



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



ESCUELA TÉCNICA
SUPERIOR INGENIEROS
INDUSTRIALES VALENCIA

TRABAJO FIN DE GRADO EN INGENIERÍA QUÍMICA



MODELIZACIÓN ENERGÉTICA DE EDIFICIOS POR MEDIO DE ESTADÍSTICA MULTIVARIANTE

AUTOR: JOAN MARTORELL GONZALEZ

TUTOR: JOSÉ MANUEL PRATS MONTALBÁN



Curso Académico: 2017-18



AGRADECIMIENTOS

Este es el punto final de todo el trabajo hecho en estos años y aunque finaliza una etapa empieza otra aún más difícil, por eso estoy tan agradecido a todas las personas que me han apoyado que lo siguen haciendo.

En primer lugar, quiero empezar por mi profesor José Manuel Prats Montalbán, quién desde el primer momento que le hablé de que quería hacer este proyecto me apoyó y se interesó para que yo pudiera realizar con autonomía este trabajo y sacará adelante todo mi potencial. Le agradezco mucho su comprensión y admiro su dedicación a sus alumnos.

En segundo lugar, no puedo olvidarme de toda mi familia y mi pareja por seguir recordándome que todo trabajo tiene su recompensa y apoyándome en todo este camino. No habría podido hacer este trabajo sin su apoyo.

Muchas gracias a todos.

RESUMEN

En las últimas décadas, el aumento de la capacidad para tomar medidas de los sistemas que se van a estudiar, ha provocado un increíble aumento de la información que se obtiene de los objetos de estudio. En consecuencia, aparecen matrices multidimensionales con decenas de variables y con millones de entradas, lo cual provoca que sea casi imposible obtener conclusiones de esa gigantesca nube de datos utilizando métodos descriptivos clásicos.

El método estadístico aplicado (Análisis de Componentes Principales, PCA) permite una condensación de los datos reduciendo el número de variables las cuales describen casi la misma información y por tanto, simplifican el modelo.

En este trabajo se expone cómo utilizar el PCA y los pasos previos, así como la manera de interpretar las nuevas variables obtenidas del modelo.

Este método nos permite el estudio de matrices de datos muy grandes donde presuponemos que las variables estudiadas tienen una correlación entre ellas y por tanto, se sabe que se están estudiando variables redundantes. De esta manera, se puede saber qué implicación tiene cada variable en el modelo.

Se han estudiado los datos de 4 estaciones meteorológicas a lo largo de un año. El fin de este estudio es verificar la validez de los datos obtenidos, y mediante el PCA, saber qué variables son las que determinan las características de las estaciones para futuros estudios. Gracias al PCA se puede observar cómo se descartan variables que parecían importantes, y aparecen rasgos del entorno que afectan a la variabilidad de las estaciones.

Palabras Clave: PCA, Eficiencia, Procesamiento, Datos, Condensación, Simplificación.

RESUM

En les últimes dècades, amb l'aument de la capacitat per a prendre mesures dels sistemes que es van a estudiar, ha provocat un increïble augment de la informació que s'obté dels objectes d'estudi. En conseqüència, apareixen matrius multidimensionals amb desenes de variables i amb milions d'entrades, el qual provoca que siga quasi impossible obtindre conclusions d'eixe gegantí núvol de dades utilitzant mètodes descriptius clàssics.

El mètode estadístic estudiat (PCA) permet una condensació de les dades reduint el nombre de variables les quals descriuen quasi la mateixa informació i per tant, simplifiquen el model.

En este treball s'exposa com utilitzar el PCA i els passos previs així com la manera d'interpretar les noves variables obtingudes del model.

Este mètode ens permet l'estudi de matrius de dades molt grans on pressuposem que les variables estudiades tenen una correlació entre elles i per tant, se sap que s'estan estudiant variables redundants. D'esta manera, es pot saber que implicació té cada variable en el model.

S'han estudiat les dades de 4 estacions meteorològiques al llarg d'un any. El fi d'este estudi és verificar la validesa de les dades obtingudes, i per mitjà del PCA, saber que variables són les que determinen les característiques de les estacions per a futurs estudis. Gràcies al PCA es pot observar com es descarten variables que pareixien importants, i apareixen característiques de l'entorn que afecten la variabilitat de les estacions.

ParaulesClau: PCA, Eficiència, Processamiento, Dades, Condensació, Simplificació.

ABSTRACT

In the last decades, with the rise in the ability to take measurements from the systems that will be studied, it cause an incredible increase in the information that we obtain from the objects that we study. In consequence, appears a multi-way matrix with around ten or more variables with millions of inputs, which cause that is almost impossible to get any conclusión about that huge cloud of data using the classic descriptive methods.

The stadistical method studied (PCA) allows for condensation the data reducing the number of variables which will describe almost the same amount of information, so it simplifies the model.

In this work it is explained how to use the PCA and the prior steps as well as the way to understand the new variables obtained from the model.

This method allows the study of really big data matrix which we presuppose that the variables studied in it have a correlation among them, therefore is known that there are redundant variables. In this way, it can be known the involvement of each variable in the model.

Has been studied the data from 4 different wether stations over a year. The purpose of this study is verify the validity of the data obtained, and using PCA, know the variables that determine the characteristics of the station for the future studies. Thanks to the PCA it can be seen how variables which seems importants to the model are discarded, and appear features from the environment that affect the variability of the stations.

Keywords:PCA, Efficiency, Processing, Data, Condensation, Simplification.

RESUMEN	2
CAPÍTULO 1. INTRODUCCIÓN, CONTENIDOS Y OBJETIVOS DEL TFG.....	9
1.1. INTRODUCCIÓN	9
1.2. CONTENIDOS.....	11
1.3 OBJETIVOS.....	11
2.1. DATOS	12
2.2 Introducción PCA.....	12
3.4.1. Algoritmo PCA.	14
3.1. PRETRATAMIENTO DE LOS DATOS	19
3.2. CENTRADO	19
3.3. ESCALADO	20
3.3.1 ESCALADO POR COLUMNAS.....	21
3.3.2 ESCALADO POR BLOQUES (LATERALES O FRONTALES).....	22
CAPITULO 4. PROCESAMIENTO DE DATOS Y ANALISIS CON PCA.....	23
4.1 PREPROCESAMIENTO	23
4.1.1 Datos faltantes.	23
4.1.2 Rango	23
4.1.3 Reestructuración de los datos.	24
4.1.4 Centrado.....	24
4.1.4 Escalado.....	24
4.2 DATOS ANÓMALOS.	25
CAPITULO 5. RESULTADOS	26
5.1 AEROPUERTO	26
5.2 ALTOCÚMULO	34
5.3 UPV.....	40
5.4 VIVEROS	47
5.5 COMPARATIVA DE ESTACIONES	52
5.5.1 Elección del número de componentes principales.	53
6.- CONCLUSIONES.....	57
7.- ANEXOS.....	58
BIBLIOGRAFÍA.....	60
8.- PRESUPUESTO	61

UPV

MEMORIA

Joan Martorell González

CAPÍTULO 1. INTRODUCCIÓN, CONTENIDOS Y

OBJETIVOS DEL TFG

1.1. INTRODUCCIÓN

En los últimos años ha habido una creciente demanda de hogares con una mejor eficiencia energética. Los ciudadanos quieren pagar menos por la electricidad o gas que consumen por mantener su hogar caliente en invierno y frío en verano. Una forma indirecta de conseguir este fin es tener una casa con mejor aislamiento, ya sea mediante ventanas más aislantes del calor o aislantes en las paredes, etc.

A causa de este nuevo movimiento, las constructoras deben plantearse cuantos vatios consumirá la vivienda que piensan construir y así ofrecer mejores casas. Los programas actuales de cálculo de vatios hora ofrecen una cifra si le especificamos los materiales de construcción. Calener y Líder son los programas más utilizados y sirven para saber si los edificios cumplen la normativa de Calificación Energética de Edificios RD 47/2007. El método general de esta ley obtiene la calificación energética (A, B hasta G) mediante el programa Calener.

El programa realiza una simulación del funcionamiento energético del edificio en cuanto al consumo de energía (medido en kWh) y las emisiones de CO₂ (medidas en kg) debidos a las instalaciones de refrigeración, calefacción, producción de agua caliente sanitaria y alumbrado. Para su funcionamiento, es necesario facilitar la descripción geométrica del mismo (superficie, número de plantas, espacios habitados, distribución interior, etcétera) aunque lo más sencillo es importarlo desde el programa Líder. Una vez que se ha descrito físicamente el edificio, se introducen los datos sobre las instalaciones citadas (iluminación, producción de ACS, calefacción y refrigeración) de forma que Calener modeliza dos edificios, objeto y otro de referencia.

El edificio de referencia va a ser idéntico al que planteaba el programa Líder, es decir, que presenta estrictamente los valores límites en cuanto a cerramientos que planteaba el CTE-HE 1-Limitación de la demanda, pero además también cumple estrictamente los valores de las exigencias HE 2-Instalaciones térmicas, HE 3-Eficiencia energética en iluminación, HE 4-Contribución de energía solar térmica a la producción de ACS y HE 5-Producción de energía fotovoltaica si da lugar. Además, debemos facilitar a Calener la ubicación geográfica del edificio. Esto supone un serio problema debido a que no tiene una muestra climática de toda España, sino sólo de las capitales de provincia.

Esto provoca que los cálculos sean mucho más imprecisos. Estos programas utilizan 12 zonas peninsulares que corresponden a la combinación de una severidad climática de invierno (A, B, C, D o E) junto con una severidad climática de verano (1, 2, 3 o 4) dando lugar a los siguientes tipos de zonas climáticas: A3, A4, B3, B4, C1, C2, C3, C4, D1, D2, D3 y E4.

Figura 1. Tabla de zonas climáticas.

		Zona Climática de Verano			
		1	2	3	4
Zona Climática de Invierno	A			Cádiz	Almería
	B			Valencia	Sevilla
	C	Bilbao	Barcelona	Granada	Toledo
	D	Vitoria	Zamora	Madrid	
	E	Burgos			

Usar esta tabla como datos base para calcular los vatios de una casa es muy impreciso ya que dentro de cada zona existe mucha variabilidad. Una manera de evitar esta mala aproximación es utilizar un sistema que detecte la temperatura exterior de la casa utilizando coordenadas para así mejorar la precisión del programa.

Para los nuevos cálculos se precisan las condiciones meteorológicas de la zona de construcción y la forma de obtener esos datos es mediante estaciones meteorológicas cercanas de acceso gratuito. El programa necesitará diversas fuentes para así contrastar toda esa información y evitar errores de cálculo que supongan un sobrecoste en la construcción de las viviendas.

En esta parte en concreto es donde se centra este trabajo: en la recolección de los datos, procesamiento, análisis y conclusiones de los mismos.

Para poder modelizar como las variables meteorológicas influyen en la temperatura de una vivienda y por tanto es su gasto en vatios se necesita saber qué variables influyen en el modelo y la mejor manera es usar métodos estadísticos que permiten construir modelos a partir de datos locales y por tanto son más precisos.

Se tiene que determinar qué variables se usarán en el modelo estadístico y para ello el primer paso es demostrar que variables son las más representativas. Hay que comprobar los patrones de variabilidad de las estaciones para establecer las similitudes y diferencias entre ellas y así poder elegir las variables más adecuadas en cada caso.

La forma adecuada de realizar este estudio es mediante una herramienta estadística de proyección sobre estructuras latentes, Análisis de Componentes Principales (PCA). El PCA permite la reducción del número de variables a estudiar perdiendo la menor cantidad de información posible. Estas nuevas variables serán combinaciones lineales de las variables iniciales que además son ortogonales entre sí, es decir, independientes. Este método, por tanto, convierte un conjunto de variables posiblemente relacionadas entre sí en un conjunto de variables no relacionadas entre sí, las componentes principales.

1.2. CONTENIDOS

Se analizarán datos obtenidos de diferentes estaciones de la ciudad de Valencia localizadas en: Aeropuerto, Universidad politécnica de Valencia, Viveros y Altocúmulo. Para ello se realizarán diferentes tipos de preprocesamiento y creación de matrices de datos para obtener las variables predominantes en la nube de datos.

1.3 OBJETIVOS

El objetivo de este proyecto consiste en utilizar el modelo estadístico de proyección sobre estructuras latentes Análisis de Componentes Principales (PCA) para

- Simplificar la enorme cantidad de datos obtenidos en las estaciones meteorológicas,
- Comprender las estructuras de correlación interna entre variables en cada estación
- Establecer las potenciales similitudes y diferencias entre las estaciones.

CAPÍTULO 2. ESTRUCTURA DE LOS DATOS

2.1. DATOS

Los datos se obtienen de 4 estaciones meteorológicas situadas en las inmediaciones de Valencia. Las estaciones son:

- Aeropuerto.
- Universidad politécnica de valencia (UPV).
- Viveros.
- Altocúmulo.

En los datos originales aparecen decenas de variables diferentes de las cuales solo se usarán 5 para el estudio. La temperatura, la precipitación, la humedad, la velocidad del viento y la dirección del viento.

Las medidas son tomadas cada 10 minutos en las 24 horas del día durante todo el año 2015. Esto suma un total de 52560 medidas por cada variable que suman 262800 datos por estación y con un total de 1051200 valores para el estudio.

Como es comprensible no se puede estudiar esta matriz de datos con los métodos triviales que se usan habitualmente. Dado que se tienen 5 variables y 5 estaciones se necesita resumir la información de la matriz de datos perdiendo el mínimo de información en el proceso.

Una manera de condensar la información cuando sabemos que las variables están correlacionadas y por tanto existen variables “redundantes” es el PCA (*Principal Component Analysis*) que es una técnica estadística de síntesis de la información que reduce las dimensiones (nº de variables) de los datos. Es decir, con un banco de datos con muchas variables, el objetivo será reducirlas a un menor número perdiendo la menor cantidad de información posible.

2.2 Introducción PCA

El Análisis de Componentes Principales, PCA (*Principal Component Analysis*) (Jackson, 1991), es un método de proyección de unas variables sobre otras, llamadas variables latentes, ortogonales entre sí, y que sirven para describir de manera óptima un proceso, y los sucesos que tienen lugar en el mismo.

Cuando se dispone de una estructura interna de datos N-way, (Prats-Montalbán, 2005), para llevar a cabo un análisis PCA es necesario desplegar la estructura hasta conseguir una matriz (estructura 2-way), por lo que a este método se le conoce como Unfold-PCA o U-PCA.

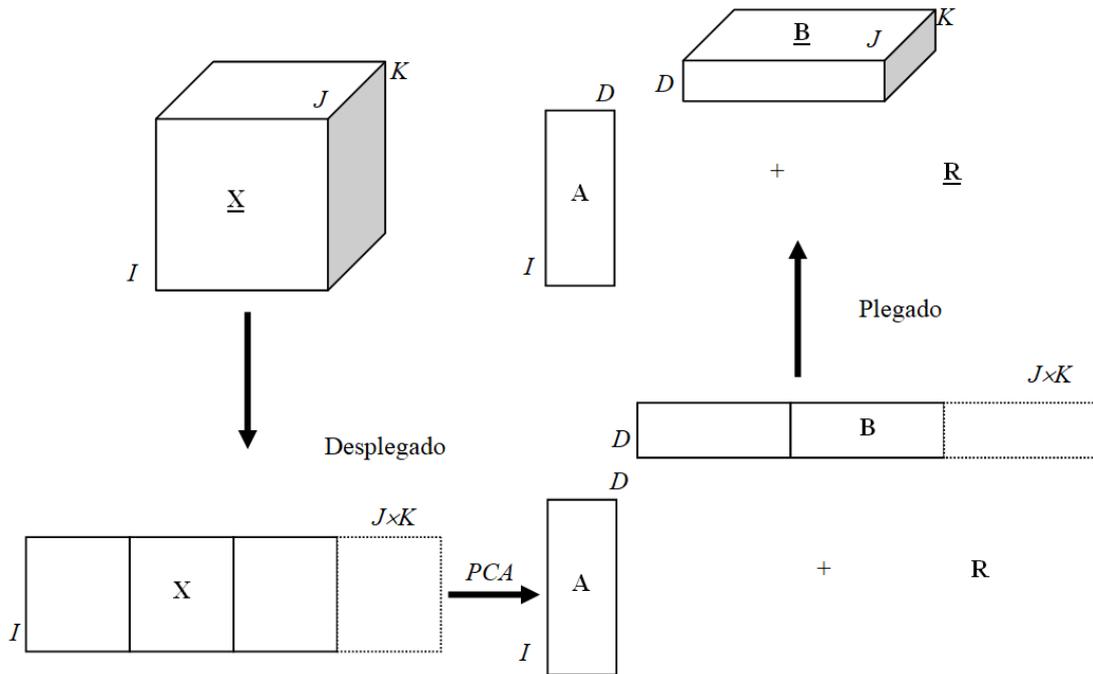


Figura 2.1 Estructura general de un modelo Unfold-PCA. (Prats-Montalbán, 2005)

El desplegado es simplemente una manera de reagrupar una estructura de datos N-way en una matriz, siendo la metodología utilizada por los métodos 2-way para poder estudiar estructuras N-way.

El modelo se construye a partir de la minimización de los cuadrados de los errores de la predicción de la estructura original de datos.

Si se tiene, como ya se ha comentado, una estructura $\underline{X}(I \times J \times K)$ y se despliega de manera que se obtiene una matriz de dimensiones $(I \times JK)$, obligando al modelo a recorrer el total de dimensiones J y K, el modelo tiene la ecuación:

$$x_{ijk} = \sum_{d=1}^D t_{id} p_{jkd} + e_{ijk} \quad \mathbf{1}$$

Lo cual equivale matricialmente a la expresión:

$$\underline{X}^{I \times JK} = \underline{T}^{I \times D} \underline{P}^{(D \times JK)} + \underline{E}^{I \times JK} \quad \mathbf{2}$$

Que constituye un modelo PCA para una matriz \underline{X} de dimensiones $(I \times JK)$.

Tradicionalmente, la matriz \underline{T} es conocida como matriz de *scores*, mientras que la matriz \underline{P} se denomina matriz de *loadings*, cuyas columnas son ortogonales. Los *loadings* pueden ser entendidos como la contribución de cada variable original cuando se calcula el componente

principal y los *scores* contienen la matriz original de datos pero con el sistema de referencia rotado.

3.4.1. Algoritmo PCA.

Los resultados del *Unfold-PCA* y su algoritmo se han llevado a cabo en base a los correspondientes al modelo PCA, asumiendo, para una mayor facilidad de interpretación, que la estructura de datos es una matriz \mathbf{X} de dimensiones $I \times J$. Asimismo, se utiliza la nomenclatura clásica correspondiente a los modelos PCA, denominando, por tanto, \mathbf{T} a la matriz de *scores* y \mathbf{P} a la matriz de *loadings*.

PCA es la descomposición de una matriz \mathbf{X} ($I \times J$) en matrices más sencillas M_a a partir de la expresión:

$$\mathbf{X} = \sum_{a=1}^A M_a \quad 3$$

Las matrices M_a son de tamaño $I \times 1$ y los vectores p_a de tamaño $J \times 1$.

