2,160	93,112			
11	,328	1,560	94,672			
12	,271	1,289	95,961			
13	,227	1,080	97,042			
14	,180	,857	97,899			
15	,151	,720	98,619			
16	,101	,479	99,098			
17	,078	,369	99,467			
18	,071	,339	99,806			
19	,025	,120	99,926			
20	,013	,061	99,988			
21	,003	,012	100,000			

Método de extracción: Análisis de Componentes principales.

**Matriz de componentes(a)**

	Componente				
	1	2	3	4	5
DJ	,344	-,666	-,118	,483	-,016
V	-,132	,467	,439	,603	,141
DV	-,582	-,114	,022	,357	-,253
Vx	-,066	,289	,330	,626	,475
T	,978	-,045	-,044	,017	,068
Tn	,968	-,099	-,069	,005	,059
Tx	,966	-,018	-,045	,004	,040
Hf	-,509	,458	,226	-,158	-,412
HR	,047	-,441	-,430	-,631	,145
Rad	,765	,461	,218	-,160	,001
P	-,168	-,409	-,188	,008	,268
ETo	,881	,359	,203	-,001	,079
VolExp	,220	,718	,145	-,502	,185
Qent	-,584	-,210	-,101	-,044	,423
Qsal	,739	-,367	-,116	,314	-,261
Filt	-,330	-,105	-,007	,075	-,124
Profundidad	,035	-,304	,724	-,302	,175
Ta	,735	,131	-,504	,242	-,105
Odis	-,591	,454	-,598	,132	-,004
pH	-,159	,169	-,677	,067	,386
Con	-,124	-,669	,637	-,077	,073

Método de extracción: Análisis de componentes principales.  
a 5 componentes extraídos

**Matriz de componentes rotados(a)**

	Componente				
	1	2	3	4	5
DJ	,147	,869	-,105	,003	,147
V	-,077	-,103	-,002	,880	-,148
DV	-,675	,196	,057	,188	-,106
Vx	,034	-,002	-,010	,874	,223
T	,931	,291	-,025	-,069	-,091
Tn	,911	,326	-,031	-,113	-,073
Tx	,919	,268	-,012	-,076	-,122
Hf	-,484	-,498	,030	,078	-,474
HR	,077	-,019	-,014	-,792	,408
Rad	,823	-,256	-,040	,116	-,336
P	-,164	,220	-,040	-,151	,451
ETo	,915	-,072	-,045	,200	-,259
VolExp	,440	-,811	,068	,030	-,123
Qent	-,477	-,110	-,024	,005	,579
Qsal	,531	,718	-,011	-,098	-,232
Filt	-,370	,060	-,003	-,007	-,018
Profundidad	,058	-,148	-,841	,016	,084
Ta	,663	,374	,523	-,066	-,156
Odis	-,494	-,267	,776	,065	,096
pH	-,031	-,068	,652	-,045	,482
Con	-,208	,237	-,870	-,027	,153

Método de extracción: Análisis de componentes principales.  
Método de rotación: Normalización Varimax con Kaiser.  
a La rotación ha convergido en 6 iteraciones.

**ANEXO 14. ANÁLISIS FACTORIAL DEL PERIODO DE  
VALIDACIÓN, SIN DATOS ANÓMALOS Y  
VARIABLES SIN TRANSFORMAR**

## PERIODO DE VALIDACIÓN

### Estadísticos descriptivos

	Media	Desviación típica	N del análisis
DJ	166,96	58,313	4573
V	1,255	,274	4573
DV	205,95	74,571	4573
Vx	5,14	1,206	4573
T	21,203	5,162	4573
Tn	16,817	5,067	4573
Tx	25,732	5,302	4573
Hf	,537	2,194	4573
HR	58,71	12,145	4573
Rad	243,36	64,330	4573
P	,36	1,569	4573
ETo	4,30	1,208	4573
VolExp	9215464,0	1592039,7	4573
Qent	8784,2	12961,6	4573
Qsal	28606,1	12005,0	4573
Filt	-803,5	1680,3	4573
Profundidad	8,639	7,488	4573
Ta	18,05	6,844	4573
Odis	7,86	4,104	4573
pH	8,36	,375	4573
Con	957,06	22,714	4573

### KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,786
	Chi-cuadrado aproximado	130410,0
Prueba de esfericidad de Bartlett	gl	210
	Sig.	,000

**Matriz de correlaciones(a)**

		DJ	V	DV	Vx	T	Tn	Tx	Hf	HR	Rad	P
Correlación	DJ	1,000	-0,049	-0,073	-0,104	0,831	0,834	0,814	-0,436	0,046	0,201	0,023
	V	-0,049	1,000	0,025	0,767	-0,063	-0,093	-0,036	0,199	-0,617	0,173	-0,149
	DV	-0,073	0,025	1,000	-0,123	-0,282	-0,319	-0,234	0,084	-0,173	-0,083	0,047
	Vx	-0,104	0,767	-0,123	1,000	-0,121	-0,138	-0,088	0,148	-0,385	-0,113	0,080
	T	0,831	-0,063	-0,282	-0,121	1,000	0,981	0,981	-0,601	-0,026	0,472	-0,130
	Tn	0,834	-0,093	-0,319	-0,138	0,981	1,000	0,937	-0,593	0,072	0,384	-0,094
	Tx	0,814	-0,036	-0,234	-0,088	0,981	0,937	1,000	-0,592	-0,100	0,510	-0,120
	Hf	-0,436	0,199	0,084	0,148	-0,601	-0,593	-0,592	1,000	-0,060	-0,315	-0,034
	HR	0,046	-0,617	-0,173	-0,385	-0,026	0,072	-0,100	-0,060	1,000	-0,360	0,232
	Rad	0,201	0,173	-0,083	-0,113	0,472	0,384	0,510	-0,315	-0,360	1,000	-0,356
	P	0,023	-0,149	0,047	0,080	-0,130	-0,094	-0,120	-0,034	0,232	-0,356	1,000
	ETo	0,448	0,277	-0,185	0,036	0,736	0,656	0,770	-0,445	-0,445	0,872	-0,302
	VolExp	-0,899	0,034	-0,054	0,067	-0,573	-0,594	-0,560	0,165	-0,087	0,115	-0,054
	Qent	-0,714	0,089	0,249	0,149	-0,802	-0,792	-0,776	0,509	0,003	-0,480	-0,084
	Qsal	0,837	-0,020	-0,238	-0,093	0,830	0,845	0,797	-0,373	0,034	0,267	-0,035
	Filt	-0,127	0,093	0,093	0,053	-0,165	-0,191	-0,145	0,117	-0,156	-0,039	-0,138
	Profundidad	-0,020	-0,010	-0,010	-0,004	-0,008	-0,009	-0,008	0,007	0,000	0,010	-0,008
	Ta	0,515	-0,042	-0,134	-0,084	0,539	0,541	0,518	-0,298	0,029	0,229	-0,032
	Odis	-0,705	0,031	0,074	0,082	-0,597	-0,604	-0,579	0,263	-0,044	-0,121	0,007
	pH	-0,456	-0,066	0,113	0,011	-0,399	-0,414	-0,376	-0,087	0,026	-0,042	0,081
Con	0,324	-0,023	0,121	-0,021	0,096	0,112	0,096	0,053	0,065	-0,208	0,045	
Sig. (Unilateral)	DJ		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,000	0,063
	V	0,000		0,045	0,000	0,000	0,000	0,008	0,000	0,000	0,000	0,000
	DV	0,000	0,045		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,001
	Vx	0,000	0,000	0,000		0,000	0,000	0,000	0,000	0,000	0,000	0,000
	T	0,000	0,000	0,000	0,000		0,000	0,000	0,000	0,040	0,000	0,000
	Tn	0,000	0,000	0,000	0,000	0,000		0,000	0,000	0,000	0,000	0,000
	Tx	0,000	0,008	0,000	0,000	0,000	0,000		0,000	0,000	0,000	0,000
	Hf	0,000	0,000	0,000	0,000	0,000	0,000	0,000		0,000	0,000	0,011
	HR	0,001	0,000	0,000	0,000	0,040	0,000	0,000	0,000		0,000	0,000
	Rad	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000		0,000
	P	0,063	0,000	0,001	0,000	0,000	0,000	0,000	0,011	0,000	0,000	
	ETo	0,000	0,000	0,000	0,007	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	VolExp	0,000	0,010	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	Qent	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,424	0,000	0,000
	Qsal	0,000	0,087	0,000	0,000	0,000	0,000	0,000	0,000	0,011	0,000	0,009
	Filt	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,004	0,000
	Profundidad	0,090	0,241	0,252	0,402	0,286	0,268	0,304	0,310	0,488	0,259	0,304
	Ta	0,000	0,002	0,000	0,000	0,000	0,000	0,000	0,000	0,025	0,000	0,017
	Odis	0,000	0,018	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,000	0,319
	pH	0,000	0,000	0,000	0,222	0,000	0,000	0,000	0,000	0,038	0,002	0,000
Con	0,000	0,064	0,000	0,075	0,000	0,000	0,000	0,000	0,000	0,000	0,001	

a Determinante = 3,90E-013

**Matriz de correlaciones (continuación)**

		ETo	VolExp	Qent	Qsal	Filt	Profundidad	Ta	Odis	pH	Con
Correlación	DJ	0,448	-0,899	-0,714	0,837	-0,127	-0,020	0,515	-0,705	-0,456	0,324
	V	0,277	0,034	0,089	-0,020	0,093	-0,010	-0,042	0,031	-0,066	-0,023
	DV	-0,185	-0,054	0,249	-0,238	0,093	-0,010	-0,134	0,074	0,113	0,121
	Vx	0,036	0,067	0,149	-0,093	0,053	-0,004	-0,084	0,082	0,011	-0,021
	T	0,736	-0,573	-0,802	0,830	-0,165	-0,008	0,539	-0,597	-0,399	0,096
	Tn	0,656	-0,594	-0,792	0,845	-0,191	-0,009	0,541	-0,604	-0,414	0,112
	Tx	0,770	-0,560	-0,776	0,797	-0,145	-0,008	0,518	-0,579	-0,376	0,096
	Hf	-0,445	0,165	0,509	-0,373	0,117	0,007	-0,298	0,263	-0,087	0,053
	HR	-0,445	-0,087	0,003	0,034	-0,156	0,000	0,029	-0,044	0,026	0,065
	Rad	0,872	0,115	-0,480	0,267	-0,039	0,010	0,229	-0,121	-0,042	-0,208
	P	-0,302	-0,054	-0,084	-0,035	-0,138	-0,008	-0,032	0,007	0,081	0,045
	ETo	1,000	-0,128	-0,649	0,527	-0,086	0,010	0,361	-0,315	-0,223	-0,118
	VolExp	-0,128	1,000	0,406	-0,708	0,047	0,026	-0,390	0,658	0,507	-0,467
	Qent	-0,649	0,406	1,000	-0,685	0,324	-0,005	-0,463	0,477	0,238	-0,008
	Qsal	0,527	-0,708	-0,685	1,000	-0,165	-0,006	0,503	-0,632	-0,528	0,210
	Filt	-0,086	0,047	0,324	-0,165	1,000	-0,012	-0,113	0,096	0,033	0,036
	Profundidad	0,010	0,026	-0,005	-0,006	-0,012	1,000	-0,693	-0,545	-0,356	0,628
	Ta	0,361	-0,390	-0,463	0,503	-0,113	-0,693	1,000	0,087	0,143	-0,476
	Odis	-0,315	0,658	0,477	-0,632	0,096	-0,545	0,087	1,000	0,672	-0,581
	pH	-0,223	0,507	0,238	-0,528	0,033	-0,356	0,143	0,672	1,000	-0,523
Con	-0,118	-0,467	-0,008	0,210	0,036	0,628	-0,476	-0,581	-0,523	1,000	
Sig. (Unilateral)	DJ	0,000	0,000	0,000	0,000	0,000	0,090	0,000	0,000	0,000	0,000
	V	0,000	0,010	0,000	0,087	0,000	0,241	0,002	0,018	0,000	0,064
	DV	0,000	0,000	0,000	0,000	0,000	0,252	0,000	0,000	0,000	0,000
	Vx	0,007	0,000	0,000	0,000	0,000	0,402	0,000	0,000	0,222	0,075
	T	0,000	0,000	0,000	0,000	0,000	0,286	0,000	0,000	0,000	0,000
	Tn	0,000	0,000	0,000	0,000	0,000	0,268	0,000	0,000	0,000	0,000
	Tx	0,000	0,000	0,000	0,000	0,000	0,304	0,000	0,000	0,000	0,000
	Hf	0,000	0,000	0,000	0,000	0,000	0,310	0,000	0,000	0,000	0,000
	HR	0,000	0,000	0,424	0,011	0,000	0,488	0,025	0,001	0,038	0,000
	Rad	0,000	0,000	0,000	0,000	0,004	0,259	0,000	0,000	0,002	0,000
	P	0,000	0,000	0,000	0,009	0,000	0,304	0,017	0,319	0,000	0,001
	ETo		0,000	0,000	0,000	0,000	0,257	0,000	0,000	0,000	0,000
	VolExp	0,000		0,000	0,000	0,001	0,040	0,000	0,000	0,000	0,000
	Qent	0,000	0,000		0,000	0,000	0,377	0,000	0,000	0,000	0,299
	Qsal	0,000	0,000	0,000		0,000	0,340	0,000	0,000	0,000	0,000
	Filt	0,000	0,001	0,000	0,000		0,217	0,000	0,000	0,012	0,008
	Profundidad	0,257	0,040	0,377	0,340	0,217		0,000	0,000	0,000	0,000
	Ta	0,000	0,000	0,000	0,000	0,000	0,000		0,000	0,000	0,000
	Odis	0,000	0,000	0,000	0,000	0,000	0,000	0,000		0,000	0,000
	pH	0,000	0,000	0,000	0,000	0,012	0,000	0,000	0,000		0,000
Con	0,000	0,000	0,299	0,000	0,008	0,000	0,000	0,000	0,000		

### Comunalidades

	Inicial	Extracción
DJ	1,000	,958
V	1,000	,894
DV	1,000	,888
Vx	1,000	,857
T	1,000	,960
Tn	1,000	,943
Tx	1,000	,931
Hf	1,000	,607
HR	1,000	,768
Rad	1,000	,845
P	1,000	,702
ETo	1,000	,952
VolExp	1,000	,940
Qent	1,000	,809
Qsal	1,000	,851
Filt	1,000	,432
Profundidad	1,000	,884
Ta	1,000	,928
Odis	1,000	,868
pH	1,000	,712
Con	1,000	,810

Método de extracción: Análisis de Componentes principales.

### Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	7,729	36,807	36,807	7,729	36,807	36,807
2	3,135	14,928	51,735	3,135	14,928	51,735
3	2,578	12,277	64,012	2,578	12,277	64,012
4	1,720	8,188	72,200	1,720	8,188	72,200
5	1,295	6,165	78,366	1,295	6,165	78,366
6	1,083	5,159	83,525	1,083	5,159	83,525
7	,884	4,208	87,733			
8	,632	3,007	90,740			
9	,370	1,763	92,503			
10	,359	1,711	94,214			
11	,279	1,331	95,545			
12	,205	,974	96,519			
13	,192	,916	97,435			
14	,179	,853	98,288			
15	,128	,607	98,895			
16	,104	,494	99,389			
17	,058	,275	99,664			
18	,035	,165	99,829			
19	,024	,113	99,942			
20	,009	,043	99,985			
21	,003	,015	100,000			

Método de extracción: Análisis de Componentes principales.

**Matriz de componentes(a) – 6 Factores**

	Componente					
	1	2	3	4	5	6
DJ	,904	-,181	-,099	,271	,131	,083
V	-,043	,207	,805	,422	-,145	,045
DV	-,241	-,126	,052	,077	,585	,681
Vx	-,135	,109	,600	,536	-,424	,017
T	,972	,103	-,020	-,043	-,019	-,038
Tn	,960	,050	-,094	-,008	-,062	-,080
Tx	,955	,126	,036	-,046	,007	,019
Hf	-,564	-,214	,197	,265	,101	-,351
HR	-,030	-,309	-,763	-,098	-,168	-,227
Rad	,464	,497	,366	-,486	,101	,059
P	-,096	-,207	-,351	,295	-,458	,479
ETo	,715	,433	,419	-,273	,003	,050
VolExp	-,678	,404	,117	-,486	-,259	-,021
Qent	-,834	-,149	,096	,104	,217	-,158
Qsal	,891	-,096	-,045	,183	,010	-,114
Filt	-,199	-,019	,217	,067	,534	-,236
Profundidad	,039	-,675	,370	-,488	-,209	,090
Ta	,514	,587	-,362	,403	,148	-,064
Odis	-,691	,596	-,171	,071	,017	-,034
pH	-,482	,566	-,276	-,056	-,044	,279
Con	,190	-,844	,218	-,030	,076	,084

Método de extracción: Análisis de componentes principales.  
a 6 componentes extraídos

**Matriz de componentes rotados(a) – 6 Factores**

	Componente					
	1	2	3	4	5	6
DJ	,961	,130	-,017	-,040	,087	,083
V	-,022	,084	-,006	,934	-,117	,013
DV	-,115	-,071	-,040	-,012	-,060	,930
Vx	-,057	-,148	,011	,888	,141	-,153
T	,816	,505	,036	-,072	,078	-,166
Tn	,833	,419	,034	-,113	,113	-,217
Tx	,790	,539	,032	-,025	,072	-,104
Hf	-,330	-,539	-,089	,231	-,371	-,093
HR	,067	-,395	,010	-,682	,243	-,288
Rad	,087	,896	,048	,073	-,165	-,029
P	,014	-,348	,010	,007	,748	,145
ETo	,385	,858	,039	,227	-,079	-,092
VolExp	-,918	,240	,073	,006	,036	-,180
Qent	-,628	-,519	-,047	,081	-,346	,131
Qsal	,886	,201	-,011	-,016	,034	-,156
Filt	-,075	-,122	,018	,058	-,614	,178
Profundidad	-,068	,114	-,926	-,025	,076	-,040
Ta	,531	,151	,787	-,041	,027	-,041
Odis	-,694	-,112	,608	,035	-,050	,009
pH	-,570	,076	,541	-,086	,224	,177
Con	,332	-,242	-,782	-,012	-,038	,170

Método de extracción: Análisis de componentes principales.  
Método de rotación: Normalización Varimax con Kaiser.  
a La rotación ha convergido en 10 iteraciones.



**Matriz de componentes(a) – 5 Factores**

	Componente				
	1	2	3	4	5
DJ	,904	-,181	-,099	,271	,131
V	-,043	,207	,805	,422	-,145
DV	-,241	-,126	,052	,077	,585
Vx	-,135	,109	,600	,536	-,424
T	,972	,103	-,020	-,043	-,019
Tn	,960	,050	-,094	-,008	-,062
Tx	,955	,126	,036	-,046	,007
Hf	-,564	-,214	,197	,265	,101
HR	-,030	-,309	-,763	-,098	-,168
Rad	,464	,497	,366	-,486	,101
P	-,096	-,207	-,351	,295	-,458
ETo	,715	,433	,419	-,273	,003
VolExp	-,678	,404	,117	-,486	-,259
Qent	-,834	-,149	,096	,104	,217
Qsal	,891	-,096	-,045	,183	,010
Filt	-,199	-,019	,217	,067	,534
Profundidad	,039	-,675	,370	-,488	-,209
Ta	,514	,587	-,362	,403	,148
Odis	-,691	,596	-,171	,071	,017
pH	-,482	,566	-,276	-,056	-,044
Con	,190	-,844	,218	-,030	,076

Método de extracción: Análisis de componentes principales.  
a 5 componentes extraídos

**Matriz de componentes rotados(a) – 5 Factores**

	Componente				
	1	2	3	4	5
DJ	,972	,012	-,062	-,039	,009
V	-,026	,140	-,009	,931	-,076
DV	-,103	-,071	-,013	-,073	-,635
Vx	-,078	-,149	,007	,894	,174
T	,863	,389	,005	-,071	,238
Tn	,872	,297	,000	-,106	,278
Tx	,841	,425	,006	-,031	,204
Hf	-,406	-,363	-,108	,270	-,320
HR	,034	-,461	-,005	-,653	,276
Rad	,174	,894	,068	,047	,076
P	,010	-,558	,030	-,022	,399
ETo	,466	,814	,042	,206	,163
VolExp	-,886	,300	,131	-,008	,218
Qent	-,690	-,339	-,040	,099	-,427
Qsal	,897	,110	-,058	-,001	,135
Filt	-,106	,067	-,002	,079	-,596
Profundidad	-,102	,117	-,913	-,040	,129
Ta	,584	,069	,759	-,030	,042
Odis	-,671	-,048	,640	,036	-,057
pH	-,516	,033	,589	-,116	,081
Con	,267	-,232	-,804	-,013	-,178

Método de extracción: Análisis de componentes principales.  
Método de rotación: Normalización Varimax con Kaiser.  
a La rotación ha convergido en 9 iteraciones.

**ANEXO 15. ANÁLISIS FACTORIAL EN EL PERIODO COMPLETO  
DE ESTUDIO, SIN DATOS ANÓMALOS Y LAS  
VARIABLES TRANSFORMADAS**

## PERIODO COMPLETO DE ESTUDIO

### Estadísticos descriptivos

	Media	Desviación típica	N del análisis
DJ_T	,0119	,9039	8974
V_T	-,3878	,4143	8974
DV_T	,4649	,7043	8974
Vx_T	-,4167	,5389	8974
T_T	,3121	,7537	8974
Tn_T	,3745	,7776	8974
Tx_T	,2126	,6957	8974
Hf_T	-1,9217	,6763	8974
HR_T	,2411	,5768	8974
Rad_T	,2705	,8326	8974
P_T	-2,0433	,4317	8974
ETo_T	-,2081	,8085	8974
VolExp_T	,1731	,9878	8974
Qent_T	-1,5429	,6170	8974
Qsal_T	,0452	,6878	8974
Filt_T	1,8589	,7554	8974
Profundidad_T	-,7198	,8877	8974
Ta_T	-,4275	1,1342	8974
Odis_T	-,0271	,9319	8974
pH_T	,6233	,3948	8974
Con_T	-,3236	,6586	8974

a) Extracción de los factores por el Método de Componentes principales, sin rotación

### KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,808
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	197776,1
	gl	210
	Sig.	,000

**Matriz de correlaciones(a)**

		DJ_T	V_T	DV_T	Vx_T	T_T	Tn_T	Tx_T	Hf_T	HR_T	Rad_T	P_T
Correlación	DJ_T	1,000	-0,037	0,011	0,034	0,324	0,340	0,302	-0,438	0,030	-0,147	0,149
	V_T	-0,037	1,000	0,118	0,693	-0,142	-0,162	-0,145	0,178	-0,634	0,092	-0,245
	DV_T	0,011	0,118	1,000	0,018	-0,453	-0,465	-0,405	0,206	-0,205	-0,385	0,142
	Vx_T	0,034	0,693	0,018	1,000	-0,058	-0,088	-0,062	0,043	-0,458	-0,033	0,001
	T_T	0,324	-0,142	-0,453	-0,058	1,000	0,980	0,980	-0,620	0,057	0,605	-0,149
	Tn_T	0,340	-0,162	-0,465	-0,088	0,980	1,000	0,936	-0,616	0,127	0,543	-0,110
	Tx_T	0,302	-0,145	-0,405	-0,062	0,980	0,936	1,000	-0,617	0,022	0,628	-0,176
	Hf_T	-0,438	0,178	0,206	0,043	-0,620	-0,616	-0,617	1,000	-0,185	-0,252	-0,056
	HR_T	0,030	-0,634	-0,205	-0,458	0,057	0,127	0,022	-0,185	1,000	-0,215	0,288
	Rad_T	-0,147	0,092	-0,385	-0,033	0,605	0,543	0,628	-0,252	-0,215	1,000	-0,405
	P_T	0,149	-0,245	0,142	0,001	-0,149	-0,110	-0,176	-0,056	0,288	-0,405	1,000
	ETo_T	0,012	0,169	-0,433	0,127	0,798	0,742	0,802	-0,366	-0,277	0,902	-0,317
	VolExp_T	-0,562	0,033	-0,267	0,054	-0,086	-0,114	-0,092	0,139	-0,033	0,402	-0,150
	Qent_T	-0,042	0,050	0,397	0,076	-0,725	-0,720	-0,709	0,318	0,026	-0,632	0,103
	Qsal_T	0,543	-0,074	-0,245	-0,075	0,719	0,734	0,699	-0,432	0,017	0,269	-0,037
	Filt_T	-0,018	0,052	0,196	0,041	-0,278	-0,290	-0,266	0,142	-0,110	-0,242	-0,012
	Profundidad_T	-0,019	0,004	-0,011	0,004	0,004	0,004	0,001	0,010	0,000	0,015	-0,009
	Ta_T	0,346	-0,091	-0,258	-0,064	0,631	0,636	0,609	-0,412	0,044	0,348	-0,061
	Odis_T	-0,387	0,061	0,166	0,058	-0,515	-0,524	-0,494	0,303	-0,045	-0,195	0,028
	pH_T	-0,228	-0,096	-0,050	-0,011	-0,104	-0,129	-0,075	-0,144	0,132	0,095	0,062
Con_T	0,266	-0,006	0,213	-0,009	-0,180	-0,165	-0,177	-0,035	0,032	-0,324	0,134	
Sig. (Unilateral)	DJ_T		0,000	0,153	0,001	0,000	0,000	0,000	0,000	0,002	0,000	0,000
	V_T	0,000		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	DV_T	0,153	0,000		0,042	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	Vx_T	0,001	0,000	0,042		0,000	0,000	0,000	0,000	0,000	0,001	0,473
	T_T	0,000	0,000	0,000	0,000		0,000	0,000	0,000	0,000	0,000	0,000
	Tn_T	0,000	0,000	0,000	0,000	0,000		0,000	0,000	0,000	0,000	0,000
	Tx_T	0,000	0,000	0,000	0,000	0,000	0,000		0,000	0,021	0,000	0,000
	Hf_T	0,000	0,000	0,000	0,000	0,000	0,000	0,000		0,000	0,000	0,000
	HR_T	0,002	0,000	0,000	0,000	0,000	0,000	0,021	0,000		0,000	0,000
	Rad_T	0,000	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000		0,000
	P_T	0,000	0,000	0,000	0,473	0,000	0,000	0,000	0,000	0,000	0,000	
	ETo_T	0,124	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	VolExp_T	0,000	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,000	0,000
	Qent_T	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,007	0,000	0,000
	Qsal_T	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,054	0,000	0,000
	Filt_T	0,041	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,137
	Profundidad_T	0,040	0,362	0,142	0,344	0,367	0,365	0,450	0,175	0,482	0,077	0,202
	Ta_T	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	Odis_T	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,004
	pH_T	0,000	0,000	0,000	0,139	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Con_T	0,000	0,280	0,000	0,187	0,000	0,000	0,000	0,001	0,001	0,000	0,000	

a Determinante = 2,63E-010

**Matriz de correlaciones (continuación)**

		ETo_T	VolExp_T	Qent_T	Qsal_T	Filt_T	Profundidad_T	Ta_T	Odis_T	pH_T	Con_T
Correlación	DJ_T	0,012	-0,562	-0,042	0,543	-0,018	-0,019	0,346	-0,387	-0,228	0,266
	V_T	0,169	0,033	0,050	-0,074	0,052	0,004	-0,091	0,061	-0,096	-0,006
	DV_T	-0,433	-0,267	0,397	-0,245	0,196	-0,011	-0,258	0,166	-0,050	0,213
	Vx_T	0,127	0,054	0,076	-0,075	0,041	0,004	-0,064	0,058	-0,011	-0,009
	T_T	0,798	-0,086	-0,725	0,719	-0,278	0,004	0,631	-0,515	-0,104	-0,180
	Tn_T	0,742	-0,114	-0,720	0,734	-0,290	0,004	0,636	-0,524	-0,129	-0,165
	Tx_T	0,802	-0,092	-0,709	0,699	-0,266	0,001	0,609	-0,494	-0,075	-0,177
	Hf_T	-0,366	0,139	0,318	-0,432	0,142	0,010	-0,412	0,303	-0,144	-0,035
	HR_T	-0,277	-0,033	0,026	0,017	-0,110	0,000	0,044	-0,045	0,132	0,032
	Rad_T	0,902	0,402	-0,632	0,269	-0,242	0,015	0,348	-0,195	0,095	-0,324
	P_T	-0,317	-0,150	0,103	-0,037	-0,012	-0,009	-0,061	0,028	0,062	0,134
	ETo_T	1,000	0,264	-0,759	0,460	-0,276	0,015	0,472	-0,334	-0,002	-0,308
	VolExp_T	0,264	1,000	-0,070	-0,445	-0,055	0,031	-0,193	0,322	0,382	-0,352
	Qent_T	-0,759	-0,070	1,000	-0,572	0,401	-0,001	-0,487	0,352	0,018	0,273
	Qsal_T	0,460	-0,445	-0,572	1,000	-0,216	-0,018	0,563	-0,508	-0,273	-0,022
	Filt_T	-0,276	-0,055	0,401	-0,216	1,000	-0,004	-0,202	0,120	-0,048	0,121
	Profundidad_T	0,015	0,031	-0,001	-0,018	-0,004	1,000	-0,545	-0,589	-0,290	0,514
	Ta_T	0,472	-0,193	-0,487	0,563	-0,202	-0,545	1,000	0,052	0,158	-0,395
	Odis_T	-0,334	0,322	0,352	-0,508	0,120	-0,589	0,052	1,000	0,578	-0,471
	pH_T	-0,002	0,382	0,018	-0,273	-0,048	-0,290	0,158	0,578	1,000	-0,434
Con_T	-0,308	-0,352	0,273	-0,022	0,121	0,514	-0,395	-0,471	-0,434	1,000	
Sig. (Unilateral)	DJ_T	0,124	0,000	0,000	0,000	0,041	0,040	0,000	0,000	0,000	0,000
	V_T	0,000	0,001	0,000	0,000	0,000	0,362	0,000	0,000	0,000	0,280
	DV_T	0,000	0,000	0,000	0,000	0,000	0,142	0,000	0,000	0,000	0,000
	Vx_T	0,000	0,000	0,000	0,000	0,000	0,344	0,000	0,000	0,139	0,187
	T_T	0,000	0,000	0,000	0,000	0,000	0,367	0,000	0,000	0,000	0,000
	Tn_T	0,000	0,000	0,000	0,000	0,000	0,365	0,000	0,000	0,000	0,000
	Tx_T	0,000	0,000	0,000	0,000	0,000	0,450	0,000	0,000	0,000	0,000
	Hf_T	0,000	0,000	0,000	0,000	0,000	0,175	0,000	0,000	0,000	0,001
	HR_T	0,000	0,001	0,007	0,054	0,000	0,482	0,000	0,000	0,000	0,001
	Rad_T	0,000	0,000	0,000	0,000	0,000	0,077	0,000	0,000	0,000	0,000
	P_T	0,000	0,000	0,000	0,000	0,137	0,202	0,000	0,004	0,000	0,000
	ETo_T		0,000	0,000	0,000	0,000	0,074	0,000	0,000	0,442	0,000
	VolExp_T	0,000		0,000	0,000	0,000	0,002	0,000	0,000	0,000	0,000
	Qent_T	0,000	0,000		0,000	0,000	0,479	0,000	0,000	0,041	0,000
	Qsal_T	0,000	0,000	0,000		0,000	0,041	0,000	0,000	0,000	0,017
	Filt_T	0,000	0,000	0,000	0,000		0,345	0,000	0,000	0,000	0,000
	Profundidad_T	0,074	0,002	0,479	0,041	0,345		0,000	0,000	0,000	0,000
	Ta_T	0,000	0,000	0,000	0,000	0,000	0,000		0,000	0,000	0,000
	Odis_T	0,000	0,000	0,000	0,000	0,000	0,000	0,000		0,000	0,000
	pH_T	0,442	0,000	0,041	0,000	0,000	0,000	0,000	0,000		0,000
Con_T	0,000	0,000	0,000	0,017	0,000	0,000	0,000	0,000	0,000		

**Inversa de la matriz de correlaciones**

DJ_T	2,524	-0,003	0,058	-0,101	-2,082	1,231	1,287	0,687	0,260	0,181	-0,160	0,283	0,334	-0,666	-0,910	0,049	Profundidad_T	0,148	Ta_T	-0,823	Odis_T	0,871	pH_T	-0,109	Con_T	-0,211
V_T	-0,003	3,277	-0,149	-1,601	2,491	-1,674	1,232	0,124	0,712	0,449	0,576	-2,272	0,452	-0,196	-0,233	0,066		0,060	0,034	0,053	0,017	-0,006				
DV_T	0,058	-0,149	1,655	0,242	1,796	-0,024	-1,412	0,146	0,479	-0,036	-0,221	0,362	0,418	-0,056	0,162	-0,013		-0,007	-0,033	0,009	0,016	-0,010				
Vx_T	-0,101	-1,601	0,242	2,568	-1,270	1,323	0,104	0,131	0,045	1,816	-0,341	-2,130	-0,113	-0,425	0,147	0,051		-0,072	0,101	-0,170	-0,066	0,021				
T_T	-2,082	2,491	1,796	-1,270	155,803	-75,309	-73,096	0,269	0,979	6,155	0,048	-14,414	-1,909	-2,263	0,981	-0,438		-0,311	-1,911	1,710	0,776	0,956				
Tn_T	1,231	-1,674	-0,024	1,323	-75,309	51,893	27,618	1,346	-2,414	1,075	0,031	1,644	0,126	1,593	-1,646	0,282		-0,440	-1,804	0,869	0,783	0,345				
Tx_T	1,287	1,232	-1,412	0,104	-73,096	27,618	48,868	1,110	-0,284	-1,686	0,758	-0,879	2,551	0,048	-1,335	0,146		0,327	1,126	-0,319	-1,094	-0,362				
Hf_T	0,687	0,124	0,146	0,131	0,269	1,346	1,110	2,506	0,095	0,595	0,154	-1,055	0,091	0,163	-0,268	0,023		-0,178	-0,327	0,083	0,810	0,518				
HR_T	0,260	0,712	0,479	0,045	0,979	-2,414	-0,284	0,095	2,682	-1,200	-0,322	3,391	-0,271	0,143	0,249	0,180		0,152	-0,057	0,516	-0,339	0,021				
Rad_T	0,181	0,449	-0,036	1,816	6,155	1,075	-1,686	0,595	-1,200	11,640	1,107	-15,243	-0,363	-1,047	-0,254	0,138		-0,215	-0,464	0,054	-0,119	-0,004				
P_T	-0,160	0,576	-0,221	-0,341	0,048	0,031	0,758	0,154	-0,322	1,107	1,518	-1,131	-0,005	0,298	-0,030	0,040		0,044	0,017	0,047	-0,125	-0,076				
ETo_T	0,283	-2,272	0,362	-2,130	-14,414	1,644	-0,879	-1,055	3,391	-15,243	-1,131	30,357	-2,212	3,810	0,949	-0,122		0,281	-0,392	1,124	0,068	0,212				
VolExp_T	0,334	0,452	0,418	-0,113	-1,909	0,126	2,551	0,091	-0,271	-0,363	-0,005	-2,212	3,144	-0,225	0,847	0,024		-0,312	0,890	-0,868	-0,329	0,312				
Qent_T	-0,666	-0,196	-0,056	-0,425	-2,263	1,593	0,048	0,163	0,143	-1,047	0,298	3,810	-0,225	3,834	1,005	-0,520		0,046	0,279	-0,459	0,300	-0,199				
Qsal_T	-0,910	-0,233	0,162	0,147	0,981	-1,646	-1,335	-0,268	0,249	-0,254	-0,030	0,949	0,847	1,005	3,874	-0,002		-0,158	-0,441	0,243	0,263	0,318				
Filt_T	0,049	0,066	-0,013	0,051	-0,438	0,282	0,146	0,023	0,180	0,138	0,040	-0,122	0,024	-0,520	-0,002	1,228		0,048	0,030	0,029	0,034	-0,001				
Profundidad_T	0,148	0,060	-0,007	-0,072	0,101	-0,170	-0,066	0,021	1,526	4,649	-1,605	-0,431	-0,166	1,650	5,763	-1,366		2,979	1,526	-1,605	-0,431	-0,166				
Ta_T	-0,823	0,034	0,009	0,016	-0,010	0,034	0,029	0,083	0,516	0,054	0,047	1,124	-0,868	-0,459	0,243	0,029		1,650	4,649	-1,605	-0,431	-0,166				
Odis_T	0,871	0,053	0,009	-0,170	1,710	0,869	-0,319	0,083	0,516	0,054	0,047	1,124	-0,868	-0,459	0,243	0,029		1,650	4,649	-1,605	-0,431	-0,166				
pH_T	-0,109	0,017	0,016	-0,066	0,776	0,783	-1,094	0,810	-0,339	-0,119	-0,125	0,058	-0,329	0,300	0,263	0,034		1,650	4,649	-1,605	-0,431	-0,166				
Con_T	-0,211	-0,006	-0,010	0,021	0,956	0,345	-0,362	0,518	0,021	-0,004	-0,076	0,212	0,312	-0,199	0,318	-0,001		-0,648	-0,166	0,964	0,438	2,398				

**Matriz anti-imagen**

Correlación	DJ_T	V_T	DV_T	Vx_T	T_T	Tn_T	Tx_T	Hf_T	HR_T	Rad_T	P_T	ETo_T	VolExp_T	Qent_T	Qsal_T	Filt_T	Profundidad_T	Ta_T	Odis_T	pH_T	Con_T
DJ_T	0,796	-0,001	0,028	-0,040	-0,105	0,108	0,116	0,273	0,100	0,033	-0,082	0,032	0,119	-0,214	-0,291	0,028	0,054	-0,240	0,228	-0,045	-0,086
V_T	-0,001	0,668	-0,064	-0,552	0,110	-0,128	0,097	0,043	0,240	0,073	0,258	-0,228	0,141	-0,055	-0,066	0,033	0,019	0,009	0,012	0,006	-0,002
DV_T	0,028	-0,064	0,897	0,117	0,112	-0,003	-0,157	0,072	0,227	-0,008	-0,140	0,051	0,183	-0,022	0,064	-0,009	-0,003	-0,012	0,003	0,008	-0,005
Vx_T	-0,040	-0,552	0,117	0,569	-0,064	0,115	0,009	0,052	0,017	0,332	-0,173	-0,241	-0,040	-0,135	0,047	0,029	-0,026	0,029	-0,044	-0,027	0,008
T_T	-0,105	0,110	0,112	-0,064	0,781	-0,838	-0,838	0,014	0,048	0,145	0,003	-0,210	-0,086	-0,093	0,040	-0,032	-0,014	-0,071	0,057	0,041	0,049
Tn_T	0,108	-0,128	-0,003	0,115	-0,838	0,822	0,548	0,118	-0,205	0,044	0,003	0,041	0,010	0,113	-0,116	0,035	-0,035	-0,116	0,050	0,071	0,031
Tx_T	0,116	0,097	-0,157	0,009	-0,838	0,548	0,822	0,100	-0,025	-0,071	0,088	-0,023	0,206	0,004	-0,097	0,019	0,027	0,075	-0,019	-0,103	-0,033
Hf_T	0,273	0,043	0,072	0,052	0,104	0,118	0,100	0,874	0,037	0,110	0,079	-0,121	0,033	0,053	-0,086	0,013	-0,065	-0,096	0,022	0,335	0,211
HR_T	0,100	0,240	0,227	0,017	0,048	-0,205	-0,025	0,037	0,682	-0,215	-0,159	0,376	-0,093	0,045	0,077	0,099	0,054	-0,016	0,131	-0,136	0,008
Rad_T	0,033	0,073	-0,008	0,332	0,145	0,044	-0,071	0,110	-0,215	0,773	0,263	-0,811	-0,060	-0,157	-0,038	0,036	-0,036	-0,063	0,007	-0,023	-0,001
P_T	-0,082	0,258	-0,140	-0,173	0,003	0,003	0,088	0,079	-0,159	0,263	0,673	-0,167	-0,002	0,124	-0,012	0,029	0,021	0,006	0,016	-0,066	-0,040
ETo_T	0,032	-0,228	0,051	-0,241	-0,210	0,041	-0,023	0,792	0,376	0,673	-0,167	0,792	-0,226	0,353	0,087	-0,020	0,030	-0,033	0,085	0,007	0,025
VolExp_T	0,119	0,141	0,183	-0,040	-0,086	-0,093	0,775	-0,065	-0,093	-0,060	-0,002	-0,226	0,775	-0,065	0,243	0,012	-0,102	0,233	-0,204	-0,122	0,113
Qent_T	-0,214	-0,055	-0,022	0,064	0,040	0,010	0,900	0,261	0,900	0,261	0,124	0,353	-0,065	0,900	0,261	-0,240	0,014	0,066	-0,098	0,100	-0,066
Qsal_T	-0,291	-0,066	0,064	0,047	0,040	-0,116	-0,097	-0,086	0,077	-0,038	-0,012	0,087	0,243	0,261	0,922	-0,001	-0,047	-0,104	0,051	0,088	0,104
Filt_T	0,028	0,033	-0,009	0,029	-0,032	0,035	0,094	0,013	0,099	0,036	0,029	-0,020	0,012	-0,240	-0,001	0,904	0,025	0,013	0,011	0,020	-0,001
Profundidad_T	0,054	0,019	-0,003	-0,026	-0,014	-0,035	0,681	-0,065	0,054	-0,036	0,021	0,030	-0,102	0,014	-0,047	0,025	0,681	0,410	0,398	-0,224	-0,243
Ta_T	-0,240	0,009	-0,012	0,029	-0,071	0,050	0,071	0,057	0,050	0,071	0,050	0,071	0,050	0,071	0,050	0,071	0,050	0,071	0,050	0,071	0,050
Odis_T	0,228	0,012	0,003	-0,044	0,057	0,050	-0,019	0,022	0,131	0,007	0,016	0,085	-0,204	0,098	0,051	0,011	0,398	-0,310	0,813	-0,373	0,259
pH_T	-0,045	0,006	0,008	-0,027	0,041	0,071	-0,103	0,335	-0,136	-0,023	-0,066	0,007	-0,122	0,100	0,088	0,020	-0,224	-0,131	0,813	0,700	0,185
Con_T	-0,086	-0,002	-0,005	0,008	0,049	0,031	-0,050	0,211	0,008	-0,001	-0,040	0,025	0,113	-0,066	0,104	-0,001	-0,243	-0,050	0,259	0,185	0,855

La diagonal principal corresponde a la medida de adecuación muestral

### Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	6,853	32,635	32,635	6,853	32,635	32,635
2	3,250	15,476	48,111	3,250	15,476	48,111
3	2,541	12,100	60,211	2,541	12,100	60,211
4	1,995	9,500	69,711	1,995	9,500	69,711
5	1,115	5,311	75,022	1,115	5,311	75,022
6	,975	4,642	79,664			
7	,797	3,795	83,459			
8	,713	3,394	86,853			
9	,440	2,097	88,950			
10	,434	2,066	91,016			
11	,360	1,716	92,732			
12	,298	1,420	94,153			
13	,269	1,282	95,435			
14	,226	1,076	96,511			
15	,202	,964	97,475			
16	,185	,882	98,357			
17	,159	,756	99,114			
18	,116	,553	99,667			
19	,041	,197	99,864			
20	,024	,115	99,979			
21	,004	,021	100,000			

Método de extracción: Análisis de Componentes principales

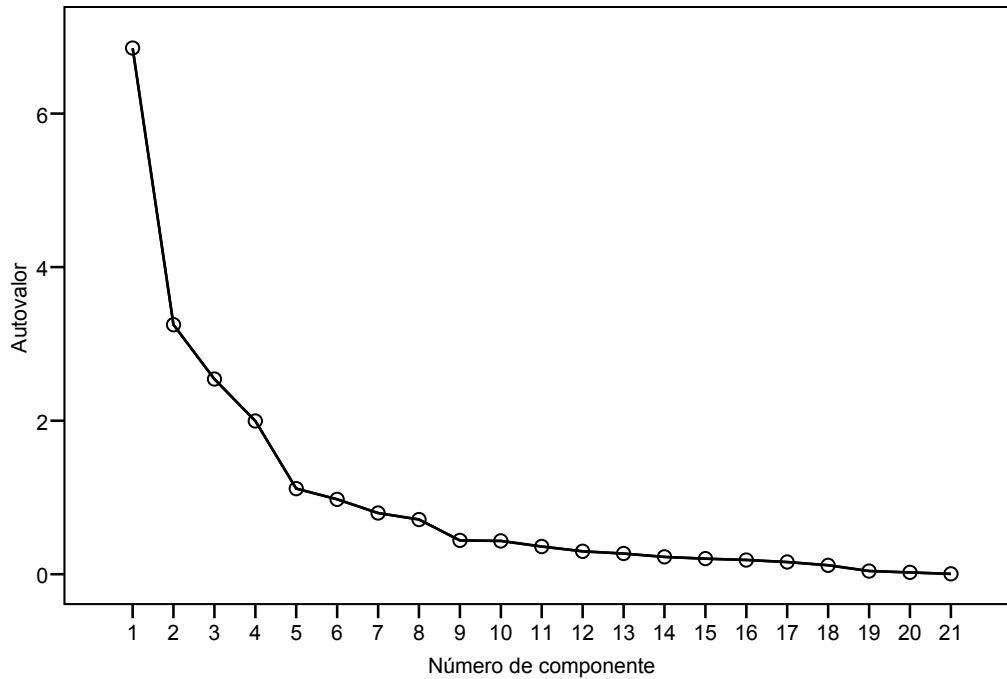
### Comunalidades

	Inicial	Extracción
DJ_T	1,000	,714
V_T	1,000	,833
DV_T	1,000	,415
Vx_T	1,000	,825
T_T	1,000	,947
Tn_T	1,000	,929
Tx_T	1,000	,914
Hf_T	1,000	,611
HR_T	1,000	,742
Rad_T	1,000	,828
P_T	1,000	,597
ETo_T	1,000	,940
VolExp_T	1,000	,800
Qent_T	1,000	,695
Qsal_T	1,000	,785
Filt_T	1,000	,226
Profundidad_T	1,000	,818
Ta_T	1,000	,853
Odis_T	1,000	,895
pH_T	1,000	,678
Con_T	1,000	,711

Método de extracción: Análisis de Componentes principales



Gráfico de sedimentación



Matriz de componentes(a)

	Componente				
	1	2	3	4	5
DJ_T	,335	-,611	-,067	,460	,109
V_T	-,107	,223	,762	,405	,163
DV_T	-,499	-,221	,086	,297	-,147
Vx_T	-,074	,160	,599	,417	,511
T_T	,972	-,051	-,015	-,008	,021
Tn_T	,956	-,097	-,063	-,012	,027
Tx_T	,955	-,032	,000	-,005	-,003
Hf_T	-,612	,197	,219	-,109	-,371
HR_T	,027	-,231	-,739	-,357	,120
Rad_T	,682	,476	,252	-,248	-,107
P_T	-,188	-,310	-,390	,083	,554
ETo_T	,842	,351	,315	-,094	,011
VolExp_T	-,071	,723	,060	-,475	,209
Qent_T	-,805	-,169	-,044	,124	,006
Qsal_T	,762	-,356	-,015	,248	-,125
Filt_T	-,359	-,098	,086	,138	-,247
Profundidad_T	-,025	-,426	,430	-,638	,210
Ta_T	,671	,165	-,328	,503	-,121
Odis_T	-,502	,647	-,342	,329	-,013
pH_T	-,087	,591	-,427	,082	,363
Con_T	-,215	-,733	,269	-,209	,104

a 5 componentes extraídos

**Correlaciones reproducidas**

		DJ_T	V_T	DV_T	Vx_T	T_T	Tn_T	Tx_T	Hf_T	HR_T	Rad_T	P_T
Correlación reproducida	DJ_T	0,714 <sup>b</sup>	-0,020	0,083	0,085	0,356	0,381	0,337	-0,431	0,049	-0,206	0,251
	V_T	-0,020	0,833 <sup>b</sup>	0,166	0,752	-0,127	-0,172	-0,112	0,171	-0,743	0,107	-0,222
	DV_T	0,083	0,166	0,415 <sup>b</sup>	0,102	-0,481	-0,469	-0,471	0,303	-0,150	-0,482	0,072
	Vx_T	0,085	0,752	0,102	0,825 <sup>b</sup>	-0,081	-0,115	-0,079	-0,028	-0,569	0,019	0,049
	T_T	0,356	-0,127	-0,481	-0,081	0,947 <sup>b</sup>	0,936	0,930	-0,615	0,054	0,634	-0,150
	Tn_T	0,381	-0,172	-0,469	-0,115	0,936	0,929 <sup>b</sup>	0,917	-0,626	0,102	0,590	-0,112
	Tx_T	0,337	-0,112	-0,471	-0,079	0,930	0,917	0,914 <sup>b</sup>	-0,589	0,035	0,637	-0,171
	Hf_T	-0,431	0,171	0,303	-0,028	-0,615	-0,626	-0,589	0,611 <sup>b</sup>	-0,229	-0,201	-0,246
	HR_T	0,049	-0,743	-0,150	-0,569	0,054	0,102	0,035	-0,229	0,742 <sup>b</sup>	-0,202	0,391
	Rad_T	-0,206	0,107	-0,482	0,019	0,634	0,590	0,637	-0,201	-0,202	0,828 <sup>b</sup>	-0,454
	P_T	0,251	-0,222	0,072	0,049	-0,150	-0,112	-0,171	-0,246	0,391	-0,454	0,597 <sup>b</sup>
	ETo_T	0,004	0,192	-0,500	0,149	0,796	0,753	0,793	-0,371	-0,257	0,843	-0,391
	VolExp_T	-0,665	0,056	-0,291	0,065	-0,099	-0,130	-0,090	0,173	-0,019	0,406	-0,157
	Qent_T	-0,105	0,066	0,472	0,061	-0,774	-0,753	-0,765	0,434	0,006	-0,672	0,234
	Qsal_T	0,574	-0,093	-0,211	-0,083	0,754	0,758	0,738	-0,520	0,011	0,298	-0,076
	Filt_T	-0,030	0,098	0,286	-0,006	-0,351	-0,347	-0,340	0,296	-0,129	-0,277	-0,061
	Profundidad_T	-0,048	0,011	-0,078	0,033	0,001	0,004	-0,007	0,017	0,033	0,024	0,032
	Ta_T	0,364	-0,101	-0,232	-0,072	0,642	0,637	0,634	-0,460	0,029	0,342	-0,075
	Odis_T	-0,390	0,069	0,178	0,067	-0,519	-0,525	-0,502	0,328	-0,030	-0,200	0,047
	pH_T	-0,285	-0,091	-0,152	0,065	-0,102	-0,105	-0,104	-0,067	0,191	0,056	0,208
Con_T	0,273	-0,003	0,215	0,026	-0,172	-0,147	-0,181	0,031	0,052	-0,388	0,203	
Residual(a)	DJ_T		-0,017	-0,072	-0,051	-0,032	-0,041	-0,035	-0,008	-0,020	0,058	-0,102
	V_T	-0,017		-0,048	-0,060	-0,015	0,011	-0,032	0,007	0,109	-0,015	-0,024
	DV_T	-0,072	-0,048		-0,083	0,027	0,004	0,066	-0,097	-0,055	0,097	0,070
	Vx_T	-0,051	-0,060	-0,083		0,024	0,027	0,017	0,071	0,111	-0,052	-0,048
	T_T	-0,032	-0,015	0,027	0,024		0,044	0,050	-0,006	0,003	-0,029	0,001
	Tn_T	-0,041	0,011	0,004	0,027	0,044		0,019	0,011	0,025	-0,048	0,002
	Tx_T	-0,035	-0,032	0,066	0,017	0,050	0,019		-0,028	-0,013	-0,009	-0,005
	Hf_T	-0,008	0,007	-0,097	0,071	-0,006	0,011	-0,028		0,044	-0,051	0,190
	HR_T	-0,020	0,109	-0,055	0,111	0,003	0,025	-0,013	0,044		-0,013	-0,103
	Rad_T	0,058	-0,015	0,097	-0,052	-0,029	-0,048	-0,009	-0,051	-0,013		0,049
	P_T	-0,102	-0,024	0,070	-0,048	0,001	0,002	-0,005	0,190	-0,103	0,049	
	ETo_T	0,009	-0,023	0,067	-0,022	0,001	-0,011	0,009	0,005	-0,020	0,059	0,074
	VolExp_T	0,103	-0,023	0,024	-0,012	0,013	0,016	-0,003	-0,034	-0,014	-0,004	0,008
	Qent_T	0,064	-0,016	-0,075	0,015	0,050	0,033	0,056	-0,116	0,020	0,040	-0,131
	Qsal_T	-0,031	0,019	-0,034	0,008	-0,035	-0,024	-0,040	0,089	0,006	-0,029	0,039
	Filt_T	0,011	-0,046	-0,089	0,048	0,073	0,057	0,074	-0,154	0,020	0,036	0,049
	Profundidad_T	0,029	-0,007	0,066	-0,028	0,002	0,000	0,009	-0,007	-0,032	-0,009	-0,041
	Ta_T	-0,017	0,010	-0,026	0,008	-0,011	-0,001	-0,025	0,049	0,015	0,007	0,013
	Odis_T	0,004	-0,008	-0,013	-0,009	0,004	0,002	0,008	-0,026	-0,015	0,005	-0,019
	pH_T	0,056	-0,005	0,103	-0,077	-0,002	-0,024	0,028	-0,076	-0,059	0,039	-0,146
Con_T	-0,007	-0,003	-0,002	-0,035	-0,008	-0,018	0,004	-0,065	-0,020	0,063	-0,069	

Método de extracción: Análisis de Componentes principales.

a Los residuos se calculan entre las correlaciones observadas y reproducidas.

