



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



ESCUELA TÉCNICA  
SUPERIOR INGENIERÍA  
GEODÉSICA, CARTOGÁFICA  
TOPOGÁFICA

# Máster Universitario en Ingeniería Geomática y Geoinformación

## TRABAJO DE FIN DE MÁSTER

Análisis espacio-temporal de la humedad del  
combustible vivo en la Comunitat Valenciana usando  
métodos estadísticos multivariados.

**Autor:** John Henry Aguilar Albán

**Tutor:** Ángel Antonio Balaguer Beser

**Cotutor:** José Luís Soriano Sancho

*septiembre de 2018*

## **Agradecimientos**

A mi tutor Ángel Antonio Balaguer, por demostrar interés y voluntad, pero sobre todo una gran paciencia, el haberme instruido y brindado sus conocimientos, ayuda y consejo me han permitido salir adelante con el estudio, además me ha mostrado una gran disposición y cooperación, sobre todo en momentos de duda, siempre se mostró muy pendiente y con mucha predisposición por ayudar hasta permitirme lograr la que considero mi meta final en esta etapa de vida.

A mi cotutor José Luís Soriano de igual forma quiero mostrar mis consideraciones, ya que me ha demostrado profesionalismo y voluntad, siempre muy pendiente en el desarrollo del trabajo, aportando con buenas e innovadoras ideas, demostrando un amplio conocimiento en el tema y siempre dispuesta a impartirlo.

Gracias a los dos por haberme permitido ser parte de este proyecto, para mí ha sido todo un honor trabajar bajo la tutela de dos grandes profesionales.

De igual forma quería presentar mi gratitud a dos personas muy especiales, que han sido pilares y se han mostrado muy comprensivos y considerados, brindándome la tranquilidad, cuidado, confianza y ánimo necesario para sacar adelante este proyecto. A mis tíos Oswaldo Aguilar y Carmen Vásconez, gracias.

A Daniela Troya ya que a pesar de la distancia siempre se hizo muy presente, y se mostró pendiente en cada paso.

Un agradeciendo especial a mi gran amigo y colega de clase Sergio Andrade, y como no olvidar a Karla Freire, Paola Hernández, Kari Orquera, Dani Ruano, Juan Pedro Carbonell, por sus ánimos, y familiaridad, haciendo que todo el camino cursado haya sido muy llevadero.

## **Dedicatoria**

Quiero dedicar primeramente a dos personas maravillosas Luis Albán y María Tapia mis abuelos, quien con su constancia y apoyo me han sabido guiar. A mis tías Cristina y Mabel por ser grandes ejemplos, pero principalmente, se los dedico todos ellos por creer en mí.

## RESUMEN

El modelamiento entre las variables meteorológicas y el porcentaje de humedad del combustible vivo busca describir e interpretar su relación, con la finalidad de comprender y predecir el contenido de agua que tiene la vegetación por medio de la interacción de factores externos como son las condiciones climáticas. El presente estudio se centra en la especie vegetal *Rosmarinus officinalis* o también llamada romero, misma que es común de la región mediterránea y catalogada como una especie influyente en la disponibilidad total del combustible vivo, éste al presentar porcentajes bajos de humedad puede ser un potencial material de combustible vivo, que en condiciones adecuadas puede favorecer el inicio y la propagación de los incendios forestales. Los datos utilizados son los obtenidos por la empresa VAERSA para la Generalitat Valenciana, información levantada en diferentes puntos dentro de la Comunitat Valenciana (Soriano, 2016). Calcular el porcentaje de humedad de una especie vegetal partiendo de la toma de muestras en campo pasa por un proceso bastante laborioso, ya que se debe planificar el posicionamiento de cada punto de muestreo en función de varios factores como son: accesibilidad, distribución homogénea de los puntos de acuerdo al área de estudio, presupuesto, tiempo, entre otros. Posterior a la obtención de las muestras, éstas deben ser analizadas en un laboratorio, obteniéndose valores muy puntuales; adicional a esto, el proceso de obtención del porcentaje de humedad de una especie vegetal podría llegar a ser costosa tomando en cuenta que son varias las especies que se deberán analizar. Es así que, evaluar el porcentaje de humedad de la especie vegetal por medio de variables meteorológicas puede ser una herramienta útil para la prevención, gestión y toma de decisiones con respecto a los incendios forestales. Para encontrar dicha relación se procede a realizar análisis estadísticos enfocados en la generación de modelos que estimen el valor de la variable de estudio, además se incursiona en el uso de métodos de agrupaciones para reducir el número de procesos optimizando el tiempo de trabajo. De igual forma, se realiza el modelado de superficie de estimación utilizando el método de interpolación kriging como herramientas y técnica alterna.