De esta manera, la función de pérdida queda establecida según la ecuación:

$$\min_A \|\mathbf{X} - \mathbf{TP}^T\|^2 \quad 4$$

Lo que se desea es predecir de manera óptima las K variables primitivas a partir de un número reducido de nuevas variables. A los vectores t_a se les llama *scores*, y son ortogonales; mientras que a los vectores p_a se les llama *loadings*, los cuales tienen la propiedad de ser ortonormales. Es decir:

$$\mathbf{X} = \sum_{a=1}^A M_a = \sum_{a=1}^A t_a p_a^T \quad 5$$

$$p_1^T p_j = \delta_{ij} \quad 6$$

$$t_i^T t_j = \delta_{ij} \lambda_i \quad 7$$

Siendo λ_i un valor propio de $\mathbf{X}^T \mathbf{X}$. Esto define la relación entre PCA y la ecuación vector propio – valor propio.

$$\lambda_a p_a^T = t_a^T \mathbf{X} \quad 8$$

$$t_a = \mathbf{X} p_a \quad 9$$

Existen tres criterios equivalentes que llevan a la explicación matemática de estos resultados:

- Subespacio que mejor se ajusta a la nube de puntos que constituyen los datos.
- Subespacio que minimiza la deformación de la nube proyectada.

- Predicción óptima de las K variables a partir de las A nuevas variables latentes (combinaciones lineales de las variables originales).

Siguiendo el tercer criterio, se trata de hallar una primera variable latente $\mathbf{t} = \mathbf{p}^T \mathbf{x}$ (y otras sucesivas) de forma que se minimice la varianza residual del ajuste x_k a la nueva variable latente (componente principal) \mathbf{t} , σ_k^{2*} , expresada como:

$$\sigma_k^{2*} = \sigma_k^2 (1 - \rho_{k,t}^2) = \sigma_k^2 \left(1 - \frac{(\sigma_{k,t}^2)^2}{\sigma_t^2 \sigma_k^2} \right) = \left(\sigma_k^2 - \frac{(\sigma_{k,t}^2)^2}{\sigma_t^2} \right) \quad 10$$

Asumiendo, sin pérdida de generalidad, que $\sigma_t^2 = 1$, sin más que normalizar el vector \mathbf{p} , se tiene que:

$$\sigma_t^2 = \mathbf{p}^T \mathbf{V} \mathbf{p} = 1 \quad 11$$

Con lo que:

$$\min \sum_k \sigma_k^{2*} = \min \sum_k \left[\sigma_k^2 - \frac{(\sigma_{k,t}^2)^2}{\sigma_t^2} \right] = \min \sum_k [\sigma_k^2 - \mathbf{p}^T \mathbf{v}_k \mathbf{v}_k^T \mathbf{p}] \quad 12$$

Ya que

$$\sigma_{k,t}^2 = \text{cov}(x_k, t) = \mathbf{p}^T \mathbf{v}_k \quad 13$$

Donde \mathbf{v}_k es la columna k-ésima de matriz de varianzas-covarianzas

$$\mathbf{V} = \frac{\mathbf{X}^T \mathbf{X}}{l-1} \quad 14$$

Como $\sum_k \sigma_k^2$ es la traza de la matriz de varianzas-covarianzas, que es constante, esta minimización equivale a la maximización de la expresión:

$$\max \sum_k [\mathbf{p}^T \mathbf{v}_k \mathbf{v}_k^T \mathbf{p}] = \max \left[\mathbf{p}^T \left(\sum_k \mathbf{v}_k \mathbf{v}_k^T \right) \mathbf{p} \right] = \max [\mathbf{p}^T \mathbf{V} \mathbf{p}] \quad 15$$

La única restricción realizada ha sido que

$$\mathbf{p}^T \mathbf{V} \mathbf{p} = 1 \quad 16$$

Para la primera componente,

$$\psi = \mathbf{p}_1^T \mathbf{V}^2 \mathbf{p}_1 - \lambda_1 (\mathbf{p}_1^T \mathbf{V} \mathbf{p}_1 - 1) \quad 17$$

$$\frac{\partial \psi}{\partial \mathbf{p}_1} = 2\mathbf{V}^2 \mathbf{p}_1 - 2\lambda_1 \mathbf{V} \mathbf{p}_1 = 0 \Rightarrow \mathbf{V}^2 \mathbf{p}_1 = \lambda_1 \mathbf{V} \mathbf{p}_1 \quad 18$$

De esta manera,

$$\max(\mathbf{p}_1^T \mathbf{V}^2 \mathbf{p}_1) = \max(\mathbf{p}_1^T \lambda_1 \mathbf{V} \mathbf{p}_1) = \max(\lambda_1 \mathbf{p}_1^T \mathbf{V} \mathbf{p}_1) = \lambda_1 \quad 19$$

Donde \mathbf{p}_1 es el vector propio de \mathbf{V} , ya que

$$\mathbf{V} \mathbf{p}_1 = \lambda_1 \mathbf{p}_1 \quad 20$$

Es decir, la primera componente o dirección principal es la asociada al vector propio asociado al mayor valor propio, el cual no representa otra cosa que la varianza de la primera componente.

Las siguientes componentes van asociadas a vectores propios ortogonales a todos los previamente obtenidos, y que a su vez van ligados a los siguientes mayores valores propios, que serán las varianzas explicadas por cada una de las componentes.

Aunque el número máximo de componentes es igual al rango la matriz de varianzas covarianzas, varias de estas componentes pueden, sin embargo, tener poco interés en la práctica si el valor propio asociado es pequeño.

Los elementos p_{ak} de cada vector \mathbf{p}_a se denominan pesos (*loadings*) de la variable k-ésima en el factor a-ésimo, y permiten obtener los valores t_{ia} de la componente a en cada observación.

Existen diversos algoritmos que se utilizan para el cálculo de las componentes principales.

Por un lado, existe la descomposición en valores singulares, SVD (*singular valuedecomposition*), pero ésta presenta problemas cuando las matrices tienen un elevado grado de colinealidad o cuando hay datos faltantes.

En su lugar, existen algoritmos secuenciales, como el NIPALS (Wold, Estimation of principal components and related models by iterative least squares 1966 a), (Wold, 1966 b) que, aunque son menos precisos, funcionan bien con datos faltantes y con estructuras de datos que presentan una elevada correlación entre sus variables.

En concreto, NIPALS (*NonlinearIterativePartialLeastSquares*), funcionan bien si los datos faltantes ocurren de manera aleatoria en la matriz de datos \mathbf{X} .

$$\mathbf{X} = \mathbf{t}_1 \mathbf{p}_1^T + \dots + \mathbf{t}_A \mathbf{p}_A^T + \mathbf{R} \quad 21$$

Este algoritmo se lleva a cabo sobre una dimensión cada vez (tA, pA), pudiendo estudiar su significación mediante diversos procedimientos. El estudio de la significación de cada una de estas dimensiones es lo que determina en número de variables latentes a retener en el modelo.

Tabla 1. Algoritmo del PCA.(Prats-Montalbán 2005)

Sea la matriz \mathbf{X} de dimensiones $I \times J$

1. Iniciar con $\mathbf{t}=\mathbf{x}_k$ (columna de \mathbf{X} con mayor varianza)

2. Regresión lineal múltiple de las columnas de \mathbf{X} sobre \mathbf{t}

$$\mathbf{X} = \mathbf{t}\mathbf{b}^T \Rightarrow \mathbf{b} = (\mathbf{t}^T \mathbf{t})^{-1} \mathbf{t}^T \mathbf{X} = \mathbf{t}^T \mathbf{X} / (\mathbf{t}^T \mathbf{t}) = \mathbf{p}^T$$

Se normaliza \mathbf{p} de manera que $\|\mathbf{p}\|=1$

$\mathbf{p}_k = \mathbf{t}^T \mathbf{x}_k / \mathbf{t}^T \mathbf{t}$ es una combinación lineal de I individuos con pesos \mathbf{t} 's

3. Regresión lineal múltiple de las filas de \mathbf{X} sobre \mathbf{p}

$$\mathbf{X}^T = \mathbf{p}\mathbf{b}^T \Rightarrow \mathbf{b} = (\mathbf{p}^T \mathbf{p})^{-1} \mathbf{p}^T \mathbf{X}^T = \mathbf{p}^T \mathbf{X}^T / (\mathbf{p}^T \mathbf{p}) = \mathbf{t}^T$$

CAPÍTULO 3. INTRODUCCIÓN AL PROCESAMIENTO DE DATOS

3.1. PRETRATAMIENTO DE LOS DATOS

La forma en que se prepara la estructura de datos antes de llevar a cabo la modelización de los mismos, independientemente del modelo que vayamos a utilizar, constituye un punto primordial del análisis, ya que va a condicionar totalmente los resultados que se obtengan.

Gurden, Westerhuis, Bro, & Smilde (2001) llegan a concluir que el tipo de datos y pretratamiento utilizado es más importante que el modelo utilizado en términos de poder predictivo, a pesar de que los modelos difieran en términos de interpretabilidad.

A pesar de la importancia dada al pretratamiento de los datos en la teoría, en la práctica las distintas actuaciones a realizar sobre los datos suelen verse simplemente como pasos previos al acto de modelización. Además, a pesar de que el escalado es necesario para muchos tipos de datos, el hecho de escalar de manera sistemática las variables a la misma varianza (la unidad), puede causar problemas cuando a variables con pequeña información se les asigna una importancia elevada.

Cuando tenemos estructuras de datos de mayor orden, las posibilidades de escalar los datos de diferentes maneras aumentan, y encontrar la manera óptima de hacerlo se complica.

Las actuaciones más comunes en cuanto al procesamiento de datos, antes de construir modelo, son el centrado y escalado de los mismos. Se van a describir las diferentes maneras en que se pueden llevar a cabo estas acciones para una estructura 2-way, si bien la extensión a estructuras más complejas es directa.

3.2. CENTRADO

Una forma sencilla de definir el proceso de centrado sería como la ubicación del origen del sistema de coordenadas del espacio de las variables originales en el centro de gravedad de la nube de puntos que constituyen el conjunto de datos.

El centrado a lo largo de un modo elimina los términos constantes del conjunto de datos. Hay diversas razones para realizar el centrado, siendo una de ellas que los modelos de regresión no contienen un término constante, sino que asumen una variación alrededor del cero.

Bro & Smilde, (2003) indican que el hecho de centrar un conjunto de datos altera la estructura de los mismos, por lo que no siempre va a interesar llevar a cabo dicha operación, sobre todo si no se consigue un modelo más simple y/o más fácilmente interpretable.

El proceso de centrado debería derivar en alguno de los puntos siguientes:

- Una reducción del rango del modelo.
- Un aumento del ajuste de los datos.
- Una eliminación específica de valores constantes.
- Evitar problemas numéricos.

El proceso correcto de centrado consiste en hacerlo siempre a lo largo de un solo modo cada vez, estando permitido el centrado secuencial a lo largo de varios modos. Un modo se puede definir como una dimensión de variación natural de los datos, tal como el tiempo, variables de un proceso, presión, velocidad del viento, etc.

Es decir, que la manera correcta y no problemática de llevar a cabo centrados a lo largo de diferentes modos (dimensiones), es hacer varios centrados consecutivos, cada uno de ellos a lo largo del modo que interese (el orden en que se realicen no importa). El hecho de preprocesar los datos eliminando la media general no es de utilidad, si esta acción se lleva a cabo de manera arbitraria.

Bro (1997) explica que el tipo de centrado más común es el que se realiza a lo largo del modo de las observaciones, el que se conoce como centrado a lo largo del primer modo, siempre que las observaciones se coloquen en dicho modo.

$$x_{ijk}^{\text{cent}} = x_{ijk} - \bar{x}_{jk} \quad 22$$

Donde

$$\bar{x}_{jk} = \frac{\sum_{i=1}^I x_{ijk}}{I} \quad 23$$

El centrado a lo largo de un modo, por ejemplo el primero, se puede realizar mediante el despliegado de la estructura de datos en una matriz I×JKL..., y el posterior centrado de esta matriz a lo largo del primer modo.

Para modelos de regresión, el centrado a lo largo del primer modo es un procedimiento ampliamente aceptado y se mantiene prácticamente igual en los métodos 2-way.

3.3. ESCALADO

El escalado a lo largo de un modo consiste en dividir todos los elementos asociados a una determinada variable en ese modo por una constante.

Las razones para llevar a cabo un escalado de los datos son:

- Cuando no se dispone de información previa, cada variable se escala a la misma varianza, de manera que todas ellas tienen la misma oportunidad de contribuir al modelo.

- Ajustar diferentes escalas, cada una de ellas propias de variables de distinta naturaleza y/o medidas.
- Acomodar el proceso de ajuste en caso de heterocedasticidad.
- Permitir que diferentes conjuntos de variables de distinto tamaño que conforman la estructura total de los mismos puedan tener el mismo peso a priori, a la hora de llevar a cabo la modelización del proceso (escalado por bloques).

En el caso de estructuras de datos 3-way, existen diferentes tipos de escalado:

- Escalado por columnas, en el cual cada vector x_{jk} queda normalizado.
- Escalado de un solo bloque (lateral).
- Escalado doble de bloques, donde se escalan tanto los bloques laterales como los frontales.

3.3.1 ESCALADO POR COLUMNAS

En el escalado por columnas, cada una de las JK columnas se divide por la raíz cuadrada de su cuadrado medio. Si se considera que el primer modo corresponde a las observaciones, la expresión correspondiente queda:

$$RMS_{jk} = \sqrt{\frac{\sum_{i=1}^I x_{ijk}^2}{I}} \quad 24$$

Y por lo tanto:

$$x_{ijk}^* = \frac{x_{ijk}}{RMS_{jk}} \quad 25$$

Si \underline{X} está centrada a lo largo de las columnas, la expresión de RMS es aproximadamente igual a la desviación típica muestral (se divide por I-1 en lugar de I). Por ello, este procedimiento es muy similar al autoescalado a varianza unitaria.

3.3.2 ESCALADO POR BLOQUES (LATERALES O FRONTALES)

Para el escalado a un solo bloque o cara, cada uno de los bloques de J o K variables se escalan multiplicando por un coeficiente que dé como resultado un cuadrado medio de valor la unidad dentro del modo elegido. En el caso de, por ejemplo, escalar dentro del segundo modo, dado que la raíz cuadrada del cuadrado medio de la variable primaria j-ésima viene dada por la ecuación:

$$RMS_j = \sqrt{\frac{\sum_{i=1}^I \sum_{k=1}^K x_{ijk}^2}{IK}} \quad 26$$

Los datos escalados a un solo bloque quedarán según la expresión:

$$x_{ijk}^* = \frac{x_{ijk}}{RMS_j} \quad 27$$

CAPITULO 4. PROCESAMIENTO DE DATOS Y

ANALISIS CON PCA

4.1 PREPROCESAMIENTO

Esta es posiblemente la parte más importante de todo el estudio, ya que un error en esta sección supone que todo lo que sigue no tendrá ningún valor.

Los datos obtenidos se tienen que adecuar y modificar para que su relevancia en el estudio sea la correcta. Para ello el primer paso es comprobar que no hay datos faltantes que, por supuesto, es el caso.

4.1.1 Datos faltantes.

Mediante la aplicación Matlab se cargan los datos y usando la función “pcambtsr” se rellenan los huecos con números estimados. Esta función coge la matriz original con huecos y usando una función interna recorre la matriz en busca de celdas vacías. Una vez encuentra un dato faltante estipula mediante interpolación el valor medio aproximado y lo introduce.

4.1.2 Rango

Todas las estaciones deben tener el mismo número de variables y repeticiones. Los datos obtenidos por las estaciones tienen un formato propio y un número de variables muy amplio. Se extrajeron las variables del estudio:

- Tiempo
- Temperatura
- Humedad
- Precipitación
- Dirección del viento
- Velocidad del viento

Las estaciones estudiadas tomaban la medida cada 10 minutos excepto la estación de altocúmulo. En el caso particular de altocúmulo la estación tomaba mediciones cada 15 minutos.

Para obtener el mismo número de mediciones para cada variable y en el mismo rango se realiza una interpolación para obtener todos los valores. Esta interpolación supuso la creación de una función que se llama interpolarsimple_2, ya que calcular a mano la interpolación simple de 6 variables y 52560 mediciones por variable no era viable.

4.1.3 Reestructuración de los datos.

Antes de utilizar el PCA es conveniente reorganizar la matriz de datos para moldear los resultados obtenidos y así poder interpretar mejor los resultados. La matriz se reorganiza en 365 filas y 720 columnas como muestra la tabla 2. Las filas corresponden a los días del año y las columnas se separan en grupos de 144 que corresponden a las variables y dentro las 144 mediciones en un día. Así se estudiará por lotes correspondiente a los días.

Tabla 2. Estructura de datos reorganizada.

Días	Temperatura			Humedad			Precipitación			V. Viento			D. Viento		
1	1	...	144	1	...	144	1	..	144	1	...	144	1	...	144
2															
...															
365															

4.1.4 Centrado

El centrado se realizará por variables. Se crea un vector de medias para cada una de las variables y cada una de las medidas intradía de las variables, dando lugar a un vector con 720 valores (144x5) donde cada valor es la media de los 365 días para cada momento del día que se toma la medida.

4.1.4 Escalado

El motivo por el cual se decide por el escalado por bloques y no por columna es debido a la estructura 3-way. Si se escala por columnas, como es el caso general del escalado en los modelos 2-way, y considerando como primero el modo de las observaciones, cada variable del segundo modo (por ejemplo medida de un proceso) para cada variable del tercer modo (por ejemplo instante de tiempo) se trata como una variable independiente. Esto significa que las relaciones 3-way entre las variables pueden quedar distorsionadas o, visto de otro modo, que la estructura multivariante presente en los datos se puede perder.

Por otro lado, el efecto de escalar todas las regiones a la misma varianza puede ser el de dar más importancia a esas variables que tienen poco contenido de información, incrementando así la cantidad aparente de variación no sistemática en los datos. Este efecto es en ocasiones más pronunciado en el escalado por columnas que en el escalado por bloques, ya que es más probable que una columna, más que un bloque entero, contenga sólo ruido de fondo. Por ello, el escalado por columnas puede verse como una manera incorrecta de escalar si nos encontramos ante una estructura N-way.

En cualquier caso, hay que tener en cuenta que la fase de preprocesado viene fuertemente determinada por el problema que se está estudiando, ya que en función del mismo, podemos obtener que preprocesados que en un principio no deberían hacerse proporcionan mejores resultados que preprocesados más “teóricos”, tales como un autoescalado a varianza unitaria.

En el caso actual, por lo anteriormente explicado se ha decidido utilizar un escalado por bloques.

4.2 DATOS ANÓMALOS.

Antes de proceder con el análisis del PCA es siempre prudente realizar un análisis más simple para comprobar que no hay errores ni datos anómalos que modifiquen la estructura global de la matriz. En caso de encontrar algún dato anómalo o erróneo se procederá a corregirlo eliminándolo y volviendo a empezar el proceso de tratamiento de datos usando la herramienta de rellenado de huecos.

Una vez en este punto la matriz de datos que en este caso son 4, una por estación, están listas para utilizar la herramienta de análisis del componente principal (PCA). En Matlab utilizamos esta función para calcular los loadings(p) y los scores(t), $[p,t] = \text{pca}(x,pc)$.

CAPITULO 5. RESULTADOS

En este apartado, se exponen los resultados obtenidos empleando los métodos explicados en la sección anterior. El apartado se dividirá en 5 secciones, 4 correspondientes a las estaciones y la 5 a una comparativa entre ellas.