Hay 55 (26,0%) residuales no redundantes con valores absolutos mayores que 0,05.

b Comunalidades reproducidas.

**Correlaciones reproducidas (continuación)**

		ETo_T	VolExp_T	Qent_T	Qsal_T	Filt_T	Profundidad_T	Ta_T	Odis_T	pH_T	Con_T
Correlación reproducida	DJ_T	0,004	-0,665	-0,105	0,574	-0,030	-0,048	0,364	-0,390	-0,285	0,273
	V_T	0,192	0,056	0,066	-0,093	0,098	0,011	-0,101	0,069	-0,091	-0,003
	DV_T	-0,500	-0,291	0,472	-0,211	0,286	-0,078	-0,232	0,178	-0,152	0,215
	Vx_T	0,149	0,065	0,061	-0,083	-0,006	0,033	-0,072	0,067	0,065	0,026
	T_T	0,796	-0,099	-0,774	0,754	-0,351	0,001	0,642	-0,519	-0,102	-0,172
	Tn_T	0,753	-0,130	-0,753	0,758	-0,347	0,004	0,637	-0,525	-0,105	-0,147
	Tx_T	0,793	-0,090	-0,765	0,738	-0,340	-0,007	0,634	-0,502	-0,104	-0,181
	Hf_T	-0,371	0,173	0,434	-0,520	0,296	0,017	-0,460	0,328	-0,067	0,031
	HR_T	-0,257	-0,019	0,006	0,011	-0,129	0,033	0,029	-0,030	0,191	0,052
	Rad_T	0,843	0,406	-0,672	0,298	-0,277	0,024	0,342	-0,200	0,056	-0,388
	P_T	-0,391	-0,157	0,234	-0,076	-0,061	0,032	-0,075	0,047	0,208	0,203
	ETo_T	0,940 <sup>b</sup>	0,260	-0,763	0,487	-0,325	0,028	0,470	-0,334	-0,004	-0,333
	VolExp_T	0,260	0,800 <sup>b</sup>	-0,125	-0,456	-0,157	0,067	-0,212	0,324	0,445	-0,377
	Qent_T	-0,763	-0,125	0,695 <sup>b</sup>	-0,523	0,318	-0,005	-0,492	0,351	0,001	0,260
	Qsal_T	0,487	-0,456	-0,523	0,785 <sup>b</sup>	-0,175	-0,058	0,597	-0,524	-0,296	0,028
	Filt_T	-0,325	-0,157	0,318	-0,175	0,226 <sup>b</sup>	-0,053	-0,186	0,136	-0,141	0,118
	Profundidad_T	0,028	0,067	-0,005	-0,058	-0,053	0,818 <sup>b</sup>	-0,575	-0,623	-0,409	0,588
	Ta_T	0,470	-0,212	-0,492	0,597	-0,186	-0,575	0,853 <sup>b</sup>	0,049	0,176	-0,472
	Odis_T	-0,334	0,324	0,351	-0,524	0,136	-0,623	0,049	0,895 <sup>b</sup>	0,594	-0,528
	pH_T	-0,004	0,445	0,001	-0,296	-0,141	-0,409	0,176	0,594	0,678 <sup>b</sup>	-0,509
Con_T	-0,333	-0,377	0,260	0,028	0,118	0,588	-0,472	-0,528	-0,509	0,711 <sup>b</sup>	
Residual(a)	DJ_T	0,009	0,103	0,064	-0,031	0,011	0,029	-0,017	0,004	0,056	-0,007
	V_T	-0,023	-0,023	-0,016	0,019	-0,046	-0,007	0,010	-0,008	-0,005	-0,003
	DV_T	0,067	0,024	-0,075	-0,034	-0,089	0,066	-0,026	-0,013	0,103	-0,002
	Vx_T	-0,022	-0,012	0,015	0,008	0,048	-0,028	0,008	-0,009	-0,077	-0,035
	T_T	0,001	0,013	0,050	-0,035	0,073	0,002	-0,011	0,004	-0,002	-0,008
	Tn_T	-0,011	0,016	0,033	-0,024	0,057	0,000	-0,001	0,002	-0,024	-0,018
	Tx_T	0,009	-0,003	0,056	-0,040	0,074	0,009	-0,025	0,008	0,028	0,004
	Hf_T	0,005	-0,034	-0,116	0,089	-0,154	-0,007	0,049	-0,026	-0,076	-0,065
	HR_T	-0,020	-0,014	0,020	0,006	0,020	-0,032	0,015	-0,015	-0,059	-0,020
	Rad_T	0,059	-0,004	0,040	-0,029	0,036	-0,009	0,007	0,005	0,039	0,063
	P_T	0,074	0,008	-0,131	0,039	0,049	-0,041	0,013	-0,019	-0,146	-0,069
	ETo_T		0,005	0,004	-0,027	0,049	-0,013	0,001	0,000	0,003	0,025
	VolExp_T	0,005		0,055	0,011	0,102	-0,036	0,019	-0,002	-0,063	0,025
	Qent_T	0,004	0,055		-0,050	0,084	0,005	0,005	0,001	0,017	0,013
	Qsal_T	-0,027	0,011	-0,050		-0,041	0,040	-0,034	0,016	0,022	-0,050
	Filt_T	0,049	0,102	0,084	-0,041		0,048	-0,016	-0,016	0,093	0,003
	Profundidad_T	-0,013	-0,036	0,005	0,040	0,048		0,029	0,034	0,119	-0,075
	Ta_T	0,001	0,019	0,005	-0,034	-0,016	0,029		0,003	-0,018	0,077
	Odis_T	0,000	-0,002	0,001	0,016	-0,016	0,034	0,003		-0,016	0,058
	pH_T	0,003	-0,063	0,017	0,022	0,093	0,119	-0,018	-0,016		0,075
Con_T	0,025	0,025	0,013	-0,050	0,003	-0,075	0,077	0,058	0,075		

Método de extracción: Análisis de Componentes principales.

a Los residuos se calculan entre las correlaciones observadas y reproducidas.

Hay 55 (26,0%) residuales no redundantes con valores absolutos mayores que 0,05.

b Comunalidades reproducidas.

b) Extracción de los factores por el Método de Componentes principales, Rotación Varimax

**Matriz de componentes rotados(a)**

	Componente				
	1	2	3	4	5
DJ_T	,172	,780	-,075	,056	,258
V_T	-,077	-,019	-,011	,894	-,164
DV_T	-,578	,246	,003	,129	-,066
Vx_T	-,015	-,025	-,009	,880	,221
T_T	,934	,264	-,001	-,061	-,030
Tn_T	,912	,292	-,010	-,107	,004
Tx_T	,918	,253	,008	-,052	-,059
Hf_T	-,586	-,288	-,021	,057	-,425
HR_T	,031	-,030	-,015	-,752	,417
Rad_T	,765	-,317	,062	,115	-,354
P_T	-,181	,127	-,032	-,127	,728
ETo_T	,896	-,125	,035	,244	-,248
VolExp_T	,145	-,874	,107	,056	-,013
Qent_T	-,824	,002	-,036	,024	,119
Qsal_T	,626	,621	-,025	-,049	-,066
Filt_T	-,413	,117	,001	,039	-,200
Profundidad_T	,052	-,176	-,884	,012	,051
Ta_T	,568	,407	,601	-,051	-,014
Odis_T	-,450	-,351	,751	,051	,057
pH_T	,027	-,442	,574	-,030	,389
Con_T	-,265	,296	-,733	-,006	,127

Método de rotación: Normalización Varimax con Kaiser

c) Extracción de los factores por el Método de la Máxima verosimilitud, incluyendo una rotación con el Método Varimax:

**Comunalidades(a)**

	Inicial	Extracción
DJ_T	,604	,513
V_T	,695	,514
DV_T	,396	,350
Vx_T	,611	,254
T_T	,994	,995
Tn_T	,981	,993
Tx_T	,980	,999
Hf_T	,601	,440
HR_T	,627	,586
Rad_T	,914	,878
P_T	,341	,144
ETo_T	,967	1,000
VolExp_T	,682	,700
Qent_T	,739	,636
Qsal_T	,742	,694
Filt_T	,186	,106
Profundidad_T	,664	,692
Ta_T	,785	,782
Odis_T	,826	,894
pH_T	,570	,471
Con_T	,583	,569

Método de extracción: Máxima verosimilitud.

a Se han encontrado una o más estimaciones de comunalidad mayores que 1 durante las iteraciones. La solución resultante deberá ser interpretada con precaución.

**Varianza total explicada (para 14 iteraciones)**

Factor	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	6,853	32,635	32,635	5,586	26,601	26,601
2	3,250	15,476	48,111	3,119	14,851	41,452
3	2,541	12,100	60,211	,451	2,149	43,600
4	1,995	9,500	69,711	2,401	11,432	55,033
5	1,115	5,311	75,022	1,655	7,882	62,915
6	,975	4,642	79,664			
7	,797	3,795	83,459			
8	,713	3,394	86,853			
9	,440	2,097	88,950			
10	,434	2,066	91,016			
11	,360	1,716	92,732			
12	,298	1,420	94,153			
13	,269	1,282	95,435			
14	,226	1,076	96,511			
15	,202	,964	97,475			
16	,185	,882	98,357			
17	,159	,756	99,114			
18	,116	,553	99,667			
19	,041	,197	99,864			
20	,024	,115	99,979			
21	,004	,021	100,000			

**Matriz factorial(a)**

	Factor				
	1	2	3	4	5
DJ_T	,017	,506	,143	-,166	,456
V_T	,164	-,473	,001	-,133	,498
DV_T	-,435	-,116	-,245	-,115	,272
Vx_T	,124	-,280	-,027	-,076	,395
T_T	,804	,577	,125	,014	-,001
Tn_T	,748	,594	,283	,031	-,009
Tx_T	,808	,585	-,056	-,002	-,001
Hf_T	-,372	-,545	-,040	-,050	,020
HR_T	-,272	,434	,226	,155	-,498
Rad_T	,900	-,179	-,096	,039	-,161
P_T	-,316	,149	,144	,026	-,019
ETo_T	1,000	-,010	,000	,000	,000
VolExp_T	,259	-,519	-,038	,233	-,555
Qent_T	-,761	-,175	-,157	-,036	,030
Qsal_T	,466	,569	,189	-,092	,330
Filt_T	-,277	-,083	-,113	-,055	,085
Profundidad_T	,015	-,015	,073	-,765	-,316
Ta_T	,476	,397	,113	,530	,323
Odis_T	-,338	-,394	-,216	,760	,001
pH_T	-,003	-,144	-,216	,585	-,248
Con_T	-,307	,123	,047	-,673	,064

Método de extracción: Máxima verosimilitud.

a Se han intentado extraer 5 factores. Convergencia=,019.

**Matriz de factores rotados(a)**

	Factor				
	1	2	3	4	5
DJ_T	,150	,690	-,107	,023	-,028
V_T	-,073	,004	-,016	,708	-,095
DV_T	-,520	,171	-,022	,155	,157
Vx_T	-,035	,064	,017	,500	-,032
T_T	,952	,262	-,018	-,093	,101
Tn_T	,939	,294	-,026	-,148	-,056
Tx_T	,924	,249	-,008	-,073	,279
Hf_T	-,535	-,292	-,024	,231	-,120
HR_T	,005	,011	,006	-,751	-,147
Rad_T	,765	-,410	,029	,311	,165
P_T	-,208	,159	-,005	-,233	-,145
ETo_T	,911	-,187	,005	,349	,118
VolExp_T	,133	-,813	,139	-,012	-,045
Qent_T	-,789	,024	-,005	-,110	,024
Qsal_T	,609	,563	-,066	,034	-,007
Filt_T	-,315	,045	-,020	,036	,059
Profundidad_T	,016	-,145	-,814	-,075	,038
Ta_T	,582	,367	,552	,037	-,046
Odis_T	-,439	-,309	,778	,014	-,012
pH_T	-,030	-,353	,553	-,163	,117
Con_T	-,276	,257	-,652	-,023	,024

Método de extracción: Máxima verosimilitud.

Método de rotación: Normalización Varimax con Kaiser.

**ANEXO 16. ANÁLISIS FACTORIAL DEL PERIODO DE CALIBRACIÓN, SIN DATOS ANÓMALOS Y VARIABLES TRANSFORMADAS**

## PERIODO DE CALIBRACIÓN

### Estadísticos descriptivos

	Media	Desviación típica	N del análisis
DJ_T	,169	1,189	4401
V_T	-,352	,461	4401
DV_T	,750	,672	4401
Vx_T	-,411	,609	4401
T_T	,063	,798	4401
Tn_T	,134	,824	4401
Tx_T	,143	,841	4401
Hf_T	-1,807	,813	4401
HR_T	,330	,671	4401
Rad_T	-,100	,782	4401
P_T	-2,021	,497	4401
ETo_T	-,624	,810	4401
VolExp_T	-,298	1,151	4401
Qent_T	-1,245	,658	4401
Qsal_T	,272	,844	4401
Filt_T	2,035	,555	4401
Profundidad_T	-,691	,909	4401
Ta_T	-,619	,978	4401
Odis_T	-,008	,964	4401
pH_T	,562	,384	4401
Con_T	-,114	,873	4401

### KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,784
	Chi-cuadrado aproximado	106403,390
Prueba de esfericidad de Bartlett	gl	210
	Sig.	,000



**Matriz de correlaciones(a)**

	DJ_T	V_T	DV_T	Vx_T	T_T	Tn_T	Tx_T	Hf_T	HR_T	Rad_T	P_T	
Correlación	DJ_T	1,000	-0,061	-0,071	0,070	0,307	0,320	0,255	-0,484	0,061	-0,167	0,189
	V_T	-0,061	1,000	0,144	0,651	-0,170	-0,183	-0,183	0,170	-0,641	0,139	-0,295
	DV_T	-0,071	0,144	1,000	0,105	-0,502	-0,497	-0,489	0,208	-0,234	-0,433	0,160
	Vx_T	0,070	0,651	0,105	1,000	-0,031	-0,070	-0,050	0,006	-0,501	0,022	-0,028
	T_T	0,307	-0,170	-0,502	-0,031	1,000	0,982	0,983	-0,610	0,099	0,661	-0,122
	Tn_T	0,320	-0,183	-0,497	-0,070	0,982	1,000	0,945	-0,610	0,146	0,620	-0,080
	Tx_T	0,255	-0,183	-0,489	-0,050	0,983	0,945	1,000	-0,572	0,082	0,691	-0,172
	Hf_T	-0,484	0,170	0,208	0,006	-0,610	-0,610	-0,572	1,000	-0,245	-0,128	-0,089
	HR_T	0,061	-0,641	-0,234	-0,501	0,099	0,146	0,082	-0,245	1,000	-0,209	0,319
	Rad_T	-0,167	0,139	-0,433	0,022	0,661	0,620	0,691	-0,128	-0,209	1,000	-0,473
	P_T	0,189	-0,295	0,160	-0,028	-0,122	-0,080	-0,172	-0,089	0,319	-0,473	1,000
	ETo_T	0,022	0,218	-0,426	0,199	0,806	0,771	0,811	-0,266	-0,298	0,907	-0,328
	VolExp_T	-0,553	0,097	-0,359	0,075	0,213	0,194	0,227	0,275	-0,050	0,568	-0,241
	Qent_T	-0,005	-0,018	0,334	0,060	-0,662	-0,670	-0,657	0,179	0,094	-0,588	0,163
	Qsal_T	0,557	-0,134	-0,194	-0,089	0,654	0,660	0,636	-0,446	0,038	0,241	-0,012
	Filt_T	-0,032	-0,034	0,195	0,015	-0,307	-0,308	-0,312	0,100	-0,048	-0,292	0,065
	Profundidad_T	-0,031	0,010	-0,038	0,010	0,029	0,029	0,028	0,004	0,003	0,060	-0,012
	Ta_T	0,413	-0,133	-0,350	-0,059	0,725	0,730	0,699	-0,469	0,058	0,435	-0,063
	Odis_T	-0,353	0,092	0,293	0,054	-0,537	-0,541	-0,514	0,343	-0,062	-0,353	0,034
	pH_T	-0,113	-0,072	0,065	-0,018	-0,077	-0,091	-0,061	-0,125	0,167	-0,166	0,071
Con_T	0,225	-0,040	0,133	-0,013	-0,132	-0,123	-0,145	-0,120	0,065	-0,236	0,150	
Sig. (Unilateral)	DJ_T		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	V_T	0,000		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	DV_T	0,000	0,000		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	Vx_T	0,000	0,000	0,000		0,018	0,000	0,000	0,336	0,000	0,075	0,031
	T_T	0,000	0,000	0,000	0,018		0,000	0,000	0,000	0,000	0,000	0,000
	Tn_T	0,000	0,000	0,000	0,000	0,000		0,000	0,000	0,000	0,000	0,000
	Tx_T	0,000	0,000	0,000	0,000	0,000	0,000		0,000	0,000	0,000	0,000
	Hf_T	0,000	0,000	0,000	0,336	0,000	0,000	0,000		0,000	0,000	0,000
	HR_T	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000		0,000	0,000
	Rad_T	0,000	0,000	0,000	0,075	0,000	0,000	0,000	0,000	0,000		0,000
	P_T	0,000	0,000	0,000	0,031	0,000	0,000	0,000	0,000	0,000	0,000	
	ETo_T	0,073	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	VolExp_T	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	Qent_T	0,358	0,112	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	Qsal_T	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,006	0,000	0,222
	Filt_T	0,018	0,012	0,000	0,158	0,000	0,000	0,000	0,000	0,001	0,000	0,000
	Profundidad_T	0,021	0,260	0,005	0,244	0,027	0,028	0,032	0,401	0,414	0,000	0,215
	Ta_T	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	Odis_T	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,013
	pH_T	0,000	0,000	0,000	0,114	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Con_T	0,000	0,004	0,000	0,189	0,000	0,000	0,000	0,000	0,000	0,000	0,000	

a Determinante = 3,01E-011

**Matriz de correlaciones (continuación)**

		ETo_T	VolExp_T	Qent_T	Qsal_T	Filt_T	Profundidad_T	Ta_T	Odis_T	pH_T	Con_T
Correlación	DJ_T	0,022	-0,553	-0,005	0,557	-0,032	-0,031	0,413	-0,353	-0,113	0,225
	V_T	0,218	0,097	-0,018	-0,134	-0,034	0,010	-0,133	0,092	-0,072	-0,040
	DV_T	-0,426	-0,359	0,334	-0,194	0,195	-0,038	-0,350	0,293	0,065	0,133
	Vx_T	0,199	0,075	0,060	-0,089	0,015	0,010	-0,059	0,054	-0,018	-0,013
	T_T	0,806	0,213	-0,662	0,654	-0,307	0,029	0,725	-0,537	-0,077	-0,132
	Tn_T	0,771	0,194	-0,670	0,660	-0,308	0,029	0,730	-0,541	-0,091	-0,123
	Tx_T	0,811	0,227	-0,657	0,636	-0,312	0,028	0,699	-0,514	-0,061	-0,145
	Hf_T	-0,266	0,275	0,179	-0,446	0,100	0,004	-0,469	0,343	-0,125	-0,120
	HR_T	-0,298	-0,050	0,094	0,038	-0,048	0,003	0,058	-0,062	0,167	0,065
	Rad_T	0,907	0,568	-0,588	0,241	-0,292	0,060	0,435	-0,353	-0,166	-0,236
	P_T	-0,328	-0,241	0,163	-0,012	0,065	-0,012	-0,063	0,034	0,071	0,150
	ETo_T	1,000	0,472	-0,715	0,416	-0,303	0,055	0,565	-0,443	-0,196	-0,212
	VolExp_T	0,472	1,000	-0,230	-0,285	-0,137	0,069	-0,017	-0,002	0,010	-0,391
	Qent_T	-0,715	-0,230	1,000	-0,556	0,318	-0,023	-0,524	0,375	0,187	0,181
	Qsal_T	0,416	-0,285	-0,556	1,000	-0,198	-0,029	0,621	-0,441	-0,118	-0,007
	Filt_T	-0,303	-0,137	0,318	-0,198	1,000	-0,013	-0,237	0,159	0,027	0,085
	Profundidad_T	0,055	0,069	-0,023	-0,029	-0,013	1,000	-0,389	-0,632	-0,270	0,627
	Ta_T	0,565	-0,017	-0,524	0,621	-0,237	-0,389	1,000	-0,080	0,115	-0,484
	Odis_T	-0,443	-0,002	0,375	-0,441	0,159	-0,632	-0,080	1,000	0,566	-0,595
	pH_T	-0,196	0,010	0,187	-0,118	0,027	-0,270	0,115	0,566	1,000	-0,431
Con_T	-0,212	-0,391	0,181	-0,007	0,085	0,627	-0,484	-0,595	-0,431	1,000	
Sig. (Unilateral)	DJ_T	0,073	0,000	0,358	0,000	0,018	0,021	0,000	0,000	0,000	0,000
	V_T	0,000	0,000	0,112	0,000	0,012	0,260	0,000	0,000	0,000	0,004
	DV_T	0,000	0,000	0,000	0,000	0,000	0,005	0,000	0,000	0,000	0,000
	Vx_T	0,000	0,000	0,000	0,000	0,158	0,244	0,000	0,000	0,114	0,189
	T_T	0,000	0,000	0,000	0,000	0,000	0,027	0,000	0,000	0,000	0,000
	Tn_T	0,000	0,000	0,000	0,000	0,000	0,028	0,000	0,000	0,000	0,000
	Tx_T	0,000	0,000	0,000	0,000	0,000	0,032	0,000	0,000	0,000	0,000
	Hf_T	0,000	0,000	0,000	0,000	0,000	0,401	0,000	0,000	0,000	0,000
	HR_T	0,000	0,000	0,000	0,006	0,001	0,414	0,000	0,000	0,000	0,000
	Rad_T	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	P_T	0,000	0,000	0,000	0,222	0,000	0,215	0,000	0,013	0,000	0,000
	ETo_T		0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	VolExp_T	0,000		0,000	0,000	0,000	0,000	0,124	0,453	0,262	0,000
	Qent_T	0,000	0,000		0,000	0,000	0,061	0,000	0,000	0,000	0,000
	Qsal_T	0,000	0,000	0,000		0,000	0,028	0,000	0,000	0,000	0,315
	Filt_T	0,000	0,000	0,000	0,000		0,196	0,000	0,000	0,039	0,000
	Profundidad_T	0,000	0,000	0,061	0,028	0,196		0,000	0,000	0,000	0,000
	Ta_T	0,000	0,124	0,000	0,000	0,000	0,000		0,000	0,000	0,000
	Odis_T	0,000	0,453	0,000	0,000	0,000	0,000	0,000		0,000	0,000
	pH_T	0,000	0,262	0,000	0,000	0,039	0,000	0,000	0,000		0,000
Con_T	0,000	0,000	0,000	0,315	0,000	0,000	0,000	0,000	0,000		

### Comunalidades

	Inicial	Extracción
DJ_T	1,000	,759
V_T	1,000	,804
DV_T	1,000	,482
Vx_T	1,000	,817
T_T	1,000	,964
Tn_T	1,000	,946
Tx_T	1,000	,932
Hf_T	1,000	,740
HR_T	1,000	,786
Rad_T	1,000	,860
P_T	1,000	,400
ETo_T	1,000	,946
VolExp_T	1,000	,846
Qent_T	1,000	,681
Qsal_T	1,000	,793
Filt_T	1,000	,155
Profundidad_T	1,000	,757
Ta_T	1,000	,874
Odis_T	1,000	,936
pH_T	1,000	,664
Con_T	1,000	,872

Método de extracción: Análisis de Componentes principales

### Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	6,985	33,263	33,263	6,985	33,263	33,263
2	3,199	15,235	48,498	3,199	15,235	48,498
3	2,701	12,861	61,359	2,701	12,861	61,359
4	2,076	9,888	71,247	2,076	9,888	71,247
5	1,051	5,006	76,252	1,051	5,006	76,252
6	,917	4,367	80,619			
7	,817	3,892	84,511			
8	,751	3,577	88,088			
9	,523	2,491	90,579			
10	,404	1,922	92,501			
11	,334	1,591	94,092			
12	,282	1,345	95,438			
13	,224	1,065	96,503			
14	,208	,990	97,492			
15	,176	,840	98,332			
16	,128	,610	98,942			
17	,087	,417	99,358			
18	,074	,351	99,709			
19	,038	,181	99,889			
20	,019	,091	99,980			
21	,004	,020	100,000			

Método de extracción: Análisis de Componentes principales

**Matriz de componentes(a)**

	Componente				
	1	2	3	4	5
DJ_T	,295	-,625	,019	,531	-,008
V_T	-,096	,560	,366	,572	,141
DV_T	-,555	-,065	,064	,349	-,210
Vx_T	-,039	,369	,286	,615	,469
T_T	,971	-,081	-,054	-,004	,113
Tn_T	,958	-,123	-,066	-,020	,095
Tx_T	,958	-,041	-,057	-,033	,084
Hf_T	-,540	,481	,130	-,229	-,384
HR_T	,033	-,550	-,369	-,563	,174
Rad_T	,746	,502	,138	-,175	-,043
P_T	-,210	-,509	-,164	-,002	,266
ETo_T	,865	,407	,169	,051	,031
VolExp_T	,241	,690	,015	-,505	,237
Qent_T	-,752	-,171	-,045	,032	,287
Qsal_T	,686	-,372	-,048	,313	-,290
Filt_T	-,376	-,094	,011	,070	,023
Profundidad_T	,036	-,195	,752	-,314	,233
Ta_T	,748	-,046	-,474	,273	-,116
Odis_T	-,564	,367	-,679	,152	-,011
pH_T	-,154	,087	-,662	,061	,437
Con_T	-,153	-,537	,742	-,070	,068

Método de extracción: Análisis de componentes principales  
a 5 componentes extraídos

**Matriz de componentes rotados(a)**

	Componente				
	1	2	3	4	5
DJ_T	,160	,116	,829	,041	,177
V_T	-,066	-,007	-,082	,879	-,141
DV_T	-,598	-,070	,232	,186	-,176
Vx_T	-,018	,025	,010	,871	,237
T_T	,953	,052	,187	-,054	,122
Tn_T	,935	,056	,209	-,097	,122
Tx_T	,948	,038	,146	-,063	,084
Hf_T	-,474	-,077	-,482	,043	-,524
HR_T	,025	,037	-,007	-,775	,428
Rad_T	,810	,027	-,336	,165	-,249
P_T	-,255	,058	,259	-,250	,449
ETo_T	,900	,046	-,122	,305	-,162
VolExp_T	,388	-,054	-,828	,079	-,009
Qent_T	-,748	,007	-,053	,015	,344
Qsal_T	,583	,007	,650	-,094	-,151
Filt_T	-,387	,007	,042	,019	,056
Profundidad_T	,018	,850	-,166	,023	,078
Ta_T	,720	-,448	,389	-,060	,016
Odis_T	-,490	-,804	-,205	,057	,064
pH_T	-,093	-,585	-,117	-,028	,547
Con_T	-,245	,867	,241	-,033	,041

Método de extracción: Análisis de componentes principales  
Método de rotación: Normalización Varimax con Kaiser

**ANEXO 17. ANÁLISIS FACTORIAL DEL PERIODO DE  
VALIDACIÓN, SIN DATOS ANÓMALOS Y  
VARIABLES TRANSFORMADAS**

## PERIODO DE VALIDACIÓN

### Estadísticos descriptivos

	Media	Desviación típica	N del análisis
DJ_T	,122	,949	4573
V_T	-,592	,474	4573
DV_T	,209	,646	4573
Vx_T	-,406	,533	4573
T_T	,450	,742	4573
Tn_T	,521	,830	4573
Tx_T	,399	,620	4573
Hf_T	-2,006	,599	4573
HR_T	-,007	,706	4573
Rad_T	,602	,766	4573
P_T	-2,024	,522	4573
ETo_T	,187	,694	4573
VolExp_T	,424	1,105	4573
Qent_T	-1,285	,963	4573
Qsal_T	-,024	,859	4573
Filt_T	1,676	,910	4573
Profundidad_T	-,643	,979	4573
Ta_T	-,244	1,244	4573
Odis_T	,019	,960	4573
pH_T	,620	,426	4573
Con_T	,415	,822	4573

### KMO y prueba de Bartlett

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,782
	Chi-cuadrado aproximado	115450,3
Prueba de esfericidad de Bartlett	gl	210
	Sig.	,000

**Matriz de correlaciones(a)**

		DJ_T	V_T	DV_T	Vx_T	T_T	Tn_T	Tx_T	Hf_T	HR_T	Rad_T	P_T
Correlación	DJ_T	1,000	-0,010	0,013	-0,070	0,777	0,770	0,765	-0,510	0,026	0,183	0,017
	V_T	-0,010	1,000	0,029	0,764	-0,038	-0,074	-0,009	0,167	-0,640	0,188	-0,178
	DV_T	0,013	0,029	1,000	-0,086	-0,203	-0,281	-0,140	0,058	-0,175	-0,045	0,067
	Vx_T	-0,070	0,764	-0,086	1,000	-0,113	-0,125	-0,083	0,125	-0,408	-0,098	0,052
	T_T	0,777	-0,038	-0,203	-0,113	1,000	0,962	0,976	-0,657	-0,030	0,433	-0,167
	Tn_T	0,770	-0,074	-0,281	-0,125	0,962	1,000	0,897	-0,638	0,092	0,317	-0,121
	Tx_T	0,765	-0,009	-0,140	-0,083	0,976	0,897	1,000	-0,629	-0,115	0,478	-0,167
	Hf_T	-0,510	0,167	0,058	0,125	-0,657	-0,638	-0,629	1,000	-0,063	-0,342	-0,012
	HR_T	0,026	-0,640	-0,175	-0,408	-0,030	0,092	-0,115	-0,063	1,000	-0,349	0,269
	Rad_T	0,183	0,188	-0,045	-0,098	0,433	0,317	0,478	-0,342	-0,349	1,000	-0,379
	P_T	0,017	-0,178	0,067	0,052	-0,167	-0,121	-0,167	-0,012	0,269	-0,379	1,000
	ETo_T	0,444	0,316	-0,149	0,085	0,732	0,620	0,773	-0,470	-0,442	0,835	-0,346
	VolExp_T	-0,848	0,015	-0,080	0,072	-0,520	-0,541	-0,509	0,132	-0,074	0,147	-0,043
	Qent_T	-0,671	0,053	0,165	0,138	-0,726	-0,716	-0,701	0,495	-0,015	-0,420	-0,087
	Qsal_T	0,749	0,060	-0,219	-0,041	0,788	0,811	0,744	-0,423	-0,021	0,256	-0,069
	Filt_T	-0,138	0,066	0,080	0,064	-0,184	-0,213	-0,165	0,138	-0,131	-0,064	-0,120
	Profundidad_T	-0,025	-0,009	-0,012	-0,004	-0,008	-0,008	-0,008	0,008	0,000	0,011	-0,008
	Ta_T	0,510	-0,012	-0,112	-0,070	0,549	0,553	0,520	-0,356	0,025	0,229	-0,054
	Odis_T	-0,624	0,001	0,041	0,065	-0,533	-0,540	-0,512	0,250	-0,032	-0,083	0,030
	pH_T	-0,387	-0,090	0,109	-0,001	-0,361	-0,395	-0,334	-0,096	0,042	-0,004	0,097
Con_T	0,346	-0,019	0,120	-0,024	0,130	0,146	0,129	0,026	0,057	-0,205	0,045	
Sig. (Unilateral)	DJ_T		0,251	0,183	0,000	0,000	0,000	0,000	0,000	0,037	0,000	0,122
	V_T	0,251		0,026	0,000	0,005	0,000	0,267	0,000	0,000	0,000	0,000
	DV_T	0,183	0,026		0,000	0,000	0,000	0,000	0,000	0,000	0,001	0,000
	Vx_T	0,000	0,000	0,000		0,000	0,000	0,000	0,000	0,000	0,000	0,000
	T_T	0,000	0,005	0,000	0,000		0,000	0,000	0,000	0,020	0,000	0,000
	Tn_T	0,000	0,000	0,000	0,000	0,000		0,000	0,000	0,000	0,000	0,000
	Tx_T	0,000	0,267	0,000	0,000	0,000	0,000		0,000	0,000	0,000	0,000
	Hf_T	0,000	0,000	0,000	0,000	0,000	0,000	0,000		0,000	0,000	0,211
	HR_T	0,037	0,000	0,000	0,000	0,020	0,000	0,000	0,000		0,000	0,000
	Rad_T	0,000	0,000	0,001	0,000	0,000	0,000	0,000	0,000	0,000		0,000
	P_T	0,122	0,000	0,000	0,000	0,000	0,000	0,000	0,211	0,000	0,000	
	ETo_T	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	VolExp_T	0,000	0,148	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,002
	Qent_T	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,161	0,000	0,000
	Qsal_T	0,000	0,000	0,000	0,003	0,000	0,000	0,000	0,000	0,078	0,000	0,000
	Filt_T	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	Profundidad_T	0,049	0,264	0,207	0,391	0,303	0,305	0,297	0,296	0,488	0,222	0,303
	Ta_T	0,000	0,201	0,000	0,000	0,000	0,000	0,000	0,000	0,044	0,000	0,000
	Odis_T	0,000	0,472	0,003	0,000	0,000	0,000	0,000	0,000	0,014	0,000	0,021
	pH_T	0,000	0,000	0,000	0,463	0,000	0,000	0,000	0,000	0,002	0,400	0,000
Con_T	0,000	0,099	0,000	0,055	0,000	0,000	0,000	0,038	0,000	0,000	0,001	

a Determinante = 1,03E-011

**Matriz de correlaciones (continuación)**

		ETo_T	VolExp_T	Qent_T	Qsal_T	Filt_T	Profundidad_T	Ta_T	Odis_T	pH_T	Con_T
Correlación	DJ_T	0,444	-0,848	-0,671	0,749	-0,138	-0,025	0,510	-0,624	-0,387	0,346
	V_T	0,316	0,015	0,053	0,060	0,066	-0,009	-0,012	0,001	-0,090	-0,019
	DV_T	-0,149	-0,080	0,165	-0,219	0,080	-0,012	-0,112	0,041	0,109	0,120
	Vx_T	0,085	0,072	0,138	-0,041	0,064	-0,004	-0,070	0,065	-0,001	-0,024
	T_T	0,732	-0,520	-0,726	0,788	-0,184	-0,008	0,549	-0,533	-0,361	0,130
	Tn_T	0,620	-0,541	-0,716	0,811	-0,213	-0,008	0,553	-0,540	-0,395	0,146
	Tx_T	0,773	-0,509	-0,701	0,744	-0,165	-0,008	0,520	-0,512	-0,334	0,129
	Hf_T	-0,470	0,132	0,495	-0,423	0,138	0,008	-0,356	0,250	-0,096	0,026
	HR_T	-0,442	-0,074	-0,015	-0,021	-0,131	0,000	0,025	-0,032	0,042	0,057
	Rad_T	0,835	0,147	-0,420	0,256	-0,064	0,011	0,229	-0,083	-0,004	-0,205
	P_T	-0,346	-0,043	-0,087	-0,069	-0,120	-0,008	-0,054	0,030	0,097	0,045
	ETo_T	1,000	-0,125	-0,595	0,540	-0,119	0,010	0,375	-0,288	-0,202	-0,081
	VolExp_T	-0,125	1,000	0,390	-0,627	0,064	0,028	-0,369	0,605	0,517	-0,482
	Qent_T	-0,595	0,390	1,000	-0,675	0,485	-0,007	-0,460	0,420	0,225	-0,040
	Qsal_T	0,540	-0,627	-0,675	1,000	-0,201	-0,003	0,504	-0,553	-0,491	0,203
	Filt_T	-0,119	0,064	0,485	-0,201	1,000	-0,012	-0,142	0,104	0,038	0,033
	Profundidad_T	0,010	0,028	-0,007	-0,003	-0,012	1,000	-0,676	-0,540	-0,310	0,634
	Ta_T	0,375	-0,369	-0,460	0,504	-0,142	-0,676	1,000	0,160	0,113	-0,466
	Odis_T	-0,288	0,605	0,420	-0,553	0,104	-0,540	0,160	1,000	0,660	-0,627
	pH_T	-0,202	0,517	0,225	-0,491	0,038	-0,310	0,113	0,660	1,000	-0,571
Con_T	-0,081	-0,482	-0,040	0,203	0,033	0,634	-0,466	-0,627	-0,571	1,000	
Sig. (Unilateral)	DJ_T	0,000	0,000	0,000	0,000	0,000	0,049	0,000	0,000	0,000	0,000
	V_T	0,000	0,148	0,000	0,000	0,000	0,264	0,201	0,472	0,000	0,099
	DV_T	0,000	0,000	0,000	0,000	0,000	0,207	0,000	0,003	0,000	0,000
	Vx_T	0,000	0,000	0,000	0,003	0,000	0,391	0,000	0,000	0,463	0,055
	T_T	0,000	0,000	0,000	0,000	0,000	0,303	0,000	0,000	0,000	0,000
	Tn_T	0,000	0,000	0,000	0,000	0,000	0,305	0,000	0,000	0,000	0,000
	Tx_T	0,000	0,000	0,000	0,000	0,000	0,297	0,000	0,000	0,000	0,000
	Hf_T	0,000	0,000	0,000	0,000	0,000	0,296	0,000	0,000	0,000	0,038
	HR_T	0,000	0,000	0,161	0,078	0,000	0,488	0,044	0,014	0,002	0,000
	Rad_T	0,000	0,000	0,000	0,000	0,000	0,222	0,000	0,000	0,400	0,000
	P_T	0,000	0,002	0,000	0,000	0,000	0,303	0,000	0,021	0,000	0,001
	ETo_T		0,000	0,000	0,000	0,000	0,241	0,000	0,000	0,000	0,000
	VolExp_T	0,000		0,000	0,000	0,000	0,029	0,000	0,000	0,000	0,000
	Qent_T	0,000	0,000		0,000	0,000	0,316	0,000	0,000	0,000	0,003
	Qsal_T	0,000	0,000	0,000		0,000	0,424	0,000	0,000	0,000	0,000
	Filt_T	0,000	0,000	0,000	0,000		0,200	0,000	0,000	0,005	0,012
	Profundidad_T	0,241	0,029	0,316	0,424	0,200		0,000	0,000	0,000	0,000
	Ta_T	0,000	0,000	0,000	0,000	0,000	0,000		0,000	0,000	0,000
	Odis_T	0,000	0,000	0,000	0,000	0,000	0,000	0,000		0,000	0,000
	pH_T	0,000	0,000	0,000	0,000	0,005	0,000	0,000	0,000		0,000
Con_T	0,000	0,000	0,003	0,000	0,012	0,000	0,000	0,000	0,000		



### Comunalidades

	Inicial	Extracción
DJ_T	1,000	,914
V_T	1,000	,896
DV_T	1,000	,891
Vx_T	1,000	,817
T_T	1,000	,940
Tn_T	1,000	,917
Tx_T	1,000	,906
Hf_T	1,000	,606
HR_T	1,000	,760
Rad_T	1,000	,825
P_T	1,000	,665
ETo_T	1,000	,937
VolExp_T	1,000	,909
Qent_T	1,000	,811
Qsal_T	1,000	,807
Filt_T	1,000	,561
Profundidad_T	1,000	,860
Ta_T	1,000	,930
Odis_T	1,000	,848
pH_T	1,000	,710
Con_T	1,000	,838

Método de extracción: Análisis de Componentes principales

### Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	7,417	35,319	35,319	7,417	35,319	35,319
2	3,165	15,073	50,391	3,165	15,073	50,391
3	2,627	12,507	62,899	2,627	12,507	62,899
4	1,706	8,126	71,024	1,706	8,126	71,024
5	1,285	6,117	77,142	1,285	6,117	77,142
6	1,148	5,465	82,607	1,148	5,465	82,607
7	,930	4,427	87,034			
8	,595	2,831	89,865			
9	,376	1,790	91,655			
10	,359	1,710	93,365			
11	,275	1,310	94,675			
12	,245	1,165	95,840			
13	,241	1,148	96,988			
14	,179	,855	97,843			
15	,142	,674	98,517			
16	,132	,628	99,145			
17	,061	,291	99,436			
18	,044	,211	99,647			
19	,037	,178	99,825			
20	,032	,151	99,976			
21	,005	,024	100,000			

Método de extracción: Análisis de Componentes principales.