Así pues, con la ayuda del programa informático MATLAB® se aplicará el algoritmo expuesto en la tabla 1, se ha podido llevar a cabo un análisis de los datos de las estaciones de acuerdo a lo explicado en estas líneas.

5.1 AEROPUERTO

En la figura 5.1 se observa una nube de datos que corresponde a visualizar los scores de la componente uno en el eje horizontal y la componente 2 en el eje vertical. Hay dos cosas que llaman la atención inmediatamente, la dispersión de los datos es mayor en el eje horizontal que corresponde a la componente principal 1 ya que todos los datos están más o menos agrupados alrededor del valor 0 del eje vertical (PC2). Esto es a causa de que el propio modelo intenta recolectar la mayor cantidad de variabilidad en la componente principal 1 y luego en la segunda y así sucesivamente.

La siguiente cosa en llamar la atención de este gráfico son dos valores muy positivos, aunque dentro de lo que podemos aceptar en la componente 1 y también extremadamente positivos para la componente 2 que como hemos comentado anteriormente son datos muy próximos al eje central 0.

Analizando con un poco más de profundidad la figura 5.1, se observa que en la PC1 los valores para el día 30 y 57 están dentro de lo que consideramos valores máximos aceptables para la dispersión de los datos. Pero si seguimos la tendencia de los datos que es más que evidente, se observa que es muy negativa para los días 1, 2 y según seguimos el gráfico se hacen positivos hasta llegar a los máximos para los días 200. Cuando los valores deberían ser negativos según la tendencia de la ecuación, para los días 30 y 57 son tan positivos como para los días alrededor del 200, por tanto algo está ocurriendo en los datos.

Si observamos los días para la componente principal 2 ya no cabe duda que son datos anómalos. Estos dos valores son más de dos veces mayores que cualquier otro valor y distorsionan toda la gráfica. Estos dos datos no se corresponden con el resto de valores y su tendencia.

Dada esta información se decide realizar el estudio eliminando estos dos valores anómalos y los resultados obtenidos se exponen a continuación.

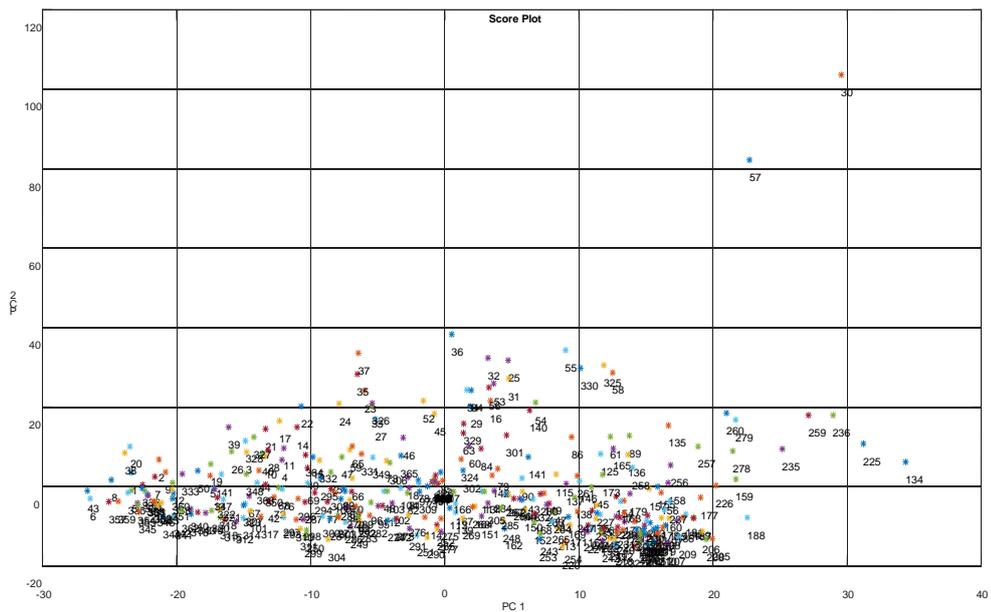


Figura 5.1. Scores de las dos primeras componentes principales.

Tabla 3. Porcentaje de variabilidad explicada por cada componente en la estación Aeropuerto

PC1	28,17%
PC2	22,10%
PC3	7,20%
PC4	4,76%

-Loadings:

Dada la estructura inicial de los datos la componente horizontal corresponde a las variables para cada momento del día. Del número 1 al 144 corresponde a la temperatura como se describe en el apartado 3.2.3 donde se pueden ver el resto de relaciones. De la 145 a la 288 corresponde a la variable humedad, de la 289 a la 432 a la precipitación, de la 433 a la 576 a la velocidad del viento y por último de la 577 a la 720 a la dirección del viento.

Internamente las variables temperatura, humedad, etc... contienen 143 valores propios de cada variable correspondientes a la medida intradía. Esto significa que tenemos valores de la PC1 para cada variable y en cada variable para cada momento del día. El día es dividido de 10 minutos en 10 minutos que empieza a las 00:00h hasta las 23:50h.

Esta forma de ordenar las variables nos permite observar si hay variaciones de las variables a lo largo del día, cosa que intuitivamente y por el caso que es se puede intuir que sí, ya que parece lógico que por la noche la temperatura influirá menos que por el día.

Dada la tabla 3 donde se muestran los porcentajes que explican las variables principales se puede comprobar que la PC3 y la PC4 contribuyen muy poco en el modelo, 7,20% y 4,76% respectivamente. Teniendo esto en consideración se presentarán las figuras de las componentes principales 3 y 4 (fig. 5.3) pero no se interpretarán dada la poca variabilidad explicada por las variables.

- Componente principal 1 (PC1) figura 5.2:
- Variable Temperatura:

Observando la PC1 en la figura 5.2 la temperatura es la variable más intensa positiva. El valor intermedio del eje horizontal corresponde a las 12:00h, por tanto medio día. Se observa como por la noche (00:00h-8:00h) la intensidad de la variable disminuye paulatinamente hasta llegar al mínimo y vuelve a subir hasta llegar al máximo sobre las 11:00h y llegados a este punto vuelve a descender de forma constante aunque más abruptamente que por la madrugada. Sobre las 19:00h se llega a otro mínimo y vuelve a subir la intensidad hasta las 21:00h que se estabiliza y conecta con el punto inicial 00:00h.

Inicialmente parece contradictorio que en el momento de más temperatura habitualmente en el día a día que es sobre las 13:00h o las 14:00h sea en el gráfico una pendiente negativa y por tanto el valor disminuya en este punto. Pero teniendo en cuenta que observamos una gráfica de variaciones lo que significa que baje la gráfica es que en esos puntos la variación de la temperatura es menor, básicamente que se estabiliza. Esto se observa también en la mañana cuando amanece y la curva aumenta rápidamente por que se está produciendo un gran cambio de temperatura.

- Variable Humedad:

En la figura 5.2 en el segundo tramo de la PC1 se observa los valores de la humedad. Se observa que tiene máximos y mínimos y estos corresponden a los momentos críticos en los que el Sol aparece en escena y desaparece, el amanecer y el atardecer. Al igual que con la temperatura, la humedad parece tener una fuerte relación con la situación del Sol. El valor de la humedad es negativo al contrario que la Temperatura. Esto se interpreta en que su relación es inversa. A mayor Temperatura menor será la humedad.

- Variable Precipitación:

En la figura 5.2 en el tercer tramo de la PC1 se observa la Precipitación. Al parecer la Precipitación no tiene ninguna variabilidad para la PC1 ya que su gráfica es prácticamente plana.

- Variable Velocidad del Viento:

En la figura 5.2 en el cuarto tramo de la PC1 se observa la Velocidad del Viento. Esta es positiva en todo el tramo, siendo muy débil por la noche de las 23:00h hasta las 11:00h y siendo bastante intensa durante el día. Así que por el día hay más variación de la velocidad del viento que por la noche que es casi nula.

- Dirección del viento:

En la figura 5.2 en el quinto tramo de la PC1 se observa la Dirección del Viento. La Dirección del Viento presenta el mismo patrón que la Temperatura y la Humedad, la variación es mayor al

amanecer y al atardecer. En el caso de la Dirección del Viento los mínimos son más próximos al cero que en la humedad y la Temperatura.

La interpretación global de la PC1 para la estación Aeropuerto es que en valores positivos tiene como variable más intensa a la Temperatura y en segundo lugar a la velocidad del viento y en valores negativos contiene a la humedad y la dirección del viento.

Por tanto, si se obtiene una PC1 positiva alta corresponderá a un momento del día con alta temperatura y velocidad del viento y baja humedad y la correspondiente dirección del viento asociada al valor de dirección del viento baja.

- Componente principal 2 (PC2) figura 5.2 segunda parte:
- Variable Temperatura:

En la figura 5.2 en la segunda gráfica y en el primer tramo se encuentra la Temperatura de la PC2. Esta es positiva aunque con un valor relativo bajo. Contiene el mismo patrón que la Temperatura de la PC1 aunque es mucho más relajado, casi es plano.

- Variable Humedad:

En la figura 5.2 en la segunda gráfica y en el segundo tramo se encuentra la Humedad de la PC2. Esta es la variable positiva más fuerte de la PC2. Es plana durante toda la noche hasta aproximadamente las 12:00h donde crece hasta las 20:00h donde vuelve a disminuir hasta su valor inicial.

- Variable Precipitación:

En la figura 5.2 en la segunda gráfica y en el tercer tramo se encuentra la Precipitación de la PC2. La Precipitación en la PC2 es básicamente 0, no contiene variabilidad en la PC2 al igual que con la PC1.

- Variable Velocidad del Viento:

En la figura 5.2 en la segunda gráfica y en el cuarto tramo se encuentra la Velocidad del Viento de la PC2. Esta es fuertemente negativa en la PC2 al contrario que la PC1. Contiene un máximo durante el amanecer y las primeras horas de la mañana. El resto del día es más o menos plana.

- Variable Dirección del Viento:

En la figura 5.2 en la segunda gráfica y en el quinto tramo se encuentra la Dirección del Viento de la PC2. La Dirección del viento es negativa en su mayor parte. Por la noche y la madrugada contiene valores muy bajos pero al amanecer hasta el medio día presenta una creciente muy pronunciada llegando al máximo sobre las 12:00h y estabilizándose durante unas horas, hasta que en el anochecer sobre las 20:00h vuelve a disminuir bruscamente.

En la figura 5.4 y las subsiguientes 5.5 y 5.6 (que son magnificaciones de zonas relevantes de la figura 5.4) corresponden a los loadings pero esta vez enfrentando la PC1 y la PC2. A primera vista se diferencian 3 zonas principales, la superior izquierda, la media derecha y la gran nube de la zona inferior.

La nube de datos situada en la zona superior izquierda representada por la figura 5.6 corresponde a variable Humedad que tiene un valor muy alto de la PC2 y un valor negativa de la PC1.

La nube de datos situada en la zona intermedia y a la derecha representada por la figura 5.5 corresponde a la variable Temperatura y presenta un valor muy alto en la PC2 y un valor pequeño en la PC1, ambos positivos.

La tercera zona, la mayor de las 3 corresponde a las otras 3 variables. En el centro, el punto (0,0) corresponde a la Precipitación que como ya se ha comentado no presenta variabilidad para las componentes principales por eso se encuentra en el 0. La zona con la PC1 negativa y la PC2 también negativa corresponde a la variable Dirección del Viento. La nube de datos que representa un donut corresponde a la Velocidad del Viento y tiene valores positivos de la PC1 y negativos de la PC2.

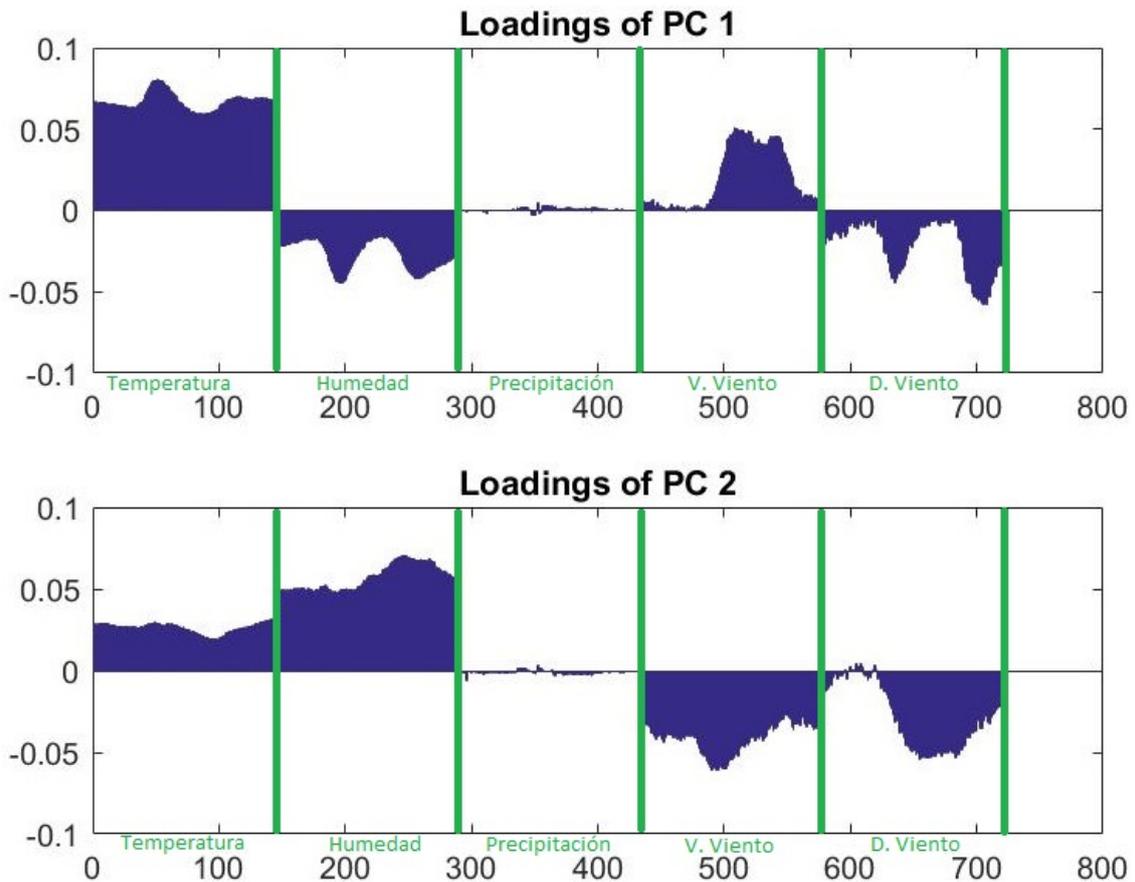


Figura 5.2. Loadings de los componentes principales 1 y 2

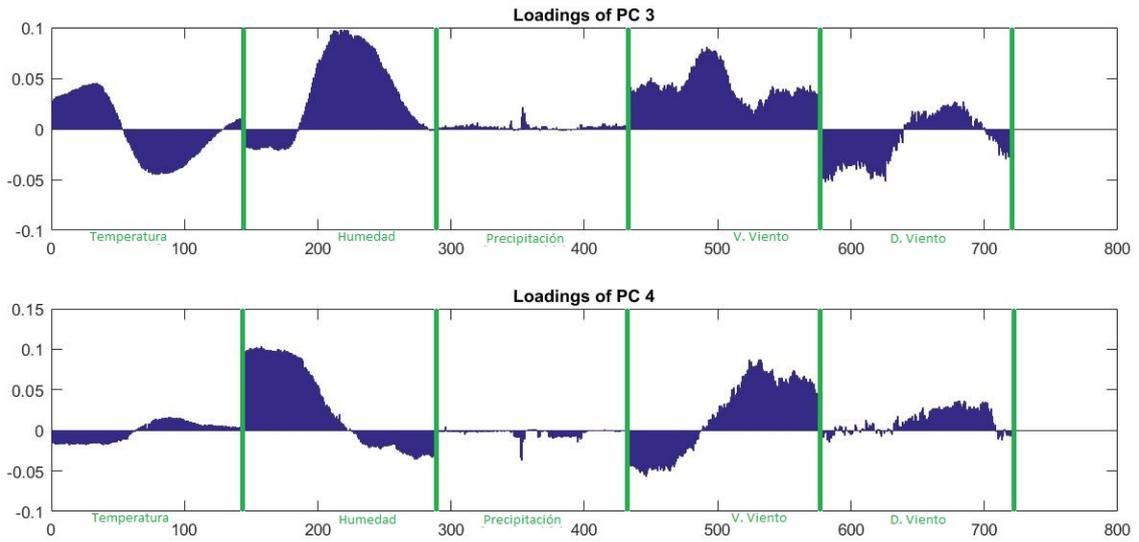


Figura 5.3. Valor de los loadings del PC3 y PC4 de la estación Aeropuerto.

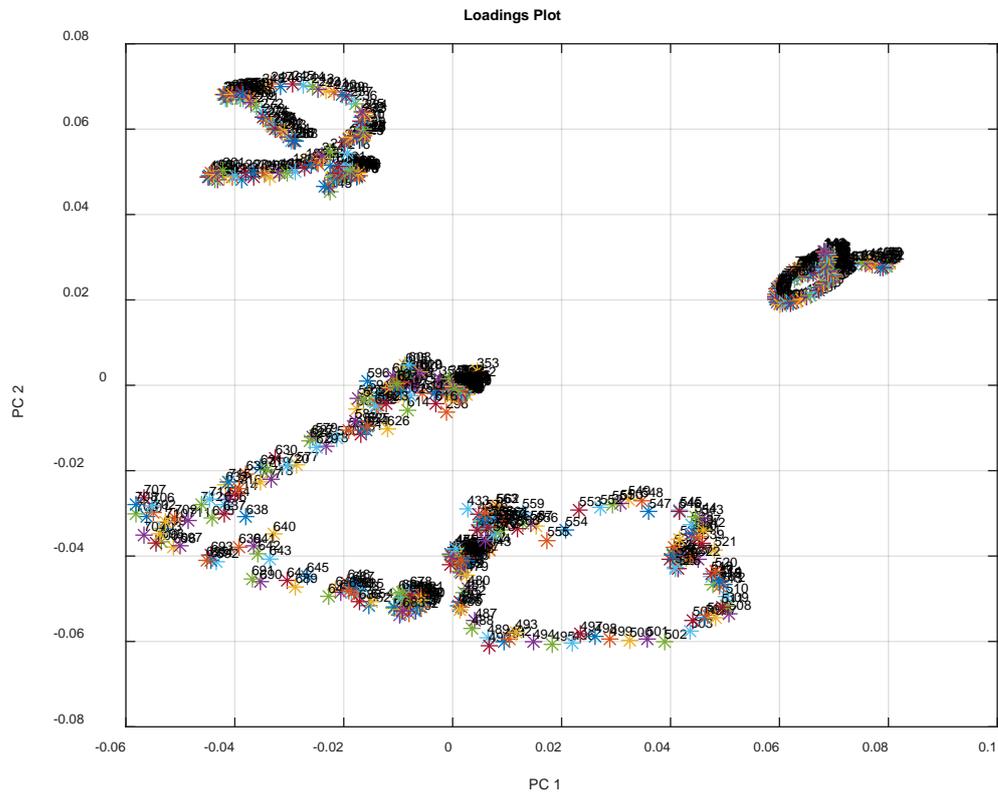


Figura 5.4. Loadings de los componentes principales.

Modelización energética de edificios por medio de estadística multivariante

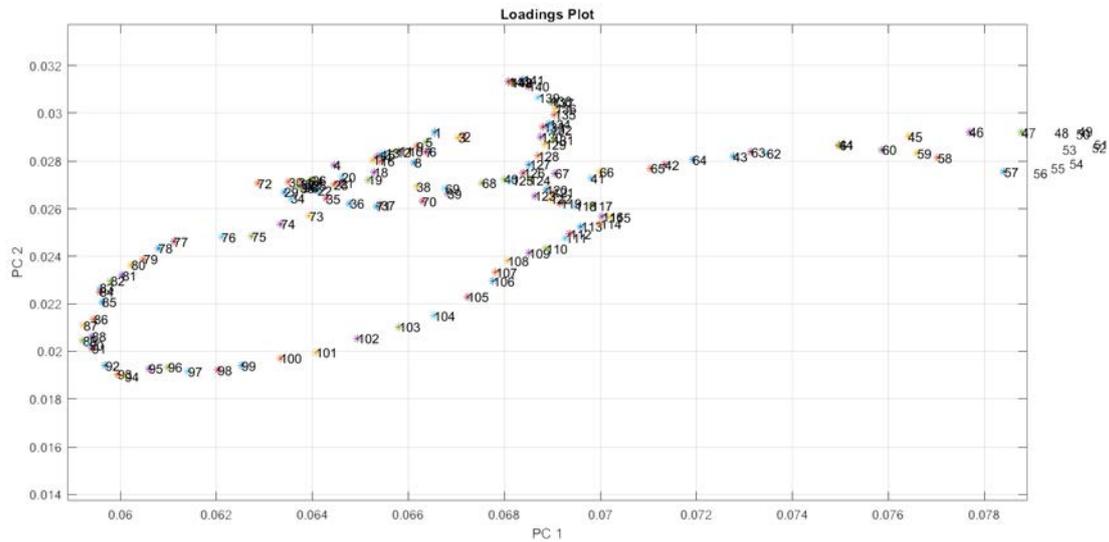


Figura 5.5. Aumento de la parte superior derecha de la figura 5.4.

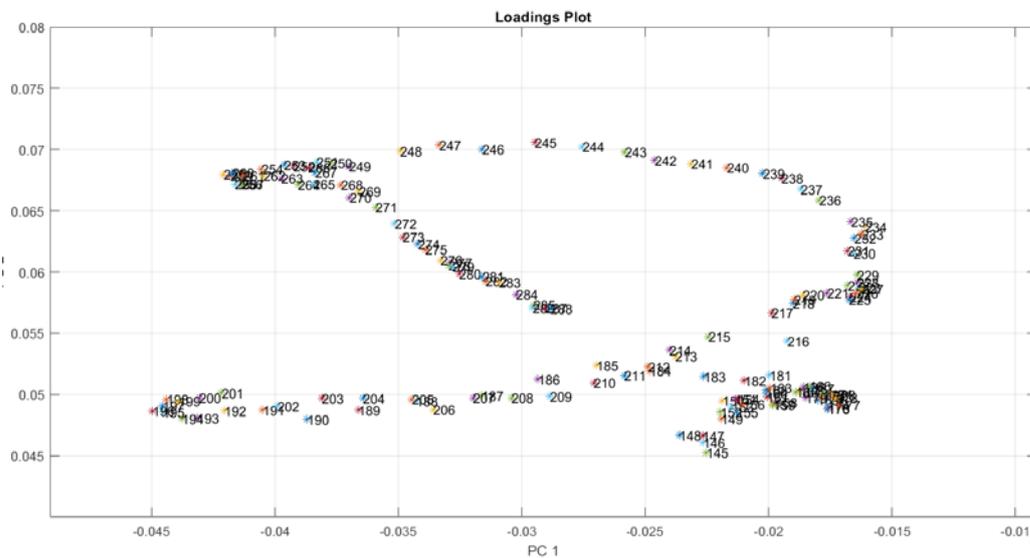


Figura 5.6. Aumento de la zona superior izquierda de la figura 5.4.

- Scores :

La figura 5.7 representa los scores de PC1 con PC2. Se puede apreciar una tendencia consistente de los datos como se observa en la figura 5.8 y no se observa ningún valor anómalo en las figuras.

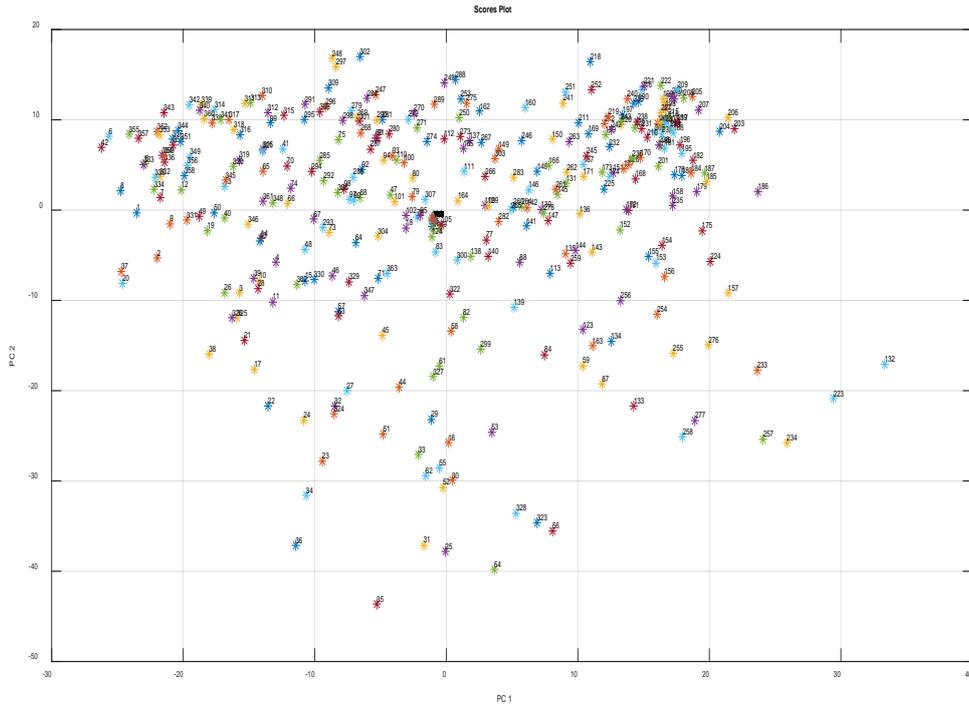


Figura 5.7. Scores de los componentes principales.

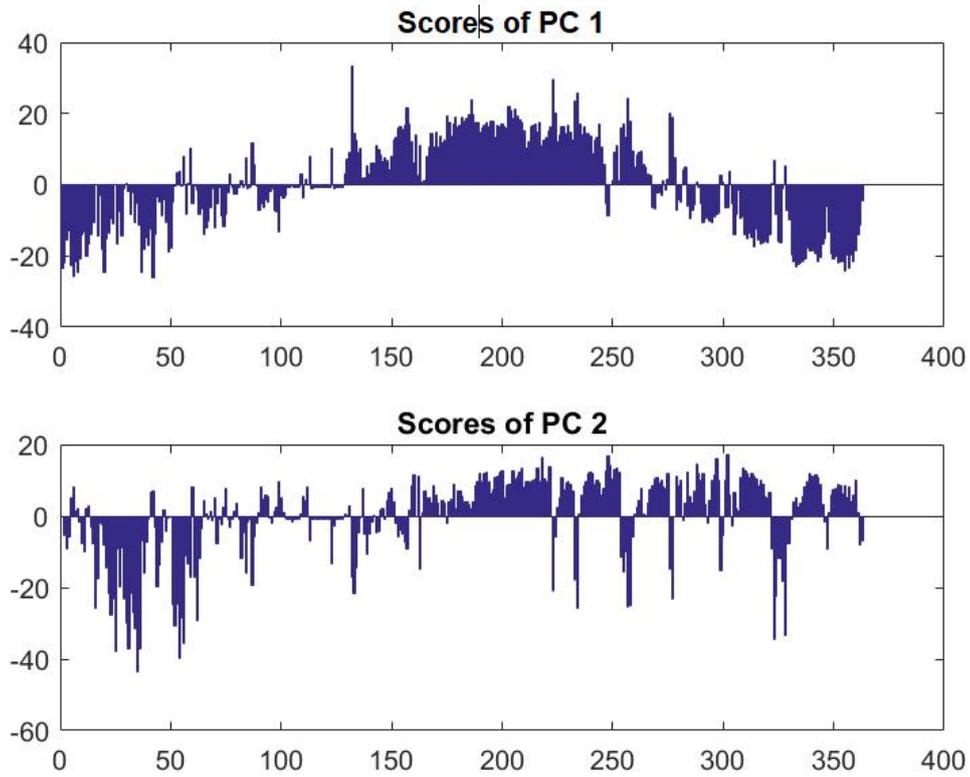


Figura 5.8. Scores de los componentes principales.

5.2 ALTOCÚMULO

Dada la tabla 4 donde se muestran los porcentajes que explican las variables principales se puede comprobar que la PC3 y la PC4 contribuyen muy poco en el modelo, 9,02% y 4,99% respectivamente. Teniendo esto en consideración se presentarán las figuras de las componentes principales 3 y 4 (figura 5.10) pero no se interpretarán dada la poca variabilidad explicada por las variables.

- Componente principal 1 (PC1) figura 5.9:
- Variable Temperatura:

Observando la PC1 en la figura 5.9 la temperatura es la variable más intensa positiva. El valor intermedio del eje horizontal corresponde a las 12:00h, por tanto medio día. Se observa como por la noche (00:00h-9:00h) la intensidad de la variable es constante y crece hasta llegar al máximo sobre las 11:00h y llegados a este punto desciende de forma constante aunque abruptamente. Sobre las 19:00h se llega al mínimo y vuelve a subir y conecta con el punto inicial 00:00h.

- Variable Humedad:

En la figura 5.9 en el segundo tramo de la PC1 se observa los valores de la humedad. Se observa que tiene máximos y mínimos y estos corresponden a los momentos críticos en los que el Sol aparece en escena y desaparece, el amanecer y el atardecer. Al igual que con la temperatura, la humedad parece tener una fuerte relación con la situación del Sol. El valor de la humedad es positivo pero menos intenso que la Temperatura.

- Variable Precipitación:

En la figura 5.9 en el tercer tramo de la PC1 se observa la Precipitación. Al parecer la Precipitación no tiene ninguna variabilidad para la PC1 ya que su gráfica es prácticamente plana, excepto en ciertos puntos muy concretos.

- Variable Velocidad del Viento:

En la figura 5.9 en el cuarto tramo de la PC1 se observa la Velocidad del Viento. La V. del Viento es negativa por la noche y justo a las 12:00h pasa por un valor 0 y se vuelve positiva hasta llegar a un máximo a las 18:00h que vuelve a disminuir hasta llegar a las 23:00h que se vuelve 0 y vuelve a hacerse negativa. Es como la función seno. Esto implica que esta variable tiene un comportamiento para este componente completamente opuesto si es de día o de noche.

- Dirección del viento:

En la figura 5.9 en el quinto tramo de la PC1 se observa la Dirección del Viento. La Dirección del Viento presenta un patrón con máximos y mínimos muy abruptos, no como en la Temperatura y la Humedad, la variación es mayor al amanecer y al atardecer. Los máximos se sitúan a las 00:00h y las 12:00h y los mínimos a las 6:00h y las 18h.

La interpretación global de la PC1 para la estación Altocúmulo es que en valores positivos tiene como variable más intensa a la Temperatura, en segundo lugar a la Humedad y de día y tarde a la Velocidad del Viento (12:00h hasta 23:30h). En valores negativos contiene a la Dirección del Viento y por la noche y de mañana (00:00h hasta 12:00h) a la Velocidad del viento.

Por tanto, si se obtiene una PC1 positiva alta corresponderá a un momento del día con alta temperatura y Humedad y la correspondiente dirección del viento asociada al valor de dirección del viento baja.

- Componente principal 2 (PC2) figura 5.9 segunda parte:
- Variable Temperatura:

En la figura 5.9 en la segunda gráfica y en el primer tramo se encuentra la Temperatura de la PC2. Esta es negativa aunque con un valor relativo bajo. Contiene un patrón diferente a la Temperatura de la PC1 y es mucho más relajado, casi es plano.

- Variable Humedad:

En la figura 5.9 en la segunda gráfica y en el segundo tramo se encuentra la Humedad de la PC2. Esta es la variable positiva más fuerte de la PC2. Es creciente durante toda la noche, la mañana y la tarde hasta aproximadamente las 21:00h donde empieza a decrecer bruscamente hasta las 23:55h donde vuelve hasta su valor inicial.

- Variable Precipitación:

En la figura 5.9 en la segunda gráfica y en el tercer tramo se encuentra la Precipitación de la PC2. La Precipitación en la PC2 es básicamente 0, no contiene variabilidad en la PC2 al igual que con la PC1.

- Variable Velocidad del Viento:

En la figura 5.9 en la segunda gráfica y en el cuarto tramo se encuentra la Velocidad del Viento de la PC2. Esta es fuertemente negativa en la PC2 y ha diferencia de la PC1 no se vuelve positiva. Contiene un máximo durante el medio día que se mantiene hasta el anochecer. Superando las 21:00 se produce una abrupta caída hasta llegar a las 00:00h que vuelve al valor inicial.

- Variable Dirección del Viento:

En la figura 5.9 en la segunda gráfica y en el quinto tramo se encuentra la Dirección del Viento de la PC2. La Dirección del viento es negativa en su mayor parte. Por la noche y la madrugada contiene valores muy bajos pero al amanecer hasta el medio día presenta una creciente muy pronunciada llegando al máximo sobre las 12:00h y estabilizándose durante unas horas, hasta que en el anochecer sobre las 20:00h vuelve a disminuir bruscamente.

- En la figura 5.11 se observan los loadings con las PC1 y PC2 enfrentadas. Se diferencian 4 zonas: En la parte central superior, representado por la figura 5.14, se muestra la nube asociada a la Humedad con valores muy positivos en la PC2 y vagamente positivos en la PC1.
- En la parte central se encuentra la Precipitación representada por la figura 5.12. Posee valores próximos a cero para ambas PC.
- En la parte inferior derecha se encuentra la Temperatura representada por la figura 5.13. Posee valores altos para la PC1 y algo negativos para la PC2.
- La zona inferior e inferior izquierda corresponde a las dos variables del Viento. La dirección del viento es negativa para ambas PC. La Velocidad del Viento es negativa para la PC2 y positiva y negativa para a la PC1.

Modelización energética de edificios por medio de estadística multivariante

La figura 5.15 muestra las dos componentes principales pero representando los scores. Es una nube bastante uniforme con una concentración en PC1=20 y PC2=0, por lo demás bastante uniforme, sin datos anómalos.

Tabla 4. Porcentaje de variabilidad explicada por cada componente en la estación Altocúmulo

PC1	23,61%
PC2	19,88%
PC3	9,02%
PC4	4,99%

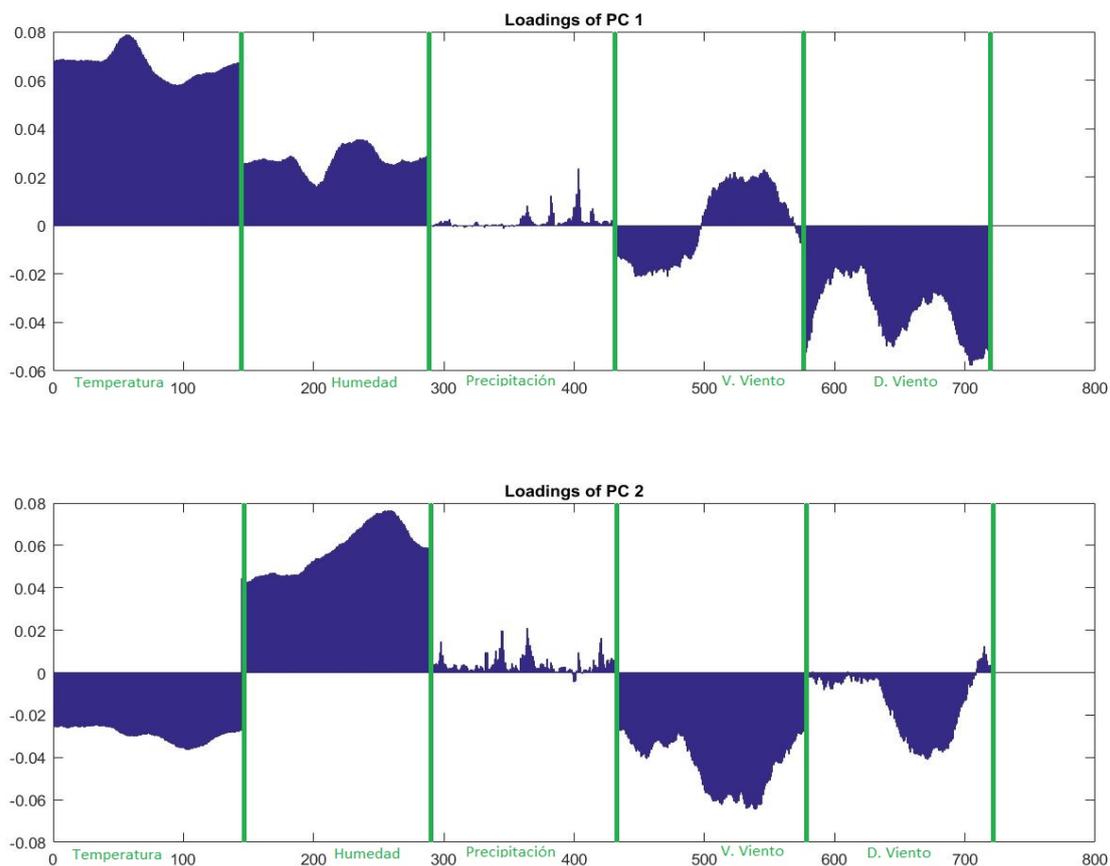


Figura 5.9 Loadings de las componentes principales

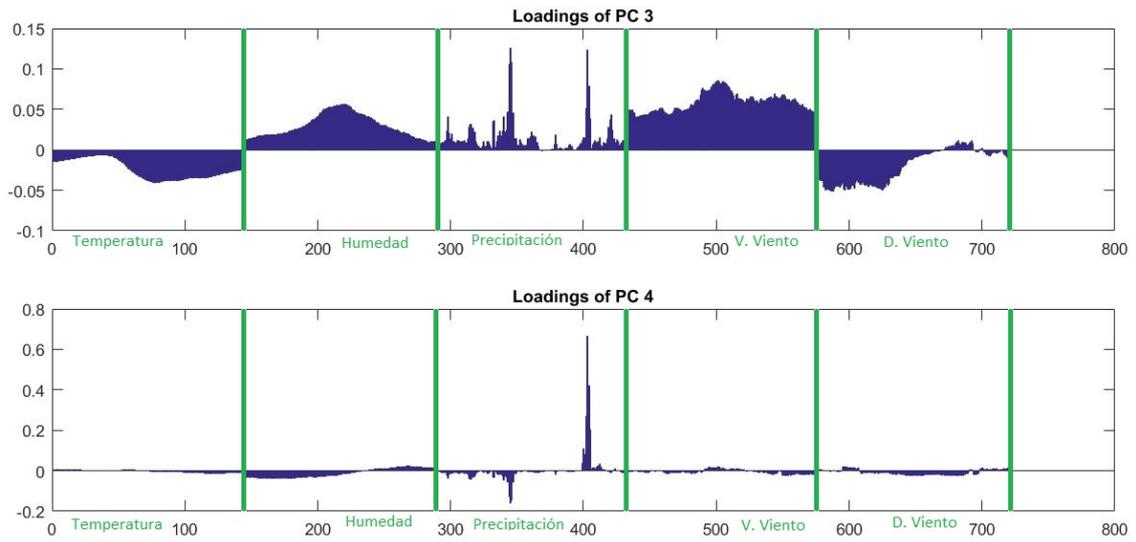


Figura 5.10 Valor de los loadings del PC3 y PC4 de la estación Altocúmulo.