**Matriz de componentes(a)**

	Componente					
	1	2	3	4	5	6
DJ_T	,871	-,170	-,099	,279	,134	,144
V_T	,011	,208	,817	,389	-,182	,034
DV_T	-,177	-,103	,069	,118	,490	,769
Vx_T	-,106	,119	,624	,493	-,397	-,013
T_T	,960	,097	-,038	-,062	,030	-,049
Tn_T	,938	,031	-,129	-,012	-,040	-,136
Tx_T	,939	,122	,027	-,068	,065	,020
Hf_T	-,616	-,223	,203	,246	,050	-,270
HR_T	-,052	-,291	-,778	-,101	-,121	-,207
Rad_T	,442	,488	,343	-,504	,097	,099
P_T	-,137	-,218	-,384	,303	-,448	,399
ETo_T	,726	,409	,414	-,265	,000	,034
VolExp_T	-,640	,419	,108	-,502	-,245	-,014
Qent_T	-,801	-,110	,140	,072	,308	-,193
Qsal_T	,868	-,061	-,012	,172	-,048	-,134
Filt_T	-,250	-,039	,214	,072	,618	-,254
Profundidad_T	,035	-,665	,337	-,503	-,201	,095
Ta_T	,530	,592	-,340	,401	,128	-,074
Odis_T	-,633	,636	-,179	,096	,021	-,020
pH_T	-,458	,579	-,263	-,091	-,044	,293
Con_T	,223	-,854	,213	-,036	,070	,087

Método de extracción: Análisis de componentes principales.  
a 6 componentes extraídos

**Matriz de componentes rotados(a)**

	Componente					
	1	2	3	4	5	6
DJ_T	,934	,043	-,072	-,018	,086	,162
V_T	-,001	,113	-,008	,937	-,080	,005
DV_T	-,088	-,037	-,035	,012	-,053	,937
Vx_T	-,063	-,133	,018	,877	,105	-,119
T_T	,851	,432	-,009	-,078	,084	-,124
Tn_T	,862	,309	-,013	-,122	,116	-,222
Tx_T	,820	,474	-,012	-,029	,070	-,046
Hf_T	-,451	-,475	-,097	,231	-,324	-,095
HR_T	,038	-,397	,009	-,703	,209	-,252
Rad_T	,134	,892	,059	,078	-,045	,003
P_T	-,025	-,472	,019	-,047	,640	,170
ETo_T	,468	,803	,023	,256	,004	-,083
VolExp_T	-,855	,334	,157	-,011	,108	-,175
Qent_T	-,657	-,363	-,003	,053	-,488	,080
Qsal_T	,860	,150	-,061	,051	,072	-,183
Filt_T	-,129	-,085	,015	,024	-,720	,135
Profundidad_T	-,129	,144	-,899	-,034	,113	-,032
Ta_T	,607	,098	,741	-,023	,027	-,043
Odis_T	-,600	-,049	,695	,021	-,037	,008
pH_T	-,511	,127	,572	-,118	,223	,202
Con_T	,280	-,227	-,825	-,013	-,070	,154

Método de extracción: Análisis de componentes principales  
Método de rotación: Normalización Varimax con Kaiser  
a La rotación ha convergido en 11 iteraciones

**ANEXO 18. ANÁLISIS CLUSTER DEL PERIODO COMPLETO DE ESTUDIO, SIN DATOS ANÓMALOS Y VARIABLES SIN TRANSFORMAR (15 CONGLOMERADOS)**

**Centros de los conglomerados finales**

Variable estandarizada	Conglomerado														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
DJ	0,804	-1,157	0,207	-1,512	0,280	-1,544	1,623	1,638	-0,457	1,299	-1,293	1,589	-0,764	1,316	0,212
V	0,084	1,663	-0,112	1,021	-0,224	-0,011	-1,099	-1,105	0,388	-0,401	1,376	-0,690	-0,937	1,691	-0,105
DV	0,698	0,923	-0,535	1,013	-0,509	0,771	1,209	-0,191	-0,591	0,890	-2,902	-2,707	-0,465	0,934	-0,623
Vx	0,164	1,670	-0,140	0,521	-0,260	-0,302	-0,341	-0,293	0,269	-0,407	2,168	0,163	-0,454	1,701	-0,136
T	-0,386	-0,768	0,897	-1,976	0,967	-1,439	-0,853	-0,876	0,255	-0,519	-1,858	-0,825	-0,309	-0,210	1,063
HR	-0,052	-1,772	0,076	-1,769	0,196	0,353	0,876	1,056	-0,798	0,309	-0,333	1,029	0,968	-1,440	0,124
ETo	-0,694	-0,040	0,820	-1,035	0,763	-1,183	-1,561	-1,575	0,994	-1,073	-1,172	-1,476	-0,115	-0,349	0,976
VolExp	-1,049	0,924	-0,243	0,406	-0,203	0,429	-0,915	-0,775	1,134	-1,289	0,906	-0,787	1,286	-0,857	0,821
Gent	0,377	0,495	-0,613	0,561	-0,595	0,733	6,325	5,426	-0,547	0,245	0,577	2,311	-0,147	0,413	-0,562
Qsal	0,003	-1,263	0,769	-1,356	0,863	-1,291	-0,740	-0,675	-0,297	0,313	-1,079	-0,215	-0,947	0,310	0,492
Profundidad	1,271	0,057	1,119	-0,621	-0,621	0,038	-0,414	1,712	-0,343	-0,513	0,099	-0,336	0,006	-0,370	1,721
Ta	-0,680	-0,775	-0,749	-1,028	1,381	-1,010	-0,538	-0,641	0,296	0,017	-1,107	-0,420	-0,353	0,090	-0,700
Odis	-1,425	0,915	-1,591	1,084	-0,121	0,999	0,431	-1,387	0,651	0,290	1,020	0,437	0,727	0,231	-1,848
pH	-1,283	0,749	-0,929	-2,836	0,085	0,395	0,276	-1,001	1,033	-0,003	-1,001	0,154	1,150	-0,121	-1,539
Con	2,426	-0,252	-0,142	-0,182	-0,653	-0,118	0,991	2,516	-0,653	0,481	-0,283	0,480	-0,454	-0,052	2,474

Distancias entre los centros de los conglomerados finales

Conglomerado	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1		5,996	3,867	5,939	5,413	5,294	7,153	5,627	5,966	3,651	6,746	5,533	5,814	4,582	3,494
2	5,996		5,857	4,243	5,718	3,693	8,290	8,789	3,622	5,473	4,770	6,867	4,705	3,805	6,776
3	3,867	5,857		6,511	3,329	5,528	8,690	7,703	4,180	4,370	6,426	5,824	4,642	4,844	2,987
4	5,939	4,243	6,511		6,789	4,187	8,411	8,597	5,855	5,762	4,954	7,204	6,085	5,231	7,210
5	5,413	5,718	3,329	6,789		5,404	8,511	8,580	2,939	3,764	6,661	5,544	3,952	4,469	5,149
6	5,294	3,693	5,528	4,187	5,404		6,895	7,387	4,271	4,154	4,951	5,464	2,938	5,204	6,522
7	7,153	8,290	8,690	8,411	8,511	6,895		3,823	8,738	6,345	9,006	5,700	7,801	7,557	9,180
8	5,627	8,789	7,703	8,597	8,580	7,387	3,823		9,017	6,644	8,611	5,449	8,040	7,893	7,393
9	5,966	3,622	4,180	5,855	2,939	4,271	8,738	9,017		4,682	5,338	5,910	2,875	4,338	5,565
10	3,651	5,473	4,370	5,762	3,764	4,154	6,345	6,644	4,682		6,612	4,391	4,353	3,612	5,649
11	6,746	4,770	6,426	4,954	6,661	4,951	9,006	8,611	5,338	6,612		5,402	5,445	5,884	7,142
12	5,533	6,867	5,824	7,204	5,544	5,464	5,700	5,449	5,910	4,391	5,402		5,106	5,802	6,692
13	5,814	4,705	4,642	6,085	3,952	2,938	7,801	8,040	2,875	4,353	5,445	5,106		5,705	5,790
14	4,582	3,805	4,844	5,231	4,469	5,204	7,557	7,893	4,338	3,612	5,884	5,802	5,705		6,130
15	3,494	6,776	2,987	7,210	5,149	6,522	9,180	7,393	5,565	5,649	7,142	6,692	5,790	6,130	

**ANOVA**

Variable estandarizada	Conglomerado Media cuadrática	gl	Error Media cuadrática	gl
DJ	558,71	14	0,13	8959
V	268,97	14	0,58	8959
DV	385,54	14	0,40	8959
Vx	237,88	14	0,63	8959
T	522,35	14	0,19	8959
HR	301,93	14	0,53	8959
ETo	487,81	14	0,24	8959
VolExp	418,17	14	0,35	8959
Qent	491,93	14	0,23	8959
Qsal	441,86	14	0,31	8959
Profundidad	319,35	14	0,50	8959
Ta	510,72	14	0,20	8959
Odis	524,58	14	0,18	8959
pH	420,00	14	0,35	8959
Con	509,59	14	0,21	8959

Conglomerado	Número de casos en cada conglomerado
1	553
2	390
3	990
4	133
5	2260
6	1184
7	76
8	27
9	763
10	967
11	67
12	100
13	740
14	418
15	306
Válidos	8974
Perdidos	0

**ANEXO 19. CENTROS DE LOS CONGLOMERADOS,  
CALIBRACIÓN, VARIABLES SIN  
TRANSFORMAR (19 CONGLOMERADOS)**

**Centros de los conglomerados finales**

Variable	Conglomerado																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
DJ	0,688	1,122	0,842	-1,376	0,865	0,861	0,650	1,152	0,000	-1,422	-1,444	-1,485	0,024	1,210	1,148	1,262	1,011	0,020	-0,949
V	1,354	-1,106	-0,038	2,192	-0,400	0,273	-0,572	-0,648	-0,263	0,287	-0,105	-0,219	0,584	1,027	-1,106	4,081	-0,544	-0,547	0,648
DV	0,462	0,906	-2,775	0,710	0,492	0,577	0,262	-3,557	-1,116	0,592	0,396	0,341	-1,603	0,685	-3,609	0,716	0,608	-0,965	0,161
Vx	1,656	-0,326	0,658	2,087	-0,419	-0,240	-0,367	0,197	-0,193	0,355	-0,360	-0,671	0,466	1,360	-0,165	2,499	-0,560	-0,460	1,237
T	0,541	-0,479	0,315	-1,274	-0,063	-0,351	0,722	-0,707	1,349	-1,035	-1,023	-1,009	1,486	-0,427	-0,528	-0,912	-0,489	1,376	-0,362
HR	-1,403	0,874	1,070	-2,216	0,302	-1,100	0,320	0,698	0,252	-0,998	0,347	0,711	-0,523	-0,892	1,330	-1,864	0,715	0,437	0,007
ETo	0,490	-1,089	-0,517	-0,067	-0,521	-0,376	-0,001	-1,076	1,513	-0,524	-0,726	-0,833	1,882	-0,561	-1,138	-0,125	-0,898	1,455	0,413
VolExp	-1,020	-0,764	-1,334	0,926	-1,083	-1,296	-0,924	-0,539	1,225	0,796	0,584	0,529	1,160	-0,067	-0,387	0,059	-1,090	1,165	-0,433
Qent	-0,544	4,841	0,048	0,338	-0,157	-0,151	-0,628	0,469	-0,801	0,174	0,282	0,310	-0,845	0,999	3,309	0,384	0,181	-0,796	0,604
Qsal	1,099	-0,633	1,187	-1,245	0,590	0,712	1,241	-0,263	0,646	-1,303	-1,169	-1,108	0,649	-0,407	-0,464	-0,406	0,076	0,669	-1,876
Profundidad	-0,131	-0,092	-0,122	0,039	1,535	-0,353	-0,432	-0,011	1,740	-0,365	1,564	-0,512	-0,396	-0,006	-0,022	0,045	-0,469	-0,527	0,301



**ANEXO 20. ANÁLISIS CLUSTER, PERIODO DE VALIDACIÓN,  
VARIABLES SIN TRANSFORMAR (19 GRUPOS),  
CENTROS PROCEDENTES DE CALIBRACIÓN**

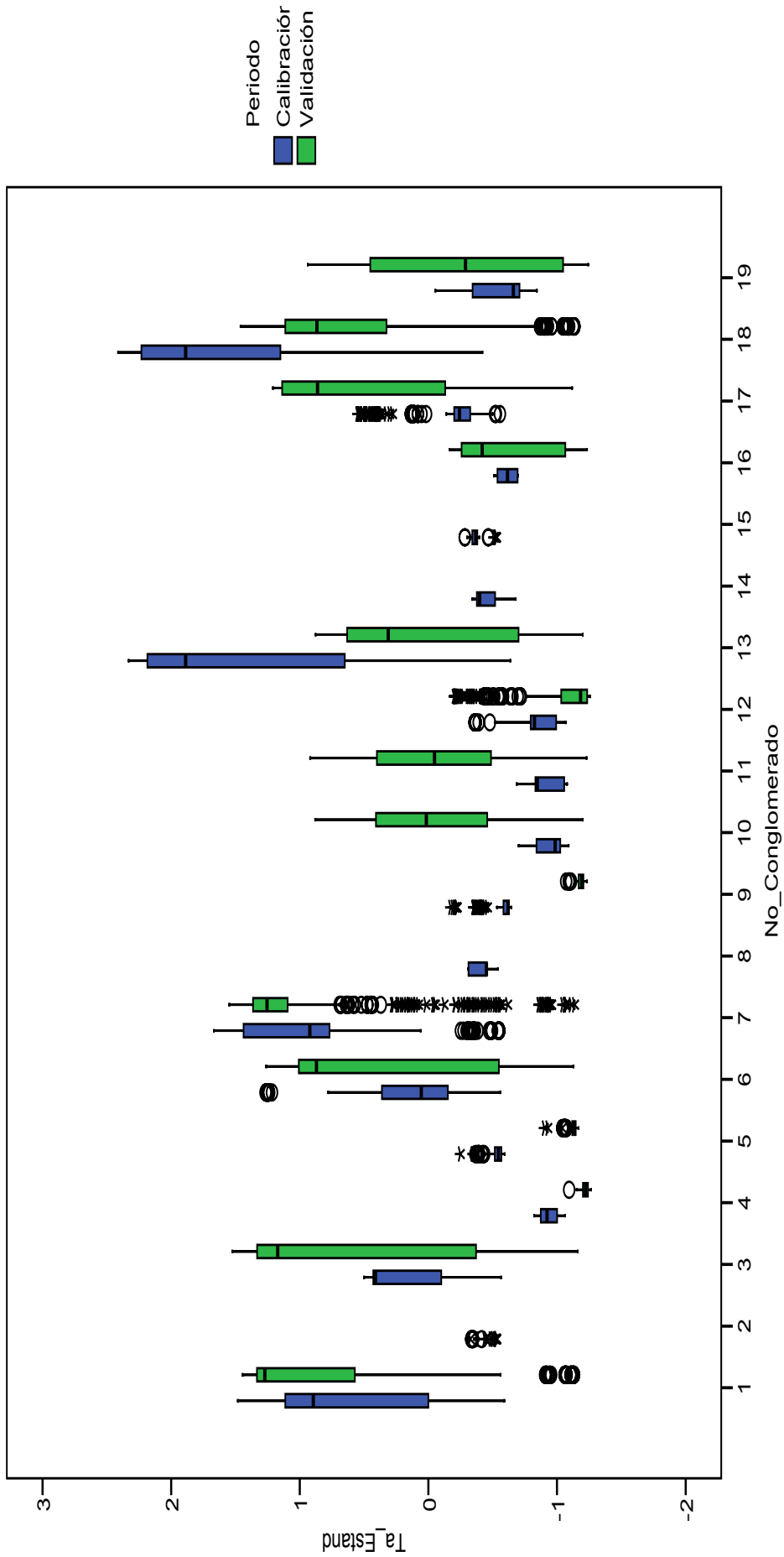
## Centros de los conglomerados finales

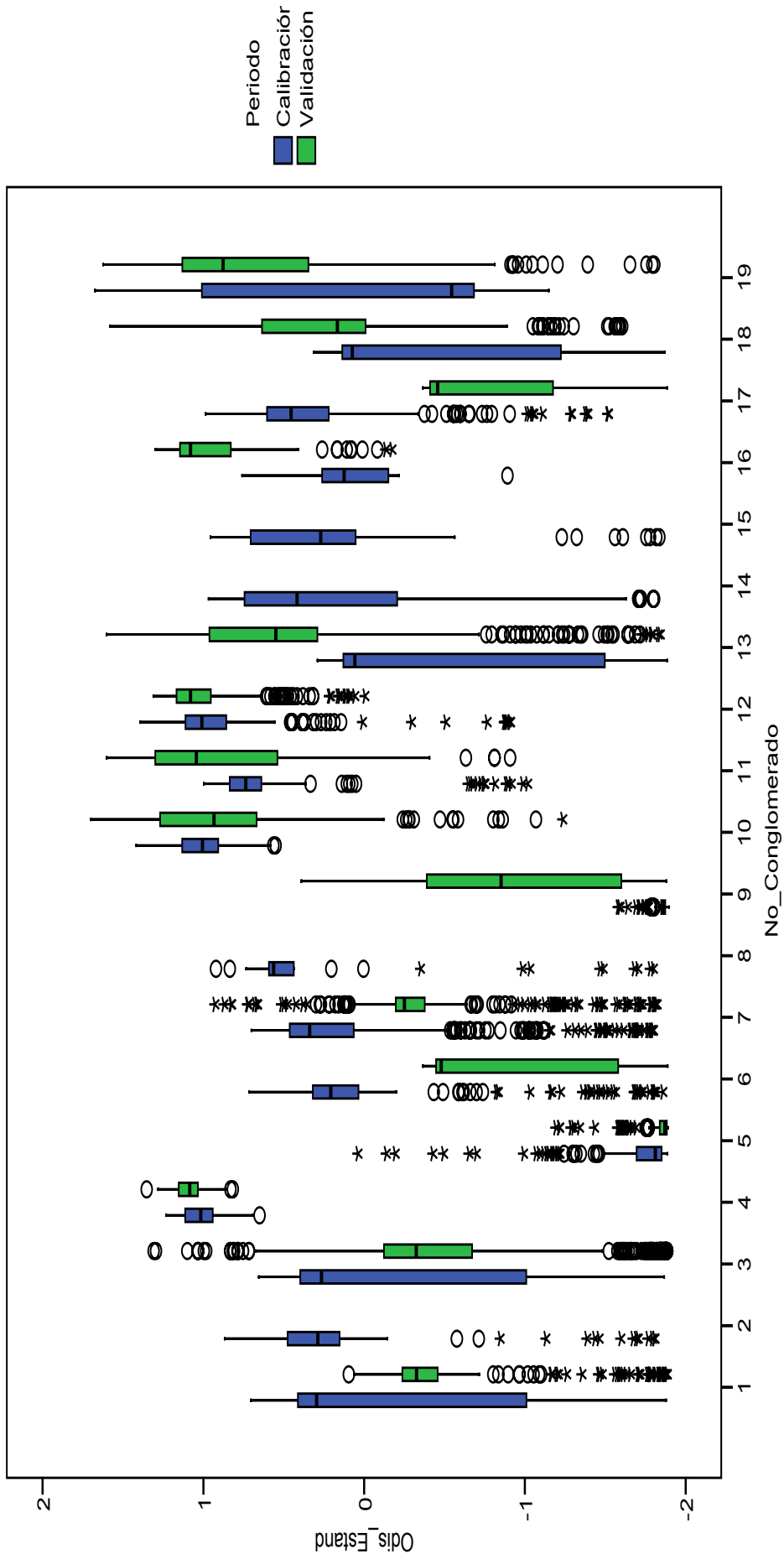
## Conglomerado

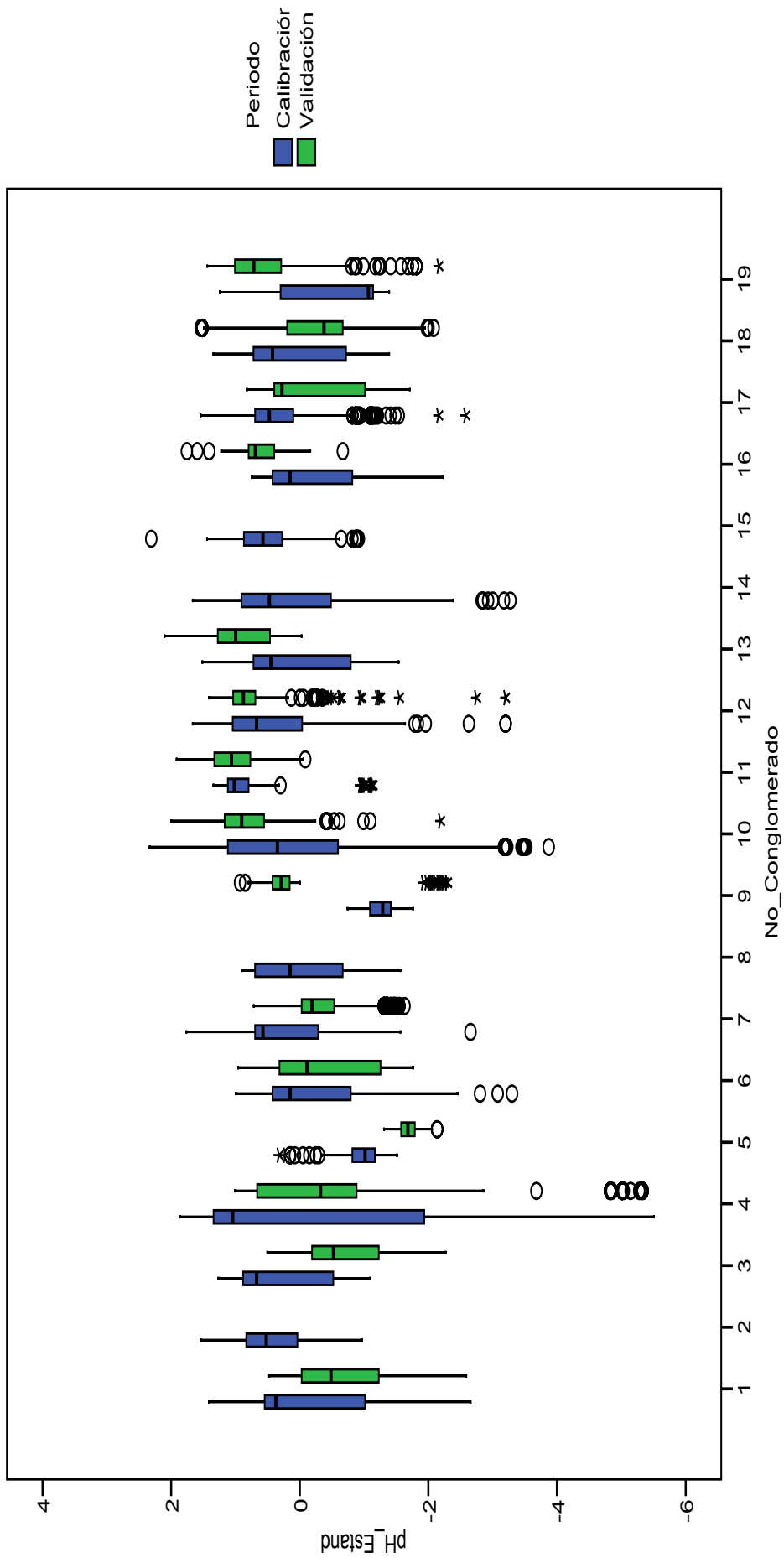
Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
DJ	0,911	.	0,666	-1,780	0,731	1,433	0,791	.	-0,476	-0,618	-0,632	-1,580	-0,363	.	.	-1,135	1,397	0,057	-0,701
V	0,175	.	0,940	1,165	-0,311	0,825	-0,521	.	-0,420	0,100	-0,910	-0,743	0,909	.	.	3,829	-0,755	-0,536	1,512
DV	0,551	.	-1,237	-0,067	-0,272	1,308	-0,559	.	-0,067	0,449	-0,362	0,755	-1,593	.	.	1,402	0,739	-0,186	1,152
Vx	-0,427	.	0,852	1,335	-0,345	0,683	-0,403	.	-0,387	-0,431	-0,335	-0,609	1,177	.	.	3,527	-0,507	-0,685	1,259
T	0,777	.	0,981	-2,436	0,828	0,378	0,831	.	-0,243	-0,422	-0,424	-1,756	0,139	.	.	-0,998	0,344	0,544	-0,270
HR	-0,875	.	-0,146	-0,879	0,373	-0,715	0,811	.	0,035	-1,025	0,974	0,934	-0,804	.	.	-1,922	0,603	0,033	-1,355
ETo	0,879	.	0,844	-1,712	0,376	0,091	-0,056	.	0,183	0,496	-0,489	-1,738	0,719	.	.	0,274	-0,592	0,772	0,456
VolExp	-0,965	.	-0,586	0,675	-0,688	-1,715	-0,780	.	0,850	0,980	0,970	0,853	0,771	.	.	1,132	-1,670	0,309	0,947
Qent	-0,577	.	-0,508	2,284	-0,489	-0,403	-0,489	.	-0,314	-0,255	-0,242	2,069	-0,232	.	.	1,720	-0,408	-0,502	0,102
Qsal	1,163	.	1,068	-1,576	0,831	0,532	0,796	.	-0,471	-0,631	-0,788	-1,468	-0,310	.	.	-0,912	0,679	0,492	-0,937
Profundidad	-0,297	.	-0,179	0,045	1,795	-0,103	-0,461	.	1,906	-0,360	-0,355	-0,001	-0,048	.	.	-0,107	-0,265	-0,403	-0,058

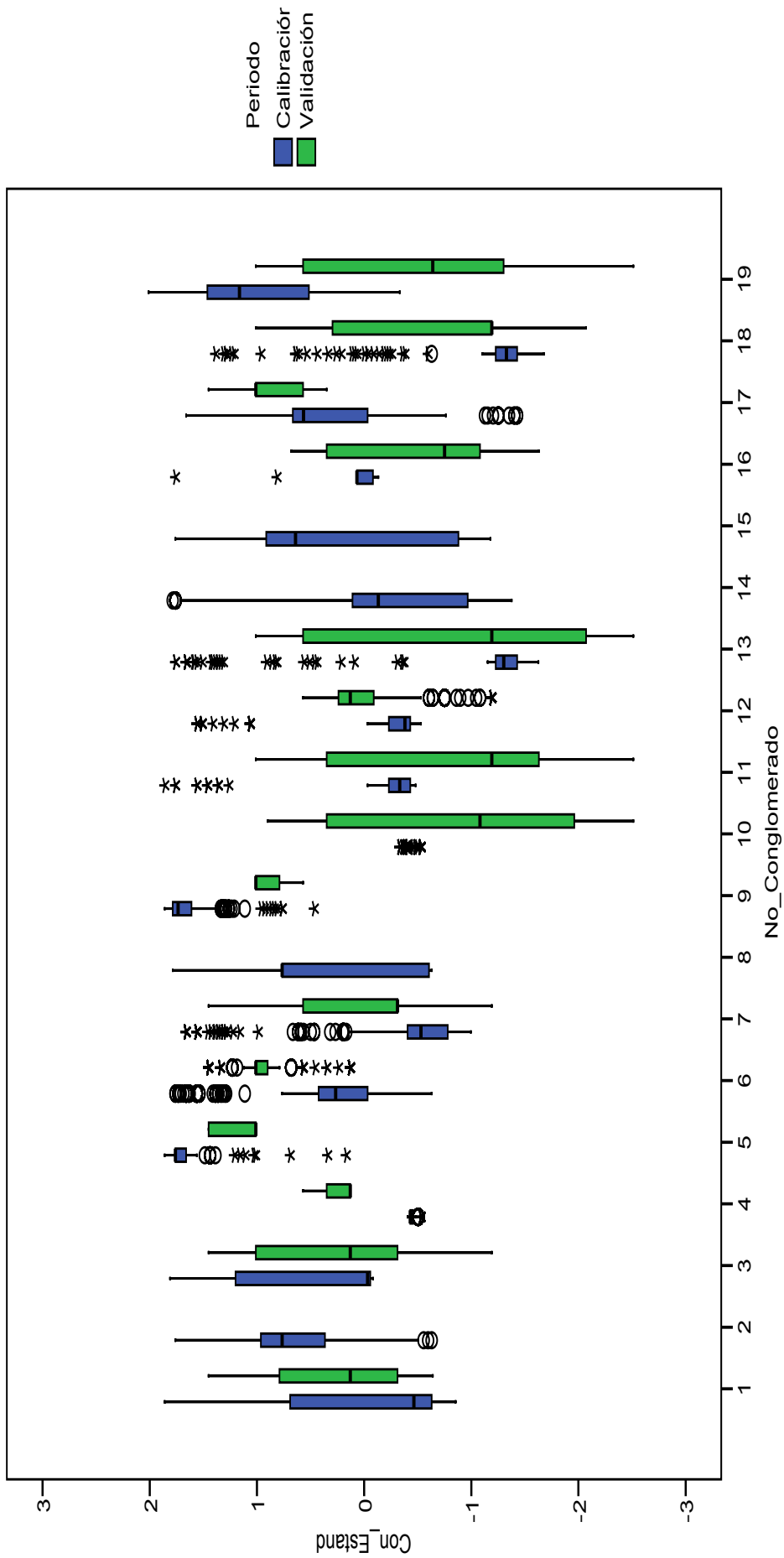
Conglomerado	Número de casos en cada conglomerado
1	245
2	0
3	334
4	198
5	305
6	271
7	569
8	0
9	217
10	311
11	395
12	446
13	264
14	0
15	0
16	61
17	272
18	475
19	210
Válidos	4573

**ANEXO 21. COMPARACIÓN DE CONGLOMERADOS EN LOS PERIODOS DE CALIBRACIÓN Y VALIDACIÓN, VARIABLES SIN TRANSFORMAR Y ESTANDARIZADAS (19 GRUPOS), GRÁFICOS DE CAJA Y BIGOTES DE VARIABLES RESPUESTA**



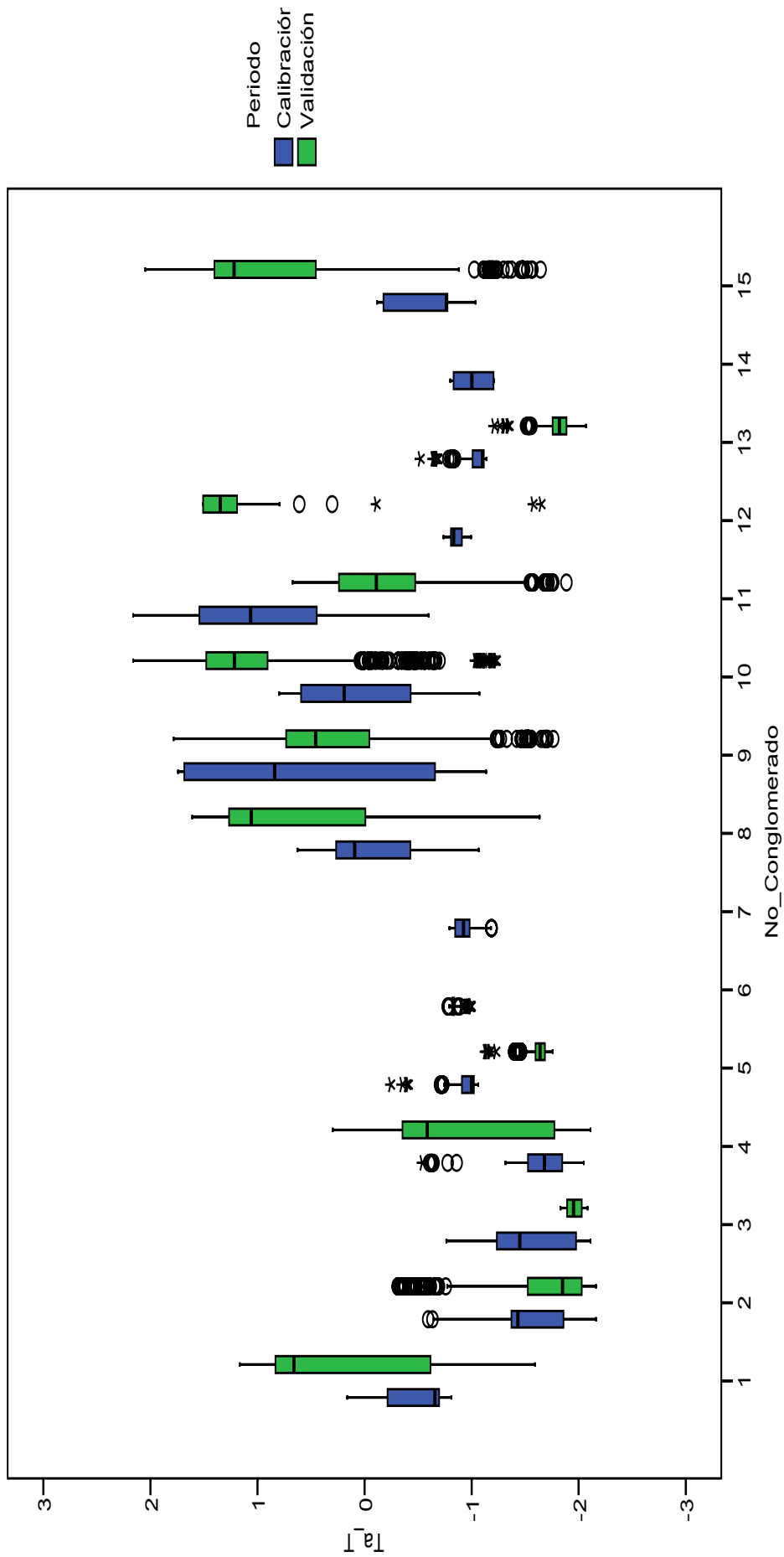




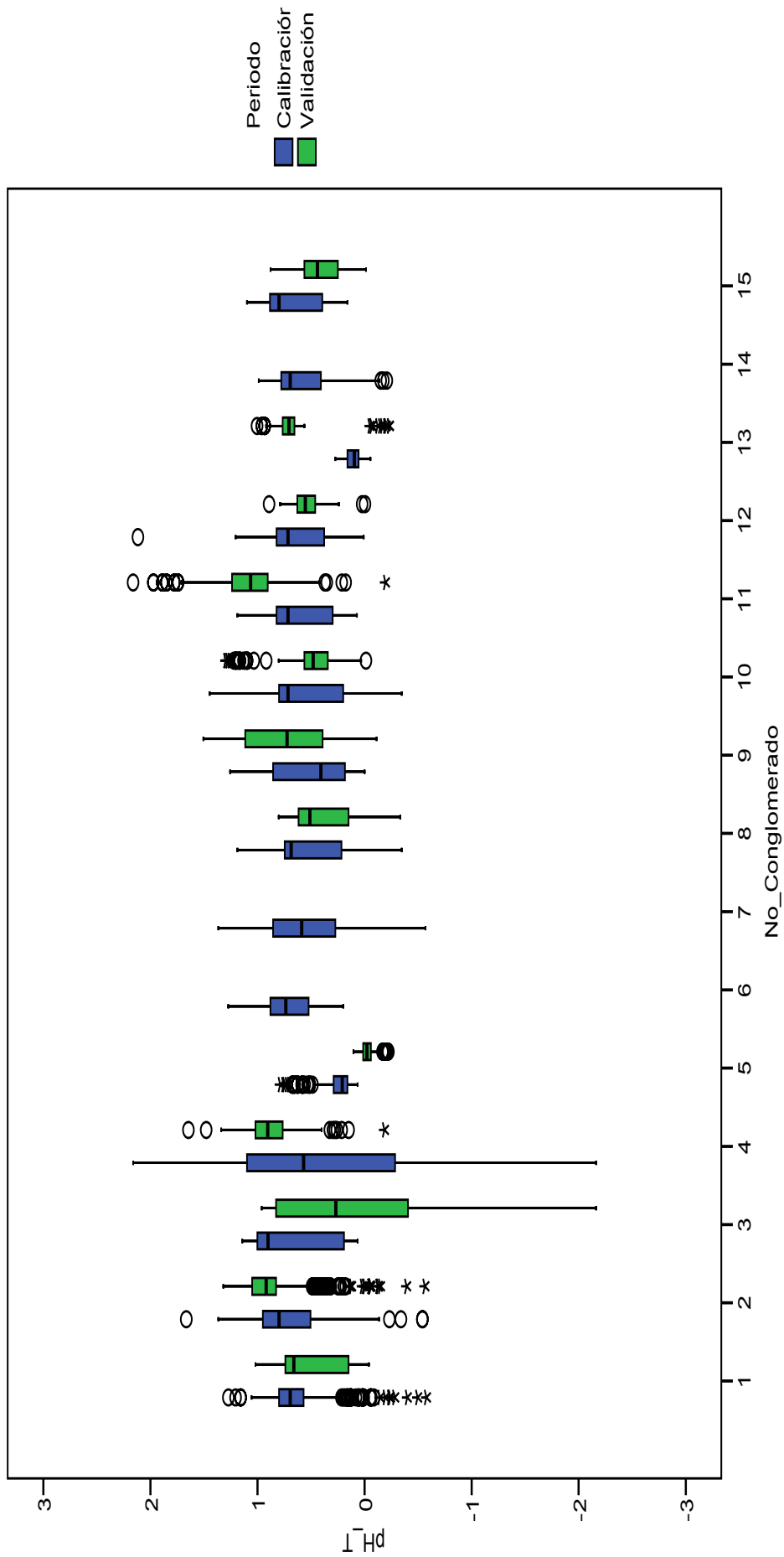


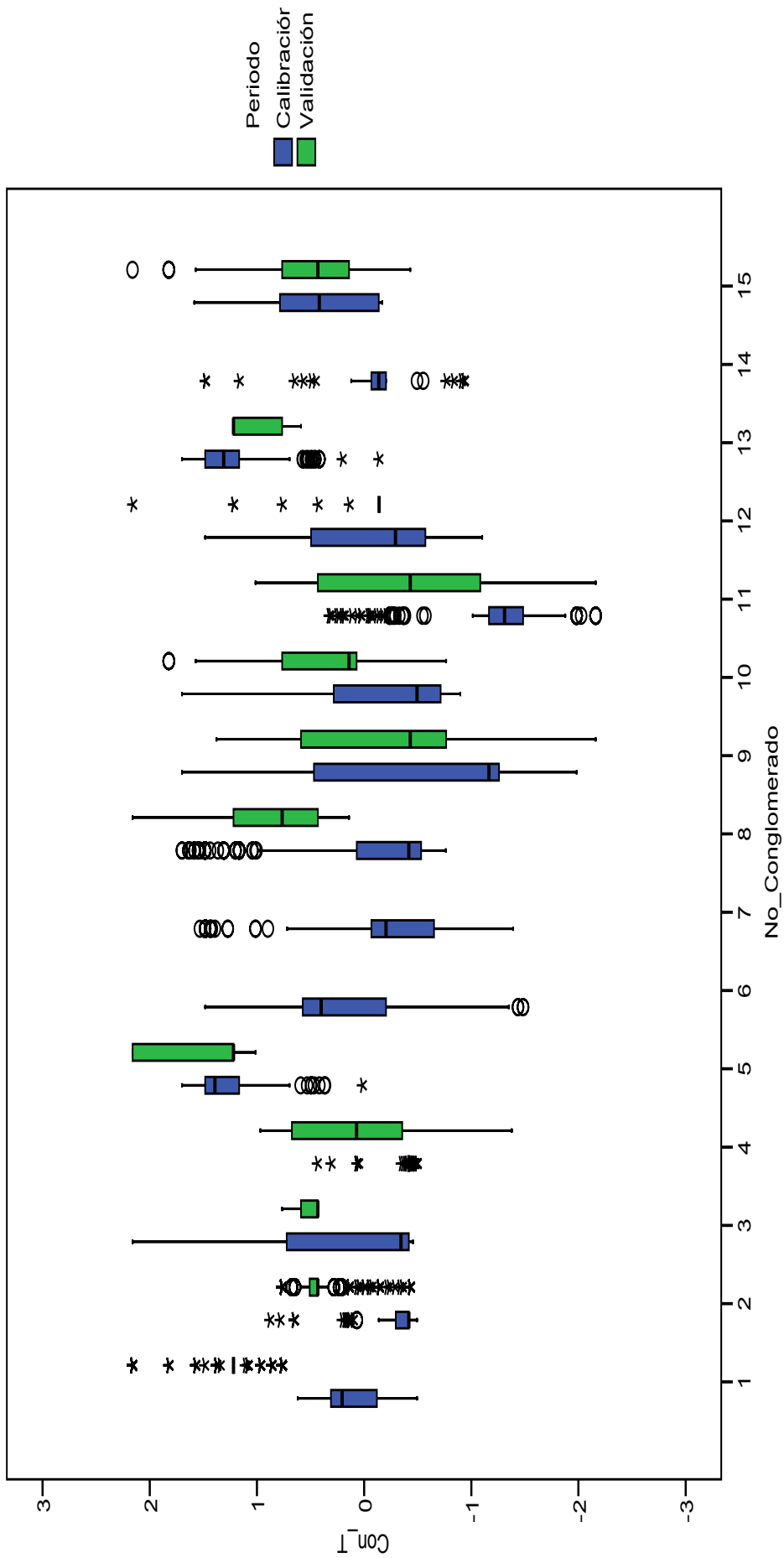


**ANEXO 22. COMPARACIÓN DE CONGLOMERADOS EN PERIODOS DE CALIBRACIÓN Y VALIDACIÓN, VARIABLES TRANSFORMADAS Y SIN ESTANDARIZAR (15 GRUPOS), GRÁFICOS DE CAJA Y BIGOTES DE VARIABLES RESPUESTA**

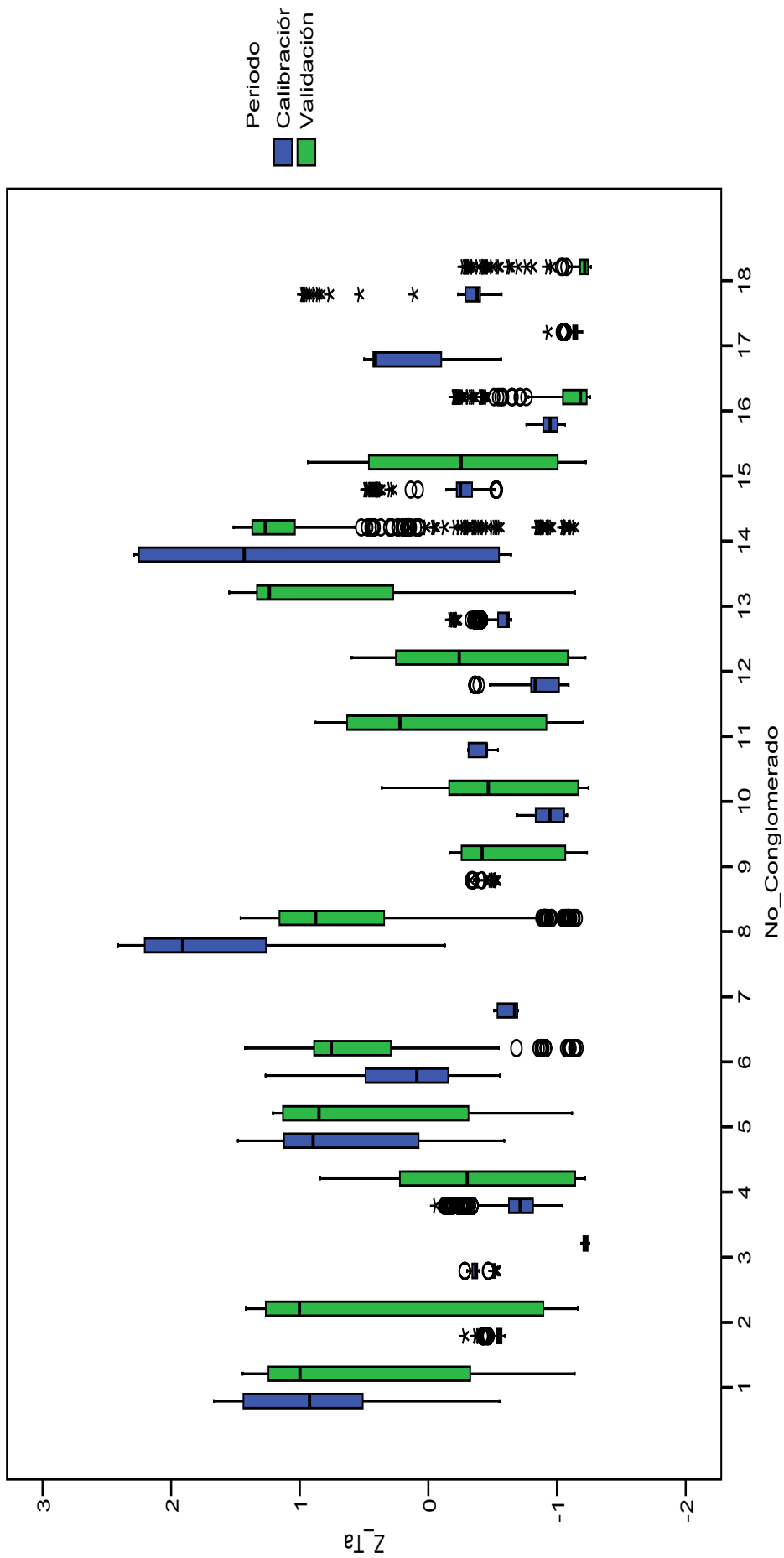


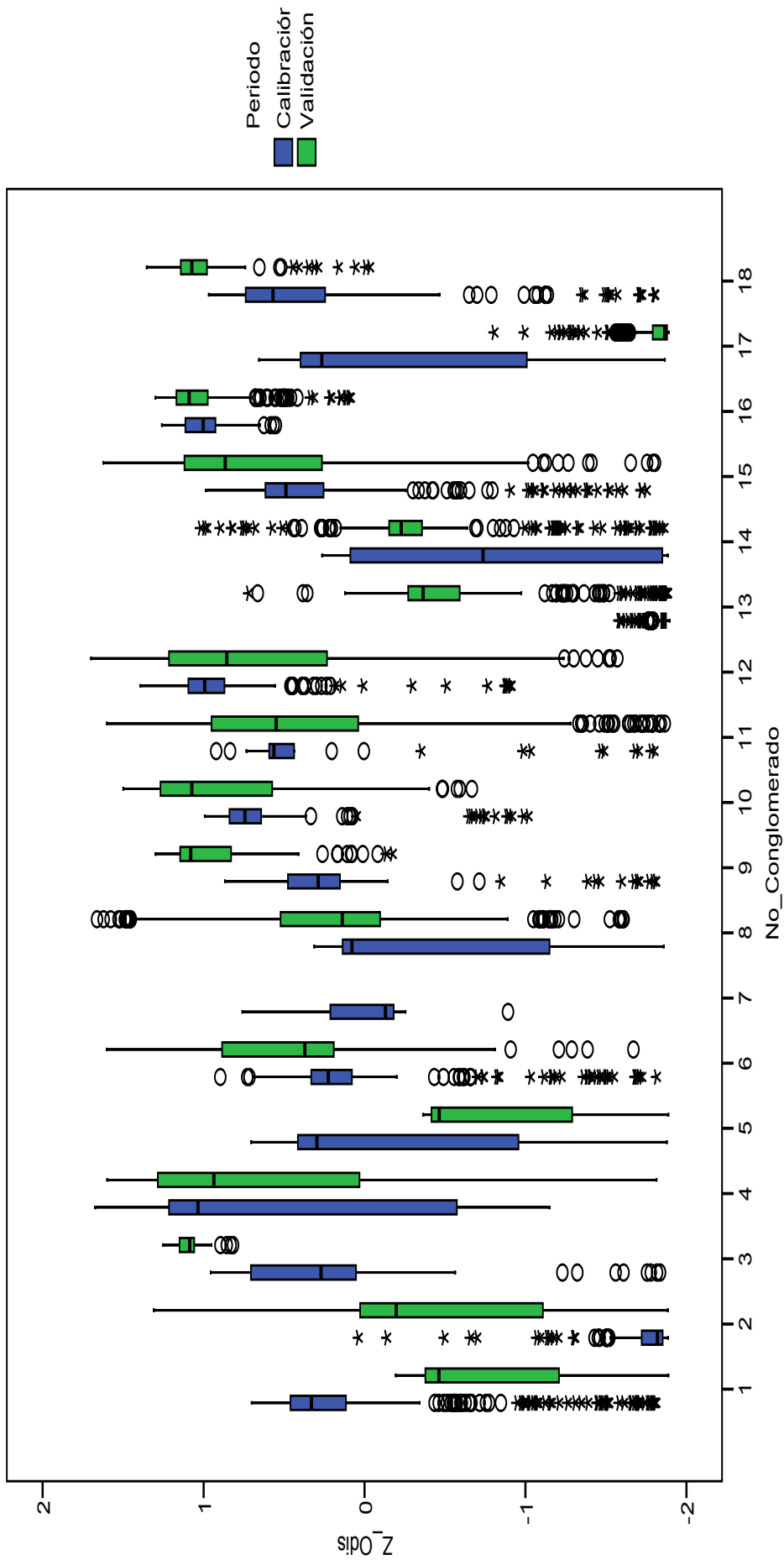




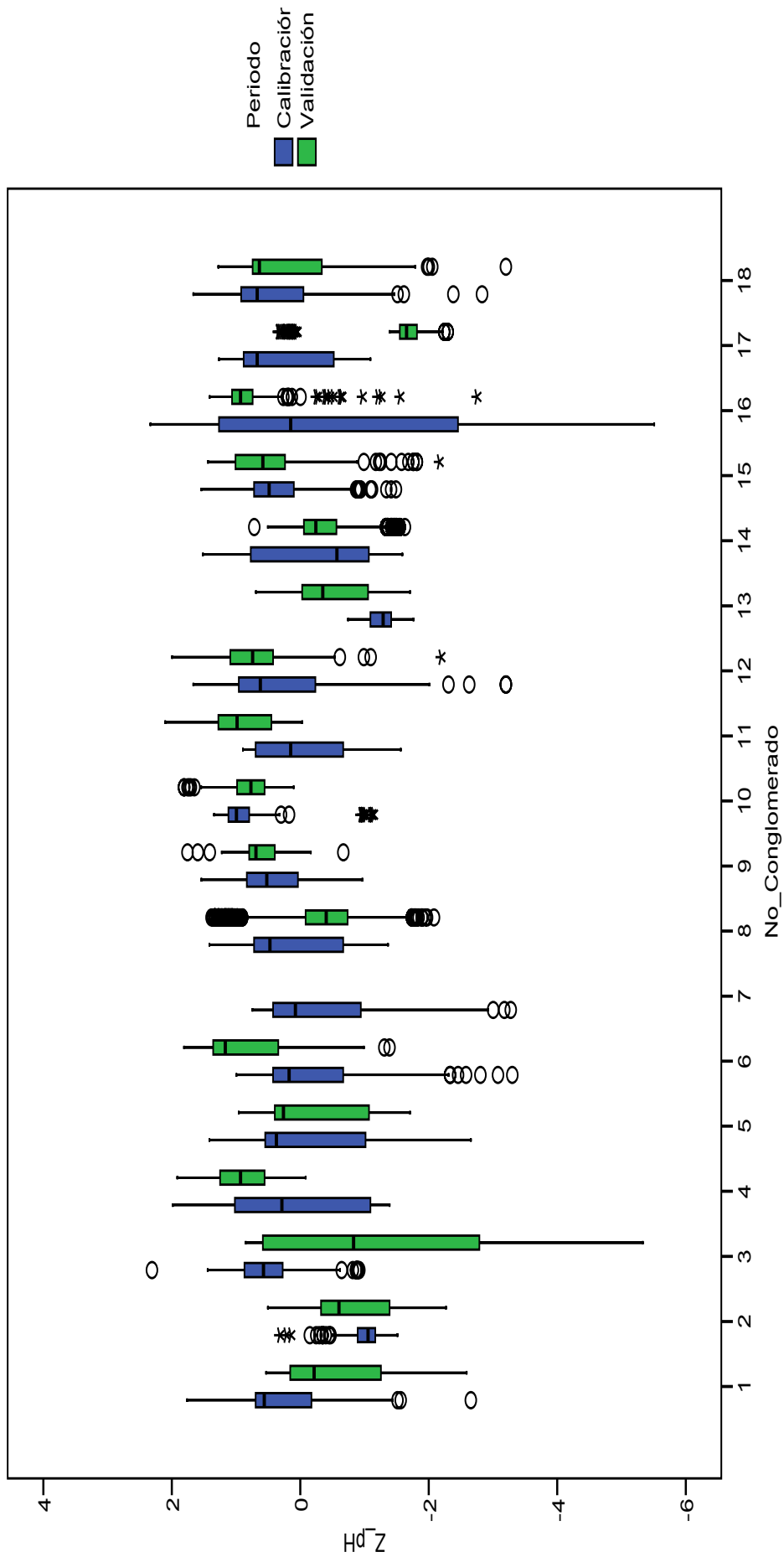


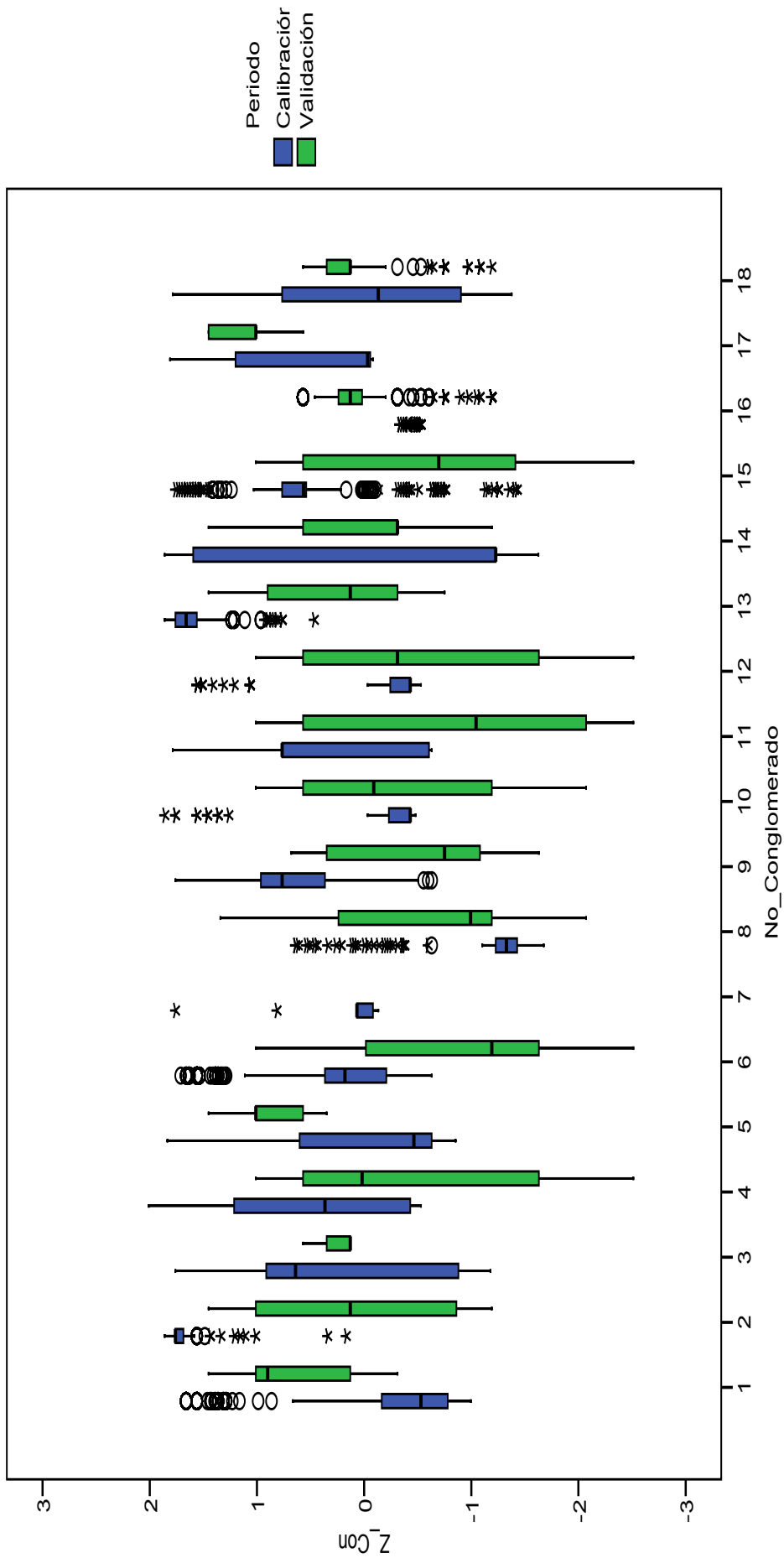
**ANEXO 23. COMPARACIÓN DE CONGLOMERADOS EN PERIODOS DE CALIBRACIÓN Y VALIDACIÓN, USANDO PARA AGRUPAR PUNTUACIONES EN COMPONENTES PRINCIPALES DE LAS VARIABLES ESTANDARIZADAS (18 GRUPOS), VARIABLES RESPUESTA ESTANDARIZADAS, GRÁFICOS DE CAJA Y BIGOTES DE VARIABLES RESPUESTA**





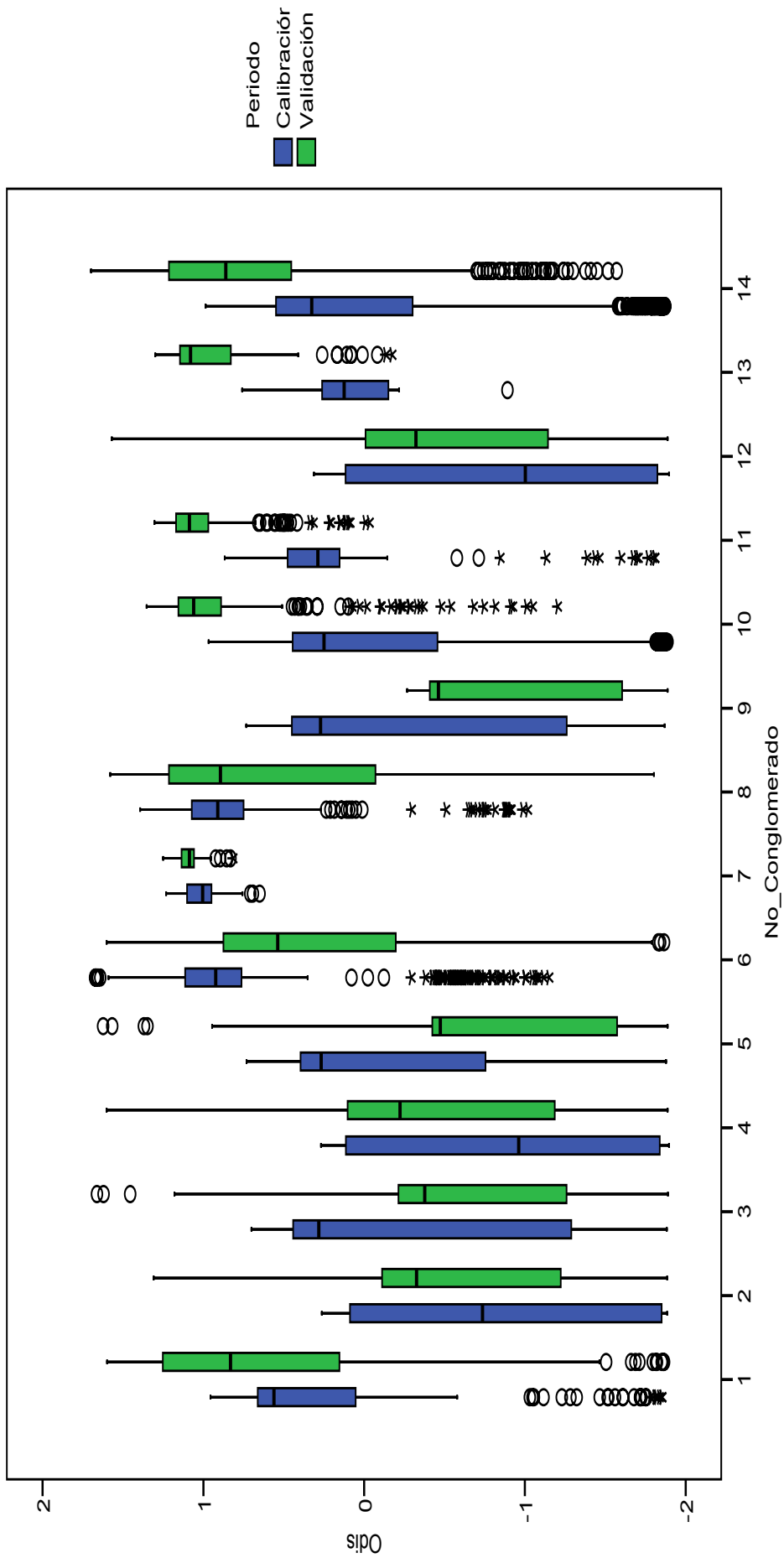


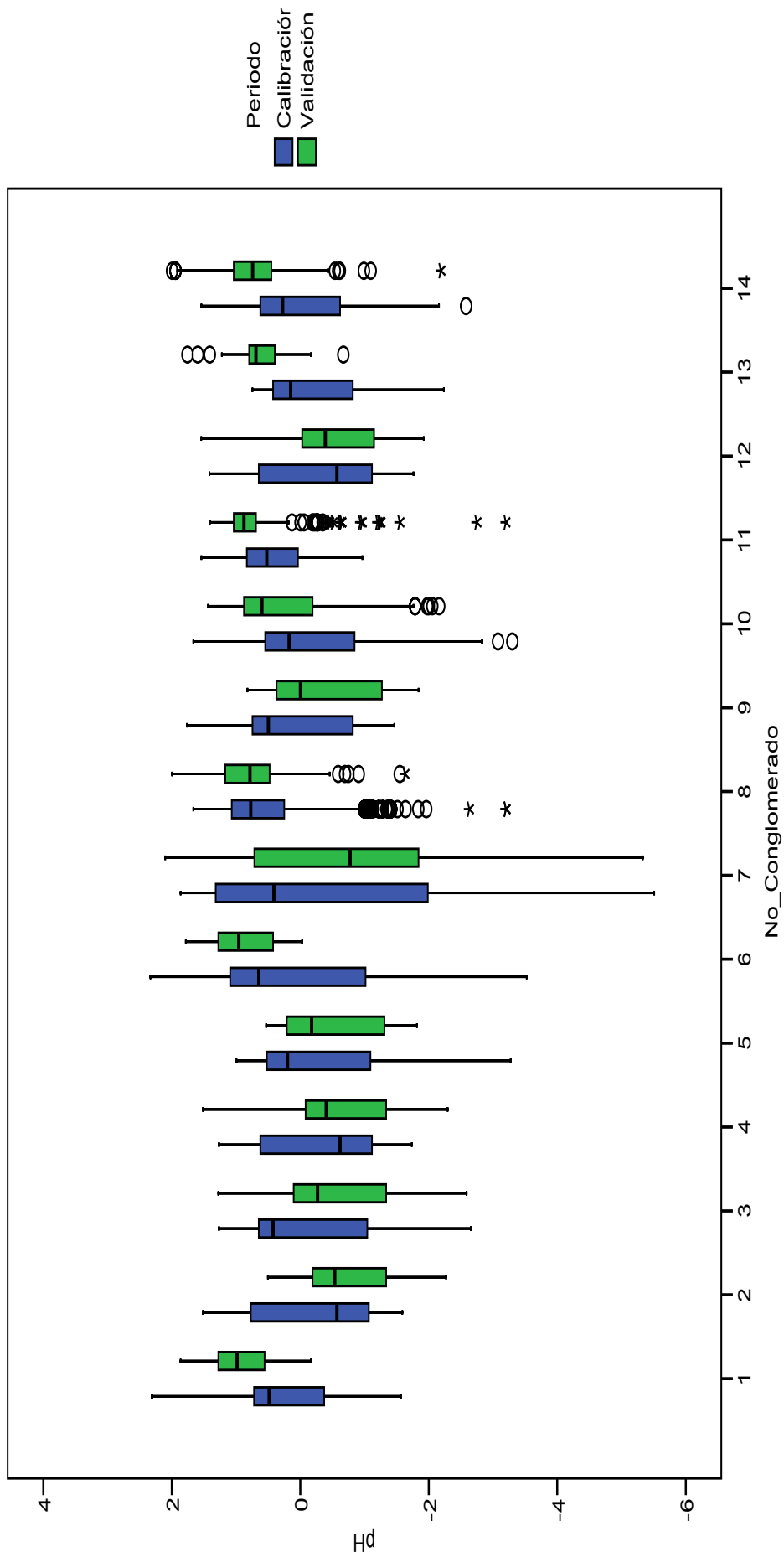


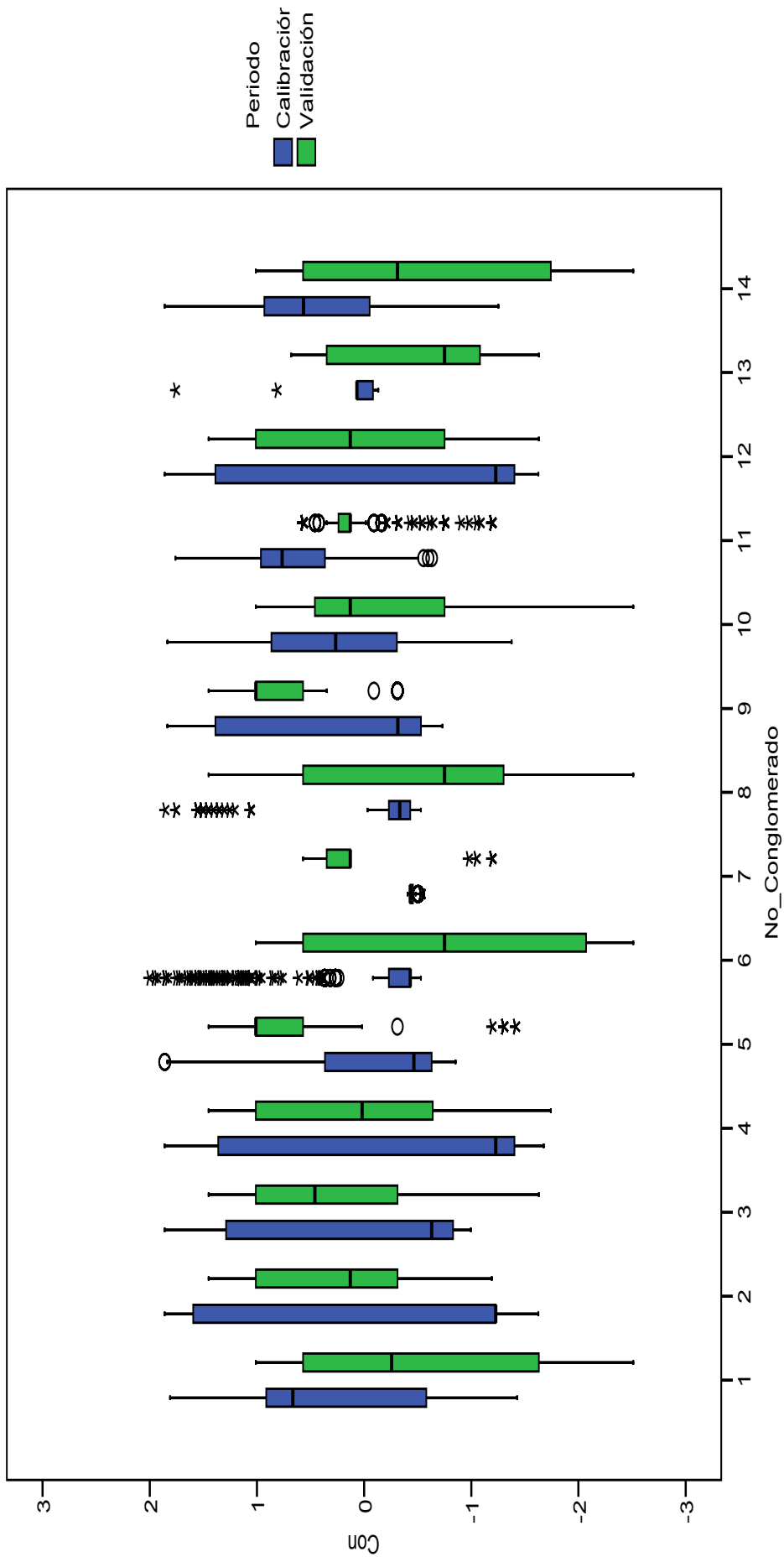


**ANEXO 24. COMPARACIÓN DE CONGLOMERADOS EN PERIODOS DE CALIBRACIÓN Y VALIDACIÓN, USANDO PARA AGRUPAR PUNTUACIONES FACTORIALES DE VARIABLES PREDICTIVAS SIN TRANSFORMAR Y ESTANDARIZADAS (14 GRUPOS), GRÁFICOS DE CAJA Y BIGOTES DE VARIABLES RESPUESTA ESTANDARIZADAS**









**ANEXO 25. ANÁLISIS CLUSTER PARA LAS VARIABLES  
ESTANDARIZADAS Y SIN TRANSFORMAR**

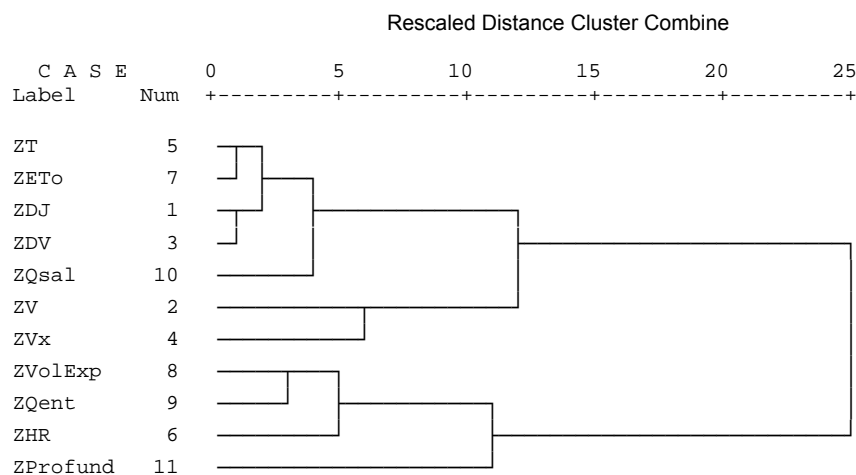


Número conglomerado	Frecuencia	Porcentaje	Porcentaje acumulado
1	474	5,3	5,3
2	92	1,0	6,3
3	358	4,0	10,3
4	324	3,6	13,9
5	559	6,2	20,1
6	514	5,7	25,9
7	993	11,1	36,9
8	45	0,5	37,4
9	490	5,5	42,9
10	631	7,0	49,9
11	636	7,1	57,0
12	953	10,6	67,6
13	491	5,5	73,1
14	143	1,6	74,7
15	50	0,6	75,3
16	107	1,2	76,4
17	780	8,7	85,1
18	991	11,0	96,2
19	343	3,8	100
Total	8974		

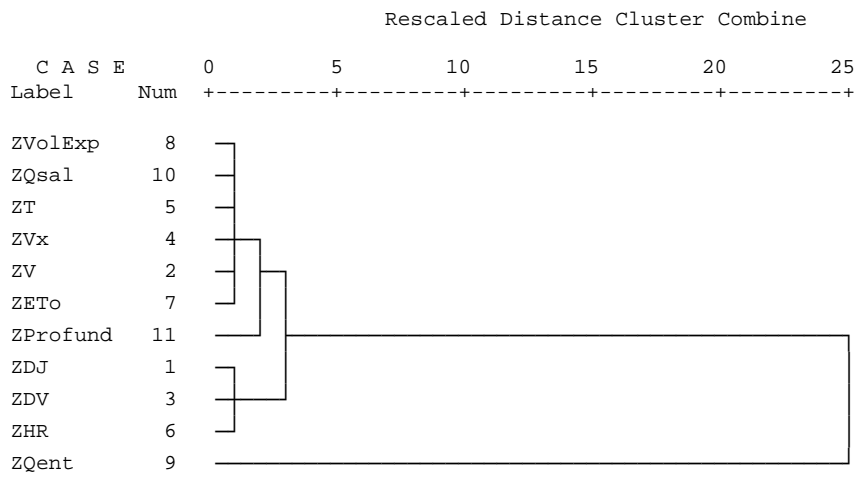
### Análisis cluster jerárquico

Para formar todos los dendogramas se usó el Método del promedio entre grupos o vinculación inter-grupos.

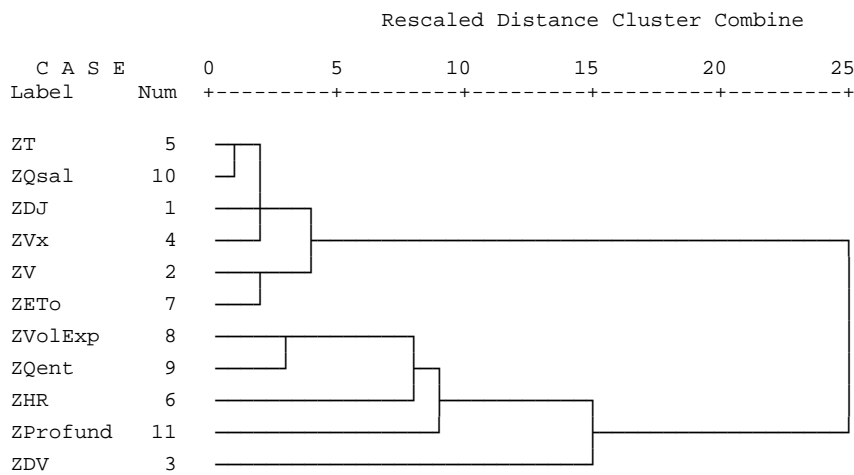
### Conglomerado 1



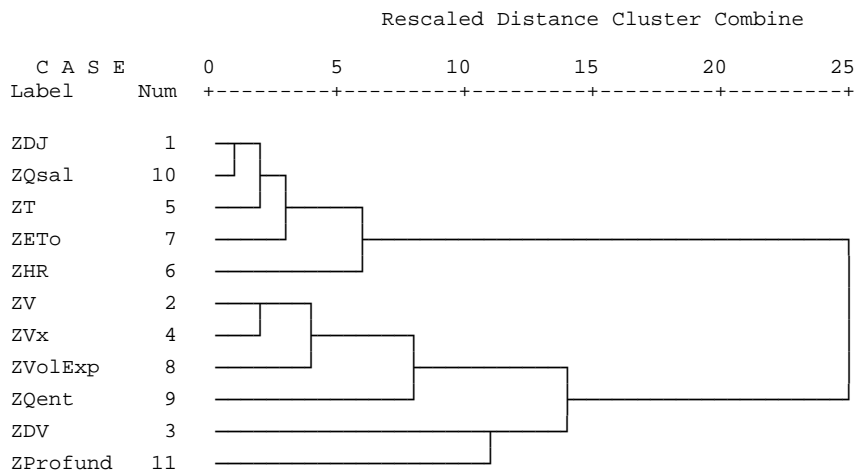
## Conglomerado 2



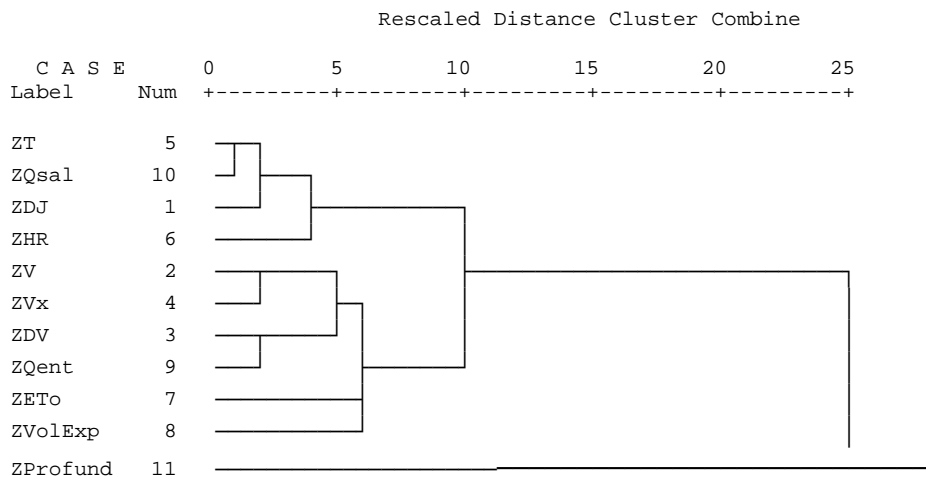
## Conglomerado 3



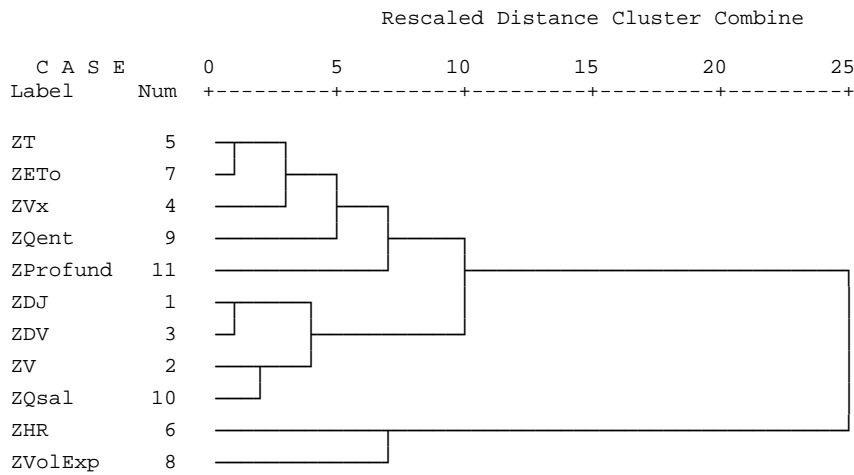
## Conglomerado 4



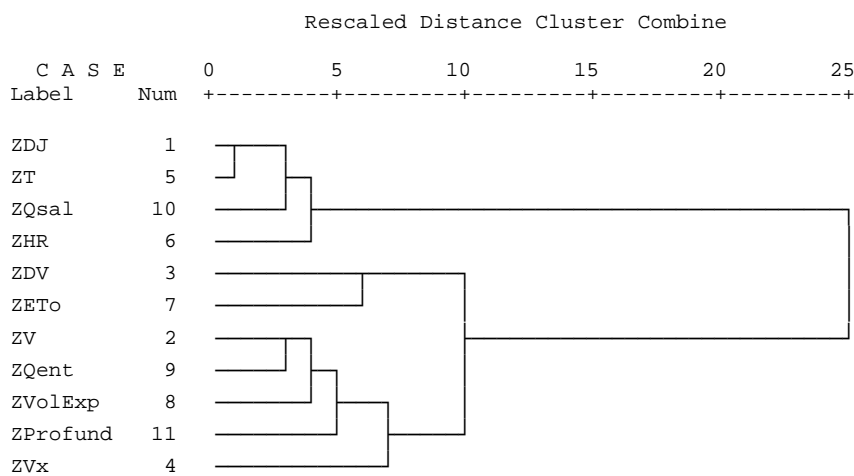
### Conglomerado 5



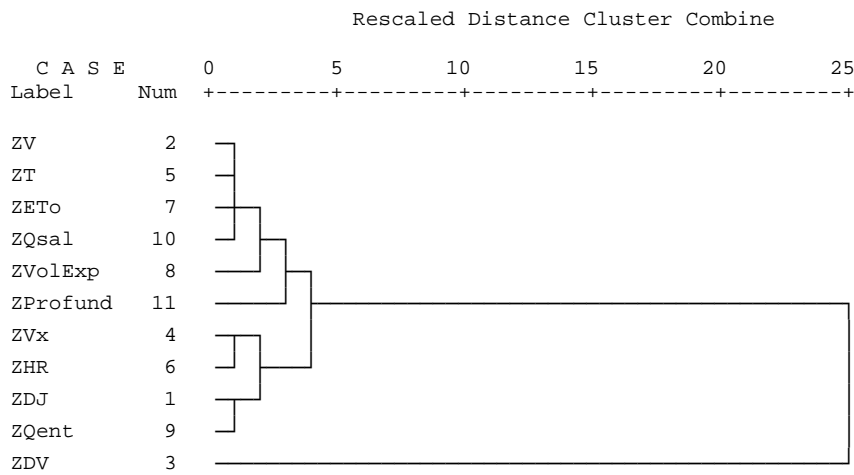
### Conglomerado 6



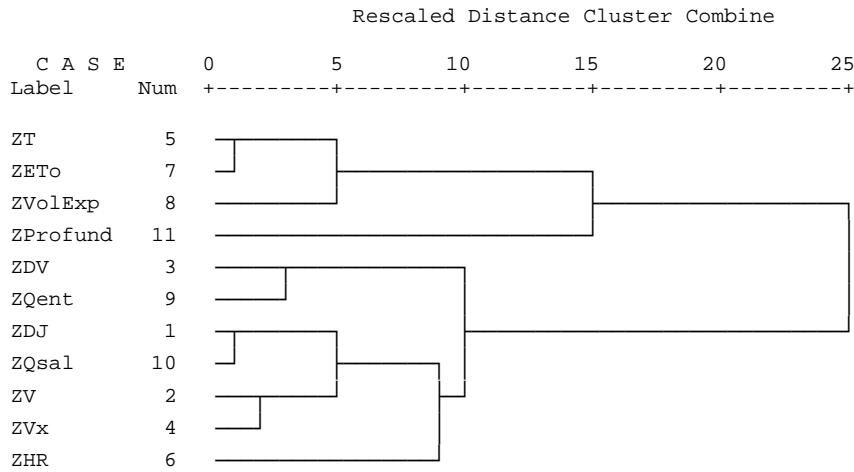
### Conglomerado 7



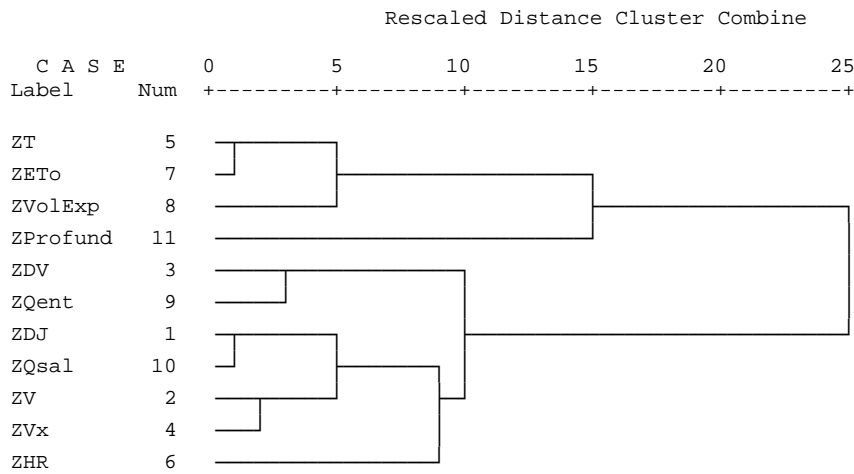
### Conglomerado 8



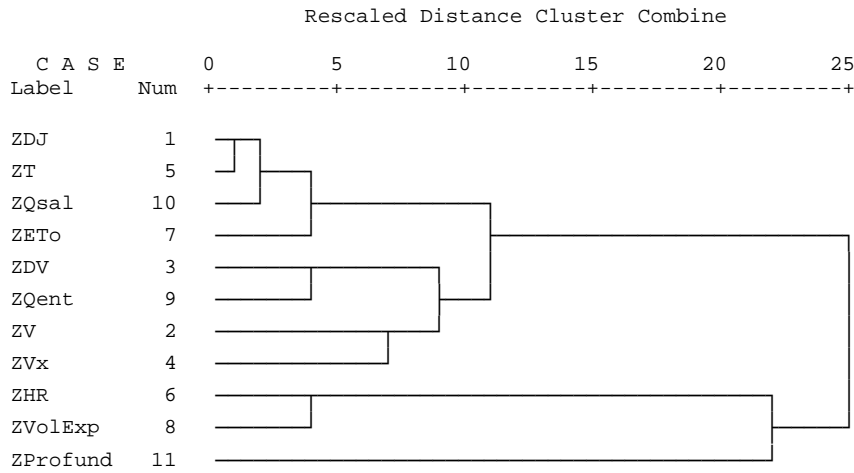
### Conglomerado 9



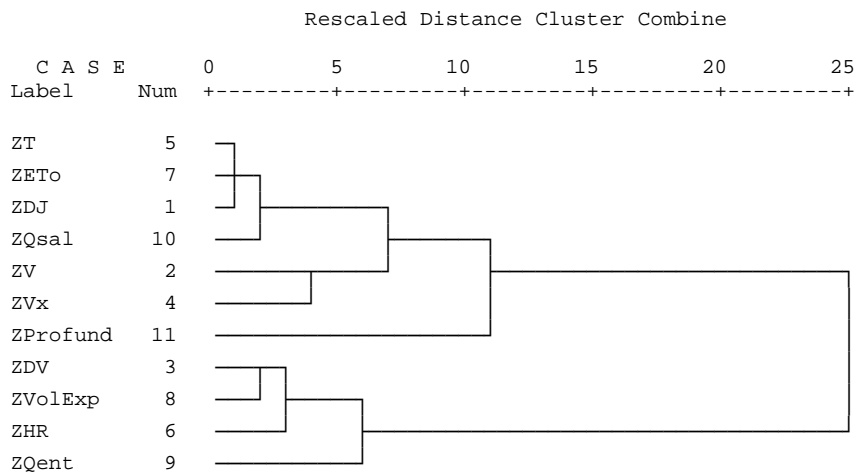
### Conglomerado 10



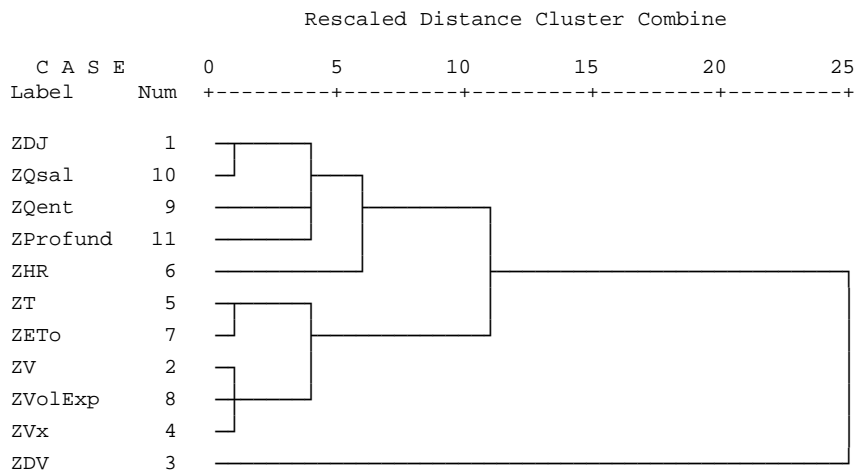
### Conglomerado 11



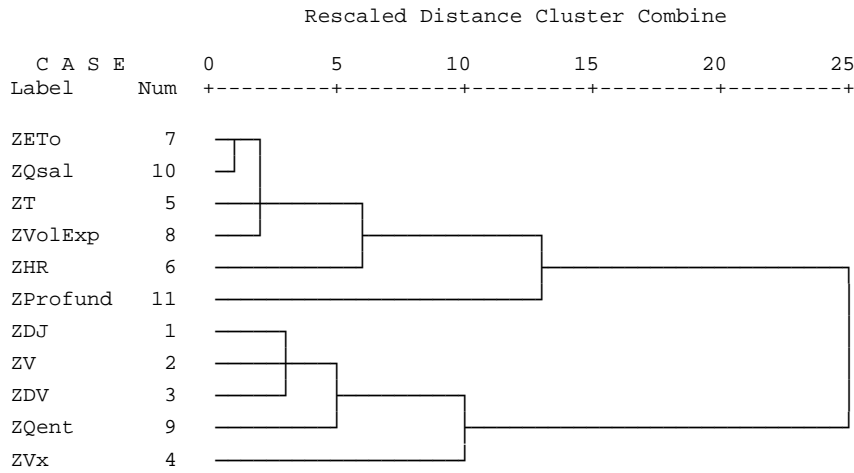
### Conglomerado 12



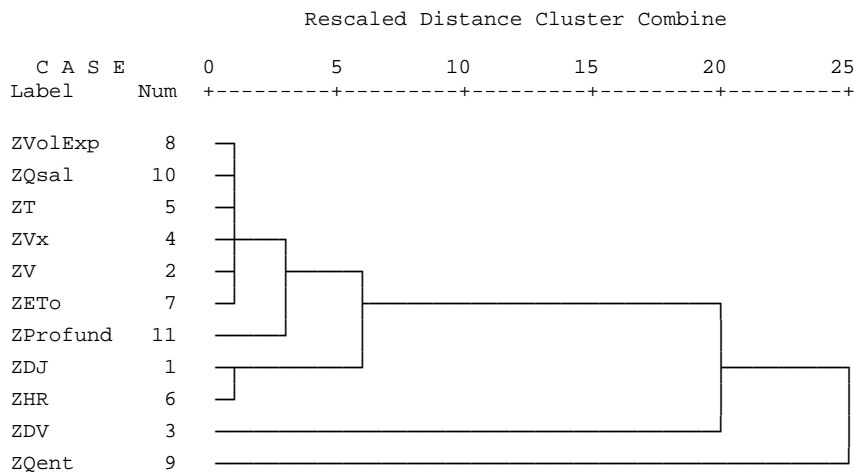
### Conglomerado 13



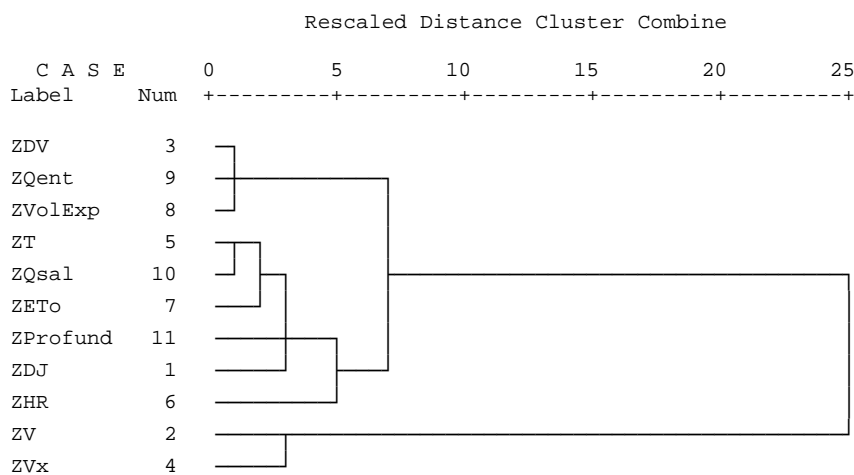
### Conglomerado 14



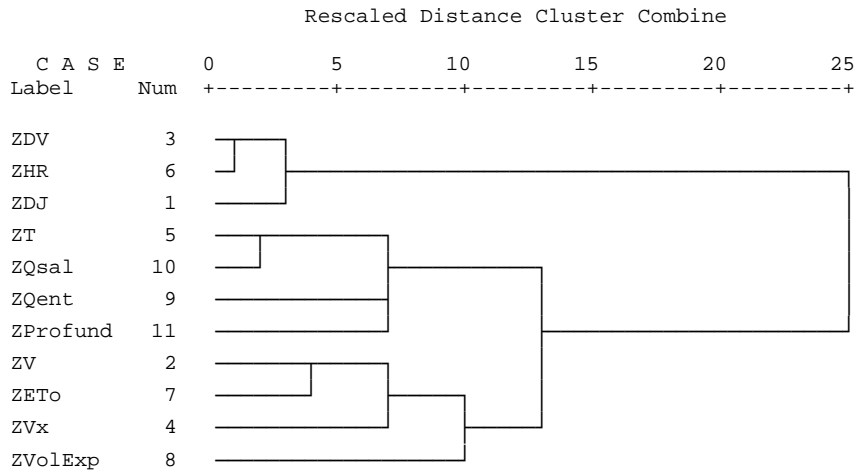
### Conglomerado 15



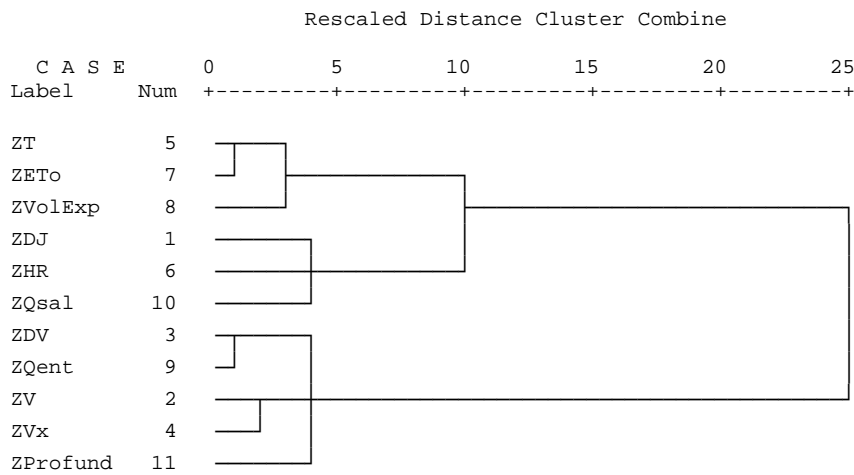
### Conglomerado 16



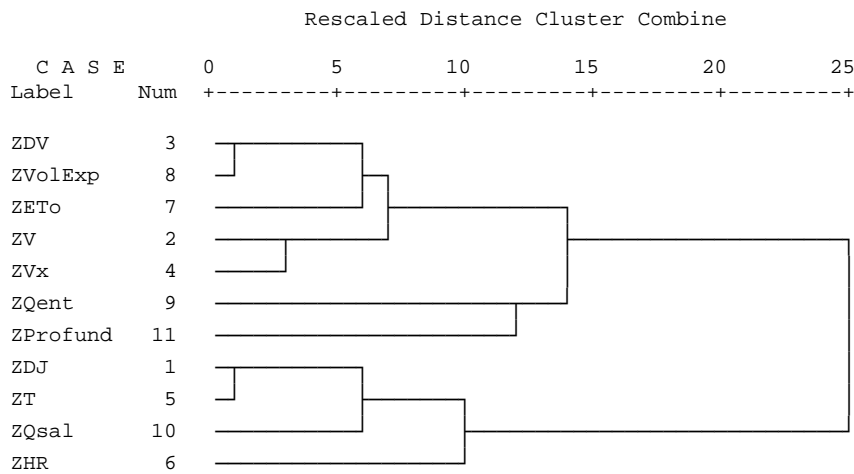
### Conglomerado 17



### Conglomerado 18



### Conglomerado 19



## **ANEXO 26. ANÁLISIS CLUSTER PARA LAS VARIABLES TRANSFORMADAS**

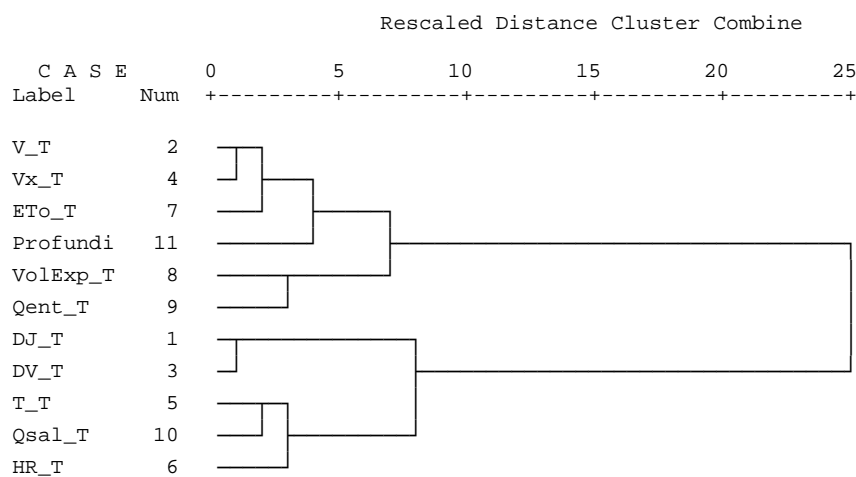


Número de conglomerado	Frecuencia	Porcentaje	Porcentaje acumulado
1	916	10,2	10,2
2	1211	13,5	23,7
3	456	5,1	28,8
4	460	5,1	33,9
5	655	7,3	41,2
6	122	1,4	42,6
7	189	2,1	44,7
8	602	6,7	51,4
9	715	8,0	59,3
10	1107	12,3	71,7
11	1349	15,0	86,7
12	103	1,1	87,9
13	624	7,0	94,8
14	71	0,8	95,6
15	394	4,4	100
Total	8974		

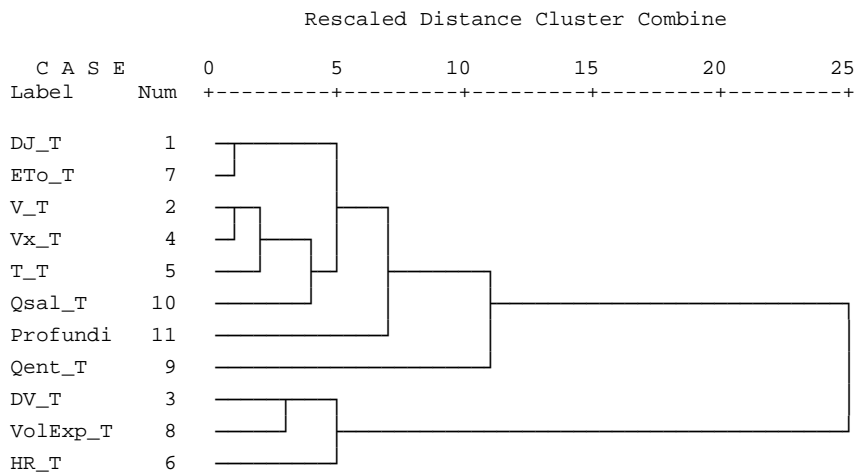
### Análisis cluster jerárquico

Para formar todos los dendogramas se usó el Método del promedio entre grupos o vinculación inter-grupos.