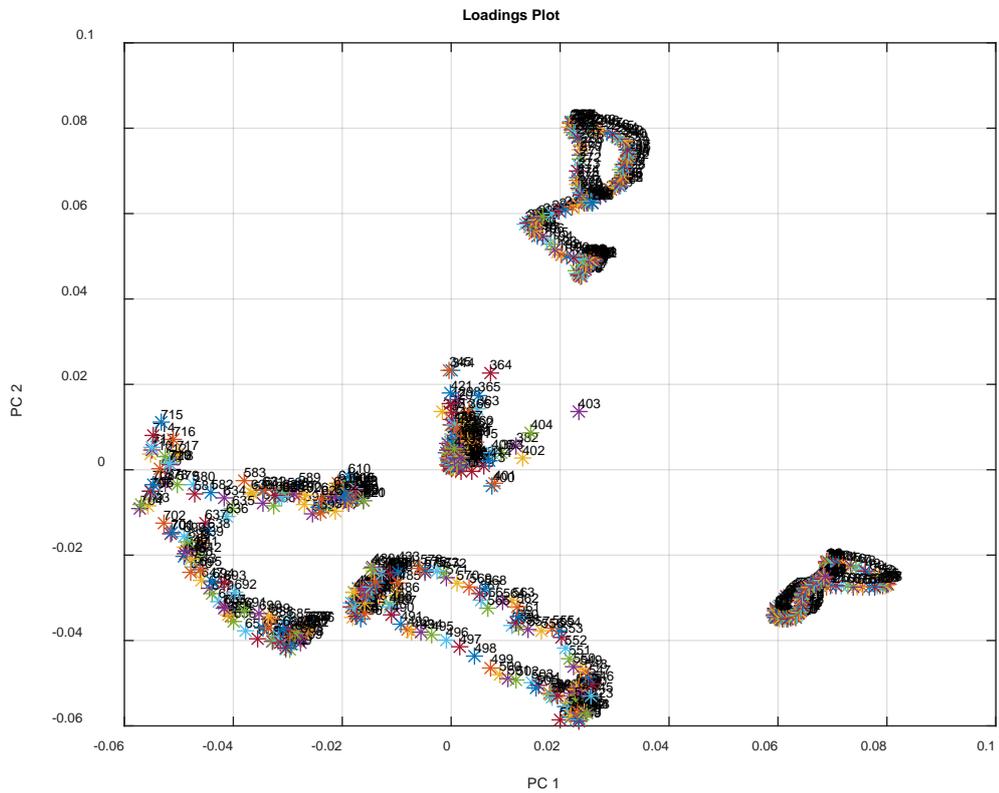


Figura 5.11 Loadings de las componentes principales.

Modelización energética de edificios por medio de estadística multivariante

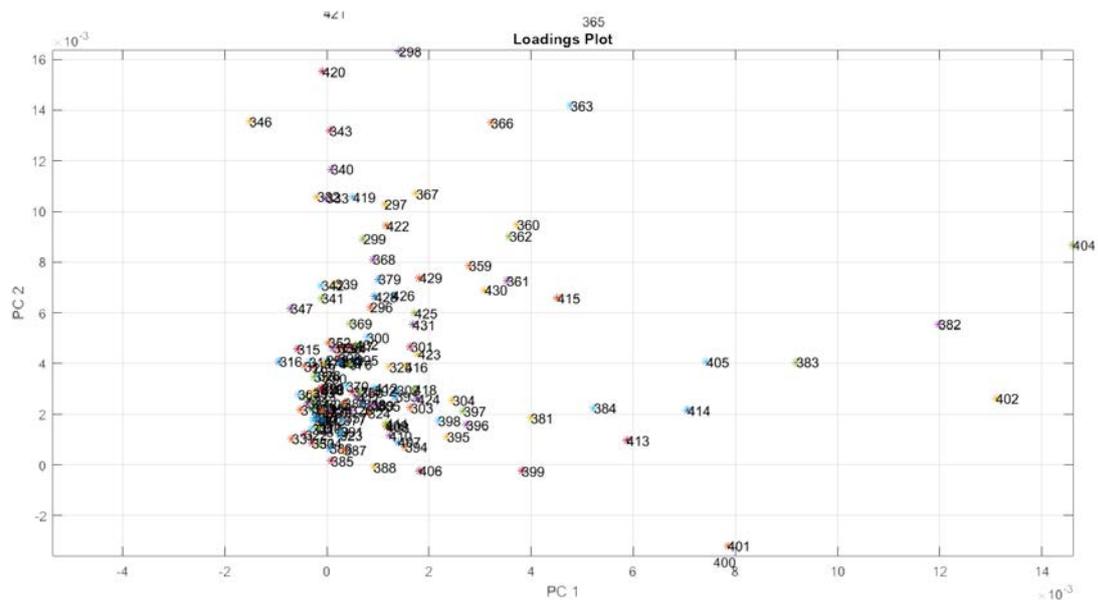


Figura 5.12 Aumento de la parte central de la figura 5.11.

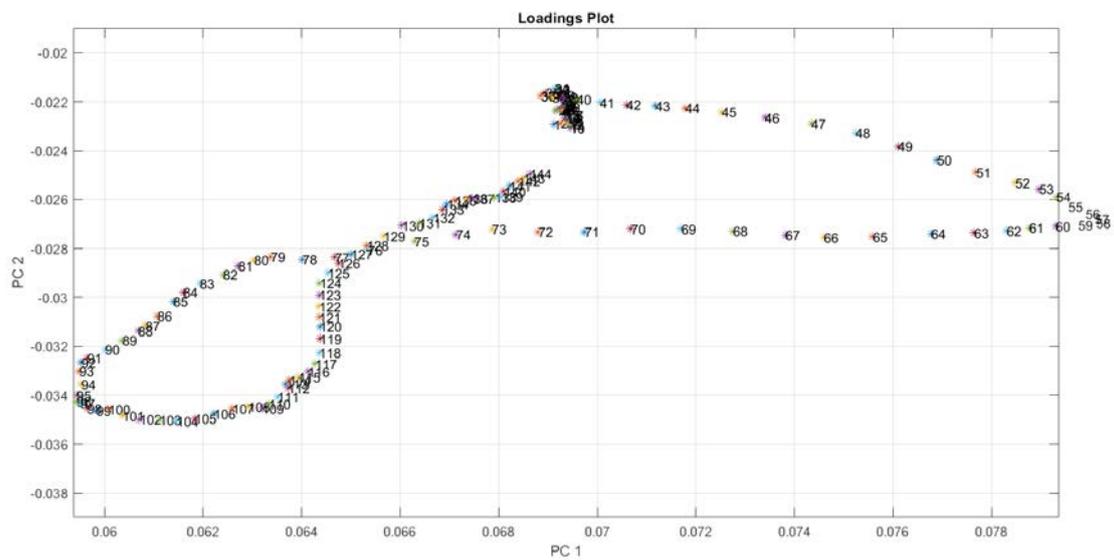


Figura 5.13 Aumento de la parte inferior derecha de la figura 5.11.

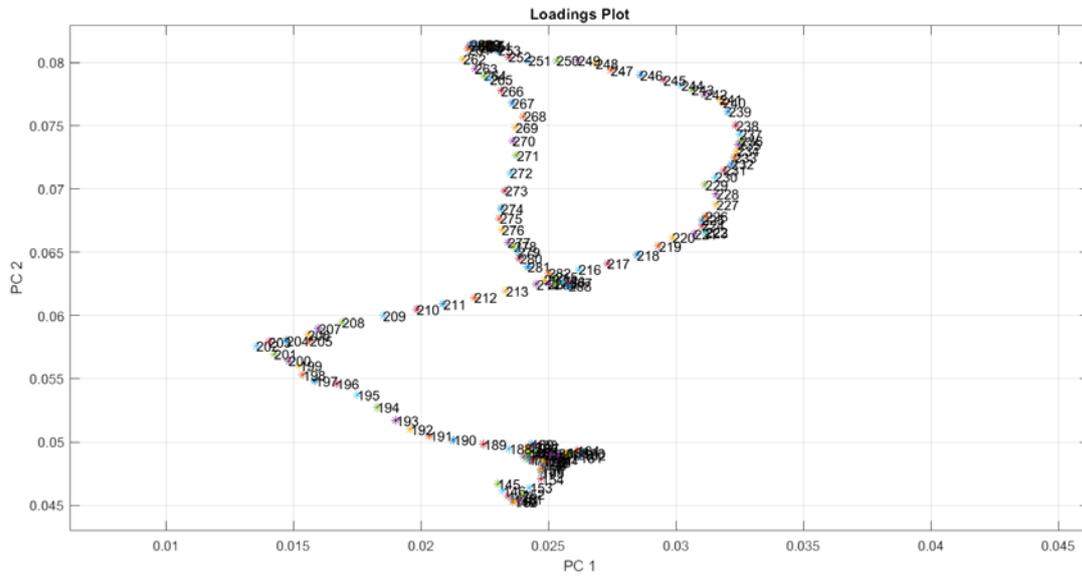


Figura 5.14 Aumento de la parte central superior de la figura 5.11.

Scores:

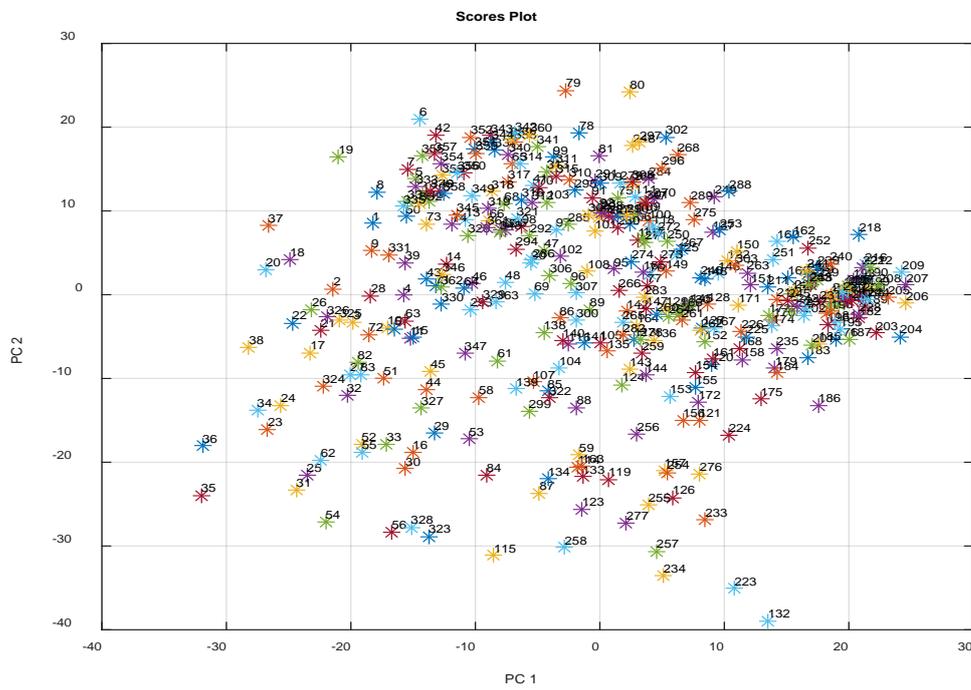


Figura 5.15 Scores de los componentes principales.

5.3 UPV

Dada la tabla 5 donde se muestran los porcentajes que explican las variables principales se puede comprobar que la PC3 y la PC4 contribuyen muy poco en el modelo, 7,98% y 5,78% respectivamente. Teniendo esto en consideración se presentarán las figuras de las componentes principales 3 y 4 (figura 5.17) pero no se interpretarán dada la poca variabilidad explicada por las variables.

Como se verá a continuación la estación meteorológica UPV se encuentra en una localización en la ciudad muy similar a la de Altocúmulo por lo que los valores de los *loadings* son un verdadero calco de Altocúmulo.

- Componente principal 1 (PC1) figura 5.16:
- Variable Temperatura:

Observando la PC1 en la figura 5.16 la temperatura es la variable más intensa positiva. El valor intermedio del eje horizontal corresponde a las 12:00h, por tanto medio día. Se observa como por la noche (00:00h-9:00h) la intensidad de la variable es constante y crece hasta llegar al máximo sobre las 11:00h y llegados a este punto desciende de forma constante aunque abruptamente. Sobre las 19:00h se llega al mínimo y vuelve a subir y conecta con el punto inicial 00:00h.

- Variable Humedad:

En la figura 5.16 en el segundo tramo de la PC1 se observa los valores de la humedad. Se observa que tiene máximos y mínimos y estos corresponden a los momentos críticos en los que el Sol aparece en escena y desaparece, el amanecer y el atardecer. Al igual que con la temperatura, la humedad parece tener una fuerte relación con la situación del Sol. El valor de la humedad es positivo pero menos intensa que la Temperatura.

- Variable Precipitación:

En la figura 5.16 en el tercer tramo de la PC1 se observa la Precipitación. Al parecer la Precipitación no tiene ninguna variabilidad para la PC1 ya que su gráfica es prácticamente plana, excepto en ciertos puntos muy concretos.

- Variable Velocidad del Viento:

En la figura 5.16 en el cuarto tramo de la PC1 se observa la Velocidad del Viento. La V. del Viento es negativa por la noche y justo a las 12:00h pasa por un valor 0 y se vuelve positiva hasta llegar a un máximo a las 18:00h que vuelve a disminuir hasta llegar a las 23:00h que se vuelve 0 y vuelve a hacerse negativa. Es como la función seno. Esto implica que esta variable tiene un comportamiento, para este componente, completamente opuesto si es de día o de noche.

- Dirección del viento:

En la figura 5.16 en el quinto tramo de la PC1 se observa la Dirección del Viento. La Dirección del Viento presenta un patrón con máximos y mínimos muy abruptos, no como en la

Temperatura y la Humedad, la variación es mayor al amanecer y al atardecer. Los máximos se sitúan a las 00:00h y las 12:00h y los mínimos a las 6:00h y las 18h.

La interpretación global de la PC1 para la estación UPV es que en valores positivos tiene como variable más intensa a la Temperatura, en segundo lugar a la Humedad y de día y tarde a la Velocidad del Viento (12:00h hasta 23:30h). En valores negativos contiene a la Dirección del Viento y por la noche y de mañana (00:00h hasta 12:00h) a la Velocidad del viento.

Por tanto, si se obtiene una PC1 positiva alta corresponderá a un momento del día con alta temperatura y Humedad y la correspondiente dirección del viento asociada al valor de dirección del viento baja.

- Componente principal 2 (PC2) figura 5.16 segunda parte:
- Variable Temperatura:

En la figura 5.16 en la segunda gráfica y en el primer tramo se encuentra la Temperatura de la PC2. Esta es negativa aunque con un valor relativo bajo. Contiene un patrón diferente a la Temperatura de la PC1 y es mucho más relajado, casi es plano.

- Variable Humedad:

En la figura 5.16 en la segunda gráfica y en el segundo tramo se encuentra la Humedad de la PC2. Esta es la variable positiva más fuerte de la PC2. Es creciente durante toda la noche, la mañana y la tarde hasta aproximadamente las 21:00h donde empieza a decrecer bruscamente hasta las 23:55h donde vuelve hasta su valor inicial.

- Variable Precipitación:

En la figura 5.16 en la segunda gráfica y en el tercer tramo se encuentra la Precipitación de la PC2. La Precipitación en la PC2 es básicamente 0, no contiene variabilidad en la PC2 al igual que con la PC1.

- Variable Velocidad del Viento:

En la figura 5.16 en la segunda gráfica y en el cuarto tramo se encuentra la Velocidad del Viento de la PC2. Esta es fuertemente negativa en la PC2 y a diferencia de la PC1 no se vuelve positiva. Contiene un máximo durante el medio día que se mantiene hasta el anochecer. Superando las 21:00 se produce una abrupta caída hasta llegar a las 00:00h que vuelve al valor inicial.

- Variable Dirección del Viento:

En la figura 5.16 en la segunda gráfica y en el quinto tramo se encuentra la Dirección del Viento de la PC2. La Dirección del viento es negativa en su mayor parte. Por la noche y la madrugada contiene valores muy bajos pero al amanecer hasta el medio día presenta una creciente muy pronunciada llegando al máximo sobre las 12:00h y estabilizándose durante unas horas, hasta que en el anochecer sobre las 20:00h vuelve a disminuir bruscamente.

En la figura 5.18 se observan los loadings con las PC1 y PC2 enfrentadas. Se diferencian 4 zonas:

Modelización energética de edificios por medio de estadística multivariante

- En la parte central superior, representado por la figura 5.20, se muestra la nube asociada a la Humedad con valores muy positivos en la PC2 y vagamente positivos en la PC1.
- En la parte central se encuentra la Precipitación representada por la figura 5.21. Posee valores próximos a cero para ambas PC.
- En la parte inferior derecha se encuentra la Temperatura representada por la figura 5.19. Posee valores altos para la PC1 y algo negativos para la PC2.
- La zona inferior e inferior izquierda, figura 5.22, corresponde a las dos variables del Viento. La dirección del viento es negativa para ambas PC. La Velocidad del Viento es negativa para la PC2 y positiva y negativa para a la PC1.

La figura 5.23 muestra las dos componentes principales pero representando los scores. Es una nube bastante uniforme con una concentración en $PC1=20$ y $Pc2=0$, por lo demás bastante uniforme, sin datos anómalos.

Tabla 5. Porcentaje de variabilidad explicada por cada componente en la estación UPV

PC1	23,17%
PC2	17.88%
PC3	7,98%
PC4	5,78%

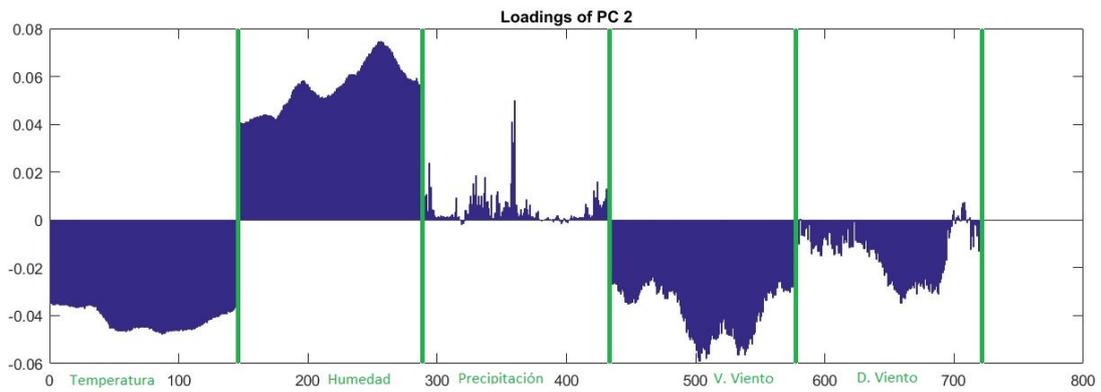
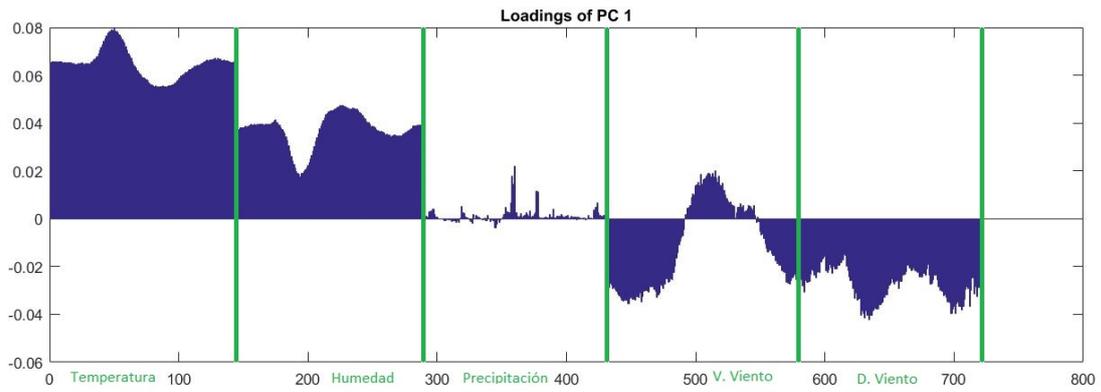


Figura 5.16 Loadings componentes principales.

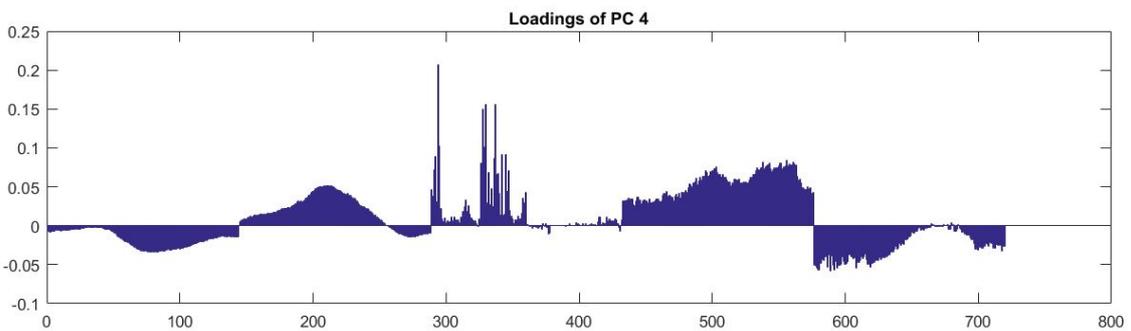
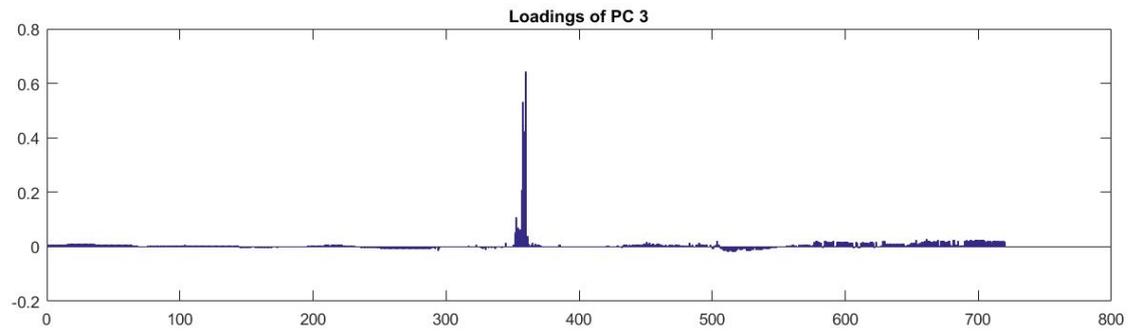


Figura 5.17 Valor de los loadings del PC3 y PC4 de la estación UPV.

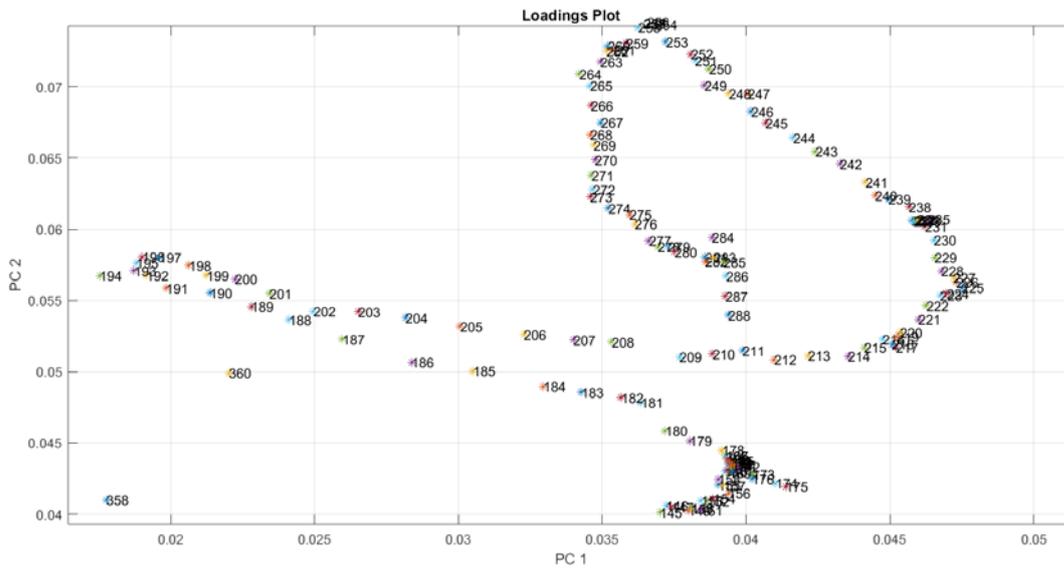


Figura 5.20 Aumento de la parte superior derecha de la figura 5.18.

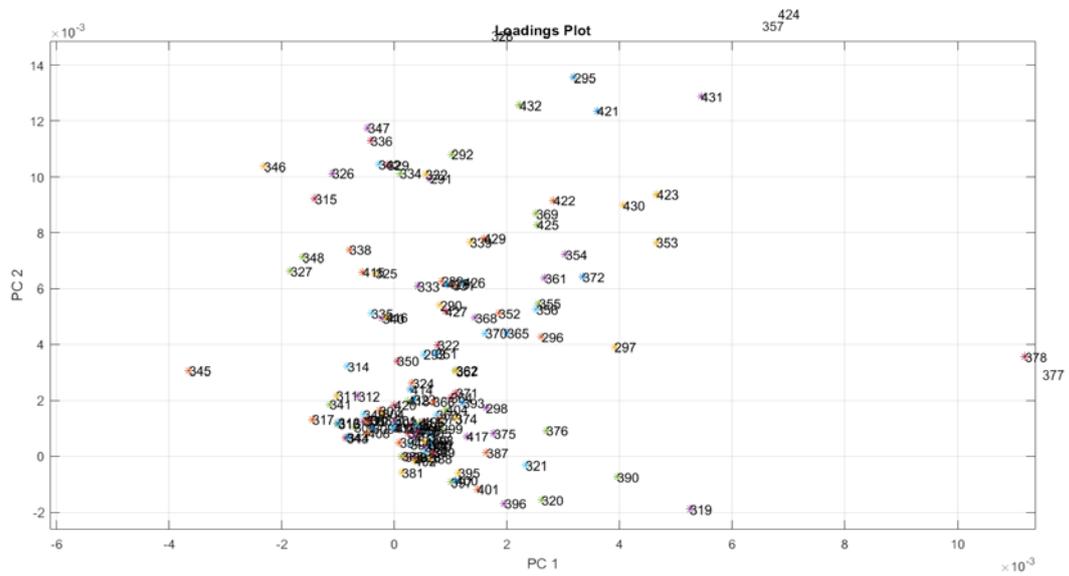


Figura 5.21 Aumento de la parte central de la figura 5.18.

Modelización energética de edificios por medio de estadística multivariante

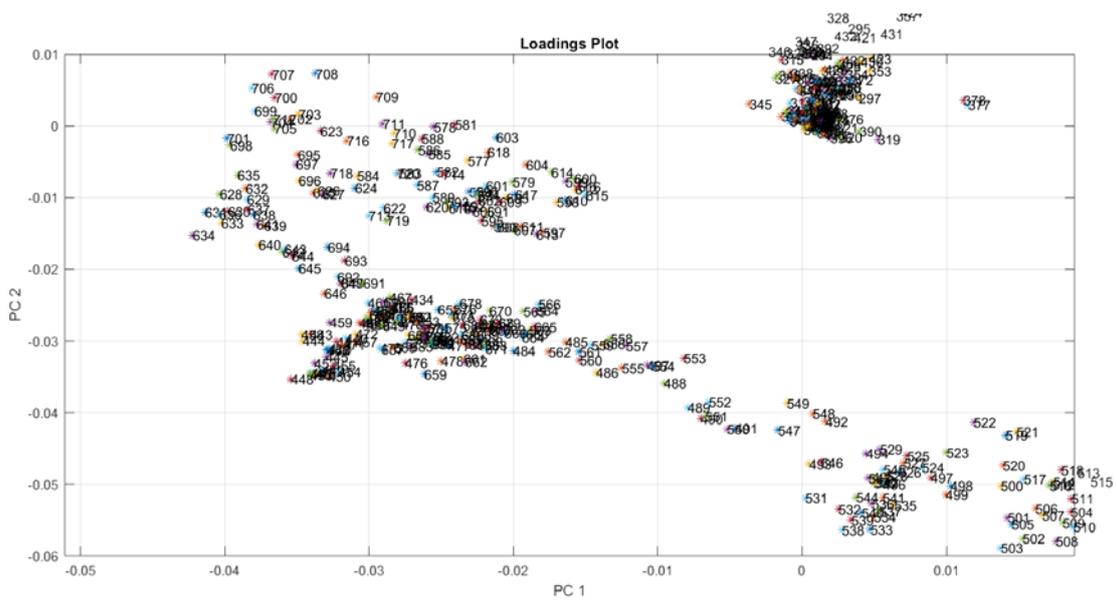


Figura 5.22 Aumento de la parte inferior izquierda de la figura 5.18

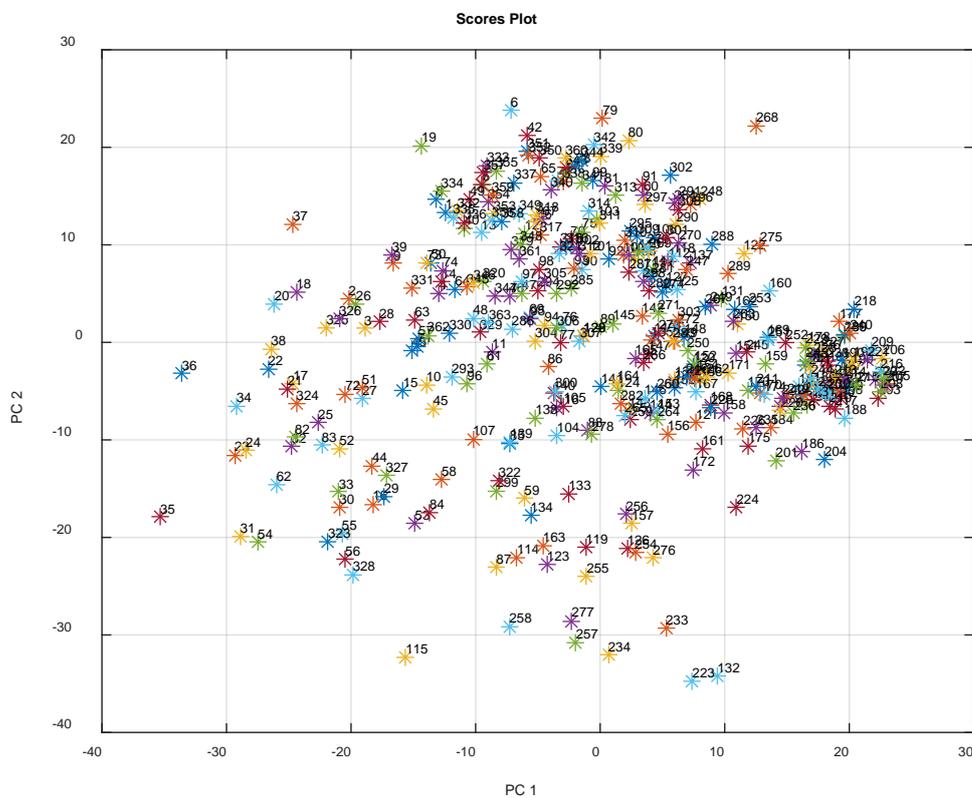


Figura 5.23 Scores de las componentes principales.

5.4 VIVEROS

Dada la tabla 6 donde se muestran los porcentajes que explican las variables principales se puede comprobar que la PC3 y la PC4 contribuyen muy poco en el modelo, 7,52% y 5,91% respectivamente. Teniendo esto en consideración se presentarán las figuras de las componentes principales 3 y 4 (figura 5.25) pero no se interpretarán dada la poca variabilidad explicada por las variables y en el caso de la estación de Viveros tampoco presenta ninguna variabilidad en las variables.

Como se verá a continuación la estación meteorológica Viveros se encuentra en una localización en la ciudad muy similar a la de Altocúmulo y UPV, un poco más en el centro de la ciudad así que los valores de los *loadings* son un prácticamente un calco de UPV y Altocúmulo.

- Componente principal 1 (PC1) figura 5.24:
- Variable Temperatura:

Observando la PC1 en la figura 5.24 la temperatura es la variable más intensa positiva. El valor intermedio del eje horizontal corresponde a las 12:00h, por tanto medio día. Se observa como por la noche (00:00h-9:00h) la intensidad de la variable es constante y crece hasta llegar al máximo sobre las 11:00h y llegados a este punto desciende de forma constante aunque abruptamente. Sobre las 19:00h se llega al mínimo y vuelve a subir y conecta con el punto inicial 00:00h.

- Variable Humedad:

En la figura 5.24 en el segundo tramo de la PC1 se observa los valores de la humedad. Se observa que tiene máximos y mínimos y estos corresponden a los momentos críticos en los que el Sol aparece en escena y desaparece, el amanecer y el atardecer. Al igual que con la temperatura, la humedad parece tener una fuerte relación con la situación del Sol. El valor de la humedad es positivo pero menos intenso que la Temperatura.

- Variable Precipitación:

En la figura 5.24 en el tercer tramo de la PC1 se observa la Precipitación. Al parecer la Precipitación no tiene ninguna variabilidad para la PC1 ya que su gráfica es prácticamente plana, excepto en ciertos puntos muy concretos.

- Variable Velocidad del Viento:

En la figura 5.24 en el cuarto tramo de la PC1 se observa la Velocidad del Viento. La V. del Viento es negativa por la noche y justo a las 12:00h pasa por un valor 0 y se vuelve positiva hasta llegar a un máximo a las 18:00h que vuelve a disminuir hasta llegar a las 23:00h que se vuelve 0 y vuelve a hacerse negativa. Es como la función seno. Esto implica que esta variable tiene un comportamiento, para este componente, completamente opuesto si es de día o de noche.

- Dirección del viento:

En la figura 5.24 en el quinto tramo de la PC1 se observa la Dirección del Viento. La Dirección del Viento presenta un patrón con máximos y mínimos muy abruptos, no como en la Temperatura y la Humedad, la variación es mayor al amanecer y al atardecer. Los máximos se sitúan a las 00:00h y las 12:00h y los mínimos a las 6:00h y las 18h.

La interpretación global de la PC1 para la estación UPV es que en valores positivos tiene como variable más intensa a la Temperatura, en segundo lugar a la Humedad y de día y tarde a la Velocidad del Viento (12:00h hasta 23:30h). En valores negativos contiene a la Dirección del Viento y por la noche y de mañana (00:00h hasta 12:00h) a la Velocidad del viento.

Por tanto, si se obtiene una PC1 positiva alta corresponderá a un momento del día con alta temperatura y Humedad y la correspondiente dirección del viento asociada al valor de dirección del viento baja.

- Componente principal 2 (PC2) figura 5.24 segunda parte:
- Variable Temperatura:

En la figura 5.24 en la segunda gráfica y en el primer tramo se encuentra la Temperatura de la PC2. Esta es negativa aunque con un valor relativo bajo. Contiene un patrón diferente a la Temperatura de la PC1 y es mucho más relajado, casi es plano.

- Variable Humedad:

En la figura 5.24 en la segunda gráfica y en el segundo tramo se encuentra la Humedad de la PC2. Esta es la variable positiva más fuerte de la PC2. Es creciente durante toda la noche, la mañana y la tarde hasta aproximadamente las 21:00h donde empieza a decrecer bruscamente hasta las 23:55h donde vuelve hasta su valor inicial.

- Variable Precipitación:

En la figura 5.24 en la segunda gráfica y en el tercer tramo se encuentra la Precipitación de la PC2. La Precipitación en la PC2 es básicamente 0, no contiene variabilidad en la PC2 al igual que con la PC1.

- Variable Velocidad del Viento:

En la figura 5.24 en la segunda gráfica y en el cuarto tramo se encuentra la Velocidad del Viento de la PC2. Esta es fuertemente negativa en la PC2 y ha diferencia de la PC1 no se vuelve positiva. Contiene un máximo durante el medio día que se mantiene hasta el anochecer. Superando las 21:00 se produce una abrupta caída hasta llegar a las 00:00h que vuelve al valor inicial.

- Variable Dirección del Viento:

En la figura 5.24 en la segunda gráfica y en el quinto tramo se encuentra la Dirección del Viento de la PC2. La Dirección del viento es negativa en su mayor parte. Por la noche y la madrugada contiene valores muy bajos pero al amanecer hasta el medio día presenta una creciente muy pronunciada llegando al máximo sobre las 12:00h y estabilizándose durante unas horas, hasta que en el anochecer sobre las 20:00h vuelve a disminuir bruscamente.

En la figura 5.26 se observan los loadings con las PC1 y PC2 enfrentadas. Se diferencian 4 zonas:

- En la parte central superior, representado por la figura 5.30, se muestra la nube asociada a la Humedad con valores muy positivos en la PC2 y vagamente positivos en la PC1.

- En la parte central se encuentra la Precipitación representada por la figura 5.27. Posee valores próximos a cero para ambas PC.
- En la parte inferior derecha se encuentra la Temperatura representada por la figura 5.28. Posee valores altos para la PC1 y algo negativos para la PC2.
- La zona inferior e inferior izquierda, figura 5.29, corresponde a las dos variables del Viento. La dirección del viento es negativa para ambas PC. La Velocidad del Viento es negativa para la PC2 y positiva y negativa para a la PC1.