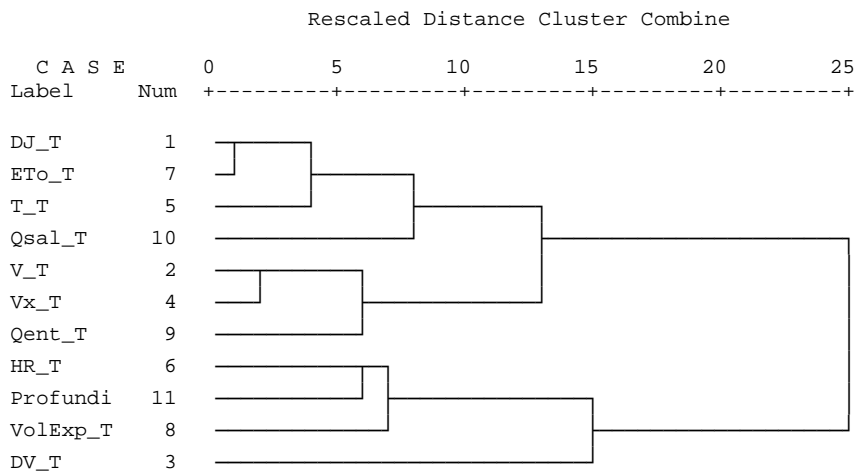
### Conglomerado 1



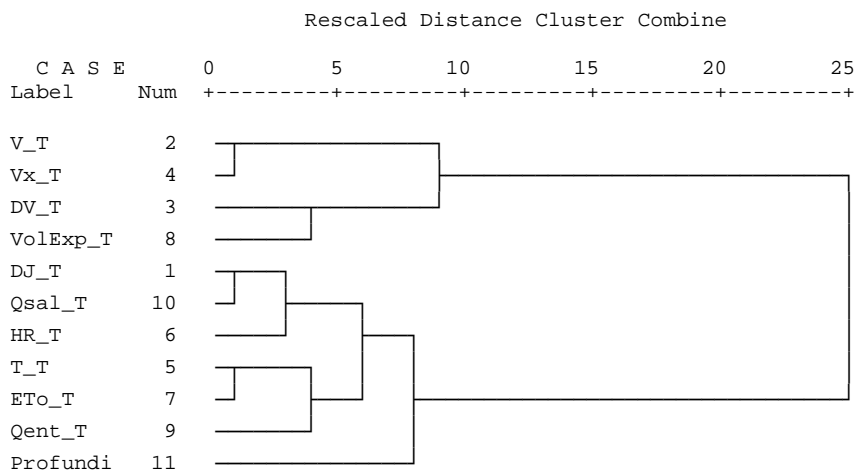
## Conglomerado 2



## Conglomerado 3

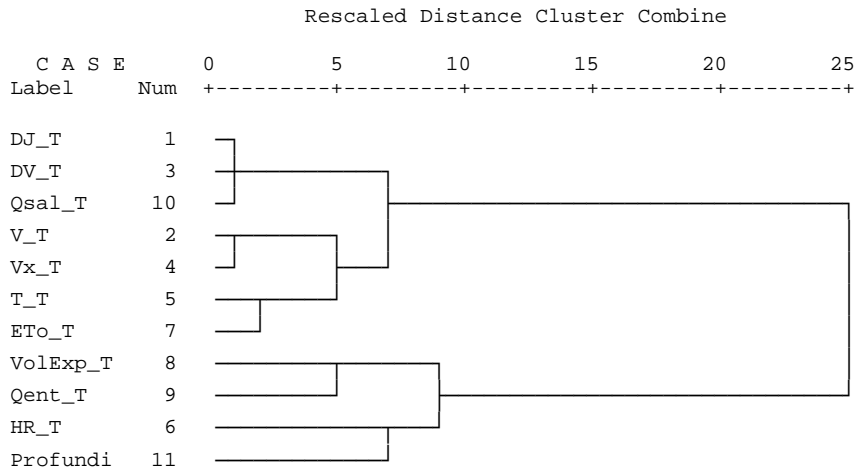


## Conglomerado 4

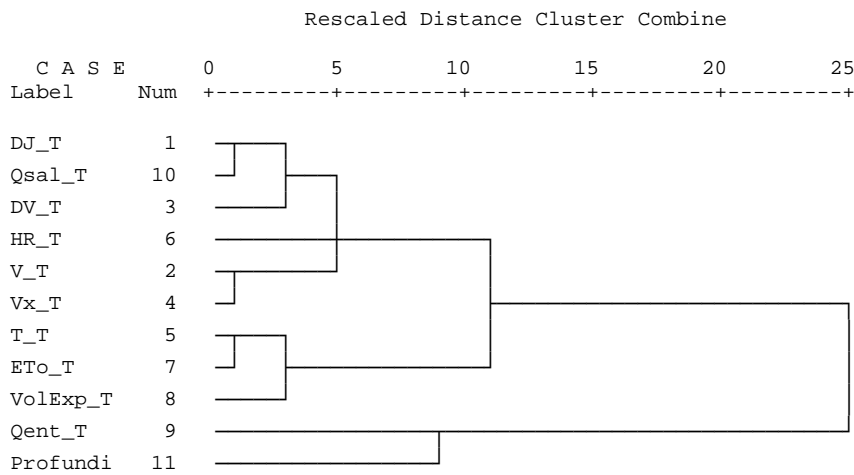




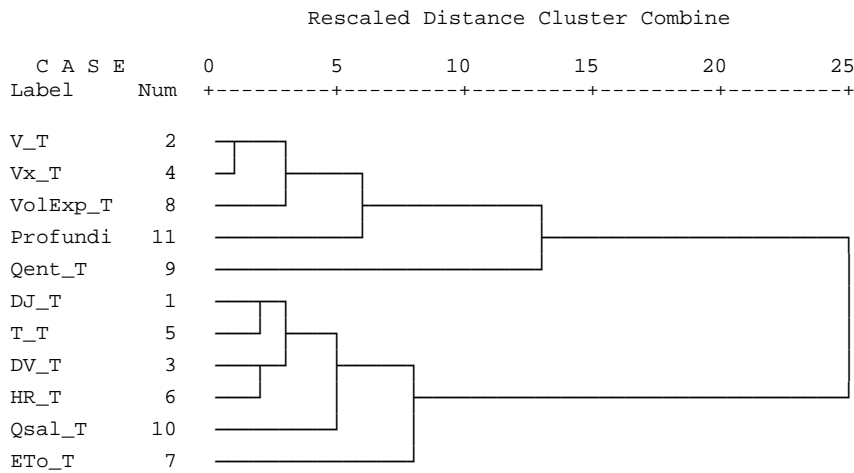
### Conglomerado 8



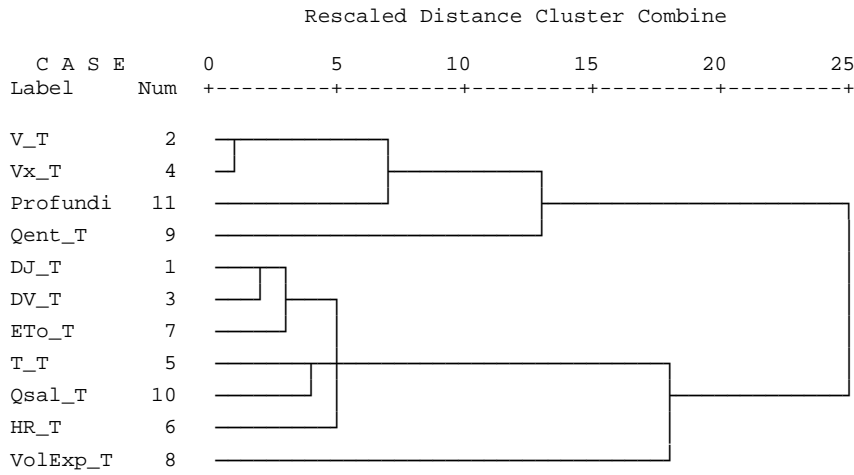
### Conglomerado 9



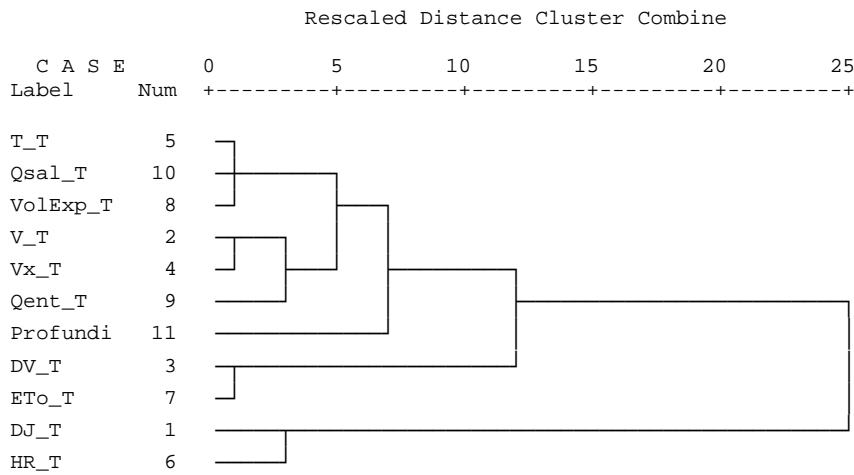
### Conglomerado 10



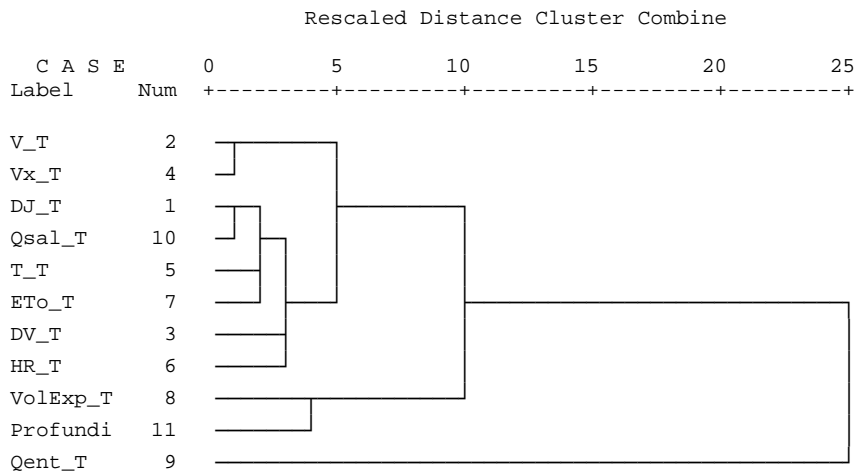
### Conglomerado 11



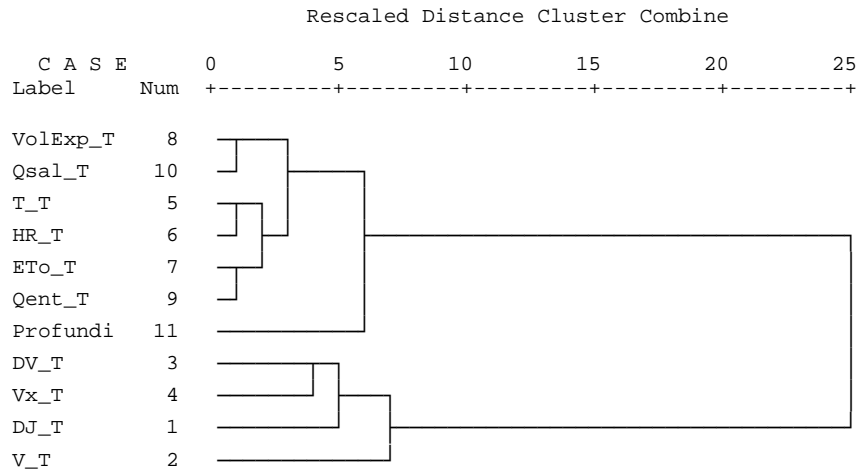
### Conglomerado 12



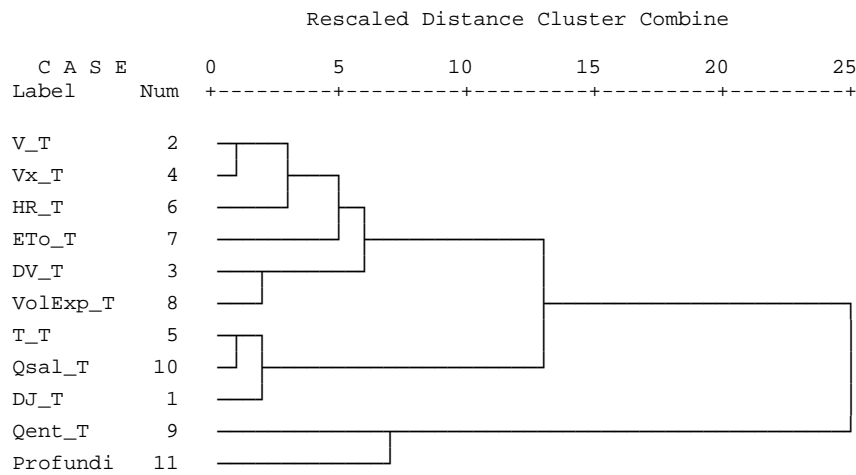
### Conglomerado 13



### Conglomerado 14



### Conglomerado 15



**ANEXO 27. ANÁLISIS DISCRIMINANTE PARA LAS VARIABLES  
SIN TRANSFORMAR, NO ESTANDARIZADAS (19  
GRUPOS), PERIODO COMPLETO DE ESTUDIO**

Estadísticos de grupo

Número de grupo	Variable	Media	Desv. ttp.	No de casos	Número de grupo	Variable	Media	Desv. ttp.	No de casos
1	DJ	254,9	37,8	474	2	DJ	345,3	1,5	92
	V	1,5	0,3	474		V	1,0	0,1	92
	DV	272,6	33,5	474		DV	331,6	14,6	92
	Vx	6,1	1,7	474		Vx	4,7	0,5	92
	T	22,9	2,8	474		T	13,7	0,7	92
	HR	43,6	8,4	474		HR	69,4	5,7	92
	ETo	4,6	1,2	474		ETo	1,0	0,0	92
	VolExp	7418094,9	488285,9	474		VolExp	7474021,7	226901,9	92
	Qent	7018,9	9390,1	474		Qent	201385,8	33851,2	92
	Qsal	40309,1	4703,6	474		Qsal	18399,4	2782,6	92
Profundidad	7,2	6,0	474	Profundidad	8,4	7,0	92		
3	DJ	212,9	33,5	358	4	DJ	53,8	14,0	324
	V	1,5	0,2	358		V	1,8	0,3	324
	DV	110,7	35,7	358		DV	246,3	124,8	324
	Vx	6,2	0,5	358		Vx	7,4	1,2	324
	T	25,8	2,2	358		T	8,6	2,2	324
	HR	57,9	9,1	358		HR	40,1	13,3	324
	ETo	5,2	1,2	358		ETo	2,4	0,6	324
	VolExp	8178245,8	951851,7	358		VolExp	10058108,0	354959,3	324
	Qent	4316,0	8167,1	358		Qent	40560,8	8440,5	324
	Qsal	41260,3	4036,2	358		Qsal	10396,0	3910,0	324
Profundidad	7,4	6,0	358	Profundidad	9,2	7,9	324		
5	DJ	256,9	56,6	559	6	DJ	280,2	32,8	514
	V	1,2	0,2	559		V	1,4	0,1	514
	DV	238,6	69,3	559		DV	305,6	25,7	514
	Vx	4,6	1,0	559		Vx	5,4	0,9	514
	T	21,4	5,3	559		T	19,1	5,0	514
	HR	62,5	8,9	559		HR	46,6	7,8	514
	ETo	3,5	1,7	559		ETo	3,4	1,3	514
	VolExp	7636202,1	927700,0	559		VolExp	6622645,9	288451,6	514
	Qent	13441,3	16681,3	559		Qent	14581,3	15359,4	514
	Qsal	35666,2	6760,2	559		Qsal	34303,3	4176,0	514
Profundidad	21,6	3,7	559	Profundidad	7,2	5,8	514		
7	DJ	244,8	39,7	993	8	DJ	349,0	9,0	45
	V	1,1	0,1	993		V	1,1	0,0	45
	DV	216,1	70,5	993		DV	13,2	11,1	45
	Vx	4,6	0,9	993		Vx	5,5	0,9	45
	T	23,8	2,4	993		T	12,2	2,0	45
	HR	65,7	6,6	993		HR	67,0	4,5	45
	ETo	3,6	1,1	993		ETo	1,0	0,0	45
	VolExp	7670999,0	689825,2	993		VolExp	7769444,4	790964,8	45
	Qent	5740,8	6208,1	993		Qent	48533,4	14085,0	45
	Qsal	38726,3	4193,0	993		Qsal	22560,0	1085,4	45
Profundidad	5,4	3,8	993	Profundidad	9,1	7,6	45		
9	DJ	177,4	37,5	490	10	DJ	81,5	50,3	631
	V	1,2	0,2	490		V	1,3	0,2	631
	DV	193,4	44,8	490		DV	274,8	43,3	631
	Vx	4,8	0,8	490		Vx	5,2	0,9	631
	T	23,1	3,3	490		T	14,5	5,5	631
	HR	60,2	9,6	490		HR	45,2	9,5	631
	ETo	4,9	0,8	490		ETo	3,6	1,6	631
	VolExp	10300212,2	476790,3	490		VolExp	10141271,0	688869,4	631
	Qent	4394,2	5603,4	490		Qent	22090,4	17420,1	631
	Qsal	28433,5	7561,6	490		Qsal	15873,8	6143,2	631
Profundidad	22,8	4,4	490	Profundidad	6,1	4,7	631		



Número de grupo	Variable	Media	Desv. típ.	No de casos	Número de grupo	Variable	Media	Desv. típ.	No de casos
11	DJ	92,4	51,8	636	12	DJ	48,7	29,7	953
	V	1,1	0,2	636		V	1,2	0,2	953
	DV	223,0	68,2	636		DV	277,7	32,7	953
	Vx	4,7	1,0	636		Vx	4,3	1,0	953
	T	15,7	4,7	636		T	11,1	2,2	953
	HR	67,4	8,6	636		HR	68,5	7,8	953
	ETo	3,0	1,4	636		ETo	1,8	0,8	953
	VolExp	10186017,3	838734,2	636		VolExp	9828097,6	800631,6	953
	Qent	19413,2	18946,2	636		Qent	39523,3	7494,0	953
	Qsal	16576,3	6517,9	636		Qsal	12088,4	5335,4	953
Profundidad	11,8	8,6	636	Profundidad	6,8	6,0	953		
13	DJ	175,8	34,8	491	14	DJ	356,1	3,9	143
	V	1,5	0,1	491		V	1,6	0,2	143
	DV	117,4	64,7	491		DV	315,8	8,9	143
	Vx	6,3	0,7	491		Vx	7,3	1,3	143
	T	24,1	2,8	491		T	14,1	2,5	143
	HR	49,7	10,2	491		HR	45,6	6,5	143
	ETo	5,5	0,6	491		ETo	2,0	0,0	143
	VolExp	10238773,9	364875,3	491		VolExp	8389804,2	109988,3	143
	Qent	4306,0	5422,9	491		Qent	67070,2	18894,5	143
	Qsal	28551,2	6657,6	491		Qsal	20947,5	650,3	143
Profundidad	7,3	6,4	491	Profundidad	9,1	7,7	143		
15	DJ	348,5	0,5	50	16	DJ	213,3	130,3	107
	V	0,9	0,1	50		V	2,5	0,3	107
	DV	9,5	9,6	50		DV	313,7	15,7	107
	Vx	5,0	0,6	50		Vx	9,2	0,8	107
	T	13,4	0,4	50		T	13,8	3,6	107
	HR	75,5	0,5	50		HR	34,1	2,6	107
	ETo	1,0	0,0	50		ETo	3,6	1,1	107
	VolExp	7969500,0	55053,3	50		VolExp	9959373,8	1226767,1	107
	Qent	147828,6	18276,3	50		Qent	37313,2	11491,0	107
	Qsal	20304,0	0,0	50		Qsal	19072,6	3227,8	107
Profundidad	9,0	7,8	50	Profundidad	8,5	7,7	107		
17	DJ	302,6	41,7	780	18	DJ	191,1	22,7	991
	V	1,1	0,2	780		V	1,1	0,1	991
	DV	293,1	36,6	780		DV	195,2	26,5	991
	Vx	4,4	1,0	780		Vx	4,4	0,6	991
	T	16,9	4,8	780		T	25,0	1,9	991
	HR	66,8	6,1	780		HR	61,4	7,2	991
	ETo	2,1	1,3	780		ETo	5,1	0,7	991
	VolExp	6875173,1	553062,7	780		VolExp	9864474,3	393407,5	991
	Qent	26266,0	21272,2	780		Qent	3326,8	4360,1	991
	Qsal	29996,2	6547,3	780		Qsal	33753,8	3912,7	991
Profundidad	5,9	4,3	780	Profundidad	5,3	3,7	991		
19	DJ	112,6	26,0	343	Total	DJ	186,9	97,6	8974
	V	1,6	0,2	343		V	1,3	0,3	8974
	DV	286,6	25,9	343		DV	235,9	79,1	8974
	Vx	6,8	0,7	343		Vx	5,2	1,4	8974
	T	17,7	3,1	343		T	19,1	6,3	8974
	HR	48,2	9,2	343		HR	58,2	12,8	8974
	ETo	4,4	1,1	343		ETo	3,6	1,7	8974
	VolExp	9631903,8	1404016,6	343		VolExp	8853793,2	1507776,4	8974
	Qent	26841,7	23950,8	343		Qent	20238,4	28660,2	8974
	Qsal	12341,6	8552,0	343		Qsal	27093,9	11742,2	8974
Profundidad	9,5	7,4	343	Profundidad	8,9	7,6	8974		

**Pruebas de igualdad de las medias de los grupos**

	Lambda de Wilks	F	gl1	gl2	Sig.
DJ	,168	2461,512	18	8955	,000
V	,393	767,037	18	8955	,000
DV	,431	657,354	18	8955	,000
Vx	,482	535,643	18	8955	,000
T	,329	1014,451	18	8955	,000
HR	,417	694,465	18	8955	,000
ETo	,447	615,423	18	8955	,000
VolExp	,209	1882,852	18	8955	,000
Qent	,226	1702,188	18	8955	,000
Qsal	,223	1731,591	18	8955	,000
Profundidad	,549	409,513	18	8955	,000

**Logaritmo de los determinantes**

No_Conglomerado	Rango	Logaritmo del determinante
1	11	70,493
2	4	.(a)
3	11	68,216
4	11	65,495
5	11	79,681
6	11	71,718
7	11	73,639
8	3	.(a)
9	11	75,228
10	11	73,800
11	11	79,013
12	11	75,576
13	11	73,249
14	7	.(a)
15	3	.(a)
16	5	.(a)
17	11	75,506
18	11	70,768
19	11	74,596
Intra-grupos combinada	11	82,567

Los rangos y logaritmos naturales de los determinantes impresos son los de las matrices de covarianzas de los grupos.

a Singular

**Resultados de la prueba(a)**

	M de Box	73519,393
	Aprox.	85,028
F	gl1	858
	gl2	25297377
	Sig.	,000

Contrasta la hipótesis nula de que las matrices de covarianzas poblacionales son iguales.

a Algunas matrices de covarianzas son singulares y el procedimiento ordinario no es válido. Los grupos no singulares se compararán con sus propias matrices de covarianzas intra-grupo combinadas. El logaritmo de su determinante es 82,408.

**Autovalores**

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	10,528(a)	47,8	47,8	,956
2	4,180(a)	19,0	66,7	,898
3	2,853(a)	12,9	79,7	,861
4	1,942(a)	8,8	88,5	,812
5	,894(a)	4,1	92,5	,687
6	,734(a)	3,3	95,9	,651
7	,379(a)	1,7	97,6	,524
8	,291(a)	1,3	98,9	,474
9	,172(a)	,8	99,7	,384
10	,044(a)	,2	99,9	,205
11	,026(a)	,1	100,0	,158

a Se han empleado las 11 primeras funciones discriminantes canónicas en el análisis.

**Lambda de Wilks**

Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1 a la 11	,000	76236,269	198	,000
2 a la 11	,002	54335,831	170	,000
3 a la 11	,012	39601,812	144	,000
4 a la 11	,046	27517,785	120	,000
5 a la 11	,136	17852,213	98	,000
6 a la 11	,258	12132,562	78	,000
7 a la 11	,448	7201,726	60	,000
8 a la 11	,617	4321,324	44	,000
9 a la 11	,797	2036,370	30	,000
10 a la 11	,934	611,112	18	,000
11	,975	227,639	8	,000

**Probabilidades previas para los grupos**

No conglomerado	Previas	Casos utilizados
1	,053	474
2	,010	92
3	,040	358
4	,036	324
5	,062	559
6	,057	514
7	,111	993
8	,005	45
9	,055	490
10	,070	631
11	,071	636
12	,106	953
13	,055	491
14	,016	143
15	,006	50
16	,012	107
17	,087	780
18	,110	991
19	,038	343
Total	1,000	8974

**Resultados de la clasificación(b,c)**

	No conglomerado	Grupo de pertenencia pronosticado																		Total	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		19
Original Recuento	1	383	0	0	0	8	42	41	0	0	0	0	0	0	0	0	0	0	0	0	474
	2	0	92	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	92
	3	0	0	312	0	2	0	0	0	3	0	0	0	18	0	0	0	0	23	0	358
	4	0	0	0	221	0	0	0	0	0	80	6	17	0	0	0	0	0	0	0	324
	5	1	0	22	0	356	11	15	0	51	0	0	0	0	6	0	0	85	12	0	559
	6	98	0	0	0	10	371	0	0	0	0	0	0	0	0	0	0	35	0	0	514
	7	19	0	49	0	16	19	770	0	0	0	0	0	0	0	0	0	81	39	0	993
	8	0	0	0	0	0	0	0	45	0	0	0	0	0	0	0	0	0	0	0	45
	9	0	0	0	0	4	0	0	0	392	7	48	0	13	0	0	0	0	24	2	490
	10	0	0	0	20	0	0	0	0	6	477	18	89	0	0	0	0	0	21	0	631
	11	0	0	0	0	0	0	0	0	2	36	373	156	0	0	0	0	0	69	0	636
	12	0	0	0	0	0	0	0	0	0	40	109	804	0	0	0	0	0	0	0	953
	13	0	0	0	0	0	0	0	0	10	0	0	0	380	0	0	0	0	88	13	491
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	143	0	0	0	0	0	143
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	50
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	96	0	0	11	107
	17	0	0	0	0	18	0	147	0	0	0	0	0	0	26	0	0	589	0	0	780
	18	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	988	0	991
	19	0	0	0	0	0	0	0	0	0	6	64	0	2	0	0	0	0	0	271	343
%	1	81	0	0	0	1,69	8,9	8,6	0	0	0	0	0	0	0	0	0	0	0	100	
	2	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	
	3	0	0	87	0	0,56	0	0	0	0,8	0	0	0	5	0	0	0	6,4	0	100	
	4	0	0	0	68	0	0	0	0	0	25	1,9	5,2	0	0	0	0	0	0	100	
	5	0,2	0	3,9	0	63,7	2	2,7	0	9,1	0	0	0	0	1,1	0	0	15	2,1	0	100
	6	19	0	0	0	1,95	72	0	0	0	0	0	0	0	0	0	0	6,8	0	0	100
	7	1,9	0	4,9	0	1,61	1,9	78	0	0	0	0	0	0	0	0	0	8,2	3,9	0	100
	8	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	100
	9	0	0	0	0	0,82	0	0	0	80	1,4	9,8	0	2,7	0	0	0	0	4,9	0,4	100
	10	0	0	0	3,2	0	0	0	0	1	76	2,9	14	0	0	0	0	0	3,3	0	100
	11	0	0	0	0	0	0	0	0	0,3	5,7	59	25	0	0	0	0	0	11	0	100
	12	0	0	0	0	0	0	0	0	0	4,2	11	84	0	0	0	0	0	0	0	100
	13	0	0	0	0	0	0	0	0	2	0	0	0	77	0	0	0	0	18	2,6	100
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	100
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	100
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	90	0	0	10	100
	17	0	0	0	0	2,31	0	19	0	0	0	0	0	0	3,3	0	0	76	0	0	100
	18	0	0	0	0	0	0	0	0	0,3	0	0	0	0	0	0	0	0	100	0	100
	19	0	0	0	0	0	0	0	0	1,7	19	0	0,6	0	0	0	0	0	0	79	100

	No conglomerado	Grupo de pertenencia pronosticado																		Total	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		19
Validación cruzada(a) Recuento	1	383	0	0	0	8	42	41	0	0	0	0	0	0	0	0	0	0	0	474	
	2	0	92	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	92	
	3	0	0	311	0	3	0	0	0	3	0	0	0	18	0	0	0	0	23	358	
	4	0	0	0	219	0	0	0	0	0	82	6	17	0	0	0	0	0	0	324	
	5	1	0	22	0	355	11	16	0	51	0	0	0	0	6	0	0	85	12	559	
	6	98	0	0	0	10	371	0	0	0	0	0	0	0	0	0	0	35	0	514	
	7	19	0	50	0	17	19	768	0	0	0	0	0	0	0	0	0	81	39	993	
	8	0	0	0	0	0	0	0	45	0	0	0	0	0	0	0	0	0	0	45	
	9	0	0	0	0	4	0	0	0	390	7	49	0	14	0	0	0	0	24	490	
	10	0	0	0	20	0	0	0	0	6	473	18	92	0	0	0	0	0	21	631	
	11	0	0	0	0	0	0	0	0	2	36	372	156	0	0	0	0	0	70	636	
	12	0	0	0	0	0	0	0	0	0	40	110	803	0	0	0	0	0	0	953	
	13	0	0	0	0	0	0	0	0	10	0	0	0	380	0	0	0	0	88	491	
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	143	0	0	0	0	143	
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	50	
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	96	0	11	107	
	17	0	0	0	0	18	0	147	0	0	0	0	0	0	27	0	0	588	0	780	
	18	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	987	991	
	19	0	0	0	0	0	0	0	0	6	66	0	2	1	0	0	0	0	268	343	
%	1	81	0	0	0	1,69	8,9	8,6	0	0	0	0	0	0	0	0	0	0	100		
	2	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100		
	3	0	0	87	0	0,84	0	0	0	0,8	0	0	0	5	0	0	0	6,4	100		
	4	0	0	0	68	0	0	0	0	0	25	1,9	5,2	0	0	0	0	0	100		
	5	0,2	0	3,9	0	63,5	2	2,9	0	9,1	0	0	0	0	1,1	0	0	15	2,1	100	
	6	19	0	0	0	1,95	72	0	0	0	0	0	0	0	0	0	0	6,8	0	100	
	7	1,9	0	5	0	1,71	1,9	77	0	0	0	0	0	0	0	0	0	8,2	3,9	100	
	8	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	100	
	9	0	0	0	0	0,82	0	0	0	80	1,4	10	0	2,9	0	0	0	0	4,9	0,4	100
	10	0	0	0	3,2	0	0	0	0	1	75	2,9	15	0	0	0	0	0	3,3	0,2	100
	11	0	0	0	0	0	0	0	0	0,3	5,7	58	25	0	0	0	0	0	11	0	100
	12	0	0	0	0	0	0	0	0	0	4,2	12	84	0	0	0	0	0	0	100	
	13	0	0	0	0	0	0	0	0	2	0	0	0	77	0	0	0	0	18	2,6	100
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	100	
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	100	
	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	90	0	10	100	
	17	0	0	0	0	2,31	0	19	0	0	0	0	0	0	3,5	0	0	75	0	100	
	18	0	0	0	0	0	0	0	0	0,4	0	0	0	0	0	0	0	0	100	0	100
	19	0	0	0	0	0	0	0	0	1,7	19	0	0,6	0,3	0	0	0	0	0	78	100

- a Cada caso se clasifica mediante las funciones derivadas a partir del resto de los casos.
- b Clasificados correctamente el 79,3% de los casos agrupados originales.
- c Clasificados correctamente el 79,1% de los casos agrupados validados mediante validación cruzada.

**ANEXO 28. ANÁLISIS DISCRIMINANTE PARA LAS VARIABLES  
SIN TRANSFORMAR, SIENDO ESTANDARIZADAS  
(19 GRUPOS), PERIODO COMPLETO DE ESTUDIO**

**a) Usando para la estandarización del periodo completo de estudio la media y la desviación estándar correspondientes al periodo de calibración.**

**Estadísticos de grupo**

No grupo	Variable	Media	Desv. Tip.	Casos	No grupo	Variable	Media	Desv. Tip.	Casos
1	Puntua(DJ)	0,385	0,308	474	2	Puntua(DJ)	1,122	0,012	92
	Puntua(V)	0,639	0,933	474		Puntua(V)	-1,106	0,153	92
	Puntua(DV)	0,079	0,470	474		Puntua(DV)	0,906	0,205	92
	Puntua(Vx)	0,603	1,098	474		Puntua(Vx)	-0,326	0,318	92
	Puntua(T)	0,925	0,436	474		Puntua(T)	-0,479	0,110	92
	Puntua(HR)	-1,043	0,626	474		Puntua(HR)	0,874	0,425	92
	Puntua(ETo)	1,138	0,695	474		Puntua(ETo)	-1,089	0,085	92
	Puntua(VolExp)	-0,807	0,372	474		Puntua(VolExp)	-0,764	0,173	92
	Puntua(Qent)	-0,718	0,269	474		Puntua(Qent)	4,841	0,968	92
	Puntua(Qsal)	1,314	0,418	474		Puntua(Qsal)	-0,633	0,247	92
	Puntua(Profundidad)	-0,245	0,774	474		Puntua(Profundidad)	-0,092	0,903	92
3	Puntua(DJ)	0,042	0,274	358	4	Puntua(DJ)	-1,255	0,114	324
	Puntua(V)	0,567	0,462	358		Puntua(V)	1,341	0,870	324
	Puntua(DV)	-2,190	0,501	358		Puntua(DV)	-0,289	1,749	324
	Puntua(Vx)	0,635	0,361	358		Puntua(Vx)	1,433	0,799	324
	Puntua(T)	1,369	0,330	358		Puntua(T)	-1,265	0,333	324
	Puntua(HR)	0,026	0,677	358		Puntua(HR)	-1,296	0,985	324
	Puntua(ETo)	1,566	0,719	358		Puntua(ETo)	-0,211	0,342	324
	Puntua(VolExp)	-0,228	0,725	358		Puntua(VolExp)	1,203	0,270	324
	Puntua(Qent)	-0,796	0,234	358		Puntua(Qent)	0,241	0,241	324
	Puntua(Qsal)	1,399	0,359	358		Puntua(Qsal)	-1,345	0,348	324
	Puntua(Profundidad)	-0,230	0,777	358		Puntua(Profundidad)	0,002	1,014	324
5	Puntua(DJ)	0,401	0,462	559	6	Puntua(DJ)	0,591	0,267	514
	Puntua(V)	-0,420	0,586	559		Puntua(V)	0,401	0,417	514
	Puntua(DV)	-0,398	0,971	559		Puntua(DV)	0,543	0,361	514
	Puntua(Vx)	-0,363	0,651	559		Puntua(Vx)	0,150	0,609	514
	Puntua(T)	0,693	0,813	559		Puntua(T)	0,344	0,762	514
	Puntua(HR)	0,366	0,662	559		Puntua(HR)	-0,817	0,582	514
	Puntua(ETo)	0,498	1,037	559		Puntua(ETo)	0,413	0,809	514
	Puntua(VolExp)	-0,641	0,706	559		Puntua(VolExp)	-1,413	0,220	514
	Puntua(Qent)	-0,535	0,477	559		Puntua(Qent)	-0,502	0,439	514
	Puntua(Qsal)	0,902	0,601	559		Puntua(Qsal)	0,780	0,371	514
	Puntua(Profundidad)	1,607	0,477	559		Puntua(Profundidad)	-0,253	0,752	514
7	Puntua(DJ)	0,302	0,324	993	8	Puntua(DJ)	1,152	0,074	45
	Puntua(V)	-0,595	0,444	993		Puntua(V)	-0,648	0,000	45
	Puntua(DV)	-0,713	0,989	993		Puntua(DV)	-3,557	0,155	45
	Puntua(Vx)	-0,364	0,598	993		Puntua(Vx)	0,197	0,595	45
	Puntua(T)	1,068	0,363	993		Puntua(T)	-0,707	0,301	45
	Puntua(HR)	0,603	0,488	993		Puntua(HR)	0,698	0,336	45
	Puntua(ETo)	0,575	0,690	993		Puntua(ETo)	-1,076	0,033	45
	Puntua(VolExp)	-0,614	0,525	993		Puntua(VolExp)	-0,539	0,602	45
	Puntua(Qent)	-0,755	0,178	993		Puntua(Qent)	0,469	0,403	45
	Puntua(Qsal)	1,174	0,373	993		Puntua(Qsal)	-0,263	0,096	45
	Puntua(Profundidad)	-0,476	0,485	993		Puntua(Profundidad)	-0,011	0,975	45
9	Puntua(DJ)	-0,247	0,305	490	10	Puntua(DJ)	-1,030	0,410	631
	Puntua(V)	-0,380	0,608	490		Puntua(V)	0,100	0,476	631
	Puntua(DV)	-1,031	0,628	490		Puntua(DV)	0,110	0,607	631
	Puntua(Vx)	-0,262	0,548	490		Puntua(Vx)	-0,009	0,605	631
	Puntua(T)	0,962	0,511	490		Puntua(T)	-0,360	0,845	631
	Puntua(HR)	0,191	0,709	490		Puntua(HR)	-0,921	0,706	631
	Puntua(ETo)	1,372	0,481	490		Puntua(ETo)	0,444	1,024	631
	Puntua(VolExp)	1,387	0,363	490		Puntua(VolExp)	1,266	0,524	631
	Puntua(Qent)	-0,794	0,160	490		Puntua(Qent)	-0,287	0,498	631
	Puntua(Qsal)	0,259	0,672	490		Puntua(Qsal)	-0,858	0,546	631
	Puntua(Profundidad)	1,755	0,562	490		Puntua(Profundidad)	-0,388	0,602	631



No grupo	Variable	Media	Desv. típica	No de casos	No grupo	Variable	Media	Desv. típica	No de casos
11	Puntua(DJ)	-0,940	0,422	636	12	Puntua(DJ)	-1,297	0,242	953
	Puntua(V)	-0,622	0,614	636		Puntua(V)	-0,490	0,675	953
	Puntua(DV)	-0,616	0,956	636		Puntua(DV)	0,151	0,459	953
	Puntua(Vx)	-0,328	0,683	636		Puntua(Vx)	-0,603	0,634	953
	Puntua(T)	-0,181	0,726	636		Puntua(T)	-0,875	0,333	953
	Puntua(HR)	0,729	0,635	636		Puntua(HR)	0,811	0,580	953
	Puntua(ETo)	0,136	0,828	636		Puntua(ETo)	-0,595	0,459	953
	Puntua(VolExp)	1,300	0,639	636		Puntua(VolExp)	1,028	0,610	953
	Puntua(Qent)	-0,364	0,542	636		Puntua(Qent)	0,211	0,214	953
	Puntua(Qsal)	-0,795	0,579	636		Puntua(Qsal)	-1,194	0,474	953
	Puntua(Profundidad)	0,340	1,113	636		Puntua(Profundidad)	-0,303	0,774	953
13	Puntua(DJ)	-0,261	0,284	491	14	Puntua(DJ)	1,210	0,032	143
	Puntua(V)	0,585	0,432	491		Puntua(V)	1,027	0,516	143
	Puntua(DV)	-2,096	0,908	491		Puntua(DV)	0,685	0,124	143
	Puntua(Vx)	0,695	0,444	491		Puntua(Vx)	1,360	0,842	143
	Puntua(T)	1,105	0,432	491		Puntua(T)	-0,427	0,383	143
	Puntua(HR)	-0,587	0,758	491		Puntua(HR)	-0,892	0,480	143
	Puntua(ETo)	1,739	0,387	491		Puntua(ETo)	-0,561	0,061	143
	Puntua(VolExp)	1,341	0,278	491		Puntua(VolExp)	-0,067	0,084	143
	Puntua(Qent)	-0,796	0,155	491		Puntua(Qent)	0,999	0,540	143
	Puntua(Qsal)	0,269	0,592	491		Puntua(Qsal)	-0,407	0,058	143
	Puntua(Profundidad)	-0,243	0,828	491		Puntua(Profundidad)	-0,006	0,994	143
15	Puntua(DJ)	1,148	0,004	50	16	Puntua(DJ)	0,045	1,063	107
	Puntua(V)	-1,106	0,154	50		Puntua(V)	3,481	0,780	107
	Puntua(DV)	-3,609	0,135	50		Puntua(DV)	0,656	0,220	107
	Puntua(Vx)	-0,165	0,365	50		Puntua(Vx)	2,646	0,509	107
	Puntua(T)	-0,528	0,062	50		Puntua(T)	-0,461	0,545	107
	Puntua(HR)	1,330	0,038	50		Puntua(HR)	-1,743	0,193	107
	Puntua(ETo)	-1,138	0,033	50		Puntua(ETo)	0,667	0,732	107
	Puntua(VolExp)	-0,387	0,042	50		Puntua(VolExp)	1,128	0,934	107
	Puntua(Qent)	3,309	0,523	50		Puntua(Qent)	0,148	0,329	107
	Puntua(Qsal)	-0,464	0,000	50		Puntua(Qsal)	-0,573	0,287	107
	Puntua(Profundidad)	-0,022	1,010	50		Puntua(Profundidad)	-0,076	0,994	107
17	Puntua(DJ)	0,774	0,340	780	18	Puntua(DJ)	-0,136	0,185	991
	Puntua(V)	-0,636	0,678	780		Puntua(V)	-0,584	0,427	991
	Puntua(DV)	0,367	0,513	780		Puntua(DV)	-1,006	0,371	991
	Puntua(Vx)	-0,520	0,670	780		Puntua(Vx)	-0,520	0,410	991
	Puntua(T)	0,010	0,743	780		Puntua(T)	1,243	0,286	991
	Puntua(HR)	0,684	0,454	780		Puntua(HR)	0,282	0,534	991
	Puntua(ETo)	-0,382	0,783	780		Puntua(ETo)	1,553	0,350	991
	Puntua(VolExp)	-1,220	0,421	780		Puntua(VolExp)	1,056	0,300	991
	Puntua(Qent)	-0,168	0,608	780		Puntua(Qent)	-0,824	0,125	991
	Puntua(Qsal)	0,398	0,582	780		Puntua(Qsal)	0,732	0,348	991
	Puntua(Profundidad)	-0,417	0,558	780		Puntua(Profundidad)	-0,492	0,480	991
19	Puntua(DJ)	-0,776	0,212	343	Total	Puntua(DJ)	-0,169	0,796	8974
	Puntua(V)	0,918	0,474	343		Puntua(V)	-0,090	0,925	8974
	Puntua(DV)	0,276	0,363	343		Puntua(DV)	-0,436	1,109	8974
	Puntua(Vx)	1,066	0,462	343		Puntua(Vx)	-0,022	0,901	8974
	Puntua(T)	0,138	0,478	343		Puntua(T)	0,341	0,960	8974
	Puntua(HR)	-0,695	0,682	343		Puntua(HR)	0,042	0,952	8974
	Puntua(ETo)	1,022	0,684	343		Puntua(ETo)	0,534	1,041	8974
	Puntua(VolExp)	0,879	1,069	343		Puntua(VolExp)	0,286	1,148	8974
	Puntua(Qent)	-0,152	0,685	343		Puntua(Qent)	-0,340	0,820	8974
	Puntua(Qsal)	-1,172	0,760	343		Puntua(Qsal)	0,140	1,044	8974
	Puntua(Profundidad)	0,043	0,953	343		Puntua(Profundidad)	-0,033	0,982	8974

**Pruebas de igualdad de las medias de los grupos**

	Lambda de Wilks	F	gl1	gl2	Sig.
Puntua(DJ)	0,168	2461,512	18	8955	0,000
Puntua(V)	0,393	767,019	18	8955	0,000
Puntua(DV)	0,431	657,351	18	8955	0,000
Puntua(Vx)	0,482	535,642	18	8955	0,000
Puntua(T)	0,329	1014,446	18	8955	0,000
Puntua(HR)	0,417	694,463	18	8955	0,000
Puntua(ETo)	0,428	664,836	18	8955	0,000
Puntua(VolExp)	0,209	1882,873	18	8955	0,000
Puntua(Qent)	0,226	1702,253	18	8955	0,000
Puntua(Qsal)	0,223	1731,541	18	8955	0,000
Puntua(Profundidad)	0,549	409,511	18	8955	0,000

**Logaritmo de los determinantes**

No_Conglomerado	Rango	Logaritmo del determinante
1	11	-29,816
2	5	.(a)
3	11	-30,551
4	11	-34,115
5	11	-19,134
6	11	-27,662
7	11	-24,960
8	3	.(a)
9	11	-23,611
10	11	-25,556
11	11	-19,674
12	11	-23,489
13	11	-25,180
14	6	.(a)
15	3	.(a)
16	5	.(a)
17	11	-23,630
18	11	-28,239
19	11	-24,650
Intra-grupos combinada	11	-16,218

Los rangos y logaritmos naturales de los determinantes impresos son los de las matrices de covarianzas de los grupos.  
a Singular

**Resultados de la prueba(a)**

	M de Box	75634,937
	Aprox.	87,475
F	gl1	858
	gl2	25297377
	Sig.	0

Contrasta la hipótesis nula de que las matrices de covarianzas poblacionales son iguales.

a Algunas matrices de covarianzas son singulares y el procedimiento ordinario no es válido. Los grupos no singulares se compararán con sus propias matrices de covarianzas intra-grupo combinadas. El logaritmo de su determinante es -15,785.