Tabla 6. Porcentaje de variabilidad explicada por cada componente en la estación Viveros

PC1	21,98%
PC2	19,18%
PC3	7,52%
PC4	5,91%

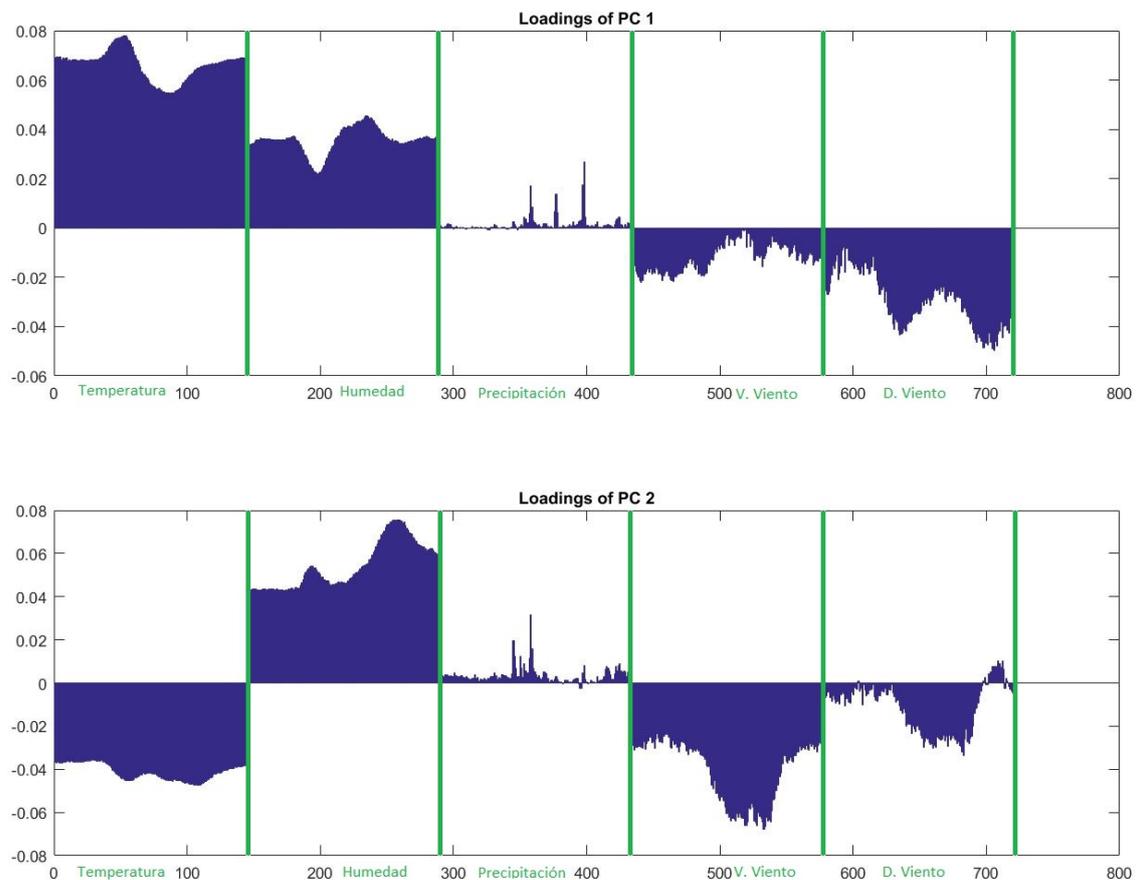


Figura 5.24 Loadings de los componentes principales.

Modelización energética de edificios por medio de estadística multivariante

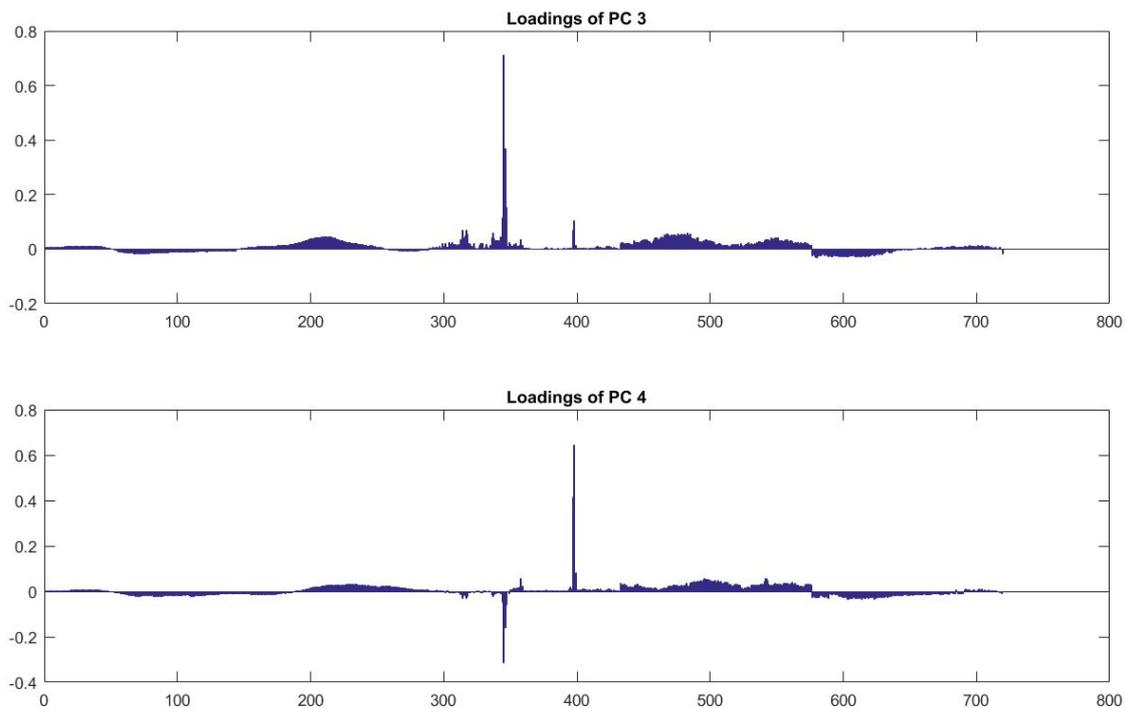


Figura 5.25 Valor de los loadings del PC3 y PC4 de la estación Viveros.

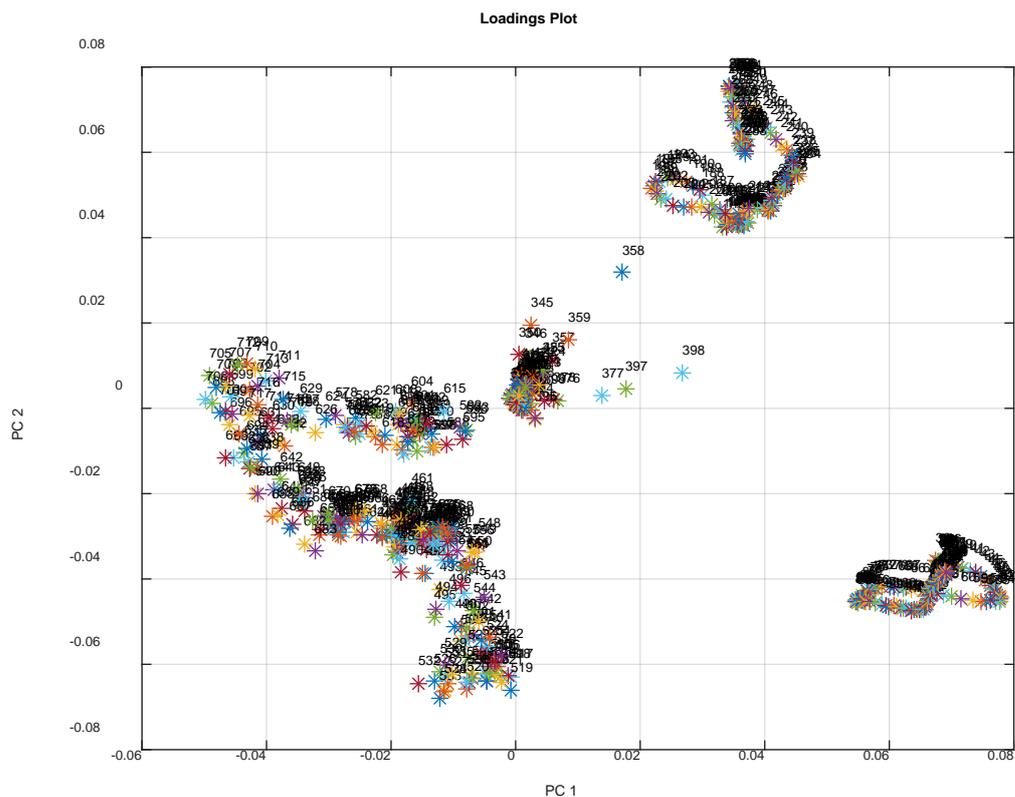


Figura 5.26 Loadings de los componentes principales.

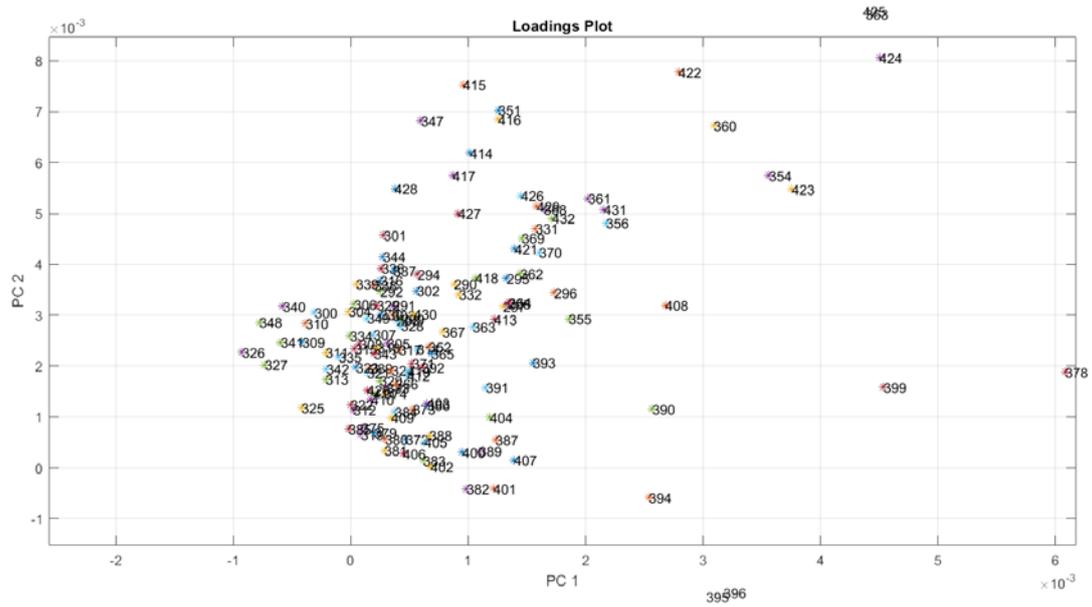


Figura 5.27 Aumento de la parte central de la figura 5.26.

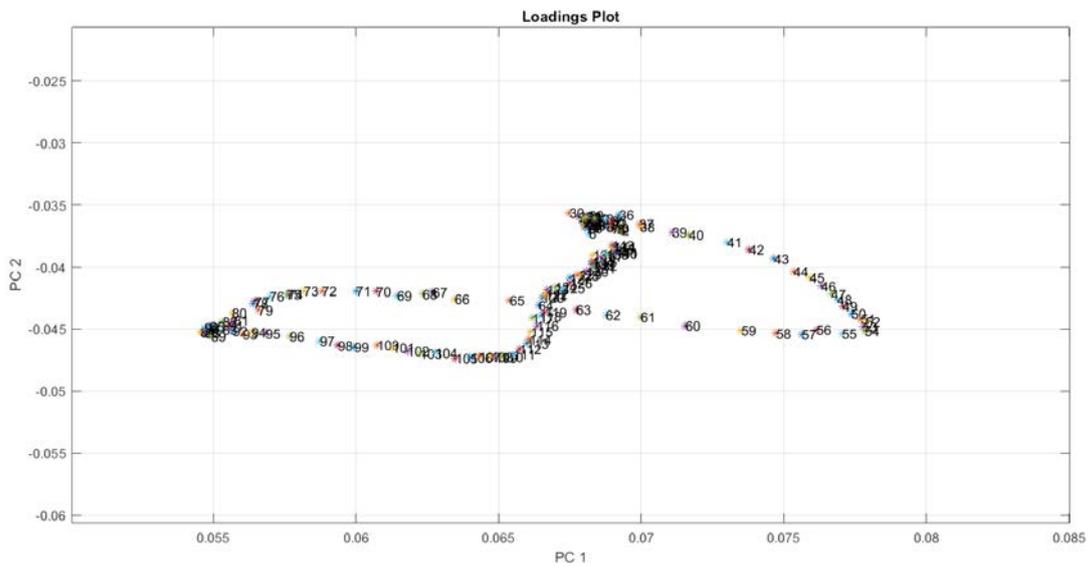


Figura 5.28 Aumento de la parte inferior derecha de la figura 5.26.

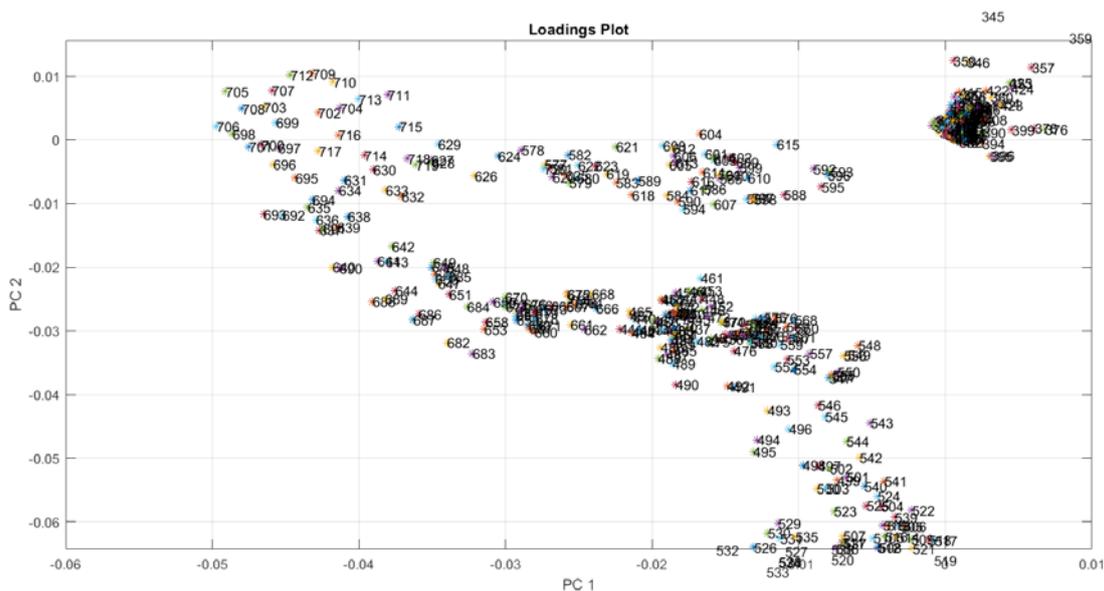


Figura 5.29 Aumento de la parte inferior izquierda de la figura 5.26.

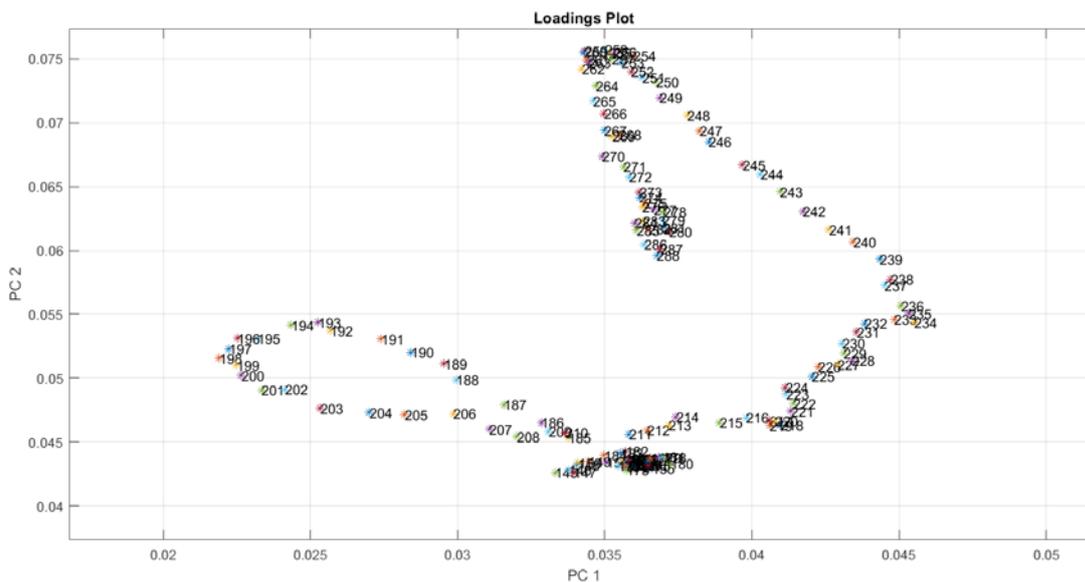


Figura 5.30 Aumento de la parte superior derecha de la figura 5.26.

5.5 COMPARATIVA DE ESTACIONES

En este apartado se estudiará en conjunto las 4 estaciones para comprobar que se parecen lo suficiente para afirmar que los datos son correctos y obtener las conclusiones en el apartado 5.

- UPV y altocúmulo: zona urbana en zona límite de la ciudad.
- Viveros: zona urbana en zona centro de Valencia.
- Aeropuerto: apartada de núcleos de población.

5.5.1 ELECCIÓN DEL NÚMERO DE COMPONENTES PRINCIPALES.

La tabla 3 muestra la variabilidad explicada por cada componente principal para la estación aeropuerto. Con los dos primeros componentes se explica el 50% de los datos y la tercera sólo explica el 7%, Esto ocurre de forma muy similar en las otras estaciones como se puede comprobar en las tablas 4,5 y 6.

Dado que las variables principales 3 y 4 para las diferentes estaciones son bastante diferentes y representan poca varianza se opta por comparar únicamente las componentes principales 1 y 2. Esto no afecta mucho al modelo ya que por la naturaleza de los datos ambientales, estos contienen mucha variabilidad y por tanto se comparan las dos primeras componentes que explican alrededor del 40% de la variabilidad.

Aún así, cabe resaltar algunos puntos de las componentes principales 3 y 4:

- La PC3 de la estación de Altocúmulo (fig. 5.10) es casi un calco de la PC4 de la estación UPV (fig. 5.17) al igual que la PC4 de la estación Altocúmulo (fig. 5.10) y la PC3 de UPV (fig. 5.17).
- La PC3 y PC4 de la estación Viveros (fig. 5.25) es casi insignificante en todas las variables excepto por unos picos muy concretos en la precipitación.
- La PC3 y PC4 de la estación Aeropuerto (fig. 5.3) no se parece a ninguna de las otras estaciones.

Las primeras figuras que se estudiarán serán:

- Figura 5.2 correspondiente a Aeropuerto.