**Autovalores**

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	10,533	47,815	47,815	0,956
2	4,158	18,876	66,691	0,898
3	2,837	12,878	79,569	0,860
4	1,937	8,793	88,362	0,812
5	0,893	4,055	92,417	0,687
6	0,733	3,327	95,743	0,650
7	0,378	1,716	97,460	0,524
8	0,286	1,296	98,756	0,471
9	0,173	0,786	99,542	0,384
10	0,074	0,337	99,879	0,263
11	0,027	0,121	100,000	0,161

a Se han empleado las 11 primeras funciones discriminantes canónicas en el análisis.

**Lambda de Wilks**

Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1 a la 11	0,00020	76369,855	198	0
2 a la 11	0,00229	54465,681	170	0
3 a la 11	0,01180	39769,424	144	0
4 a la 11	0,04528	27724,205	120	0
5 a la 11	0,13298	18073,227	98	0
6 a la 11	0,25176	12355,556	78	0
7 a la 11	0,43625	7431,059	60	0
8 a la 11	0,60120	4558,092	44	0
9 a la 11	0,77285	2308,161	30	0
10 a la 11	0,90673	877,061	18	0
11	0,97395	236,439	8	0

**Probabilidades previas para los grupos**

No Conglomerado	Previas	Casos en el análisis
1	0,053	474
2	0,010	92
3	0,040	358
4	0,036	324
5	0,062	559
6	0,057	514
7	0,111	993
8	0,005	45
9	0,055	490
10	0,070	631
11	0,071	636
12	0,106	953
13	0,055	491
14	0,016	143
15	0,006	50
16	0,012	107
17	0,087	780
18	0,110	991
19	0,038	343
Total	1	8974

**Resultados de la clasificación(b,c)**

	No grupo	Grupo de pertenencia pronosticado																		Total		
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		19	
Original Recuento	1	382				11	41	40													474	
	2		92																		92	
	3			312		2				3				6					35		358	
	4				238						65	4	17								324	
	5			21		358	11	16		49					5			86	13		559	
	6	101				10	368											35			514	
	7	4		42		16	19	824										49	39		993	
	8								45												45	
	9					4				394	5	48		13						24	2	490
	10				19					6	487	21	77							21		631
	11									3	33	355	158							75	12	636
	12										39	108	791								15	953
	13									10				380						88	13	491
	14														143							143
	15															50						50
	16																96				11	107
	17					22		163									26		569			780
	18									5										986		991
	19									4	61		2								276	343
%	1	80,6				2,3	8,6	8,4													100	
	2		100																		100	
	3			87,2		0,6				0,8				1,7					9,8		100	
	4				73,5						20,1	1,2	5,2								100	
	5			3,8		64,0	2,0	2,9		8,8					0,9			15,4	2,3		100	
	6	19,6				1,9	71,6											6,8			100	
	7	0,4		4,2		1,6	1,9	83,0										4,9	3,9		100	
	8								100												100	
	9					0,8				80,4	1,0	9,8		2,7						4,9	0,4	100
	10				3,0					1,0	77,2	3,3	12,2							3,3		100
	11									0,5	5,2	55,8	24,8							11,8	1,9	100
	12										4,1	11,3	83,0								1,6	100
	13									2,0				77,4						17,9	2,6	100
	14														100							100
	15															100						100
	16																89,7				10,3	100
	17					2,8		20,9									3,3		72,9			100
	18									0,5										99,5		100
	19									1,2	17,8		0,6								80,5	100

### Resultados de la clasificación(a)

	No grupo	Grupo de pertenencia pronosticado																	Total			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17		18	19	
Validación cruzada (a) Recuento	1	382				11	41	40													474	
	2		92																		92	
	3			311		3				3				6					35		358	
	4				236						67	4	17								324	
	5			21		355	13	17		49					5			86	13		559	
	6	102				11	366											35			514	
	7	5		43		16	19	822										49	39		993	
	8								45												45	
	9					4				393	5	48		13						25	2	490
	10				20					6	483	21	80							21		631
	11									3	33	353	158							76	13	636
	12										39	108	789								17	953
	13									10				380						88	13	491
	14														143							143
	15															50						50
	16																96				11	107
	17					22		163									26		569			780
	18									5										986		991
	19									4	62		2								275	343
%	1	80,6				2,3	8,6	8,4													100	
	2		100																		100	
	3			86,9		0,8				0,8				1,7					9,8		100	
	4				72,8						20,7	1,2	5,2								100	
	5			3,8		63,5	2,3	3,0		8,8					0,9			15,4	2,3		100	
	6	19,8				2,1	71,2											6,8			100	
	7	0,5		4,3		1,6	1,9	82,8										4,9	3,9		100	
	8								100												100	
	9					0,8				80,2	1,0	9,8		2,7						5,1	0,4	100
	10				3,2					1,0	76,5	3,3	12,7							3,3		100
	11									0,5	5,2	55,5	24,8							11,9	2,0	100
	12										4,1	11,3	82,8								1,8	100
	13									2,0				77,4						17,9	2,6	100
	14														100							100
	15															100						100
	16																89,7				10,3	100
	17					2,8		20,9								3,3		72,9				100
	18									0,5										99,5		100
	19									1,2	18,1		0,6								80,2	100

- a En la validación cruzada, cada caso se clasifica a partir del resto de los casos.
- b Clasificados correctamente el 79,6% de los casos agrupados originales.
- c Clasificados correctamente el 79,4% de los casos agrupados validados mediante validación cruzada.

**b) Efectuando previamente la estandarización de los periodos de calibración y validación por separado, y luego uniéndolos para realizar el análisis.**

**Estadísticos de grupo**

No grupo	Variable	Media	Desv. Tip.	Casos	No grupo	Variable	Media	Desv. Tip.	Casos
1	Puntua(DJ)	0,803	0,198	474	2	Puntua(DJ)	1,122	0,012	92
	Puntua(V)	0,745	0,877	474		Puntua(V)	-1,106	0,153	92
	Puntua(DV)	0,508	0,287	474		Puntua(DV)	0,906	0,205	92
	Puntua(Vx)	0,580	1,137	474		Puntua(Vx)	-0,326	0,318	92
	Puntua(T)	0,663	0,269	474		Puntua(T)	-0,479	0,109	92
	Puntua(HR)	-1,130	0,610	474		Puntua(HR)	0,874	0,425	92
	Puntua(ETo)	0,753	0,466	474		Puntua(ETo)	-1,085	0,000	92
	Puntua(VolExp)	-0,991	0,268	474		Puntua(VolExp)	-0,764	0,173	92
	Puntua(Qent)	-0,561	0,239	474		Puntua(Qent)	4,841	0,968	92
	Puntua(Qsal)	1,132	0,347	474		Puntua(Qsal)	-0,633	0,247	92
	Puntua(Profundidad)	-0,217	0,781	474		Puntua(Profundidad)	-0,092	0,903	92
3	Puntua(DJ)	0,678	0,358	358	4	Puntua(DJ)	-1,623	0,211	324
	Puntua(V)	0,874	0,572	358		Puntua(V)	1,564	0,789	324
	Puntua(DV)	-1,340	0,596	358		Puntua(DV)	0,235	1,538	324
	Puntua(Vx)	0,839	0,458	358		Puntua(Vx)	1,627	0,818	324
	Puntua(T)	0,936	0,272	358		Puntua(T)	-1,984	0,672	324
	Puntua(HR)	-0,065	0,748	358		Puntua(HR)	-1,399	0,962	324
	Puntua(ETo)	0,692	0,689	358		Puntua(ETo)	-1,072	0,906	324
	Puntua(VolExp)	-0,636	0,577	358		Puntua(VolExp)	0,772	0,195	324
	Puntua(Qent)	-0,471	0,208	358		Puntua(Qent)	1,528	1,103	324
	Puntua(Qsal)	1,076	0,333	358		Puntua(Qsal)	-1,447	0,358	324
	Puntua(Profundidad)	-0,175	0,803	358		Puntua(Profundidad)	0,042	1,035	324
5	Puntua(DJ)	0,792	0,317	559	6	Puntua(DJ)	1,163	0,311	514
	Puntua(V)	-0,351	0,632	559		Puntua(V)	0,564	0,523	514
	Puntua(DV)	0,075	0,637	559		Puntua(DV)	0,962	0,503	514
	Puntua(Vx)	-0,378	0,705	559		Puntua(Vx)	0,247	0,711	514
	Puntua(T)	0,423	0,624	559		Puntua(T)	0,033	0,554	514
	Puntua(HR)	0,341	0,693	559		Puntua(HR)	-0,897	0,590	514
	Puntua(ETo)	-0,061	0,701	559		Puntua(ETo)	-0,142	0,450	514
	Puntua(VolExp)	-0,868	0,536	559		Puntua(VolExp)	-1,517	0,265	514
	Puntua(Qent)	-0,338	0,404	559		Puntua(Qent)	-0,284	0,360	514
	Puntua(Qsal)	0,722	0,530	559		Puntua(Qsal)	0,617	0,362	514
	Puntua(Profundidad)	1,677	0,499	559		Puntua(Profundidad)	-0,221	0,775	514
7	Puntua(DJ)	0,731	0,242	993	8	Puntua(DJ)	1,152	0,074	45
	Puntua(V)	-0,543	0,483	993		Puntua(V)	-0,648	0,000	45
	Puntua(DV)	-0,209	0,643	993		Puntua(DV)	-3,557	0,155	45
	Puntua(Vx)	-0,388	0,642	993		Puntua(Vx)	0,197	0,595	45
	Puntua(T)	0,784	0,229	993		Puntua(T)	-0,707	0,301	45
	Puntua(HR)	0,602	0,500	993		Puntua(HR)	0,698	0,336	45
	Puntua(ETo)	-0,045	0,581	993		Puntua(ETo)	-1,085	0,000	45
	Puntua(VolExp)	-0,841	0,391	993		Puntua(VolExp)	-0,539	0,602	45
	Puntua(Qent)	-0,549	0,236	993		Puntua(Qent)	0,469	0,403	45
	Puntua(Qsal)	0,986	0,415	993		Puntua(Qsal)	-0,263	0,096	45
	Puntua(Profundidad)	-0,448	0,493	993		Puntua(Profundidad)	-0,011	0,975	45
9	Puntua(DJ)	-0,211	0,328	490	10	Puntua(DJ)	-1,026	0,439	631
	Puntua(V)	-0,332	0,641	490		Puntua(V)	0,195	0,478	631
	Puntua(DV)	-0,651	0,804	490		Puntua(DV)	0,522	0,356	631
	Puntua(Vx)	-0,279	0,620	490		Puntua(Vx)	-0,032	0,646	631
	Puntua(T)	0,644	0,851	490		Puntua(T)	-0,733	0,596	631
	Puntua(HR)	0,156	0,759	490		Puntua(HR)	-1,011	0,723	631
	Puntua(ETo)	0,930	0,830	490		Puntua(ETo)	0,090	0,578	631
	Puntua(VolExp)	1,059	0,344	490		Puntua(VolExp)	0,887	0,222	631
	Puntua(Qent)	-0,585	0,404	490		Puntua(Qent)	-0,037	0,363	631
	Puntua(Qsal)	0,151	0,739	490		Puntua(Qsal)	-0,972	0,443	631
	Puntua(Profundidad)	1,814	0,575	490		Puntua(Profundidad)	-0,362	0,611	631

No grupo	Variable	Media	Desv. típica	No de casos	No grupo	Variable	Media	Desv. típica	No de casos
11	Puntua(DJ)	-0,940	0,470	636	12	Puntua(DJ)	-1,530	0,203	953
	Puntua(V)	-0,605	0,623	636		Puntua(V)	-0,464	0,707	953
	Puntua(DV)	-0,075	0,636	636		Puntua(DV)	0,535	0,451	953
	Puntua(Vx)	-0,345	0,773	636		Puntua(Vx)	-0,642	0,703	953
	Puntua(T)	-0,651	0,458	636		Puntua(T)	-1,359	0,518	953
	Puntua(HR)	0,736	0,658	636		Puntua(HR)	0,815	0,623	953
	Puntua(ETo)	-0,534	0,636	636		Puntua(ETo)	-1,255	0,619	953
	Puntua(VolExp)	0,824	0,341	636		Puntua(VolExp)	0,680	0,326	953
	Puntua(Qent)	-0,044	0,497	636		Puntua(Qent)	1,133	0,940	953
	Puntua(Qsal)	-0,933	0,512	636		Puntua(Qsal)	-1,276	0,487	953
	Puntua(Profundidad)	0,372	1,099	636		Puntua(Profundidad)	-0,273	0,805	953
13	Puntua(DJ)	-0,184	0,264	491	14	Puntua(DJ)	1,210	0,032	143
	Puntua(V)	0,759	0,488	491		Puntua(V)	1,027	0,516	143
	Puntua(DV)	-1,598	0,779	491		Puntua(DV)	0,685	0,124	143
	Puntua(Vx)	0,849	0,573	491		Puntua(Vx)	1,360	0,842	143
	Puntua(T)	0,761	0,732	491		Puntua(T)	-0,427	0,383	143
	Puntua(HR)	-0,674	0,807	491		Puntua(HR)	-0,892	0,480	143
	Puntua(ETo)	1,255	0,756	491		Puntua(ETo)	-0,445	0,000	143
	Puntua(VolExp)	0,951	0,289	491		Puntua(VolExp)	-0,067	0,084	143
	Puntua(Qent)	-0,515	0,488	491		Puntua(Qent)	0,999	0,540	143
	Puntua(Qsal)	0,133	0,660	491		Puntua(Qsal)	-0,407	0,058	143
	Puntua(Profundidad)	-0,209	0,855	491		Puntua(Profundidad)	-0,006	0,994	143
15	Puntua(DJ)	1,148	0,004	50	16	Puntua(DJ)	-0,104	1,198	107
	Puntua(V)	-1,106	0,154	50		Puntua(V)	3,937	0,699	107
	Puntua(DV)	-3,609	0,135	50		Puntua(DV)	1,107	0,398	107
	Puntua(Vx)	-0,165	0,365	50		Puntua(Vx)	3,085	0,716	107
	Puntua(T)	-0,528	0,062	50		Puntua(T)	-0,961	0,456	107
	Puntua(HR)	1,330	0,038	50		Puntua(HR)	-1,897	0,180	107
	Puntua(ETo)	-1,085	0,000	50		Puntua(ETo)	-0,051	0,366	107
	Puntua(VolExp)	-0,387	0,042	50		Puntua(VolExp)	0,671	0,535	107
	Puntua(Qent)	3,309	0,523	50		Puntua(Qent)	1,146	0,817	107
	Puntua(Qsal)	-0,464	0,000	50		Puntua(Qsal)	-0,695	0,342	107
	Puntua(Profundidad)	-0,022	1,010	50		Puntua(Profundidad)	-0,042	1,011	107
17	Puntua(DJ)	1,146	0,227	780	18	Puntua(DJ)	0,038	0,139	991
	Puntua(V)	-0,618	0,682	780		Puntua(V)	-0,542	0,449	991
	Puntua(DV)	0,654	0,387	780		Puntua(DV)	-0,592	0,533	991
	Puntua(Vx)	-0,541	0,697	780		Puntua(Vx)	-0,568	0,476	991
	Puntua(T)	-0,198	0,519	780		Puntua(T)	0,977	0,510	991
	Puntua(HR)	0,676	0,465	780		Puntua(HR)	0,244	0,576	991
	Puntua(ETo)	-0,819	0,440	780		Puntua(ETo)	1,068	0,530	991
	Puntua(VolExp)	-1,293	0,464	780		Puntua(VolExp)	0,755	0,499	991
	Puntua(Qent)	-0,025	0,490	780		Puntua(Qent)	-0,655	0,244	991
	Puntua(Qsal)	0,286	0,473	780		Puntua(Qsal)	0,584	0,341	991
	Puntua(Profundidad)	-0,398	0,573	780		Puntua(Profundidad)	-0,467	0,492	991
19	Puntua(DJ)	-0,797	0,348	343	Total	Puntua(DJ)	0,000	1,000	8974
	Puntua(V)	1,177	0,637	343		Puntua(V)	0,000	1,000	8974
	Puntua(DV)	0,768	0,589	343		Puntua(DV)	0,000	1,000	8974
	Puntua(Vx)	1,251	0,548	343		Puntua(Vx)	0,000	1,000	8974
	Puntua(T)	-0,305	0,334	343		Puntua(T)	0,000	1,000	8974
	Puntua(HR)	-0,827	0,779	343		Puntua(HR)	0,000	1,000	8974
	Puntua(ETo)	0,448	0,654	343		Puntua(ETo)	0,000	1,000	8974
	Puntua(VolExp)	0,412	0,699	343		Puntua(VolExp)	0,000	1,000	8974
	Puntua(Qent)	0,297	0,797	343		Puntua(Qent)	0,000	1,000	8974
	Puntua(Qsal)	-1,301	0,664	343		Puntua(Qsal)	0,000	1,000	8974
	Puntua(Profundidad)	0,081	0,968	343		Puntua(Profundidad)	0,000	1,000	8974



**Pruebas de igualdad de las medias de los grupos**

	Lambda de Wilks	F	gl1	gl2	Sig.
Puntua(DJ)	0,099	4553,108	18	8955	0,000
Puntua(V)	0,363	874,649	18	8955	0,000
Puntua(DV)	0,377	822,869	18	8955	0,000
Puntua(Vx)	0,468	565,033	18	8955	0,000
Puntua(T)	0,269	1349,535	18	8955	0,000
Puntua(HR)	0,413	707,466	18	8955	0,000
Puntua(ETo)	0,367	859,115	18	8955	0,000
Puntua(VolExp)	0,162	2581,634	18	8955	0,000
Puntua(Qent)	0,291	1214,708	18	8955	0,000
Puntua(Qsal)	0,222	1747,332	18	8955	0,000
Puntua(Profundidad)	0,549	409,300	18	8955	0,000

**Logaritmo de los determinantes**

No_Conglomerado	Rango	Logaritmo del determinante
1	11	-28,439
2	5	.(a)
3	11	-28,154
4	11	-23,451
5	11	-18,107
6	11	-26,367
7	11	-24,397
8	2	.(a)
9	11	-19,225
10	11	-24,278
11	11	-17,985
12	11	-19,604
13	11	-19,772
14	6	.(a)
15	3	.(a)
16	5	.(a)
17	11	-21,580
18	11	-25,717
19	11	-22,039
Intra-grupos combinada	11	-15,719

Los rangos y logaritmos naturales de los determinantes impresos son los de las matrices de covarianzas de los grupos.

a Singular

### Resultados de la prueba(a)

	M de Box	58651,1
	Aprox.	67,832
F	gl1	858
	gl2	25297377
	Sig.	0,000

Contrasta la hipótesis nula de que las matrices de covarianzas poblacionales son iguales.

a Algunas matrices de covarianzas son singulares y el procedimiento ordinario no es válido. Los grupos no singulares se compararán con sus propias matrices de covarianzas intra-grupo combinadas. El logaritmo de su determinante es -15,785.

### Autovalores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	11,988	52,068	52,068	0,961
2	3,590	15,593	67,661	0,884
3	2,847	12,367	80,028	0,860
4	1,777	7,717	87,746	0,800
5	0,973	4,227	91,973	0,702
6	0,688	2,988	94,961	0,638
7	0,582	2,530	97,491	0,607
8	0,290	1,261	98,752	0,474
9	0,220	0,956	99,708	0,425
10	0,043	0,187	99,895	0,203
11	0,024	0,105	100	0,153

a Se han empleado las 11 primeras funciones discriminantes canónicas en el análisis.

### Lambda de Wilks

Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1 a la 11	0,000	77384,540	198	0
2 a la 11	0,002	54416,135	170	0
3 a la 11	0,011	40765,352	144	0
4 a la 11	0,041	28695,538	120	0
5 a la 11	0,113	19546,962	98	0
6 a la 11	0,223	13458,458	78	0
7 a la 11	0,376	8768,736	60	0
8 a la 11	0,595	4657,171	44	0
9 a la 11	0,767	2373,917	30	0
10 a la 11	0,936	591,353	18	5,78E-114
11	0,977	212,980	8	1,17E-41

**Probabilidades previas para los grupos**

No Conglomerado	Previas	Casos en el análisis
1	0,053	474
2	0,010	92
3	0,040	358
4	0,036	324
5	0,062	559
6	0,057	514
7	0,111	993
8	0,005	45
9	0,055	490
10	0,070	631
11	0,071	636
12	0,106	953
13	0,055	491
14	0,016	143
15	0,006	50
16	0,012	107
17	0,087	780
18	0,110	991
19	0,038	343
Total	1	8974

**Resultados de la clasificación(b,c)**

	No grupo	Grupo de pertenencia pronosticado																		Total		
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		19	
Original Recuento	1	381	0	0	0	9	64	20	0	0	0	0	0	0	0	0	0	0	0	0	474	
	2	0	92	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	92	
	3	0	0	334	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	21	358	
	4	0	0	0	308	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	13	324
	5	2	0	11	0	402	12	17	0	34	0	0	0	0	15	0	0	55	11	0	559	
	6	38	0	0	0	1	475	0	0	0	0	0	0	0	0	0	0	0	0	0	514	
	7	36	0	40	0	9	0	908	0	0	0	0	0	0	0	0	0	0	0	0	993	
	8	0	0	0	0	0	0	0	45	0	0	0	0	0	0	0	0	0	0	0	45	
	9	0	0	0	0	0	0	0	0	378	17	56	0	15	0	0	0	0	22	2	490	
	10	0	0	0	0	0	0	0	0	3	577	10	18	0	0	0	0	0	21	2	631	
	11	0	0	0	0	0	0	0	0	2	35	455	82	0	0	0	0	0	60	2	636	
	12	0	0	0	0	0	0	0	0	0	26	49	878	0	0	0	0	0	0	0	953	
	13	0	0	0	11	0	0	0	0	13	0	0	0	379	0	0	0	0	88	0	491	
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	143	0	0	0	0	0	143	
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	50	
	16	0	0	0	11	0	0	0	0	0	0	0	0	0	10	86	0	0	0	0	107	
	17	0	0	0	0	10	0	19	0	0	0	0	0	0	43	0	0	708	0	0	780	
	18	0	0	0	0	0	0	13	0	6	0	0	0	0	0	0	0	0	972	0	991	
	19	0	0	0	0	0	0	0	0	0	37	0	0	0	0	0	0	0	0	306	343	
%	1	80,4	0	0	0	1,9	13,5	4,2	0	0	0	0	0	0	0	0	0	0	0	100		
	2	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100		
	3	0	0	93,3	0	0	0	0,3	0	0,3	0	0	0	0,3	0	0	0	0	5,9	0	100	
	4	0	0	0	95,1	0	0	0	0	0	0,9	0	0	0	0	0	0	0	0	4,0	100	
	5	0,4	0	2,0	0	71,9	2,1	3,0	0	6,1	0	0	0	0	2,7	0	0	9,8	2,0	0	100	
	6	7,4	0	0	0	0,2	92,4	0	0	0	0	0	0	0	0	0	0	0	0	0	100	
	7	3,6	0	4,0	0	0,9	0	91,4	0	0	0	0	0	0	0	0	0	0	0	0	100	
	8	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	100	
	9	0	0	0	0	0	0	0	0	77,1	3,5	11,4	0	3,1	0	0	0	0	4,5	0,4	100	
	10	0	0	0	0	0	0	0	0	0,5	91,4	1,6	2,9	0	0	0	0	0	3,3	0,3	100	
	11	0	0	0	0	0	0	0	0	0,3	5,5	71,5	12,9	0	0	0	0	0	9,4	0,3	100	
	12	0	0	0	0	0	0	0	0	0	2,7	5,1	92,1	0	0	0	0	0	0	0	100	
	13	0	0	0	2,2	0	0	0	0	2,6	0	0	0	77,2	0	0	0	0	17,9	0	100	
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	100	
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	100	
	16	0	0	0	10,3	0	0	0	0	0	0	0	0	0	9,3	0	80,4	0	0	0	100	
	17	0	0	0	0	1,3	0	2,4	0	0	0	0	0	0	5,5	0	0	90,8	0	0	100	
	18	0	0	0	0	0	0	1,3	0	0,6	0	0	0	0	0	0	0	0	98,1	0	100	
	19	0	0	0	0	0	0	0	0	0	10,8	0	0	0	0	0	0	0	0	89,2	100	

### Resultados de la clasificación(a)

	No grupo	Grupo de pertenencia pronosticado																		Total		
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		19	
Validación cruzada (a) Recuento	1	381	0	0	0	9	64	20	0	0	0	0	0	0	0	0	0	0	0	0	474	
	2	0	92	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	92	
	3	0	0	334	0	0	0	1	0	1	0	0	0	1	0	0	0	0	21	0	358	
	4	0	0	0	308	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	13	324
	5	2	0	13	0	398	12	17	0	34	0	0	0	0	15	0	0	56	12	0	559	
	6	38	0	0	0	1	475	0	0	0	0	0	0	0	0	0	0	0	0	0	0	514
	7	39	0	40	0	9	0	905	0	0	0	0	0	0	0	0	0	0	0	0	0	993
	8	0	0	0	0	0	0	0	45	0	0	0	0	0	0	0	0	0	0	0	0	45
	9	0	0	0	0	0	0	0	0	377	17	56	0	15	0	0	0	0	23	2	490	
	10	0	0	0	0	0	0	0	0	3	577	10	18	0	0	0	0	0	21	2	631	
	11	0	0	0	0	0	0	0	0	2	36	454	82	0	0	0	0	0	60	2	636	
	12	0	0	0	0	0	0	0	0	0	26	49	878	0	0	0	0	0	0	0	0	953
	13	0	0	0	11	0	0	0	0	14	0	0	0	378	0	0	0	0	88	0	491	
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	143	0	0	0	0	0	0	143
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	50
	16	0	0	0	11	0	0	0	0	0	0	0	0	0	14	0	82	0	0	0	0	107
	17	0	0	0	0	10	0	19	0	0	0	0	0	0	46	0	0	705	0	0	0	780
	18	0	0	0	0	0	0	13	0	6	0	0	0	0	0	0	0	0	972	0	0	991
	19	0	0	0	0	0	0	0	0	0	39	0	0	0	0	0	0	0	0	0	304	343
%	1	80,4	0	0	0	1,9	13,5	4,2	0	0	0	0	0	0	0	0	0	0	0	0	100	
	2	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	
	3	0	0	93,3	0	0	0	0,3	0	0,3	0	0	0	0,3	0	0	0	0	5,9	0	100	
	4	0	0	0	95,1	0	0	0	0	0	0,9	0	0	0	0	0	0	0	0	4,0	100	
	5	0,4	0	2,3	0	71,2	2,1	3,0	0	6,1	0	0	0	0	2,7	0	0	10,0	2,1	0	100	
	6	7,4	0	0	0	0,2	92,4	0	0	0	0	0	0	0	0	0	0	0	0	0	100	
	7	3,9	0	4,0	0	0,9	0	91,1	0	0	0	0	0	0	0	0	0	0	0	0	100	
	8	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	100	
	9	0	0	0	0	0	0	0	0	76,9	3,5	11,4	0	3,1	0	0	0	0	4,7	0,4	100	
	10	0	0	0	0	0	0	0	0	0,5	91,4	1,6	2,9	0	0	0	0	0	3,3	0,3	100	
	11	0	0	0	0	0	0	0	0	0,3	5,7	71,4	12,9	0	0	0	0	0	9,4	0,3	100	
	12	0	0	0	0	0	0	0	0	0	2,7	5,1	92,1	0	0	0	0	0	0	0	100	
	13	0	0	0	2,2	0	0	0	0	2,9	0	0	0	77,0	0	0	0	0	17,9	0	100	
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	100	
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	100	
	16	0	0	0	10,3	0	0	0	0	0	0	0	0	0	13,1	0	76,6	0	0	0	100	
	17	0	0	0	0	1,3	0	2,4	0	0	0	0	0	0	5,9	0	0	90,4	0	0	100	
	18	0	0	0	0	0	0	1,3	0	0,6	0	0	0	0	0	0	0	0	98,1	0	100	
	19	0	0	0	0	0	0	0	0	0	11,4	0	0	0	0	0	0	0	0	88,6	100	

a En la validación cruzada, cada caso se clasifica a partir del resto de los casos.

b Clasificados correctamente el 87,8% de los casos agrupados originales.

c Clasificados correctamente el 87,6% de los casos agrupados validados mediante validación cruzada.

**ANEXO 29. ANÁLISIS DISCRIMINANTE DE VARIABLES  
TRANSFORMADAS (15 GRUPOS), PERIODO  
COMPLETO DE ESTUDIO**

**Estadísticos de grupo**

No del grupo	Variable	Media	Desv. típ.	No casos utilizados	No del grupo	Variable	Media	Desv. típ.	No casos utilizados
1	DJ_T	1,417	0,334	916	2	DJ_T	-1,521	0,380	1211
	V_T	-0,581	0,415	916		V_T	-0,591	0,489	1211
	DV_T	1,098	0,395	916		DV_T	0,856	0,334	1211
	Vx_T	-0,606	0,477	916		Vx_T	-0,637	0,523	1211
	T_T	0,053	0,447	916		T_T	-0,761	0,352	1211
	HR_T	0,414	0,539	916		HR_T	0,547	0,625	1211
	ETo_T	-0,838	0,622	916		ETo_T	-1,127	0,442	1211
	VolExp_T	-1,858	0,271	916		VolExp_T	0,741	0,572	1211
	Qent_T	-1,273	0,414	916		Qent_T	-0,248	0,778	1211
	Qsal_T	0,439	0,339	916		Qsal_T	-1,008	0,684	1211
Profundidad_T	-1,027	0,644	916	Profundidad_T	-0,947	0,794	1211		
3	DJ_T	-1,362	0,476	456	4	DJ_T	-1,140	0,398	460
	V_T	-0,240	0,289	456		V_T	0,233	0,588	460
	DV_T	0,510	1,068	456		DV_T	1,201	0,288	460
	Vx_T	-0,178	0,667	456		Vx_T	0,435	0,728	460
	T_T	-0,746	0,582	456		T_T	-0,614	0,730	460
	HR_T	0,315	0,509	456		HR_T	-1,168	0,548	460
	ETo_T	-0,965	0,523	456		ETo_T	-0,360	0,543	460
	VolExp_T	0,238	0,541	456		VolExp_T	1,055	0,675	460
	Qent_T	-0,620	0,741	456		Qent_T	-0,591	0,648	460
	Qsal_T	-0,868	0,421	456		Qsal_T	-0,802	0,329	460
Profundidad_T	0,216	0,742	456	Profundidad_T	-0,851	0,821	460		
5	DJ_T	0,860	0,400	655	6	DJ_T	1,631	0,062	122
	V_T	-0,602	0,366	655		V_T	-0,828	0,081	122
	DV_T	0,446	0,639	655		DV_T	1,491	0,424	122
	Vx_T	-0,579	0,446	655		Vx_T	-0,553	0,230	122
	T_T	0,603	0,647	655		T_T	-0,289	0,066	122
	HR_T	0,259	0,568	655		HR_T	0,972	0,305	122
	ETo_T	-0,107	0,915	655		ETo_T	-1,741	0,178	122
	VolExp_T	-0,860	0,937	655		VolExp_T	-0,771	0,293	122
	Qent_T	-1,537	0,470	655		Qent_T	0,933	0,890	122
	Qsal_T	0,628	0,425	655		Qsal_T	-0,188	0,156	122
Profundidad_T	0,675	0,593	655	Profundidad_T	-0,797	0,816	122		
7	DJ_T	1,916	0,111	189	8	DJ_T	1,007	0,280	602
	V_T	-0,078	0,288	189		V_T	-0,114	0,354	602
	DV_T	1,149	0,181	189		DV_T	0,911	0,322	602
	Vx_T	-0,205	0,480	189		Vx_T	0,051	0,507	602
	T_T	-0,445	0,308	189		T_T	0,615	0,359	602
	HR_T	-0,055	0,370	189		HR_T	-0,578	0,533	602
	ETo_T	-1,133	0,244	189		ETo_T	0,091	0,538	602
	VolExp_T	-0,191	0,063	189		VolExp_T	-1,231	0,490	602
	Qent_T	-0,760	0,376	189		Qent_T	-1,655	0,398	602
	Qsal_T	-0,042	0,033	189		Qsal_T	0,908	0,391	602
Profundidad_T	-0,686	0,893	189	Profundidad_T	-0,878	0,725	602		

No del grupo	Variable	Media	Desv. típ.	No casos utilizados	No del grupo	Variable	Media	Desv. típ.	No casos utilizados
9	DJ_T	0,045	0,167	715	10	DJ_T	0,701	0,273	1107
	V_T	-0,451	0,390	715		V_T	-0,741	0,241	1107
	DV_T	-0,039	0,541	715		DV_T	0,368	0,377	1107
	Vx_T	-0,400	0,548	715		Vx_T	-0,706	0,256	1107
	T_T	0,802	0,460	715		T_T	0,881	0,280	1107
	HR_T	-0,435	0,531	715		HR_T	0,434	0,387	1107
	ETo_T	0,859	0,438	715		ETo_T	0,091	0,507	1107
	VolExp_T	0,817	0,341	715		VolExp_T	-0,505	0,575	1107
	Qent_T	-1,779	0,455	715		Qent_T	-1,803	0,338	1107
	Qsal_T	0,215	0,326	715		Qsal_T	0,987	0,627	1107
Profundidad_T	-0,992	0,609	715	Profundidad_T	-0,989	0,655	1107		
11	DJ_T	-0,097	0,319	1349	12	DJ_T	1,382	0,619	103
	V_T	-0,606	0,365	1349		V_T	-0,854	0,223	103
	DV_T	0,065	0,386	1349		DV_T	-1,540	0,432	103
	Vx_T	-0,500	0,343	1349		Vx_T	-0,552	0,174	103
	T_T	0,664	0,585	1349		T_T	-0,127	0,503	103
	HR_T	0,292	0,570	1349		HR_T	0,983	0,372	103
	ETo_T	0,341	0,410	1349		ETo_T	-1,744	0,377	103
	VolExp_T	1,294	0,493	1349		VolExp_T	-0,225	0,287	103
	Qent_T	-1,663	0,561	1349		Qent_T	-0,334	0,743	103
	Qsal_T	0,059	0,749	1349		Qsal_T	0,152	0,373	103
Profundidad_T	-1,120	0,547	1349	Profundidad_T	-0,837	0,867	103		
13	DJ_T	-0,070	0,287	624	14	DJ_T	2,005	0,144	71
	V_T	-0,535	0,357	624		V_T	1,236	0,941	71
	DV_T	0,072	0,448	624		DV_T	1,177	0,076	71
	Vx_T	-0,448	0,391	624		Vx_T	1,481	0,673	71
	T_T	0,666	0,535	624		T_T	-0,434	0,298	71
	HR_T	0,172	0,638	624		HR_T	-0,483	0,467	71
	ETo_T	0,408	0,400	624		ETo_T	-0,725	0,192	71
	VolExp_T	1,254	0,472	624		VolExp_T	-0,151	0,063	71
	Qent_T	-1,699	0,571	624		Qent_T	-0,756	0,203	71
	Qsal_T	0,144	0,679	624		Qsal_T	-0,075	0,036	71
Profundidad_T	0,884	0,706	624	Profundidad_T	-0,672	0,907	71		
15	DJ_T	0,877	0,353	394	Total	DJ_T	0,145	1,074	8974
	V_T	-0,344	0,230	394		V_T	-0,475	0,483	8974
	DV_T	-0,632	0,450	394		DV_T	0,474	0,712	8974
	Vx_T	-0,078	0,229	394		Vx_T	-0,409	0,572	8974
	T_T	0,947	0,358	394		T_T	0,261	0,794	8974
	HR_T	0,267	0,401	394		HR_T	0,158	0,709	8974
	ETo_T	0,230	0,620	394		ETo_T	-0,211	0,855	8974
	VolExp_T	-0,625	0,649	394		VolExp_T	0,070	1,184	8974
	Qent_T	-1,630	0,415	394		Qent_T	-1,266	0,828	8974
	Qsal_T	0,802	0,394	394		Qsal_T	0,121	0,864	8974
Profundidad_T	-0,994	0,605	394	Profundidad_T	-0,667	0,945	8974		



**Pruebas de igualdad de las medias de los grupos**

	Lambda de Wilks	F	gl1	gl2	Sig.
DJ_T	,095	6065,185	14	8959	,000
V_T	,642	357,388	14	8959	,000
DV_T	,436	827,821	14	8959	,000
Vx_T	,653	339,442	14	8959	,000
T_T	,364	1116,009	14	8959	,000
HR_T	,568	486,523	14	8959	,000
ETo_T	,381	1037,908	14	8959	,000
VolExp_T	,209	2428,384	14	8959	,000
Qent_T	,443	805,878	14	8959	,000
Qsal_T	,394	983,045	14	8959	,000
Profundidad_T	,525	578,326	14	8959	,000

**Logaritmo de los determinantes**

Número del Grupo asignado en el CA	Rango	Logaritmo del determinante
1	11	-25,440
2	11	-20,903
3	11	-22,114
4	11	-23,439
5	11	-22,238
6	6	.(a)
7	8	.(a)
8	11	-27,002
9	11	-28,079
10	11	-27,652
11	11	-24,989
12	5	.(a)
13	11	-22,902
14	3	.(a)
15	11	-30,164
Intra-grupos combinada	11	-18,689

Los rangos y logaritmos naturales de los determinantes impresos son los de las matrices de covarianzas de los grupos.

a Singular

### Resultados de la prueba(a)

	M de Box	52618,545
	Aprox.	79,221
F	gl1	660
	gl2	28874007
	Sig.	,000

Contrasta la hipótesis nula de que las matrices de covarianzas poblacionales son iguales.

a Algunas matrices de covarianzas son singulares y el procedimiento ordinario no es válido. Los grupos no singulares se compararán con sus propias matrices de covarianzas intra-grupo combinadas. El logaritmo de su determinante es -18,635.

### Autovalores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	11,316(a)	56,8	56,8	,959
2	3,372(a)	16,9	73,7	,878
3	2,272(a)	11,4	85,1	,833
4	1,113(a)	5,6	90,6	,726
5	,939(a)	4,7	95,4	,696
6	,547(a)	2,7	98,1	,595
7	,176(a)	,9	99,0	,387
8	,121(a)	,6	99,6	,328
9	,066(a)	,3	99,9	,249
10	,014(a)	,1	100,0	,119
11	,002(a)	,0	100,0	,050

a Se han empleado las 11 primeras funciones discriminantes canónicas en el análisis.

### Lambda de Wilks

Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1 a la 11	,001	66081,661	154	,000
2 a la 11	,008	43584,032	130	,000
3 a la 11	,034	30366,715	108	,000
4 a la 11	,110	19744,925	88	,000
5 a la 11	,233	13041,918	70	,000
6 a la 11	,452	7106,663	54	,000
7 a la 11	,700	3198,583	40	,000
8 a la 11	,823	1744,328	28	,000
9 a la 11	,922	723,038	18	,000
10 a la 11	,983	149,579	10	,000
11	,998	22,209	4	,000

**Probabilidades previas para los grupos**

Número del Grupo	Probabilidad previa	Casos utilizados en el análisis
1	,102	916
2	,135	1211
3	,051	456
4	,051	460
5	,073	655
6	,014	122
7	,021	189
8	,067	602
9	,080	715
10	,123	1107
11	,150	1349
12	,011	103
13	,070	624
14	,008	71
15	,044	394
Total	1,000	8974

**Resultados de la clasificación(b,c)**

	Número del grupo	Grupo de pertenencia pronosticado														Total	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14		15
Original Recuento	1	893	0	0	0	22	0	0	0	0	1	0	0	0	0	0	916
	2	0	1100	48	18	0	0	0	0	0	0	41	0	4	0	0	1211
	3	0	78	362	16	0	0	0	0	0	0	0	0	0	0	0	456
	4	0	37	20	403	0	0	0	0	0	0	0	0	0	0	0	460
	5	194	0	0	0	279	0	0	11	27	57	0	0	73	0	14	655
	6	0	0	0	0	0	107	15	0	0	0	0	0	0	0	0	122
	7	0	0	0	0	0	0	189	0	0	0	0	0	0	0	0	189
	8	0	0	0	0	13	0	0	499	0	90	0	0	0	0	0	602
	9	0	0	0	0	0	0	0	0	585	0	122	0	8	0	0	715
	10	13	0	0	0	65	0	0	0	26	927	66	0	0	0	10	1107
	11	0	0	0	0	0	0	0	0	73	50	1225	0	1	0	0	1349
	12	0	0	0	0	0	0	0	0	0	0	0	95	0	0	8	103
	13	0	0	2	4	0	0	0	0	18	6	53	0	541	0	0	624
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	71	0	71
	15	0	0	0	0	2	0	0	0	40	60	6	20	0	0	266	394
%	1	97,5	0	0	0	2,4	0	0	0	0	0,1	0	0	0	0	100	
	2	0	90,8	4,0	1,5	0	0	0	0	0	0	3,4	0	0,3	0	100	
	3	0	17,1	79,4	3,5	0	0	0	0	0	0	0	0	0	0	100	
	4	0	8,0	4,3	87,6	0	0	0	0	0	0	0	0	0	0	100	
	5	29,6	0	0	0	42,6	0	0	1,7	4,1	8,7	0	0	11,1	0	2,1	100
	6	0	0	0	0	0	87,7	12,3	0	0	0	0	0	0	0	0	100
	7	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	100
	8	0	0	0	0	2,2	0	0	82,9	0	15,0	0	0	0	0	0	100
	9	0	0	0	0	0	0	0	0	81,8	0	17,1	0	1,1	0	0	100
	10	1,2	0	0	0	5,9	0	0	0	2,3	83,7	6,0	0	0	0	0,9	100
	11	0	0	0	0	0	0	0	0	5,4	3,7	90,8	0	0,1	0	0	100
	12	0	0	0	0	0	0	0	0	0	0	0	92,2	0	0	7,8	100
	13	0	0	0,3	0,6	0	0	0	0	2,9	1,0	8,5	0	86,7	0	0	100
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	100
	15	0	0	0	0	0,5	0	0	0	10,2	15,2	1,5	5,1	0	0	67,5	100

	No del grupo	Grupo de pertenencia pronosticado														Total	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14		15
Validación cruzada (a) Recuento	1	891	0	0	0	23	0	0	0	0	2	0	0	0	0	0	916
	2	0	1100	48	18	0	0	0	0	0	0	41	0	4	0	0	1211
	3	0	80	360	16	0	0	0	0	0	0	0	0	0	0	0	456
	4	0	37	21	402	0	0	0	0	0	0	0	0	0	0	0	460
	5	194	0	0	0	278	0	0	11	27	58	0	0	73	0	14	655
	6	0	0	0	0	0	107	15	0	0	0	0	0	0	0	0	122
	7	0	0	0	0	0	0	189	0	0	0	0	0	0	0	0	189
	8	0	0	0	0	14	0	0	498	0	90	0	0	0	0	0	602
	9	0	0	0	0	0	0	0	0	585	0	122	0	8	0	0	715
	10	13	0	0	0	65	0	0	0	26	926	66	0	0	0	11	1107
	11	0	0	0	0	0	0	0	0	73	50	1225	0	1	0	0	1349
	12	0	0	0	0	0	0	0	0	0	0	0	95	0	0	8	103
	13	0	0	2	4	0	0	0	0	18	6	54	0	540	0	0	624
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	71	0	71
	15	0	0	0	0	2	0	0	0	41	60	6	20	0	0	265	394
%	1	97,3	0	0	0	2,5	0	0	0	0	0,2	0	0	0	0	0	100
	2	0	90,8	4,0	1,5	0	0	0	0	0	0	3,4	0	0,3	0	0	100
	3	0	17,5	78,9	3,5	0	0	0	0	0	0	0	0	0	0	0	100
	4	0	8,0	4,6	87,4	0	0	0	0	0	0	0	0	0	0	0	100
	5	29,6	0	0	0	42,4	0	0	1,7	4,1	8,9	0	0	11,1	0	2,1	100
	6	0	0	0	0	0	87,7	12,3	0	0	0	0	0	0	0	0	100
	7	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	100
	8	0	0	0	0	2,3	0	0	82,7	0	15,0	0	0	0	0	0	100
	9	0	0	0	0	0	0	0	0	81,8	0	17,1	0	1,1	0	0	100
	10	1,2	0	0	0	5,9	0	0	0	2,3	83,6	6,0	0	0	0	1,0	100
	11	0	0	0	0	0	0	0	0	5,4	3,7	90,8	0	0,1	0	0	100
	12	0	0	0	0	0	0	0	0	0	0	0	92,2	0	0	7,8	100
	13	0	0	0,3	0,6	0	0	0	0	2,9	1,0	8,7	0	86,5	0	0	100
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	100
	15	0	0	0	0	0,5	0	0	0	10,4	15,2	1,5	5,1	0	0	67,3	100

- a Cada caso se clasifica mediante las funciones derivadas a partir del resto de los casos.  
b Clasificados correctamente el 84,0% de los casos agrupados originales.  
c Clasificados correctamente el 83,9% de los casos agrupados validados mediante validación cruzada.

**ANEXO 30. ANÁLISIS DISCRIMINANTE DE PUNTUACIONES EN  
COMPONENTES PRINCIPALES (18 GRUPOS),  
PERIODO COMPLETO DE ESTUDIO**

**Estadísticos de grupo**

No de grupo	Puntuación en la componente	Media	Desv. típ.	No de casos	No de grupo	Puntuación en la componente	Media	Desv. típ.	No de casos
1	ComPrin_1	1,486	0,545	811	2	ComPrin_1	0,864	1,176	444
	ComPrin_2	1,178	0,731	811		ComPrin_2	1,332	0,776	444
	ComPrin_3	1,326	0,597	811		ComPrin_3	0,234	1,032	444
	ComPrin_4	-0,358	0,577	811		ComPrin_4	1,218	0,695	444
	ComPrin_5	-0,366	0,701	811		ComPrin_5	-0,383	0,775	444
	ComPrin_6	-0,040	0,459	811		ComPrin_6	-0,038	0,557	444
	ComPrin_7	0,029	0,383	811		ComPrin_7	0,239	0,654	444
	ComPrin_8	-0,058	0,367	811		ComPrin_8	-0,084	0,434	444
	ComPrin_9	-0,084	0,236	811		ComPrin_9	0,036	0,266	444
	ComPrin_10	-0,160	0,180	811		ComPrin_10	0,034	0,305	444
	ComPrin_11	-0,052	0,095	811		ComPrin_11	-0,040	0,123	444
3	ComPrin_1	-2,679	1,792	117	4	ComPrin_1	-1,349	0,454	516
	ComPrin_2	2,200	0,364	117		ComPrin_2	-1,255	0,441	516
	ComPrin_3	-0,902	0,203	117		ComPrin_3	-0,964	0,386	516
	ComPrin_4	2,258	1,507	117		ComPrin_4	-0,179	0,731	516
	ComPrin_5	2,555	1,800	117		ComPrin_5	0,110	0,832	516
	ComPrin_6	-0,896	0,694	117		ComPrin_6	-0,006	0,483	516
	ComPrin_7	-0,940	0,613	117		ComPrin_7	0,289	0,676	516
	ComPrin_8	-0,554	0,395	117		ComPrin_8	-0,075	0,446	516
	ComPrin_9	0,129	0,245	117		ComPrin_9	-0,447	0,597	516
	ComPrin_10	-0,418	0,157	117		ComPrin_10	0,062	0,355	516
	ComPrin_11	-0,039	0,137	117		ComPrin_11	0,095	0,175	516
5	ComPrin_1	1,206	0,565	550	6	ComPrin_1	-0,155	0,560	511
	ComPrin_2	-1,423	0,744	550		ComPrin_2	0,058	0,867	511
	ComPrin_3	2,146	0,673	550		ComPrin_3	1,031	1,142	511
	ComPrin_4	0,159	0,663	550		ComPrin_4	-0,154	0,540	511
	ComPrin_5	0,050	0,611	550		ComPrin_5	-0,487	0,387	511
	ComPrin_6	-0,063	0,441	550		ComPrin_6	0,424	0,535	511
	ComPrin_7	-0,113	0,567	550		ComPrin_7	-0,352	0,585	511
	ComPrin_8	-0,037	0,424	550		ComPrin_8	-0,153	0,402	511
	ComPrin_9	-0,141	0,312	550		ComPrin_9	0,109	0,326	511
	ComPrin_10	-0,205	0,211	550		ComPrin_10	-0,002	0,309	511
	ComPrin_11	0,053	0,101	550		ComPrin_11	0,091	0,143	511
7	ComPrin_1	-1,300	0,275	69	8	ComPrin_1	2,020	0,912	1135
	ComPrin_2	-3,297	0,906	69		ComPrin_2	-0,310	0,688	1135
	ComPrin_3	2,547	0,464	69		ComPrin_3	-0,832	0,490	1135
	ComPrin_4	0,276	1,006	69		ComPrin_4	-0,721	0,422	1135
	ComPrin_5	0,799	0,540	69		ComPrin_5	-0,115	0,744	1135
	ComPrin_6	-0,452	0,339	69		ComPrin_6	-0,021	0,459	1135
	ComPrin_7	-0,648	0,272	69		ComPrin_7	0,110	0,426	1135
	ComPrin_8	1,076	0,448	69		ComPrin_8	0,063	0,370	1135
	ComPrin_9	0,688	0,172	69		ComPrin_9	0,132	0,305	1135
	ComPrin_10	0,533	0,197	69		ComPrin_10	0,047	0,230	1135
	ComPrin_11	-0,176	0,179	69		ComPrin_11	0,018	0,163	1135
9	ComPrin_1	-2,835	0,427	153	10	ComPrin_1	-1,894	0,531	372
	ComPrin_2	3,659	1,474	153		ComPrin_2	-0,402	0,833	372
	ComPrin_3	0,864	0,526	153		ComPrin_3	-1,401	0,673	372
	ComPrin_4	0,575	0,726	153		ComPrin_4	1,244	0,669	372
	ComPrin_5	1,819	1,482	153		ComPrin_5	-0,373	0,797	372
	ComPrin_6	1,738	0,970	153		ComPrin_6	0,280	0,715	372
	ComPrin_7	0,730	0,312	153		ComPrin_7	-0,259	0,694	372
	ComPrin_8	0,473	0,518	153		ComPrin_8	0,111	0,570	372
	ComPrin_9	-0,104	0,226	153		ComPrin_9	0,039	0,424	372
	ComPrin_10	-0,138	0,455	153		ComPrin_10	-0,083	0,418	372
	ComPrin_11	0,102	0,170	153		ComPrin_11	-0,017	0,164	372

No de grupo	Puntuación en la componente	Media	Desv. típ.	No de casos	No de grupo	Puntuación en la componente	Media	Desv. típ.	No de casos
11	ComPrin_1	-0,021	0,635	314	12	ComPrin_1	-1,472	0,580	930
	ComPrin_2	1,812	0,486	314		ComPrin_2	0,295	0,655	930
	ComPrin_3	-1,449	0,541	314		ComPrin_3	-1,203	0,690	930
	ComPrin_4	0,603	0,635	314		ComPrin_4	-0,944	0,580	930
	ComPrin_5	0,352	1,270	314		ComPrin_5	-0,441	0,651	930
	ComPrin_6	-0,404	0,930	314		ComPrin_6	-0,127	0,398	930
	ComPrin_7	-0,503	0,637	314		ComPrin_7	-0,209	0,386	930
	ComPrin_8	-0,090	0,617	314		ComPrin_8	0,048	0,504	930
	ComPrin_9	-0,098	0,499	314		ComPrin_9	-0,019	0,350	930
	ComPrin_10	-0,003	0,270	314		ComPrin_10	-0,131	0,350	930
	ComPrin_11	0,041	0,097	314		ComPrin_11	0,009	0,132	930
13	ComPrin_1	2,275	0,513	604	14	ComPrin_1	1,666	0,451	479
	ComPrin_2	-0,127	0,756	604		ComPrin_2	-1,513	0,602	479
	ComPrin_3	-0,689	0,644	604		ComPrin_3	-0,011	0,461	479
	ComPrin_4	1,448	0,610	604		ComPrin_4	0,438	0,751	479
	ComPrin_5	0,073	0,612	604		ComPrin_5	-0,311	0,513	479
	ComPrin_6	-0,106	0,496	604		ComPrin_6	0,102	0,492	479
	ComPrin_7	-0,001	0,375	604		ComPrin_7	0,204	0,420	479
	ComPrin_8	0,052	0,365	604		ComPrin_8	0,057	0,258	479
	ComPrin_9	-0,036	0,416	604		ComPrin_9	0,047	0,389	479
	ComPrin_10	0,047	0,194	604		ComPrin_10	0,114	0,269	479
	ComPrin_11	-0,006	0,161	604		ComPrin_11	-0,018	0,110	479
15	ComPrin_1	-0,931	0,653	576	16	ComPrin_1	-3,050	1,097	620
	ComPrin_2	2,178	0,767	576		ComPrin_2	-2,336	1,107	620
	ComPrin_3	0,622	0,469	576		ComPrin_3	0,132	0,692	620
	ComPrin_4	-0,425	0,612	576		ComPrin_4	-0,099	0,715	620
	ComPrin_5	-0,155	0,672	576		ComPrin_5	-0,081	0,838	620
	ComPrin_6	0,144	0,660	576		ComPrin_6	0,111	0,452	620
	ComPrin_7	-0,048	0,536	576		ComPrin_7	0,114	0,540	620
	ComPrin_8	0,149	0,411	576		ComPrin_8	-0,281	0,549	620
	ComPrin_9	0,038	0,455	576		ComPrin_9	-0,016	0,354	620
	ComPrin_10	0,383	0,300	576		ComPrin_10	-0,021	0,199	620
	ComPrin_11	0,002	0,114	576		ComPrin_11	-0,052	0,257	620
17	ComPrin_1	1,432	0,586	373	18	ComPrin_1	-2,827	1,620	400
	ComPrin_2	-0,452	0,875	373		ComPrin_2	0,214	1,067	400
	ComPrin_3	-0,328	0,733	373		ComPrin_3	1,378	0,425	400
	ComPrin_4	-0,641	0,836	373		ComPrin_4	-0,081	0,795	400
	ComPrin_5	1,694	0,544	373		ComPrin_5	0,452	0,924	400
	ComPrin_6	-0,184	0,668	373		ComPrin_6	-0,408	0,702	400
	ComPrin_7	0,124	0,418	373		ComPrin_7	0,357	0,684	400
	ComPrin_8	-0,017	0,330	373		ComPrin_8	0,114	0,583	400
	ComPrin_9	0,015	0,388	373		ComPrin_9	0,310	0,395	400
	ComPrin_10	-0,037	0,279	373		ComPrin_10	0,133	0,328	400
	ComPrin_11	-0,027	0,068	373		ComPrin_11	-0,110	0,220	400
Total	ComPrin_1	0,000	1,956	8974					
	ComPrin_2	0,000	1,541	8974					
	ComPrin_3	0,000	1,264	8974					
	ComPrin_4	0,000	1,007	8974					
	ComPrin_5	0,000	0,968	8974					
	ComPrin_6	0,000	0,634	8974					
	ComPrin_7	0,000	0,575	8974					
	ComPrin_8	0,000	0,473	8974					
	ComPrin_9	0,000	0,404	8974					
	ComPrin_10	0,000	0,320	8974					
	ComPrin_11	0,000	0,160	8974					



**Pruebas de igualdad de las medias de los grupos**

	Lambda de Wilks	F	gl1	gl2	Sig.
ComPrin_1	,163	2698,309	17	8956	,000
ComPrin_2	,258	1511,557	17	8956	,000
ComPrin_3	,267	1443,252	17	8956	,000
ComPrin_4	,429	700,189	17	8956	,000
ComPrin_5	,629	310,857	17	8956	,000
ComPrin_6	,758	167,949	17	8956	,000
ComPrin_7	,799	132,659	17	8956	,000
ComPrin_8	,873	76,820	17	8956	,000
ComPrin_9	,845	96,721	17	8956	,000
ComPrin_10	,771	156,265	17	8956	,000
ComPrin_11	,890	65,332	17	8956	,000

**Logaritmo de los determinantes**

Número inicial de casos	Rango	Logaritmo del determinante
1	11	-24,099
2	11	-18,835
3	10	.(a)
4	11	-22,270
5	11	-23,364
6	11	-21,554
7	7	.(a)
8	11	-21,857
9	11	-98,016
10	11	-18,360
11	11	-21,633
12	11	-21,254
13	11	-20,848
14	11	-27,662
15	11	-22,227
16	11	-18,306
17	11	-21,883
18	11	-15,992
Intra-grupos combinada	11	-15,989

Los rangos y logaritmos naturales de los determinantes impresos son los de las matrices de covarianzas de los grupos.

a Singular

### Resultados de la prueba (a)

M de Box		59671,251
Aprox.		59,722
F	gl1	990
	gl2	13931320
	Sig.	,000

Contrasta la hipótesis nula de que las matrices de covarianzas poblacionales son iguales.

a Algunas matrices de covarianzas son singulares y el procedimiento ordinario no es válido. Los grupos no singulares se compararán con sus propias matrices de covarianzas intra-grupo combinadas. El logaritmo de su determinante es -16,065.

### Autovalores

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	6,410(a)	39,3	39,3	,930
2	3,664(a)	22,4	61,7	,886
3	2,951(a)	18,1	79,8	,864
4	1,455(a)	8,9	88,7	,770
5	,527(a)	3,2	91,9	,588
6	,479(a)	2,9	94,8	,569
7	,448(a)	2,7	97,6	,556
8	,172(a)	1,1	98,6	,383
9	,132(a)	,8	99,4	,341
10	,068(a)	,4	99,9	,253
11	,022(a)	,1	100,0	,147

a Se han empleado las 11 primeras funciones discriminantes canónicas en el análisis.

### Lambda de Wilks

Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1 a la 11	,001	66019,040	187	,000
2 a la 11	,005	48076,892	160	,000
3 a la 11	,022	34281,764	135	,000
4 a la 11	,086	21973,333	112	,000
5 a la 11	,211	13928,543	91	,000
6 a la 11	,323	10135,145	72	,000
7 a la 11	,477	6630,264	55	,000
8 a la 11	,691	3316,012	40	,000
9 a la 11	,809	1895,412	27	,000
10 a la 11	,916	787,684	16	,000
11	,978	195,193	7	,000

**Probabilidades previas para los grupos**

Número inicial de casos	Probabilidades previas	Casos utilizados en el análisis
1	,090	811
2	,049	444
3	,013	117
4	,057	516
5	,061	550
6	,057	511
7	,008	69
8	,126	1135
9	,017	153
10	,041	372
11	,035	314
12	,104	930
13	,067	604
14	,053	479
15	,064	576
16	,069	620
17	,042	373
18	,045	400
Total	1,000	8974

**Resultados de la clasificación (b,c)**

	No de grupo	Grupo de pertenencia pronosticado																	Total	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17		18
Original Recuento	1	800	11																	811
	2	12	298			9	33		22			6		33		22		9		444
	3			117																117
	4				437						18		41				20			516
	5	25				502									22			1		550
	6	39			47	20	359		16						10	17		3		511
	7							69												69
	8		1						1079					36	13	6				1135
	9									147						6				153
	10				19						309	6	8				26		4	372
	11		7	11								296								314
	12				43		15		2		4	2	847			16		1		930
	13			28					26					528	22					604
	14								58					14	406			1		479
	15		1				11									536				576
	16										7		54				557		2	620
	17	4			8	2			46			3	8	14	6	1		281		373
	18	15	3		25											18			339	400
%	1	98,6	1,4																	100
	2	2,7	67,1			2,0	7,4		5,0			1,4		7,4		5,0		2,0		100
	3			100																100
	4				84,7						3,5		7,9				3,9			100
	5	4,5				91,3									4			0,2		100
	6	7,6			9,2	3,9	70,3		3,1						2,0	3,3		0,6		100
	7							100												100
	8		0,1						95,1					3,2	1,1	0,5				100
	9									96,1						3,9				100
	10				5,1						83,1	1,6	2,2				7,0		1,1	100
	11		2,2	3,5								94,3								100
	12				4,6		1,6		0,2		0,4	0,2	91,1			1,7		0,1		100
	13								4,3					87,4	3,6					100
	14								12,1					2,9	84,8			0,2		100
	15		0,2				1,9						4,9			93,1				100
	16										1,1		8,7				89,8		0,3	100
	17	1,1			2,1	0,5			12,3			0,8	2,1	3,8	1,6	0,3		75,3		100
	18	3,8	0,8		6,3											4,5			84,8	100

	No de grupo	Grupo de pertenencia pronosticado																	Total	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17		18
Validación cruzada (a) Recuento	1	800	11																	811
	2	12	297			9	33		22			6		34		22			9	444
	3			117																117
	4				434						18		44				20			516
	5	25				502									22			1		550
	6	39			47	20	359		16						10	17		3		511
	7							69												69
	8		1						1078					36	13	7				1135
	9									146						7				153
	10				19						309	6	8				26		4	372
	11		8	11								295								314
	12				43		15		2		4	2	847			16		1		930
	13			28					29					525	22					604
	14								60					15	403			1		479
	15		1				11									536				576
	16				1						8						555		2	620
	17	4			8	3			46			3	8	14	6	1		280		373
	18	15	3		25											18			339	400
%	1	98,6	1,4																100	
	2	2,7	66,9			2,0	7,4		5,0			1,4		7,7		5,0		2,0	100	
	3			100															100	
	4				84,1						3,5		8,5				3,9		100	
	5	4,5				91,3									4			0,2	100	
	6	7,6			9,2	3,9	70,3		3,1						2,0	3,3		0,6	100	
	7							100											100	
	8		0,1						95,0					3,2	1,1	0,6			100	
	9									95,4							4,6		100	
	10				5,1						83,1	1,6	2,2				7,0		1,1	100
	11		2,5	3,5								93,9								100
	12				4,6		1,6		0,2		0,4	0,2	91,1			1,7		0,1	100	
	13								4,8					86,9	3,6				100	
	14								12,5					3,1	84,1			0,2	100	
	15		0,2				1,9									93,1			100	
	16				0,2						1,3						89,5		0,3	100
	17	1,1			2,1	0,8			12,3			0,8	2,1	3,8	1,6	0,3		75,1		100
	18	3,8	0,8		6,3											4,5			84,8	100

a En la validación cruzada, cada caso se clasifica mediante las funciones derivadas a partir del resto de los casos.

b Clasificados correctamente el 88,1% de los casos agrupados originales.

c Clasificados correctamente el 87,9% de los casos agrupados validados mediante validación cruzada.

**ANEXO 31. ANÁLISIS DISCRIMINANTE DE PUNTUACIONES  
FACTORIALES (3 FACTORES Y 14 GRUPOS),  
PERIODO COMPLETO DE ESTUDIO**

**Estadísticos de grupo**

Número de grupos	Factor	Media	Desv. típ.	No de casos
1	1	-0,610	0,312	510
	2	-1,208	0,335	510
	3	-0,193	0,596	510
2	1	0,929	0,319	363
	2	0,982	0,439	363
	3	-1,064	0,384	363
3	1	0,785	0,251	676
	2	0,022	0,371	676
	3	0,886	0,330	676
4	1	0,814	0,473	1286
	2	-0,758	0,263	1286
	3	-0,339	0,320	1286
5	1	0,798	0,432	506
	2	1,435	0,648	506
	3	1,295	0,446	506
6	1	-0,651	0,378	825
	2	0,717	0,424	825
	3	-1,211	0,343	825
7	1	-1,220	0,606	224
	2	2,088	0,540	224
	3	-1,498	0,505	224
8	1	-0,913	0,290	764
	2	-0,670	0,371	764
	3	-1,146	0,343	764
9	1	0,738	0,367	513
	2	-0,804	0,305	513
	3	1,060	0,377	513
10	1	-0,689	0,665	760
	2	0,892	0,516	760
	3	1,143	0,326	760
11	1	-1,790	0,171	491
	2	-1,038	0,537	491
	3	0,767	0,507	491
12	1	1,199	0,513	970
	2	-0,006	0,311	970
	3	-0,704	0,287	970
13	1	-0,843	0,403	107
	2	3,484	0,451	107
	3	0,699	0,328	107
14	1	-0,559	0,324	979
	2	-0,180	0,464	979
	3	0,712	0,513	979
Total	1	0,000	1,000	8974
	2	0,000	1,000	8974
	3	0,000	1,000	8974

**Pruebas de igualdad de las medias de los grupos**

	Lambda de Wilks	F	gl1	gl2	Sig.
Factor 1	,175	3253,243	13	8960	,000
Factor 2	,173	3283,915	13	8960	,000
Factor 3	,157	3690,765	13	8960	,000

**Logaritmo de los determinantes**

Número de Conglomerados obtenido en CA	Rango	Logaritmo del determinante
1	3	-6,156
2	3	-6,582
3	3	-7,066
4	3	-6,807
5	3	-5,089
6	3	-7,057
7	3	-5,243
8	3	-6,650
9	3	-6,672
10	3	-4,664
11	3	-6,578
12	3	-6,295
13	3	-7,979
14	3	-6,153
Intra-grupos combinada	3	-5,380

Los rangos y logaritmos naturales de los determinantes impresos son los de las matrices de covarianzas de los grupos.

**Resultados de la prueba**

M de Box	8601,044
Aprox.	109,968
F	78
gl1	7381690
gl2	,000
Sig.	

Contrasta la hipótesis nula de que las matrices de covarianzas poblacionales son iguales.

**Autovalores**

Función	Autovalor	% de varianza	% acumulado	Correlación canónica
1	6,505(a)	42,5	42,5	,931
2	4,836(a)	31,6	74,0	,910
3	3,974(a)	26,0	100,0	,894

a Se han empleado las 3 primeras funciones discriminantes canónicas en el análisis.



**Lambda de Wilks**

Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1 a la 3	,005	48262,604	39	,000
2 a la 3	,034	30194,569	24	,000
3	,201	14381,277	11	,000

**Probabilidades previas para los grupos**

Número de grupo	Probabilidades previas	Casos utilizados en el análisis
1	,057	510
2	,040	363
3	,075	676
4	,143	1286
5	,056	506
6	,092	825
7	,025	224
8	,085	764
9	,057	513
10	,085	760
11	,055	491
12	,108	970
13	,012	107
14	,109	979
Total	1,000	8974

**Resultados de la clasificación (b,c)**

	No de grupo	Grupo de pertenencia pronosticado													Total	
		1	2	3	4	5	6	7	8	9	10	11	12	13		14
Original Recuento	1	399							59			52				510
	2		320									43				363
	3			559						117						676
	4	4			1154							128				1286
	5			6		477					23					506
	6						788	14	23							825
	7						34	190								224
	8						12		752							764
	9			82	31					395					5	513
	10			44		44					616				56	760
	11	22										467			2	491
	12				125		20						825			970
	13													107		107
	14						29		21		38				891	979
%	1	78,2						11,6			10,2				100	
	2		88,2									11,8			100	
	3			82,7					17,3						100	
	4	0,3			89,7							10,0			100	
	5			1,2		94,3				4,5					100	
	6						95,5	1,7	2,8						100	
	7						15,2	84,8							100	
	8						1,6		98,4						100	
	9			16,0	6,0					77,0					1,0	100
	10			5,8		5,8					81,1				7,4	100
	11	4,5										95,1			0,4	100
	12				12,9		2,1						85,1			100
	13													100		100
	14						3,0		2,1		3,9				91,0	100

	No de grupo	Grupo de pertenencia pronosticado														Total
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Validación cruzada (a) Recuento	1	398							59			53				510
	2		320									43				363
	3			559						117						676
	4	4			1153							128		1		1286
	5			6		477					23					506
	6						787	15	23							825
	7						34	190								224
	8						12		752							764
	9			82	31					395					5	513
	10			44		44					616				56	760
	11	22										467			2	491
	12		1		126		20						823			970
	13													107		107
	14						29		21		39				890	979
%	1	78,0							11,6			10,4				100
	2		88,2									11,8				100
	3			82,7						17,3						100
	4	0,3			89,7							10,0		0,1		100
	5			1,2		94,3					4,5					100
	6						95,4	1,8	2,8							100
	7						15,2	84,8								100
	8						1,6		98,4							100
	9			16,0	6,0					77,0					1,0	100
	10			5,8		5,8					81,1				7,4	100
	11	4,5										95,1			0,4	100
	12		0,1		13,0		2,1						84,8			100
	13													100		100
	14						3,0		2,1		4,0				90,9	100

a Cada caso se clasifica mediante las funciones derivadas a partir del resto de los casos.

b Clasificados correctamente el 88,5% de los casos agrupados originales.

c Clasificados correctamente el 88,4% de los casos agrupados validados mediante validación cruzada.

**ANEXO 32. ANÁLISIS DE CORRELACIÓN CANÓNICA, CON 17  
VARIABLES PREDICTIVAS EN SU ESCALA  
ORIGINAL**

## Análisis Canónico

Resumen del Análisis

Variables en serie 1:

DJ, V, DV, Vx, T, Tn, Tx, Hf, HR, Rad, P, ETo, VolExp, Qent, Qsal, Filt, Profundidad

Variables en serie 2:

Ta, Odis, Ph, Con

Número de casos completos: 4401

### Correlaciones Canónicas

Número	Autovalor	Canónica Correlación	Wilks Lambda	Chi-Cuadrado	G.I.	P-Valor
1	0,896446	0,946809	0,0137515	18813,9	68	0,0000
2	0,703161	0,838547	0,132796	8861,13	48	0,0000
3	0,428104	0,654297	0,447367	3530,4	30	0,0000
4	0,217747	0,466633	0,782253	1077,84	14	0,0000

### Coefficientes para las Variables Canónicas de la Primera serie

DJ	0,304264	-0,0702927	0,082606	-0,236832
V	-0,0276439	-0,0420981	0,0949846	0,262851
DV	-0,0199393	-0,00239246	-0,00447209	0,0858118
Vx	-0,045199	0,0217572	-0,0583784	-0,0712968
T	0,539116	0,119	-0,463028	0,22811
Tn	0,123938	0,350086	-0,26978	-1,2746
Tx	-0,286091	-0,128381	0,668938	1,33589
Hf	-0,0209047	0,246337	-0,0907551	-0,916975
HR	0,0823389	-0,102975	-0,0900482	0,250719
Rad	0,183142	-0,0277103	-0,186208	-0,373472
P	0,0134344	-0,0627029	-0,0812639	0,12574
ETo	0,214414	0,000798271	-0,286013	-0,512449
VolExp	-0,165675	0,235588	1,26412	0,609458
Qent	-0,0384391	-0,0551667	0,0253396	0,188839
Qsal	0,0928459	0,173228	0,371925	0,229593
Filt	0,00277595	-0,0196346	-0,0339945	-0,0667668
Profundidad	0,285316	-0,855746	0,342407	-0,204331

### Coefficientes para las Variables Canónicas de la Segunda serie

Ta	0,648028	0,543355	-1,16754	-0,0472823
Odis	-0,740516	0,361833	-1,3763	-0,394182
pH	0,160005	-0,398293	0,187775	1,18231
Con	0,220593	-0,563638	-1,62206	-0,00984443

Este procedimiento encuentra las combinaciones lineales de dos grupos de variables que tienen la correlación más alta entre ellos. En este caso, se ha formado 4 series de combinaciones lineales. La primera serie de combinaciones lineales es

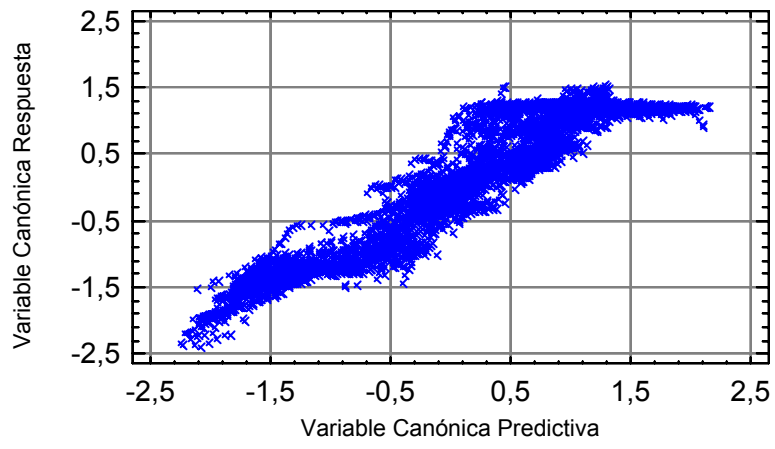
$$0,304264 \cdot DJ - 0,0276439 \cdot V - 0,0199393 \cdot DV - 0,045199 \cdot Vx + 0,539116 \cdot T + 0,123938 \cdot Tn - 0,286091 \cdot Tx - 0,0209047 \cdot Hf + 0,0823389 \cdot HR + 0,183142 \cdot Rad + 0,0134344 \cdot P + 0,214414 \cdot ETo - 0,165675 \cdot VolExp - 0,0384391 \cdot Qent + 0,0928459 \cdot Qsal + 0,00277595 \cdot Filt + 0,285316 \cdot Profundidad$$

Y

$$0,648028 \cdot Ta - 0,740516 \cdot Odis + 0,160005 \cdot pH + 0,220593 \cdot Con$$

Donde las variables han sido estandarizadas primero substrayendo sus medias y dividiendo por sus desviaciones típicas. La tabla muestra la correlación estimada entre cada grupo de variables canónicas. Dado que 4 de los p-valores es menor de 0.05, esos grupos tienen correlaciones estadísticamente significativas al 95% de nivel de confianza.

### Variables Canónicas en el Periodo de Calibración



**ANEXO 33. ANÁLISIS DE CORRELACIÓN CANÓNICA, CON 11  
VARIABLES PREDICTIVAS EN SU ESCALA  
ORIGINAL**

## Análisis Canónico

Resumen del Análisis

Variables en serie 1:

DJ, V, DV, Vx, T, HR, ETo, VolExp, Qent, Qsal, Profundidad

Variables en serie 2:

Ta, Odis, pH, Con

Número de casos completos: 4401

### Correlaciones Canónicas

Número	Autovalor	Canónica Correlación	Wilks Lambda	Chi-Cuadrado	G.I.	P-Valor
1	0,892434	0,944687	0,0174482	17781,1	44	0,0000
2	0,685543	0,827975	0,162209	7988,49	30	0,0000
3	0,41794	0,646483	0,515837	2907,35	18	0,0000
4	0,113773	0,337303	0,886227	530,477	8	0,0000

Coefficientes para las Variables Canónicas de la Primera serie

DJ	0,305928	-0,0898232	0,0421831	-0,287779
V	0,00619051	-0,0346218	0,0821152	0,409576
DV	-0,0185978	-0,0153291	-0,0168351	0,199777
Vx	-0,073658	-0,00972828	-0,0571881	0,110165
T	0,408906	0,130424	0,000124128	1,81004
HR	0,11726	-0,140252	-0,158113	0,542226
ETo	0,375032	-0,0308727	-0,438027	-1,49703
VolExp	-0,161625	0,367439	1,20744	0,401369
Qent	-0,034636	-0,0758131	0,0404518	0,454832
Qsal	0,0719432	0,266439	0,369639	-0,0648378
Profundidad	0,288186	-0,868897	0,363616	-0,0812949

Coefficientes para las Variables Canónicas de la Segunda serie

Ta	0,636553	0,517912	-1,1729	-0,177293
Odis	-0,756466	0,308884	-1,34105	-0,512638
pH	0,17017	-0,305527	0,103977	1,21829
Con	0,204637	-0,59019	-1,61247	-0,0852958

Este procedimiento encuentra las combinaciones lineales de dos grupos de variables que tienen la correlación más alta entre ellos. En este caso, se ha formado 4 series de combinaciones lineales. La primera serie de combinaciones lineales es

$$0,305928 \cdot DJ + 0,00619051 \cdot V - 0,0185978 \cdot DV - 0,073658 \cdot Vx + 0,408906 \cdot T + 0,11726 \cdot HR + 0,375032 \cdot ETo - 0,161625 \cdot VolExp - 0,034636 \cdot Qent + 0,0719432 \cdot Qsal + 0,288186 \cdot Profundidad$$

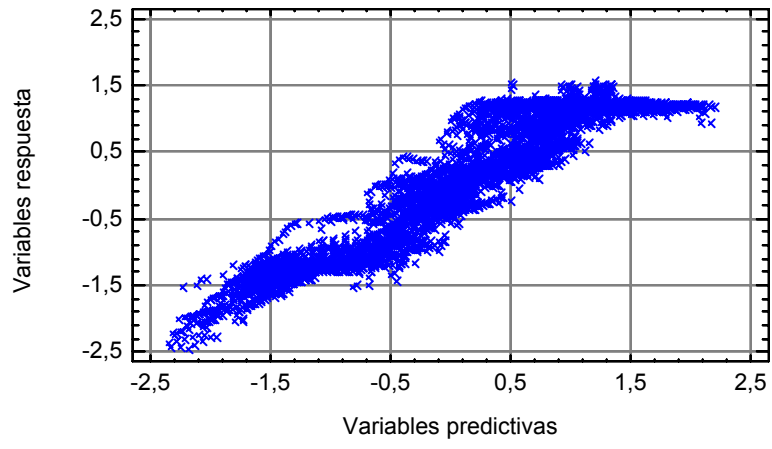
Y

$$0,636553 \cdot Ta - 0,756466 \cdot Odis + 0,17017 \cdot pH + 0,204637 \cdot Con$$

Donde las variables han sido estandarizadas primero substrayendo sus medias y dividiendo por sus desviaciones típicas. La tabla muestra la correlación estimada entre cada grupo de variables canónicas. Dado que 4 de los p-valores es menor de 0.05, esos grupos tienen correlaciones estadísticamente significativas al 95% de nivel de confianza.



### Variables Canónicas



**ANEXO 34. ANÁLISIS DE CORRELACIÓN CANÓNICA, CON 17  
VARIABLES PREDICTIVAS ESTANDARIZADAS**

## Análisis Canónico

Resumen del Análisis

Variables en serie 1:

ZDJ, ZV, ZDV, ZVx, ZT, ZTn, ZTx, ZHf, ZHR, ZRad, ZP, ZETo, ZVolExp, ZQent, ZQsal, ZFilt, ZProfundidad

Variables en serie 2:

ZTa, ZOdis, ZpH, ZCon

Número de casos completos: 4401

### Correlaciones Canónicas

Número	Autovalor	Canónica Correlación	Wilks Lambda	Chi-Cuadrado	G.I.	P-Valor
1	0,904494	0,951049	0,0122793	19310,9	68	0,0000
2	0,704042	0,839072	0,128571	9003,04	48	0,0000
3	0,434364	0,659063	0,434423	3659,27	30	0,0000
4	0,231973	0,481636	0,768027	1158,39	14	0,0000

### Coefficientes para las Variables Canónicas de la Primera serie

ZDJ	0,272633	-0,0853333	0,120717	-0,145515
ZV	-0,0854055	-0,0563987	0,170572	0,370544
ZDV	-0,0103936	0,00112355	-0,0156338	0,0600345
ZVx	-0,0559591	0,0158472	-0,0446145	-0,0300857
ZT	0,31868	0,0632276	-0,162413	0,751575
ZTn	0,0564557	0,313672	-0,188462	-0,950595
ZTx	-0,250424	-0,115229	0,631036	1,09856
ZHf	-0,011932	0,251164	-0,11904	-0,876824
ZHR	0,123647	-0,0860781	-0,138145	0,135877
ZRad	-0,095708	-0,122349	0,171304	0,347837
ZP	-0,0250993	-0,074738	-0,0286943	0,214129
ZETo	0,752064	0,171227	-0,979889	-1,75258
ZVolExp	-0,246889	0,206596	1,35838	0,664944
ZQent	-0,0156209	-0,0472966	-0,00193201	0,119155
ZQsal	0,114434	0,176609	0,346753	0,137499
ZFilt	0,00165795	-0,0196199	-0,0341488	-0,0586808
ZProfundidad	0,273012	-0,860242	0,331521	-0,218713

### Coefficientes para las Variables Canónicas de la Segunda serie

ZTa	0,671568	0,545498	-1,15396	0,0210641
ZOdis	-0,710067	0,382981	-1,4012	-0,337266
ZpH	0,134663	-0,409633	0,242198	1,17165
ZCon	0,238783	-0,553555	-1,62143	0,0660953

Este procedimiento encuentra las combinaciones lineales de dos grupos de variables que tienen la correlación más alta entre ellos. En este caso, se ha formado 4 series de combinaciones lineales. La primera serie de combinaciones lineales es

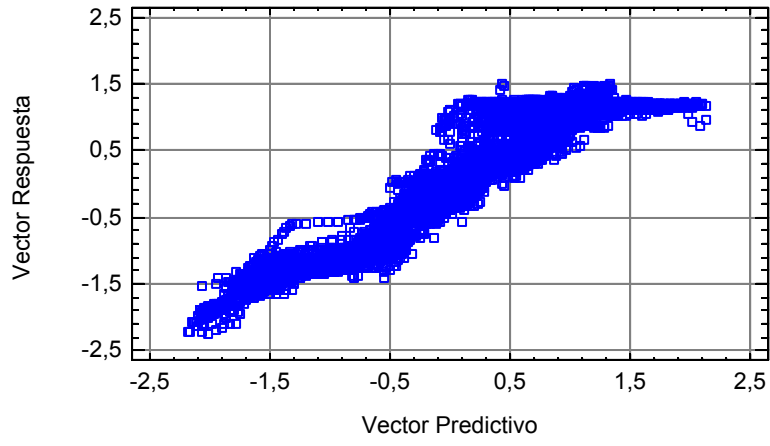
$$0,272633*ZDJ - 0,0854055*ZV - 0,0103936*ZDV - 0,0559591*ZVx + 0,31868*ZT + 0,0564557*ZTn - 0,250424*ZTx - 0,011932*ZHf + 0,123647*ZHR - 0,095708*ZRad - 0,0250993*ZP + 0,752064*ZETo - 0,246889*ZVolExp - 0,0156209*ZQent + 0,114434*ZQsal + 0,00165795*ZFilt + 0,273012*ZProfundidad$$

Y

$$0,671568*ZTa - 0,710067*ZOdis + 0,134663*ZpH + 0,238783*ZCon$$

Donde las variables han sido estandarizadas primero substrayendo sus medias y dividiendo por sus desviaciones típicas. La tabla muestra la correlación estimada entre cada grupo de variables canónicas. Dado que 4 de los p-valores es menor de 0.05, esos grupos tienen correlaciones estadísticamente significativas al 95% de nivel de confianza.

### Variables Canónicas



**ANEXO 35. ANÁLISIS DE CORRELACIÓN CANÓNICA, CON 16  
VARIABLES PREDICTIVAS ESTANDARIZADAS**

## Análisis Canónico

Resumen del Análisis

Variables en serie 1:

ZV, ZDV, ZVx, ZT, ZTn, ZTx, ZHf, ZHR, ZRad, ZP, ZETo, ZVolExp, ZQent, ZQsal, ZFilt, ZProfundidad

Variables en serie 2:

ZTa, ZOdis, ZpH, ZCon

Número de casos completos: 4401

### Correlaciones Canónicas

Número	Autovalor	Canónica Correlación	Wilks Lambda	Chi-Cuadrado	G.I.	P-Valor
1	0,888754	0,942737	0,0144282	18605,2	64	0,0000
2	0,702735	0,838293	0,129695	8965,85	45	0,0000
3	0,432815	0,657887	0,436296	3640,8	28	0,0000
4	0,23077	0,480385	0,76923	1151,65	13	0,0000

Coefficientes para las Variables Canónicas de la Primera serie

ZV	-0,0838986	-0,0551825	0,17334	0,368866
ZDV	-0,03772	0,0105656	-0,028211	0,0757902
ZVx	-0,0693106	0,0217486	-0,051991	-0,0214816
ZT	0,677751	-0,0658852	0,00714095	0,555952
ZTn	-0,0541823	0,352256	-0,248008	-0,88621
ZTx	-0,50511	-0,0234373	0,5199	1,23667
ZHf	-0,0169983	0,255278	-0,129209	-0,872115
ZHR	0,124527	-0,0902762	-0,135151	0,135398
ZRad	-0,242082	-0,0713798	0,106942	0,427207
ZP	-0,0298414	-0,0732598	-0,0293633	0,217089
ZETo	0,950195	0,0867622	-0,893896	-1,86081
ZVolExp	-0,389282	0,264273	1,29593	0,734255
ZQent	0,0628955	-0,0738968	0,0351601	0,0754598
ZQsal	0,247545	0,130787	0,408679	0,0614406
ZFilt	-0,00127224	-0,0188533	-0,0355367	-0,057187
ZProfundidad	0,255128	-0,86664	0,341321	-0,230359

Coefficientes para las Variables Canónicas de la Segunda serie

ZTa	0,69656	0,523291	-1,14932	0,0257865
ZOdis	-0,688102	0,392857	-1,41293	-0,322156
ZpH	0,129993	-0,414543	0,253684	1,16801
ZCon	0,243333	-0,567943	-1,61543	0,073958

Este procedimiento encuentra las combinaciones lineales de dos grupos de variables que tienen la correlación más alta entre ellos. En este caso, se ha formado 4 series de combinaciones lineales. La primera serie de combinaciones lineales es

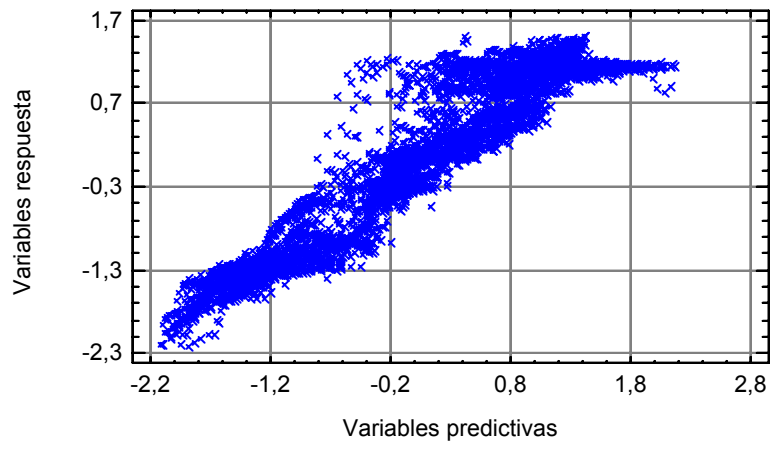
$$-0,0838986*ZV - 0,03772*ZDV - 0,0693106*ZVx + 0,677751*ZT - 0,0541823*ZTn - 0,50511*ZTx - 0,0169983*ZHf + 0,124527*ZHR - 0,242082*ZRad - 0,0298414*ZP + 0,950195*ZETo - 0,389282*ZVolExp + 0,0628955*ZQent + 0,247545*ZQsal - 0,00127224*ZFilt + 0,255128*ZProfundidad$$

Y

$$0,69656*ZTa - 0,688102*ZOdis + 0,129993*ZpH + 0,243333*ZCon$$

Donde las variables han sido estandarizadas primero substrayendo sus medias y dividiendo por sus desviaciones típicas. La tabla muestra la correlación estimada entre cada grupo de variables canónicas. Dado que 4 de los p-valores es menor de 0.05, esos grupos tienen correlaciones estadísticamente significativas al 95% de nivel de confianza.

Variables Canónicas



**ANEXO 36. ANÁLISIS DE CORRELACIÓN CANÓNICA, CON 9  
VARIABLES PREDICTIVAS ESTANDARIZADAS**



## Análisis Canónico

Resumen del Análisis

VARIABLES EN SERIE 1:

ZT, ZTn, ZTx, ZHf, ZHR, ZETo, ZVolExp, ZQsal, ZProfundidad

VARIABLES EN SERIE 2:

ZTa, ZOdis, ZpH, ZCon

Número de casos completos: 4401

### Correlaciones Canónicas

Número	Autovalor	Canónica Correlación	Wilks Lambda	Chi-Cuadrado	G.I.	P-Valor
1	0,878535	0,937302	0,016546	18018,4	36	0,0000
2	0,698269	0,835625	0,13622	8757,38	24	0,0000
3	0,426065	0,652736	0,451461	3493,61	14	0,0000
4	0,213395	0,461947	0,786605	1054,45	6	0,0000

### Coefficientes para las Variables Canónicas de la Primera serie

ZT	0,887309	-0,229647	-0,301244	0,297692
ZTn	-0,0299283	0,41887	-0,199611	-0,9486
ZTx	-0,53691	0,156156	0,546816	0,960649
ZHf	-0,0158312	0,269957	-0,142189	-0,985355
ZHR	0,166896	-0,104247	-0,171385	0,0672761
ZETo	0,521978	-0,0156019	-0,584138	-1,08298
ZVolExp	-0,345362	0,288191	1,29268	0,699026
ZQsal	0,24061	0,149963	0,432042	0,081627
ZProfundidad	0,252772	-0,871791	0,345578	-0,220588

### Coefficientes para las Variables Canónicas de la Segunda serie

ZTa	0,693675	0,519758	-1,15285	-0,0149347
ZOdis	-0,697042	0,392105	-1,40205	-0,350143
ZpH	0,143933	-0,403277	0,225959	1,17599
ZCon	0,233976	-0,569055	-1,61764	0,0391144

Este procedimiento encuentra las combinaciones lineales de dos grupos de variables que tienen la correlación más alta entre ellos. En este caso, se ha formado 4 series de combinaciones lineales. La primera serie de combinaciones lineales es

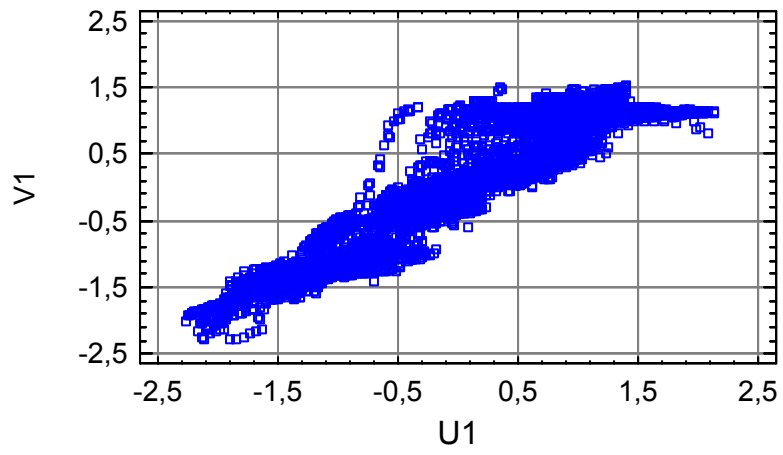
$$0,887309 \cdot ZT - 0,0299283 \cdot ZTn - 0,53691 \cdot ZTx - 0,0158312 \cdot ZHf + 0,166896 \cdot ZHR + 0,521978 \cdot ZETo - 0,345362 \cdot ZVolExp + 0,24061 \cdot ZQsal + 0,252772 \cdot ZProfundidad$$

Y

$$0,693675 \cdot ZTa - 0,697042 \cdot ZOdis + 0,143933 \cdot ZpH + 0,233976 \cdot ZCon$$

Donde las variables han sido estandarizadas primero substrayendo sus medias y dividiendo por sus desviaciones típicas. La tabla muestra la correlación estimada entre cada grupo de variables canónicas. Dado que 4 de los p-valores es menor de 0,05, esos grupos tienen correlaciones estadísticamente significativas al 95% de nivel de confianza.