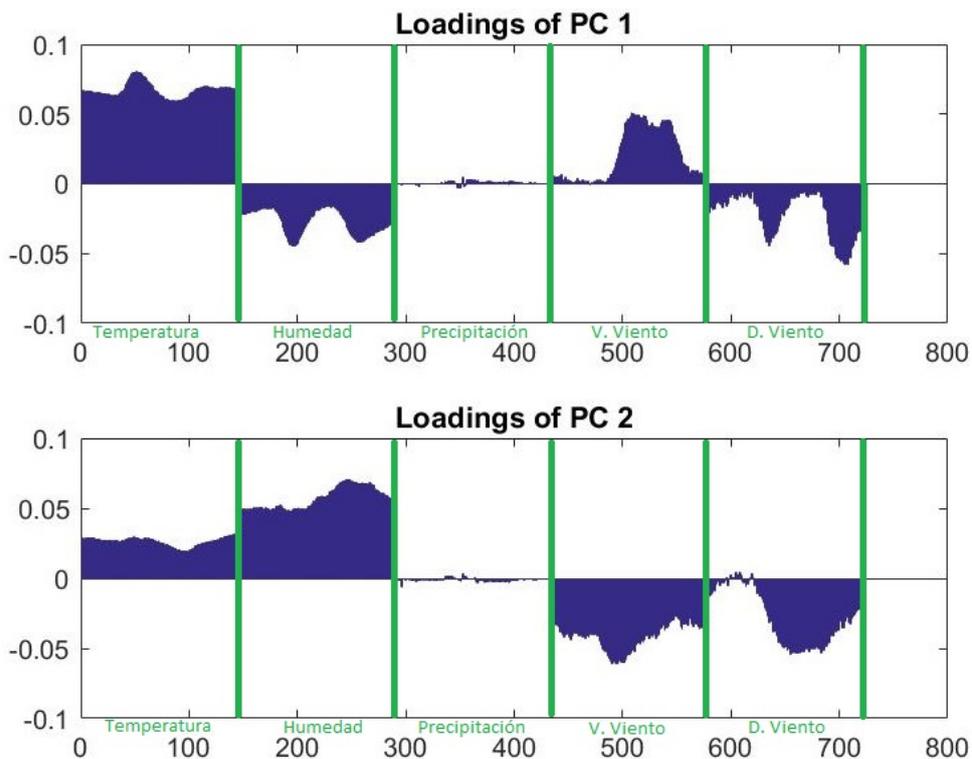


Figura 5.2 Loadings de los componentes principales

-Figura 5.9 correspondiente a Altocúmulo.

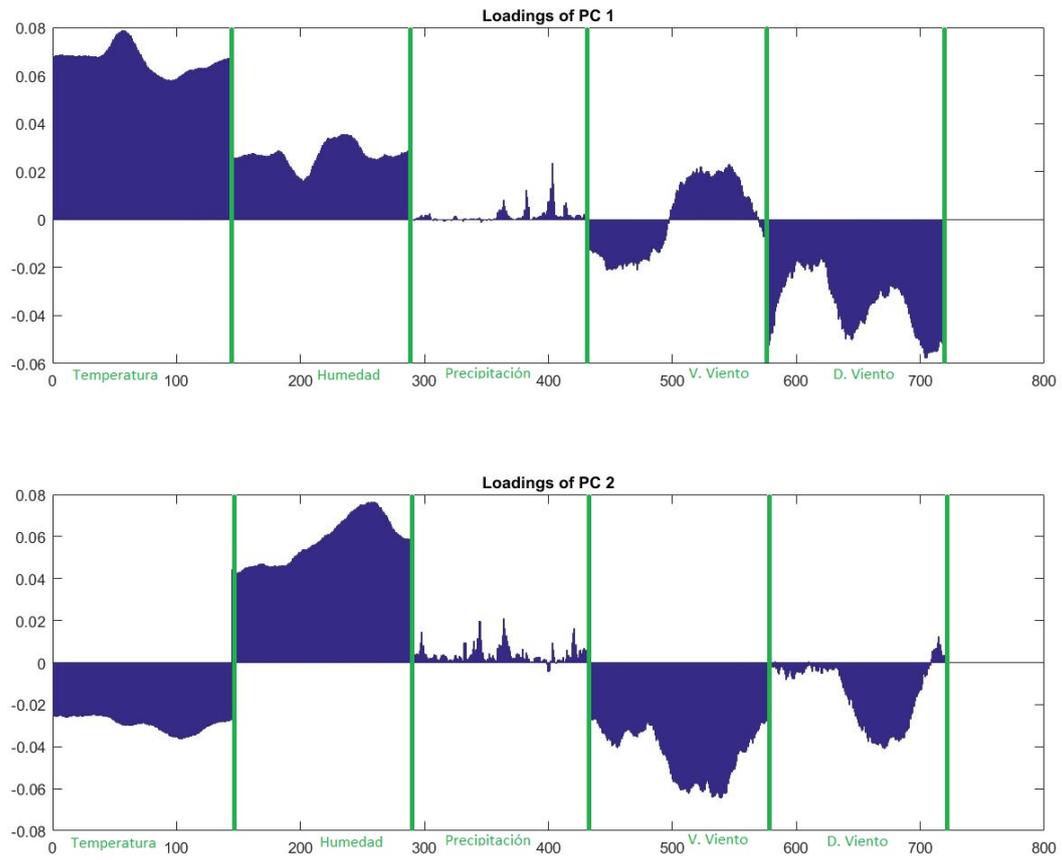


Figura 5.9 Loadings de las componentes principales

-Figura 5.16 correspondiente a UPV.

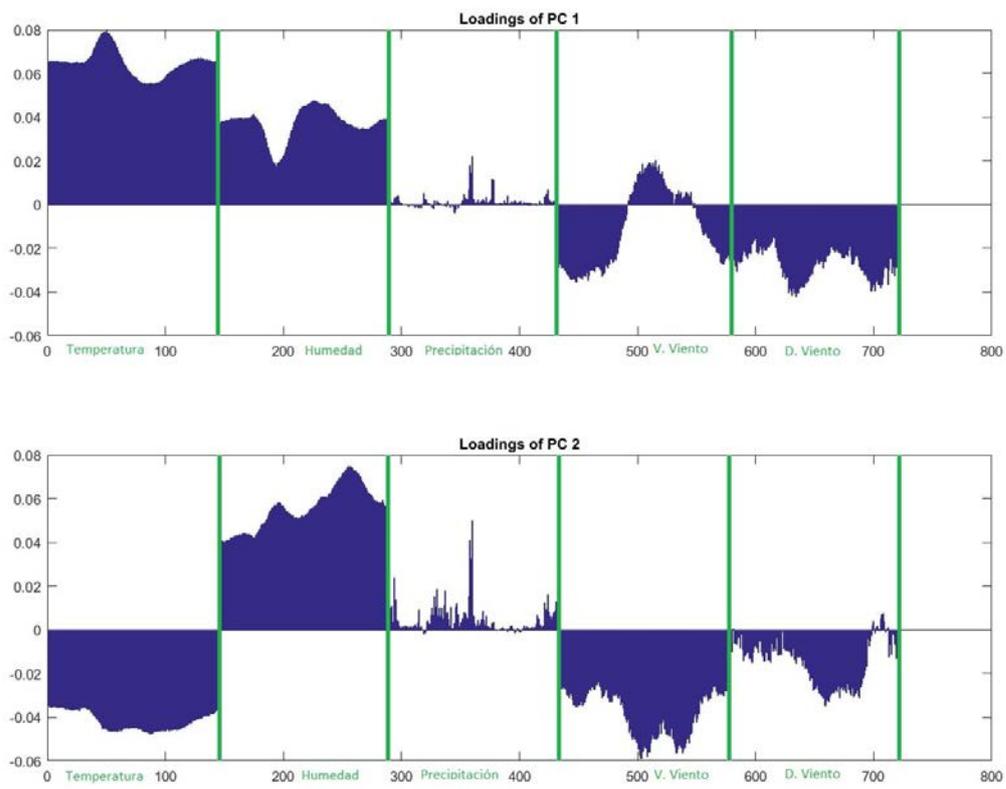


Figura 5.16 Loadings componentes principales.

-Figura 5.24 correspondiente a Viveros.

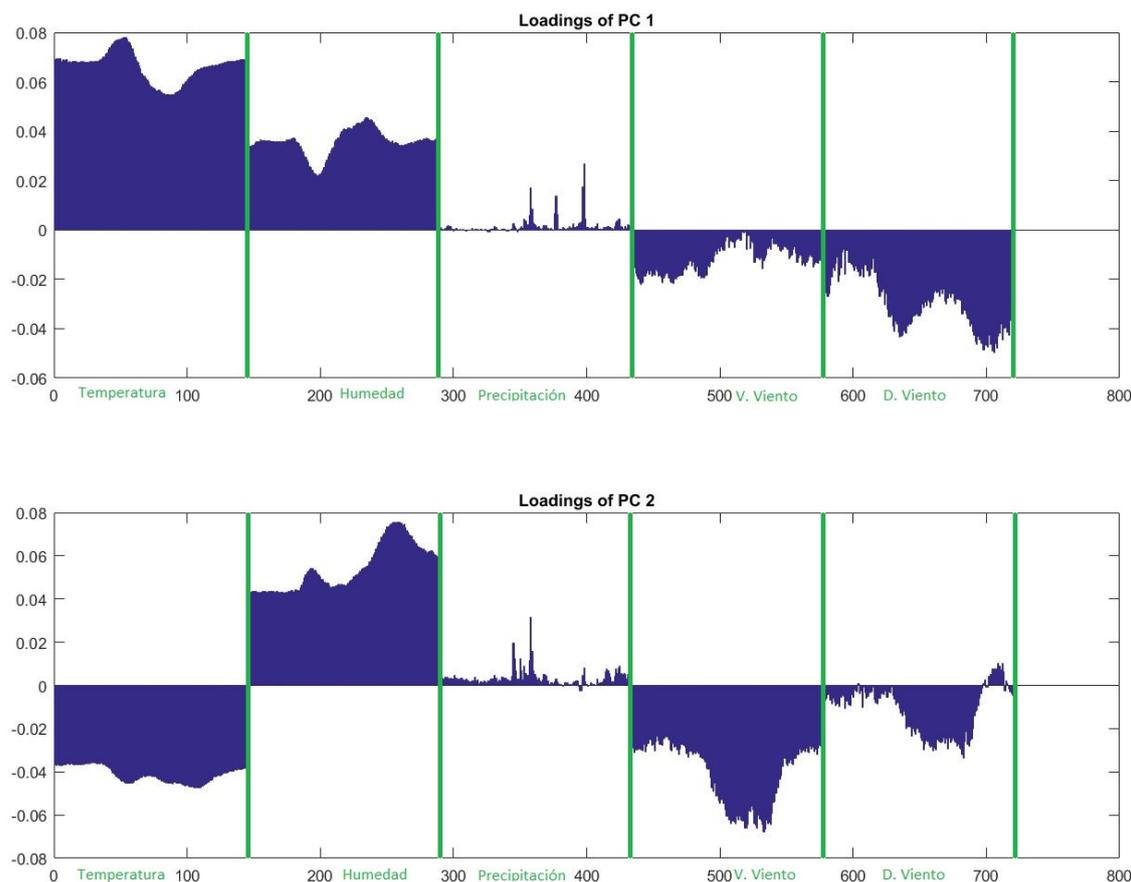


Figura 5.25 Loadings de los componentes principales.

Lo primero que se observa es que todas las estaciones son muy similares. Pero una vez nos fijamos más, lo primero que destaca es la gran diferencia en la PC2 de Aeropuerto, ya que, es la única de las cuatro estaciones que la temperatura es positiva. En la misma estación se puede observar otra diferencia frente a las otras tres, la humedad es negativa en el PC1.

Cabe destacar la gran similitud entre Altocúmulo, UPV y Viveros. Sus gráficas son prácticamente calcadas y esto es debido a su situación en la ciudad ya que son sometidas a las mismas condiciones ambientales. Las únicas diferencias apreciables a simple vista es que la intensidad de la Temperatura en la PC2 de Altocúmulo es menor a la de UPV y Viveros. UPV y Viveros tiene una velocidad del viento cíclica y básicamente igual, en cambio, Viveros presenta una que es siempre negativa.

Aeropuerto es la estación que más se diferencia del resto al estar alejada de núcleos urbanos. La no influencia de núcleos urbanos próximos da lugar a una mayor influencia del viento (velocidad y dirección) sobre el modelo. De forma menor, también se observa un mayor peso de la humedad.

Las cuatro estaciones ponen de manifiesto que la variabilidad de las precipitaciones es inapreciable, de una intensidad muy baja en comparación con las otras variables. En futuros estudios debería considerarse la opción de no incluir la precipitación en el modelo.

6.- CONCLUSIONES

A la vista de los resultados presentados en los puntos anteriores, se puede confirmar que de forma general, los resultados han sido positivos y se ha conseguido cumplir los objetivos de un modo aceptable.

Existen tres tipos de estaciones en este estudio, centro urbano, límite urbano y fuera del núcleo urbano. Esto ha sido constatado en los resultados, apartado 5.5, donde la mayor diferencia entre las estaciones es el viento. Lo cual tiene mucho sentido ya que, aun estando separadas, en el marco de la meteorología están cercanas y prácticamente todas las mediciones deberían ser la misma. Se demuestra que esto no es cierto del todo, ya que, en las mediciones se aprecian diferencias significativas debidas a su situación. La más apartada del núcleo urbano presenta unos vientos más fuertes que las del centro urbano.

Un estudio previo, más sencillo y práctico siempre es recomendable para localizar posibles errores en el preprocesamiento de datos, errores humanos o en las mediciones. En este caso el estudio previo desvelo que la presión era la misma para las estaciones y no suponía ninguna mejora para el modelo.

Se ha podido observar en el apartado 5.5.1 la gran similitud entre las estaciones y sus diferencias. A la vista de estos resultados se debería eliminar la precipitación de futuros estudios.

Dada la naturaleza de los datos de las estaciones ha sido necesaria la realización de código propio en Matlab por él autor para poder adecuar estos datos a las necesidades de las funciones existentes.

Para finalizar, se ha justificado la necesidad de considerar el número de variables a incluir en el estudio en función del efecto de éstas sobre el modelo, ya que no necesariamente una mayor cantidad de ellas aumenta de forma significativa la robustez del mismo y sí lo complica.

7.- ANEXOS

Aquí se anexan todas las funciones utilizadas:

- Interpolarsimple_2: Sirve para calcular medias entre dos filas:

```
function M=interpolarsimple_2(A)

[n,~]=size(A);%obtenemos el número de filas.%Esta variable controla si estamos por debajo
           %o por arriba de n calculando m
a=1;%redefinimos A y usamos el vector N
N=A;% puesto que el nuevo vector es 6/4 veces mayor que el original calculamos
    % la longitud del nuevo vector
d=(6*n/4)-1;%definimos el nuevo vector
M=zeros([d+1,1]);%n es la variable que nos mueve por el vector N
n=1;%m es la variable que nos mueve por el vector M
m=1;
d=d-1;%este es el bucle general
    for i=1:d
        %este primer if nos posiciona, nos indica si estamos en la fila(n)
        %donde M y N son iguales, que esto sucede cuando la fila del vector
        %N es impar
        if rem(n,2)==1
            M(m)=N(n);
            n=n+1;
            m=m+1;
        else
            % cuando n es par hay dos posibles opciones, a=1 o a=2.
            % cuando a=1 significa que estamos calculando por detras de n
            %por ejemplo en las horas estaríamos en n=2 que son 15 min y
            %calculamos por detras, 10 min
            if a==1
                M(m)= N(n)*2/3 + N(n-1)/3;
                a=a+1;
                m=m+1;
            end
            %cuando a=2 estamos calculando por delante de
            %por ejemplo en las horas estaríamos en n=2 que son 15min y
            %calculamos por delante, 20min
            if a==2
                if n==size(N,1)
                    break
                else
                    M(m)= N(n)*2/3 + N(n+1)/3;
                    a=a-1;
                    m=m+1;
                    n=n+1;
                end
                %cuando el bucle pasa por este if pasamos al siguiente n
                %que será impar y todo vuelve a empezar
            end
        end
    end
end
end
end
```

- Organi: Función que organiza la tabla X que corresponde a 5 columnas (las variables) y 52560 tomas de datos en una tabla M de 365x720.

```

]function M=Organi(X)

a=365; %365 dias
b=24*6; %144 intervalos en un dia 24 horas y intervalos de 10 min cada hora.
c=5; %variables.
d=b*c; %720
M=zeros(a,d);

]   for u=1:c %variables
]       for p=1:a %dia son 365
]           for i=1:b %intervalos en un dia hasta 144
]               n=144*(p-1)+i; %esto permite proseguir en la matriz X
]               m=144*(u-1)+i;
]               M(p,m)=X(n,u);
]           end
]       end
]   end

end

```

BIBLIOGRAFÍA

Bro, R. (1997). PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems* , 149-171.

Bro, R., & Smilde, A. (2003). Centring and Scaling in Component. *Journal of Chemometrics* , 16-33.

Gurden, S., Wsterhuis, J., Bro, R., & Smilde, A. (2001). A comparison of multiway regression and scaling methods. *Chemometrics and Intelligent Laboratory Systems* (59), 121-136.

Jackson, J. (1991). *A User's guide to Principal Components*. New York: Wiley.

Prats-Montalbán, J. (2005). *Control Estadístico de Procesos mediante Análisis Multivariantes de Imágenes*. (A. J. Riquelme, Ed.) Valencia.

Wold, H. (1966 a). *Estimation of principal components and related models by iterative least squares*. New York: Academic Press.

Wold, H. (1966 b). *Nonlinear estimation by iterative least squares procedures*. New York: John Wiley.

8.- PRESUPUESTO

Para este proyecto se han tenido en cuenta gastos de personal. El autor del proyecto ha trabajado para el mismo alrededor de 400 horas. Como ingeniero, éste debería cobrar una cantidad de 30€/h, así la suma llegaría a los 6000€ en total. José Manuel Prats-Montalbán se ha considerado como consultor, con un coste de 60€/h. Él a su vez ha trabajado en el proyecto 150h.

El software empleado ha sido el siguiente: Microsoft Office que estaba previamente instalado y Matlab versión de estudiante y del cual los códigos de los programas son propiedad del Grupo de Ingeniería en Estadística Multivariante, Universitat Politècnica de València, Valencia, España.

Por tanto, se han tenido en cuenta dos divisiones del presupuesto:

CUADRO MANO DE OBRA					
Nº	Ud.	Descripción	Cantidad	Precio por unidad(€)	Total (€)
1.1	h	Ingeniero	400,00	30,00	12000,00
1.2	h	Consultor	50,00	60,00	3000,00
1.3	h	Reuniones con tutor	40,00	50,00	2000,00
				Suma total	17000,00

CUADRO MATERIALES					
Nº	Ud.	Descripción	Cantidad	Precio por unidad(€)	Total (€)
2.1	Ud.	Software	2	0	0
				Suma total	0

Finalmente, para calcular el presupuesto final se han considerado un 20% de beneficio industrial, 10% de gastos generales y un 21% de IVA.

Modelización energética de edificios por medio de estadística multivariante

Capítulo	Descripción	Importe (€)
1	Mano de obra	17000,00
2	Materiales	0,00
	Presupuesto de ejecución por material	17000,00
	20% Beneficio Industrial	3400,00
	10% Gastos Generales	1700,00
	Presupuesto de ejecución por contrata	22100,00
	21% IVA	4641,00
	Presupuesto total	26741,00

El coste total que hay que invertir en este proyecto es de DIECINUEVE MIL SEISCIENTOS SESENTA Y DOS EUROS CON CINCUENTA CÉNTIMOS.