### Primer par de Variables Canónicas



**ANEXO 37. PLSR DESARROLLADO DENTRO DEL PERIODO DE  
CALIBRACIÓN (17 VARIABLES PREDICTIVAS)**

## **Partial Least Squares**

Number of dependent variables: 1

Ta

Number of independent variables: 17

DJ, V, DV, Vx, T, Tn, Tx, Hf, HR, Rad, P, ETo, VolExp, Qent, Qsal, Filt, Profundidad

Number of complete cases: 4401

Number of components extracted: 7

### **Analysis of Variance for Ta**

Source	Sum of Squares	Df	Mean Square	F-ratio	P-Value
Model	98827,0	7	14118,1	1831,53	0,0
Residual	33862,9	4393	7,70839		
Total (corr.)	132690,	4400			

### **Model for Ta**

Component	% Variation in Y	R-Squared
1	58,9331	58,9331
2	13,1242	72,0573
3	1,91718	73,9745
4	0,155507	74,13
5	0,164368	74,2944
6	0,083094	74,3775
7	0,102163	74,4796

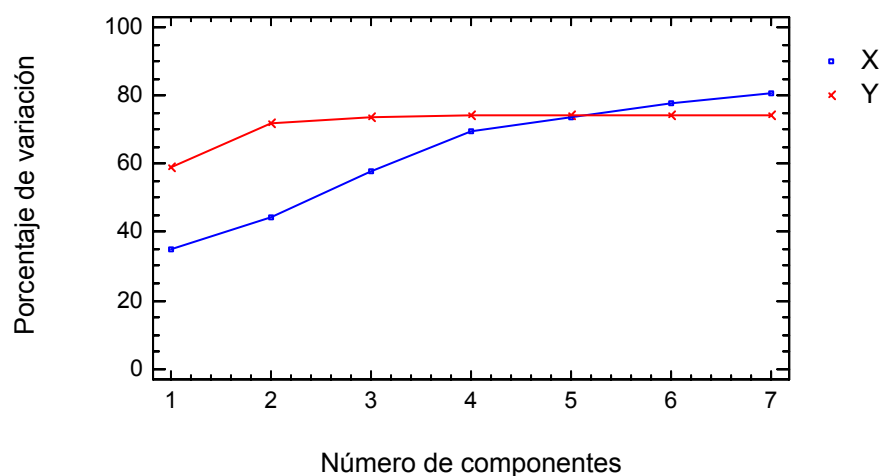
### **Independent and Dependent Variables**

Component	% Variation in X	Cumulative % of X	% Variation in Y	Cumulative % of Y
1	34,9421	34,9421	58,9331	58,9331
2	9,70516	44,6473	13,1242	72,0573
3	13,4685	58,1158	1,91718	73,9745
4	11,3855	69,5013	0,155507	74,13
5	4,37061	73,872	0,164368	74,2944
6	3,90933	77,7813	0,083094	74,3775
7	2,95417	80,7355	0,102163	74,4796

### **The StatAdvisor**

The output shows the result of fitting models for 1 dependent variables based on 17 predictor variables. To fit the model, 7 components have been extracted. The P-Values in the ANOVA table indicate that 1 of the models are significant at the 5.0% significance level.

Gráfico comparativo del modelo



**Regression Coefficients**

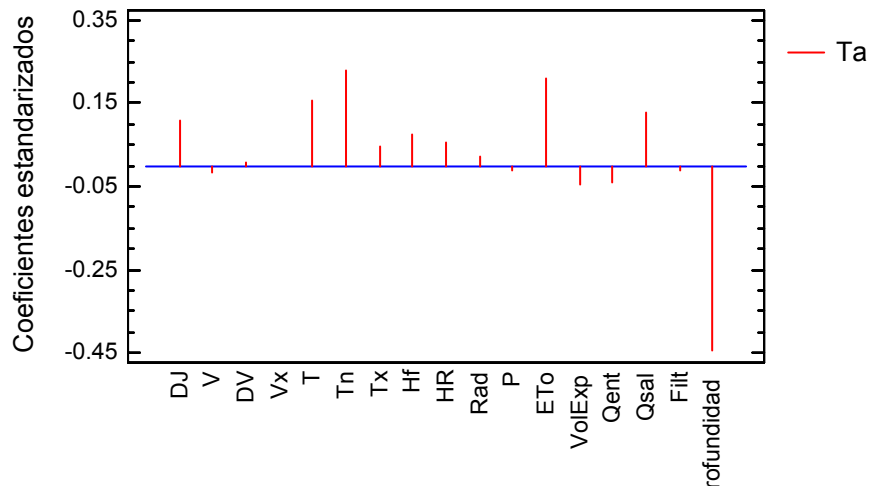
Standardized Coefficients

	<i>Ta</i>
Constant	0,0
DJ	0,109638
V	-0,016443
DV	0,00724376
Vx	-0,00219588
T	0,157697
Tn	0,228678
Tx	0,043938
Hf	0,0756468
HR	0,057989
Rad	0,0219019
P	-0,0120057
ETo	0,211395
VolExp	-0,0444486
Qent	-0,0384057
Qsal	0,128394
Filt	-0,010624
Profundidad	-0,441343

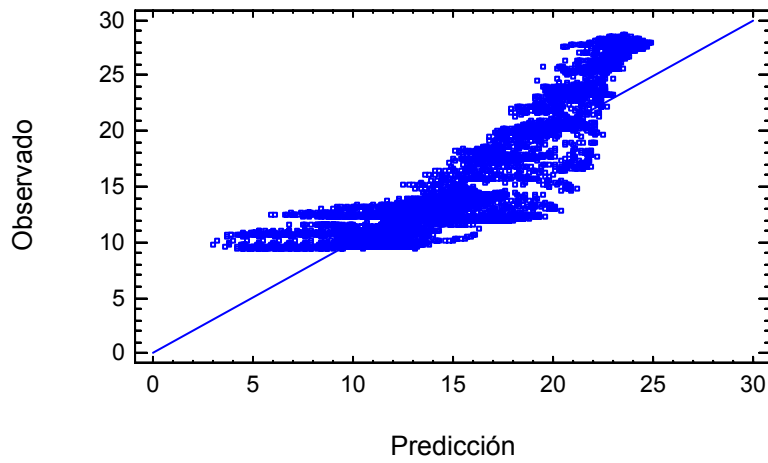
Unstandardized Coefficients

	<i>Ta</i>
Constant	8,21165
DJ	0,004911
V	-0,275504
DV	0,00055768
Vx	-0,00793173
T	0,132786
Tn	0,198842
Tx	0,0369386
Hf	0,109653
HR	0,0236497
Rad	0,00163213
P	-0,0274099
ETo	0,758365
VolExp	-1,85832E-7
Qent	-0,00000603205
Qsal	0,00006267
Filt	-0,0000645576
Profundidad	-0,312097

Coeficientes PLSR



Observaciones y Predicciones para Ta



## **Partial Least Squares**

Number of dependent variables: 1

Odis

Number of independent variables: 17

DJ, V, DV, Vx, T, Tn, Tx, Hf, HR, Rad, P, ETo, VolExp, Qent, Qsal, Filt, Profundidad

Number of complete cases: 4401

Number of components extracted: 14

### **Analysis of Variance for Odis**

<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-ratio</i>	<i>P-Value</i>
Model	63563,9	14	4540,28	1075,59	0,0
Residual	18514,1	4386	4,22119		
Total (corr.)	82078,0	4400			

### **Model for Odis**

<i>Component</i>	<i>% Variation in Y</i>	<i>R-Squared</i>
1	47,3968	47,3968
2	26,1711	73,5679
3	2,29779	75,8657
4	0,507514	76,3732
5	0,407192	76,7804
6	0,207851	76,9883
7	0,0877983	77,076
8	0,102892	77,1789
9	0,100952	77,2799
10	0,0887232	77,3686
11	0,0447219	77,4133
12	0,00926014	77,4226
13	0,00753942	77,4301
14	0,013098	77,4432

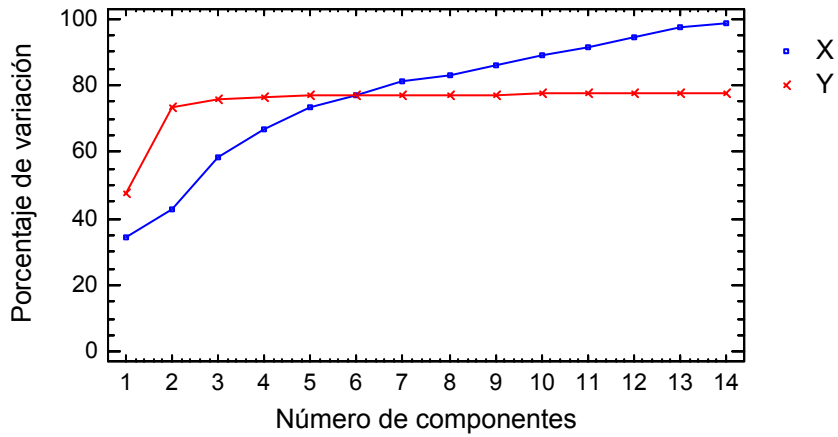
### **Independent and Dependent Variables**

<i>Component</i>	<i>% Variation in X</i>	<i>Cumulative % of X</i>	<i>% Variation in Y</i>	<i>Cumulative % of Y</i>
1	34,2642	34,2642	47,3968	47,3968
2	8,82656	43,0908	26,1711	73,5679
3	15,0595	58,1503	2,29779	75,8657
4	8,47104	66,6213	0,507514	76,3732
5	6,72819	73,3495	0,407192	76,7804
6	3,67659	77,0261	0,207851	76,9883
7	3,87891	80,905	0,0877983	77,076
8	2,2392	83,1442	0,102892	77,1789
9	2,91264	86,0569	0,100952	77,2799
10	2,91355	88,9704	0,0887232	77,3686
11	2,22927	91,1997	0,0447219	77,4133
12	3,48397	94,6837	0,00926014	77,4226
13	2,97932	97,663	0,00753942	77,4301
14	0,941691	98,6047	0,013098	77,4432

### **The StatAdvisor**

The output shows the result of fitting models for 1 dependent variables based on 17 predictor variables. To fit the model, 14 components have been extracted. The P-Values in the ANOVA table indicate that 1 of the models are significant at the 5,0% significance level.

Gráfico comparativo del modelo



**Regression Coefficients**

Standardized Coefficients

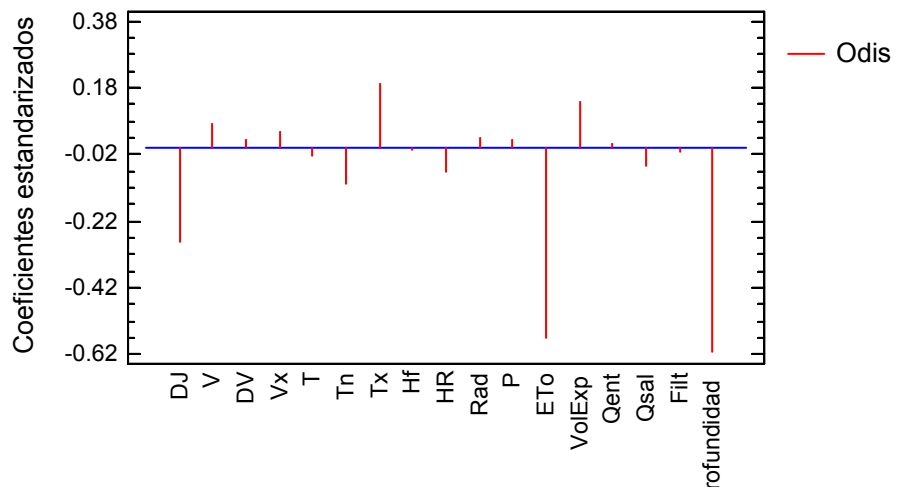
	<i>Odis</i>
Constant	0,0
DJ	-0,279619
V	0,0702126
DV	0,0242216
Vx	0,0501549
T	-0,0241526
Tn	-0,106816
Tx	0,193173
Hf	-0,00670741
HR	-0,0741299
Rad	0,0296855
P	0,0253633
ETo	-0,571932
VolExp	0,138546
Qent	0,0125015
Qsal	-0,0512261
Filt	-0,0111546
Profundidad	-0,614221

Unstandardized Coefficients

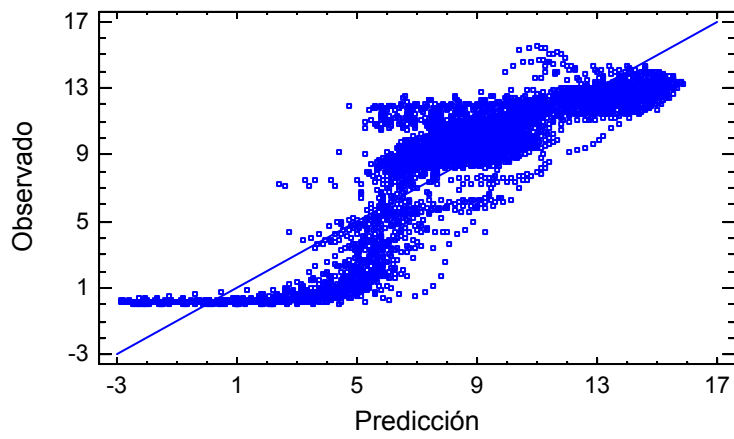
	<i>Odis</i>
Constant	11,5234
DJ	-0,0098508
V	0,925246
DV	0,00146662
Vx	0,142484
T	-0,0159951
Tn	-0,0730491
Tx	0,127727
Hf	-0,00764679
HR	-0,0237776
Rad	0,00173985
P	0,0455428
ETo	-1,6137
VolExp	4,55566E-7
Qent	0,00000154428
Qsal	-0,0000196653
Filt	-0,00005331
Profundidad	-0,341611



Coeficientes PLSR



Observaciones y Predicciones



## **Partial Least Squares**

Number of dependent variables: 1

pH

Number of independent variables: 17

DJ, V, DV, Vx, T, Tn, Tx, Hf, HR, Rad, P, ETo, VolExp, Qent, Qsal, Filt, Profundidad

Number of complete cases: 4401

Number of components extracted: 11

### **Analysis of Variance for pH**

<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-ratio</i>	<i>P-Value</i>
Model	205,818	11	18,7108	161,587	0,0
Residual	508,22	4389	0,115794		
Total (corr.)	714,038	4400			

### **Model for pH**

<i>Component</i>	<i>% Variation in Y</i>	<i>R-Squared</i>
1	11,2767	11,2767
2	9,73981	21,0166
3	2,97619	23,9927
4	1,30994	25,3027
5	0,862639	26,1653
6	0,939449	27,1048
7	0,425079	27,5299
8	0,708498	28,2384
9	0,460499	28,6989
10	0,05605	28,7549
11	0,0696674	28,8246

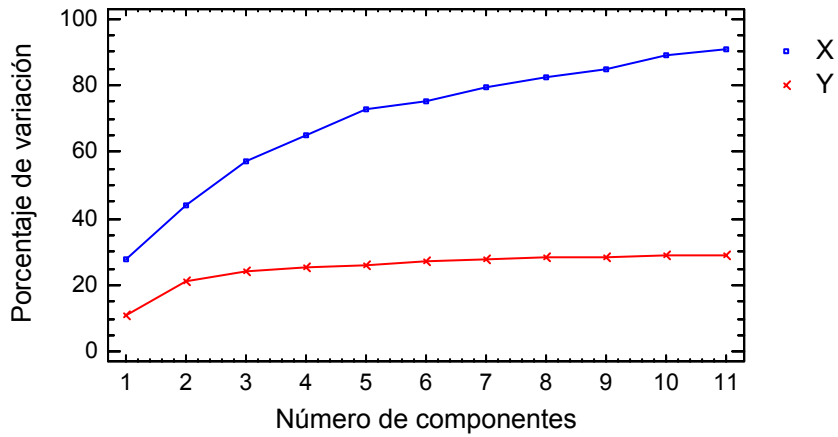
### **Independent and Dependent Variables**

<i>Component</i>	<i>% Variation in X</i>	<i>Cumulative % of X</i>	<i>% Variation in Y</i>	<i>Cumulative % of Y</i>
1	27,7952	27,7952	11,2767	11,2767
2	16,0414	43,8366	9,73981	21,0166
3	13,3591	57,1957	2,97619	23,9927
4	8,00312	65,1988	1,30994	25,3027
5	7,82276	73,0216	0,862639	26,1653
6	2,45978	75,4814	0,939449	27,1048
7	3,86153	79,3429	0,425079	27,5299
8	2,88304	82,226	0,708498	28,2384
9	2,32803	84,554	0,460499	28,6989
10	4,63474	89,1887	0,05605	28,7549
11	1,68388	90,8726	0,0696674	28,8246

### **The StatAdvisor**

The output shows the result of fitting models for 1 dependent variables based on 17 predictor variables. To fit the model, 11 components have been extracted. The P-Values in the ANOVA table indicate that 1 of the models are significant at the 5,0% significance level.

Gráfico comparativo del modelo



**Regression Coefficients**

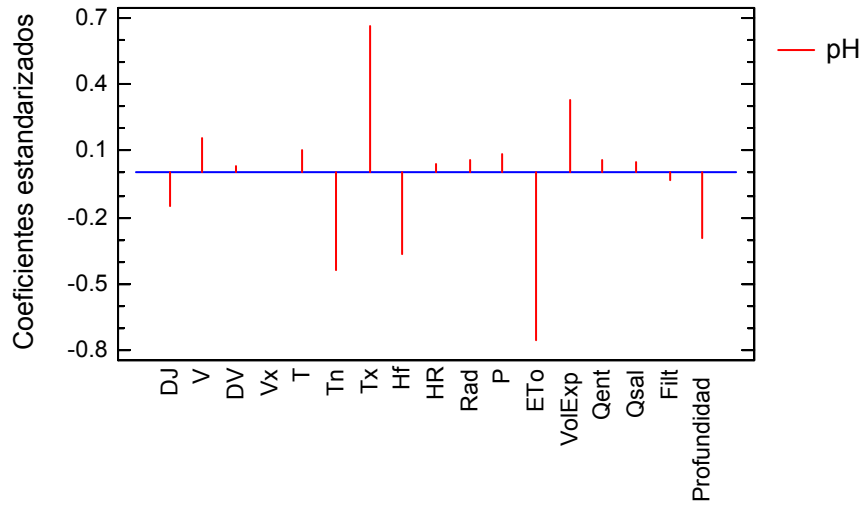
Standardized Coefficients

	<i>pH</i>
Constant	0,0
DJ	-0,14902
V	0,15766
DV	0,0314621
Vx	0,000590769
T	0,103308
Tn	-0,437354
Tx	0,657646
Hf	-0,364084
HR	0,0395135
Rad	0,0566685
P	0,0855655
ETo	-0,75028
VolExp	0,330298
Qent	0,055344
Qsal	0,0509874
Filt	-0,0331481
Profundidad	-0,296255

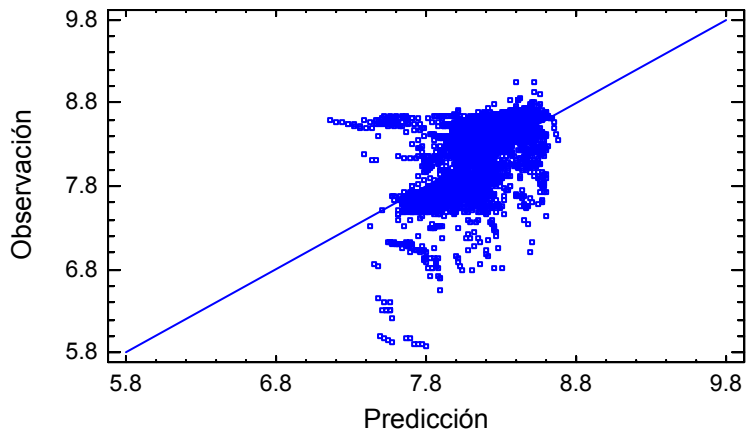
Unstandardized Coefficients

	<i>pH</i>
Constant	6,96525
DJ	-0,000489663
V	0,193781
DV	0,000177685
Vx	0,000156537
T	0,00638118
Tn	-0,0278971
Tx	0,0405578
Hf	-0,0387145
HR	0,00118214
Rad	0,000309782
P	0,0143305
ETo	-0,197446
VolExp	1,013E-7
Qent	6,37647E-7
Qsal	0,00000182566
Filt	-0,0000147761
Profundidad	-0,0153681

Coeficientes PLSR



Observaciones y Predicciones



## **Partial Least Squares**

Number of dependent variables: 1

Con

Number of independent variables: 17

DJ, V, DV, Vx, T, Tn, Tx, Hf, HR, Rad, P, ETo, VolExp, Qent, Qsal, Filt, Profundidad

Number of complete cases: 4401

Number of components extracted: 8

### **Analysis of Variance for Con**

<i>Source</i>	<i>Sum of Squares</i>	<i>Df</i>	<i>Mean Square</i>	<i>F-ratio</i>	<i>P-Value</i>
Model	2,89641E7	8	3,62051E6	1034,37	0,0
Residual	1,5373E7	4392	3500,22		
Total (corr.)	4,4337E7	4400			

### **Model for Con**

<i>Component</i>	<i>% Variation in Y</i>	<i>R-Squared</i>
1	39,565	39,565
2	16,5202	56,0852
3	5,74515	61,8303
4	1,36797	63,1983
5	1,83961	65,0379
6	0,151462	65,1894
7	0,0847997	65,2742
8	0,0528898	65,327

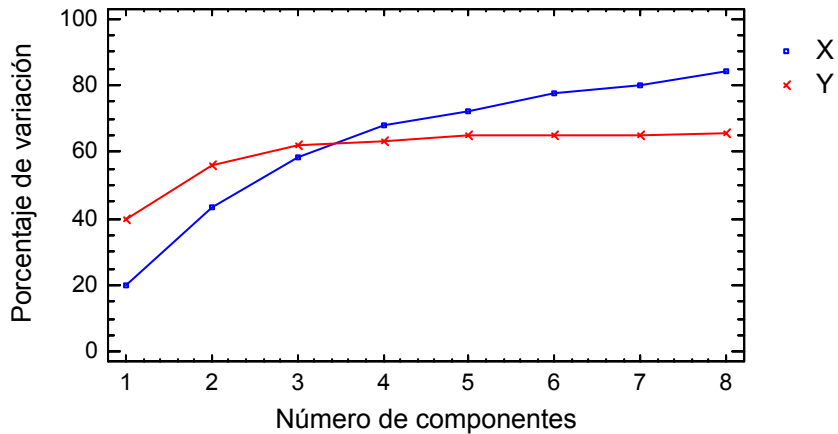
### **Independent and Dependent Variables**

<i>Component</i>	<i>% Variation in X</i>	<i>Cumulative % of X</i>	<i>% Variation in Y</i>	<i>Cumulative % of Y</i>
1	20,2893	20,2893	39,565	39,565
2	23,06	43,3492	16,5202	56,0852
3	14,9767	58,326	5,74515	61,8303
4	9,65923	67,9852	1,36797	63,1983
5	4,49641	72,4816	1,83961	65,0379
6	5,16544	77,6471	0,151462	65,1894
7	2,33894	79,986	0,0847997	65,2742
8	4,12395	84,11	0,0528898	65,327

### **The StatAdvisor**

The output shows the result of fitting models for 1 dependent variables based on 17 predictor variables. To fit the model, 8 components have been extracted. The P-Values in the ANOVA table indicate that 1 of the models are significant at the 5.0% significance level.

Gráfico comparativo del modelo



**Regression Coefficients**

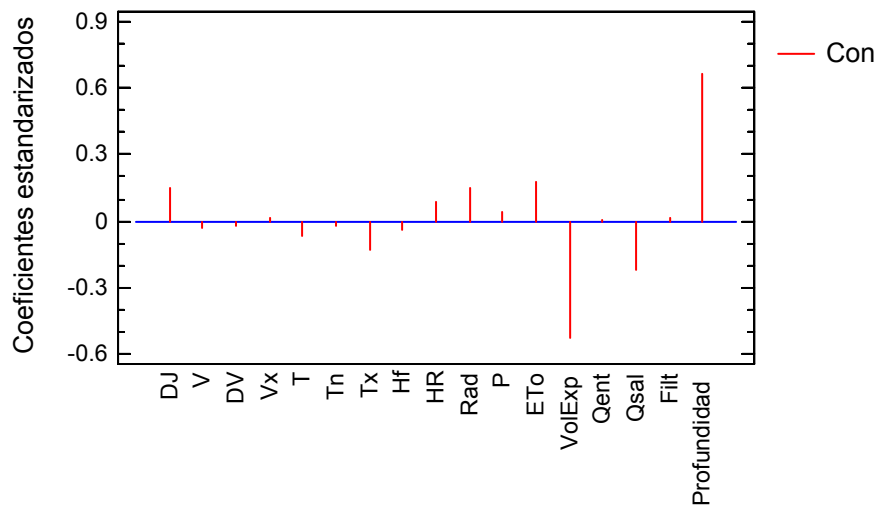
Standardized Coefficients

	Con
Constant	0,0
DJ	0,153485
V	-0,034166
DV	-0,0175249
Vx	0,0120414
T	-0,0622402
Tn	-0,0186619
Tx	-0,13243
Hf	-0,0405804
HR	0,0873054
Rad	0,147363
P	0,0446914
ETo	0,175285
VolExp	-0,523791
Qent	0,00260884
Qsal	-0,216399
Filt	0,0192388
Profundidad	0,664889

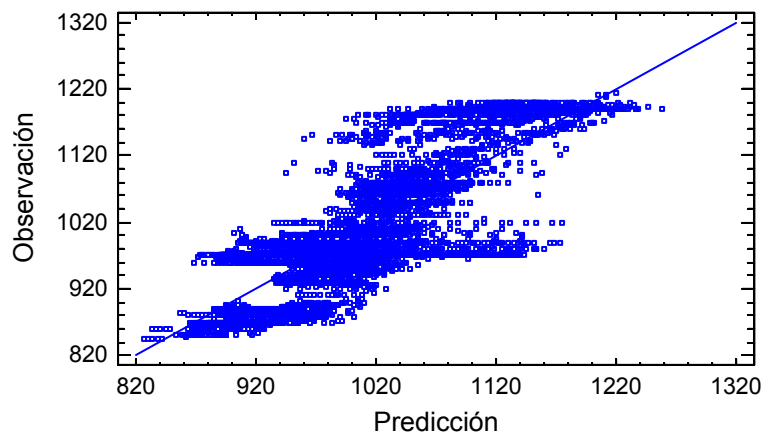
Unstandardized Coefficients

	Con
Constant	1277,51
DJ	0,125672
V	-10,4642
DV	-0,0246626
Vx	0,795063
T	-0,957994
Tn	-0,296621
Tx	-2,03513
Hf	-1,07525
HR	0,650858
Rad	0,200736
P	1,86513
ETo	11,4946
VolExp	-0,0000400299
Qent	0,00000748997
Qsal	-0,00193079
Filt	0,00213699
Profundidad	8,59461

Coeficientes PLSR



Observaciones y Predicciones



**ANEXO 38. PLSR EN PERIODO DE CALIBRACIÓN – MODELO  
REDUCIDO EN TÉRMINOS LINEALES**



## Partial Least Squares

Number of dependent variables: 1

Ta

Number of independent variables: 5

Profundidad, Qsal, ETo, Tn, T

Number of complete cases: 4401

Number of components extracted: 2

### Analysis of Variance for Ta

Source	Sum of Squares	Df	Mean Square	F-ratio	P-Value
Model	97823,2	2	48911,6	6169,58	0,0
Residual	34866,8	4398	7,92787		
Total (corr.)	132690,	4400			

### Model for Ta

Component	% Variation in Y	R-Squared
1	64,1674	64,1674
2	9,55568	73,7231

### Independent and Dependent Variables

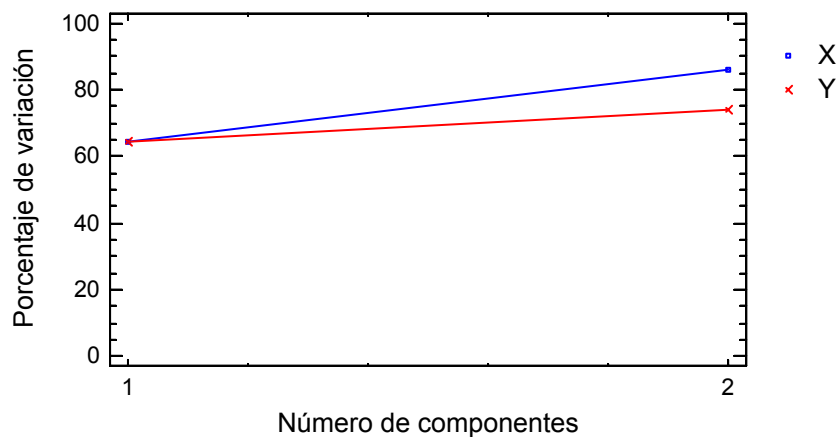
Component	% Variation in X	Cumulative % of X	% Variation in Y	Cumulative % of Y
1	64,4255	64,4255	64,1674	64,1674
2	21,7332	86,1587	9,55568	73,7231

### The StatAdvisor

The output shows the result of fitting models for 1 dependent variables based on 5 predictor variables. To fit the model, 2 components have been extracted. The P-Values in the ANOVA table indicate that 1 of the models are significant at the 5,0% significance level.

To test the predictive ability of the model, select Analysis Options and choose a cross-validation method.

Gráfico comparativo del modelo



### Regression Coefficients

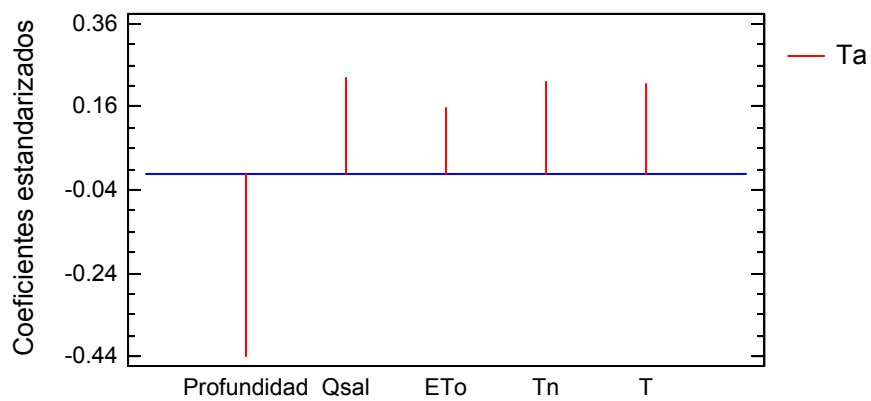
Standardized Coefficients

	<i>Ta</i>
Constant	0.0
Profundidad	-0,439158
Qsal	0,22994
ETo	0,156148
Tn	0,220213
T	0,214914

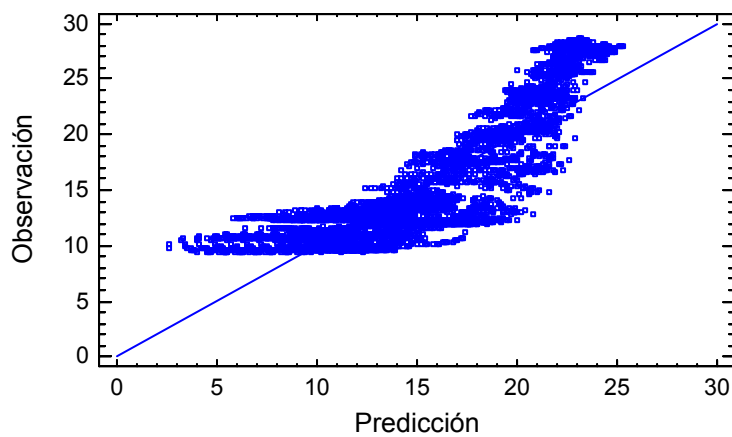
Unstandardized Coefficients

	<i>Ta</i>
Constant	8,3202
Profundidad	-0,310551
Qsal	0,000112235
ETo	0,56017
Tn	0,191481
T	0,180964

Coeficientes PLSR



Observaciones y Predicciones



## Partial Least Squares

Number of dependent variables: 1

Odis

Number of independent variables: 4

ETo, Profundidad, Tx, VolExp

Number of complete cases: 4401

Number of components extracted: 3

### Analysis of Variance for Odis

Source	Sum of Squares	Df	Mean Square	F-ratio	P-Value
Model	58192,5	3	19397,5	3570,81	0,0
Residual	23885,6	4397	5,43224		
Total (corr.)	82078,0	4400			

### Model for Odis

Component	% Variation in Y	R-Squared	Mean Square PRESS	Prediction R-Squared
1	59,7785	59,7785		
2	10,9339	70,7124		
3	0,186603	70,899		

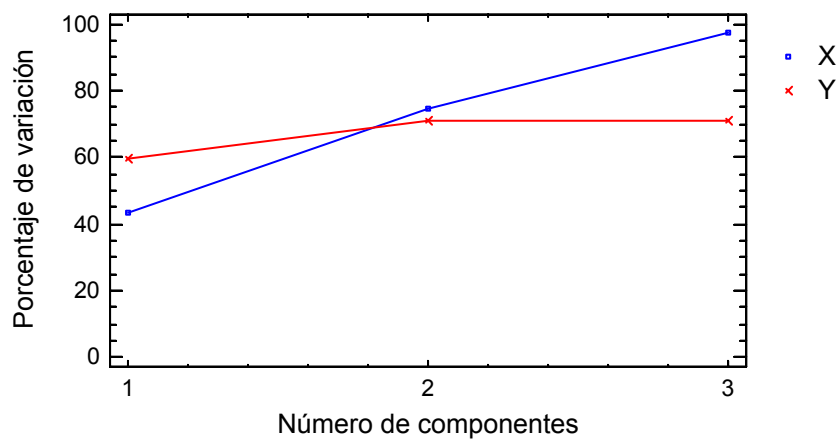
### Independent and Dependent Variables

Component	% Variation in X	Cumulative % of X	% Variation in Y	Cumulative % of Y
1	43,5215	43,5215	59,7785	59,7785
2	31,0199	74,5414	10,9339	70,7124
3	22,7136	97,2551	0,186603	70,899

### The StatAdvisor

The output shows the result of fitting models for 1 dependent variables based on 4 predictor variables. To fit the model, 3 components have been extracted. The P-Values in the ANOVA table indicate that 1 of the models are significant at the 5,0% significance level.

Gráfico comparativo del modelo



### Regression Coefficients

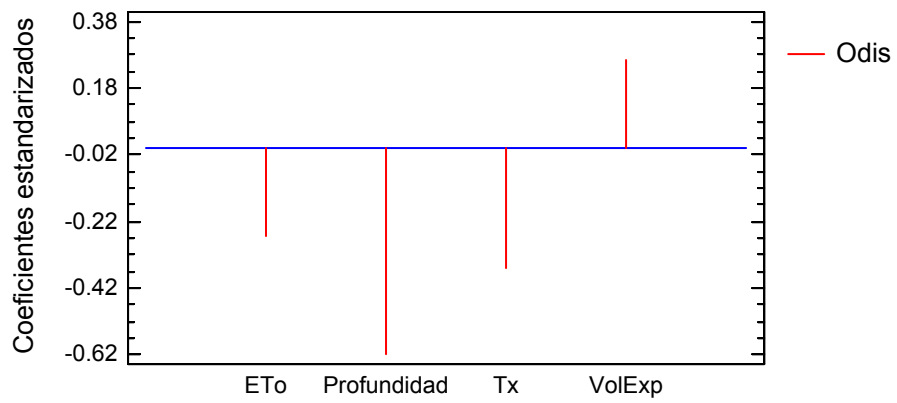
#### Standardized Coefficients

	<i>Odis</i>
Constant	0,0
ETo	-0,266579
Profundidad	-0,615005
Tx	-0,360309
VolExp	0,264212

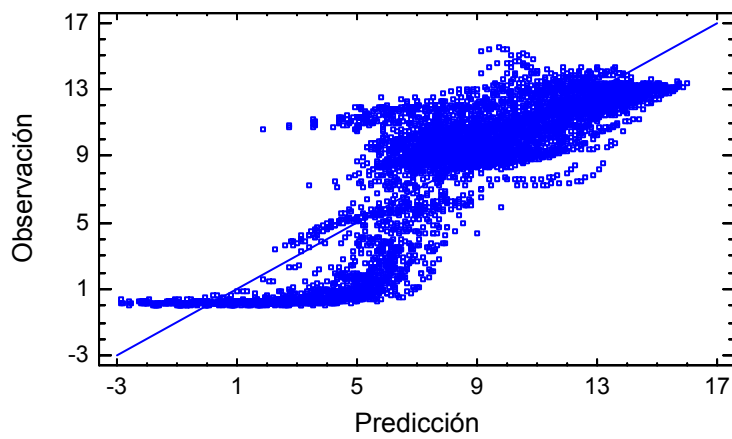
#### Unstandardized Coefficients

	<i>Odis</i>
Constant	11,1268
ETo	-0,752148
Profundidad	-0,342047
Tx	-0,238237
VolExp	8,68778E-7

Coeficientes PLSR



Observaciones y Predicciones



## Partial Least Squares

Number of dependent variables: 1

pH

Number of independent variables: 6

ETo, Tx, Tn, VolExp, Hf, Profundidad

Number of complete cases: 4401

Number of components extracted: 6

### Analysis of Variance for pH

Source	Sum of Squares	Df	Mean Square	F-ratio	P-Value
Model	189,666	6	31,6111	264,887	0,0
Residual	524,372	4394	0,119338		
Total (corr.)	714,038	4400			

### Model for pH

Component	% Variation in Y	R-Squared
1	14,9773	14,9773
2	5,32004	20,2974
3	3,68042	23,9778
4	0,795309	24,7731
5	1,36584	26,139
6	0,423547	26,5625

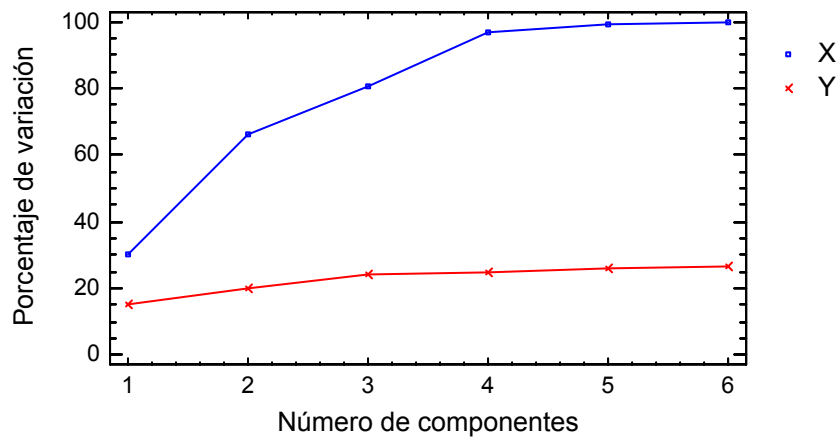
### Independent and Dependent Variables

Component	% Variation in X	Cumulative % of X	% Variation in Y	Cumulative % of Y
1	30,2273	30,2273	14,9773	14,9773
2	36,2233	66,4507	5,32004	20,2974
3	14,0116	80,4623	3,68042	23,9778
4	16,4176	96,8799	0,795309	24,7731
5	2,27296	99,1529	1,36584	26,139
6	0,847149	100,0	0,423547	26,5625

### The StatAdvisor

The output shows the result of fitting models for 1 dependent variables based on 6 predictor variables. To fit the model, 6 components have been extracted. The P-Values in the ANOVA table indicate that 1 of the models are significant at the 5,0% significance level.

Gráfico comparativo del modelo



### Regression Coefficients

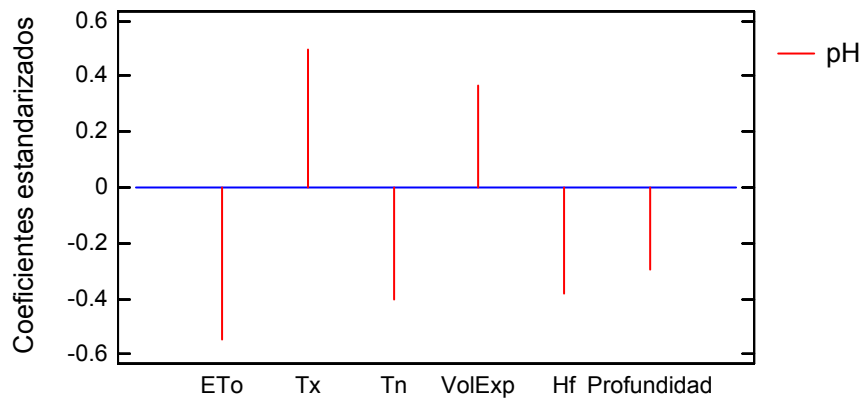
Standardized Coefficients

	<i>pH</i>
Constant	0,0
ETo	-0,549642
Tx	0,495561
Tn	-0,406796
VolExp	0,36436
Hf	-0,384376
Profundidad	-0,297632

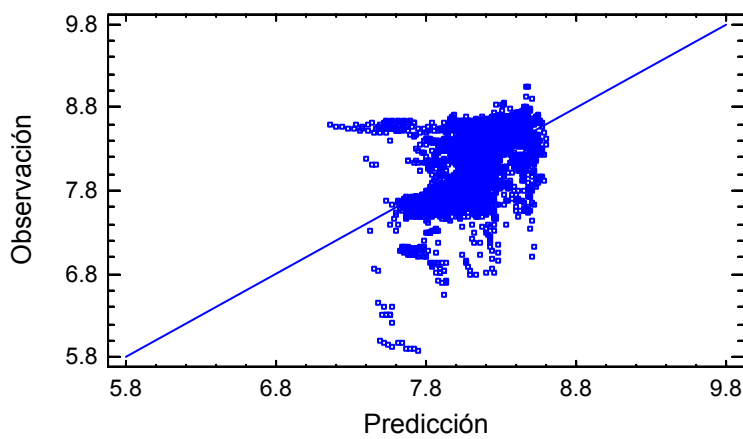
Unstandardized Coefficients

	<i>pH</i>
Constant	7,43143
ETo	-0,144645
Tx	0,0305618
Tn	-0,0259479
VolExp	1,11747E-7
Hf	-0,0408722
Profundidad	-0,0154395

Coeficientes PLSR



Observaciones y Predicciones



## Partial Least Squares

Number of dependent variables: 1

Con

Number of independent variables: 3

Profundidad, VolExp, ETo

Number of complete cases: 4401

Number of components extracted: 3

### Analysis of Variance for Con

Source	Sum of Squares	Df	Mean Square	F-ratio	P-Value
Model	2,71112E7	3	9,03706E6	2306,76	0,0
Residual	1,72259E7	4397	3917,64		
Total (corr.)	4,4337E7	4400			

### Model for Con

Component	% Variation in Y	R-Squared
1	58,3916	58,3916
2	2,29018	60,6818
3	0,466104	61,1479

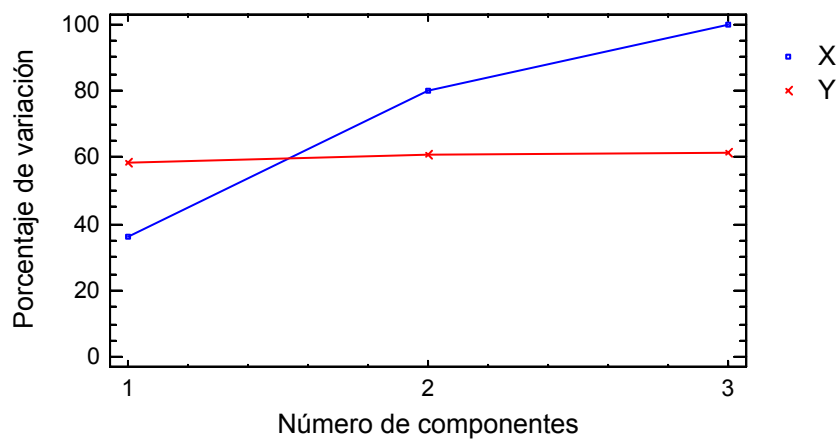
### Independent and Dependent Variables

Component	% Variation in X	Cumulative % of X	% Variation in Y	Cumulative % of Y
1	36,3792	36,3792	58,3916	58,3916
2	43,6009	79,9801	2,29018	60,6818
3	20,0199	100,0	0,466104	61,1479

### The StatAdvisor

The output shows the result of fitting models for 1 dependent variables based on 3 predictor variables. To fit the model, 3 components have been extracted. The P-Values in the ANOVA table indicate that 1 of the models are significant at the 5,0% significance level.

Gráfico comparativo del modelo



### Regression Coefficients

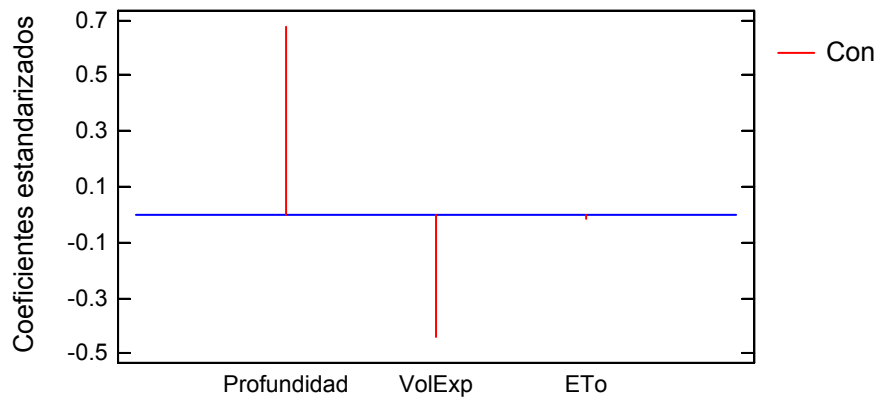
#### Standardized Coefficients

	<i>Con</i>
Constant	0,0
Profundidad	0,67449
VolExp	-0,439858
ETo	-0,0130319

#### Unstandardized Coefficients

	<i>Con</i>
Constant	1220,88
Profundidad	8,71871
VolExp	-0,0000336155
ETo	-0,854586

Coeficientes PLSR



Observaciones y Predicciones