Palabras claves: variables meteorológicas, incendios forestales, análisis estadísticos, humedad del combustible vivo, kriging

## **ABSTRACT**

Modeling between the meteorological variables and the living fuel moisture percentage seeks to describe and interpret their relationship to understand and predict the vegetation percentage of water through the interaction of external factors such as weather conditions. The present study puts its effort on the specie *Rosmarinus officinalis* or also called Romero. Romero is distributed along the Mediterranean Region and cataloged as an important specie related with live fuel total availability. Romero presents low levels of moisture, a characteristic that catalogs it as a potential material of live fuel that in appropriate conditions can favor the initiation and propagation of forest fires. The data used are those obtained by the company VAERSA for the Generalitat Valenciana, information raised at different places within the Valencian Community (Soriano, 2016). The calculation of the moisture percentage of a plant species starting from the sampling in the field goes through a rather laborious process. Initially, the positioning of the sampling points should be planned according to several factors such as: accessibility, homogeneous distribution of the points according to the study area, budget, time, among others. After obtaining the samples, they must be analyzed in a laboratory, obtaining very specific values; In addition to this, the process of obtaining the percentage of moisture of a plant species could become expensive considering that there are several species that should be analyzed. Thus, assessing the moisture content of the plant specie by meteorological variables can be a useful tool for prevention, management and decision making regarding forest fires. To find this relationship, statistical analyzes were performed focused on the generation of models that estimate the value of the study variable, in addition to the use of grouping methods to reduce the number of processes optimizing work time. In the same way, estimation surface modeling is performed using the kriging interpolation method as tools and alternative technique.

Keywords: meteorological variables, forest fires, statistical analysis, live fuel moisture, kriging

## CONTENIDO

1. DATOS GENERALES.....	1
1.1 Introducción .....	1
1.2 Objetivos .....	4
1.2.1 Objetivo general .....	4
1.2.2 Objetivos específicos.....	4
1.3 Ubicación de la zona de estudio .....	5
1.4 Datos .....	12
1.4.1 Humedad del combustible vivo .....	12
1.4.2 Humedad del combustible vivo especie vegetal romero .....	13
1.4.3 Estaciones meteorológicas.....	13
1.4.3.1 Cálculo de variables meteorológicas.....	18
1.5 Diagrama de flujo de trabajo .....	21
2. ANÁLISIS ESTADÍSTICO DE LA VARIABLE HUMEDAD .....	22
2.1 Análisis descriptivo de la variable humedad de la especie romero por mes. ....	29
2.2 Análisis descriptivo de la variable humedad de la especie romero por estación del año .....	35
3. ANÁLISIS DE CONGLOMERADOS.....	39
3.1 Desarrollo del análisis de conglomerados.....	40
3.1.1 Identificación de número de conglomerados óptimos.....	43
3.1.2 Análisis de los conglomerados con las diferentes variables. ....	45
3.1.2.1 Análisis con los datos de humedad de la especie vegetal romero por estación del año. .	45
3.1.2.2 Análisis con los estadísticos de humedad de la especie vegetal romero por mes. ....	53
3.2 Valoración del porcentaje de humedad en función de los percentiles .....	59
3.2.1 Análisis de conglomerados por estación del año y cálculo de percentiles.....	62
4. ANÁLISIS DE REGRESIÓN LINEAL MÚLTIPLE .....	65
4.1 Variables utilizadas en el análisis de regresión lineal múltiple .....	67
4.2 Diagrama de flujo del análisis de regresión lineal múltiple .....	68
4.3 Desarrollo del análisis de regresión lineal múltiple .....	70
4.3.1 Selección del modelo de regresión lineal múltiple.....	70
4.3.1.1 Selección de mejores modelos por grupos de variables .....	70
4.3.1.2 Modelos de regresión con todas las variables. ....	74
4.3.1.3 Modelos de regresión por grupo de variables. ....	76

5. INTERPOLACIÓN ESPACIAL .....	78
5.1 Análisis descriptivo de los datos de mayo del 2018.....	79
5.2 Semivariograma .....	83
5.2.1 Semivariograma experimental .....	84
5.2.2 Semivariograma teórico.....	86
5.2.3 Kriging .....	87
5.2.3.1 Mapa de la superficie de estimación del porcentaje de humedad final.....	91
5.2.3.2 Estimaciones con puntos externos .....	93
6. RESULTADOS .....	95
6.1 Análisis de conglomerados .....	95
6.2 Análisis de percentiles como clasificadores.....	105
6.3 Análisis de regresión lineal múltiple.....	113
6.4 Interpolación espacial.....	125
7. CONCLUSIONES .....	133
7.1 Análisis de conglomerados .....	134
7.2 Análisis de percentiles como clasificadores.....	135
7.3 Análisis de regresión lineal múltiple.....	136
7.4 Interpolación espacial.....	137
8. BIBLIOGRAFIA.....	137
9. ANEXOS.....	140

## LISTADO DE FIGURAS

Figura 1.3-1 Porcentaje de área clasificado de la cobertura vegetal de la Comunidad Valenciana. ....	5
Figura 1.3-2 Zona de estudio y ubicación de los puntos de muestreo. ....	7
Figura 1.3-3 Cobertura vegetal de la Comunidad Valenciana. ....	8
Figura 1.3-4 Subtipos Fitoclimáticos de la Comunidad Valenciana. ....	11
Figura 1.4.3-1 Ubicación de Estaciones Meteorológica en la Comunidad Valenciana. ....	15
Figura 1.4.3-2 Selección de Estaciones Meteorológica por el área de influencia. ....	16
Figura 1.4.3-3 Distribución de las estaciones meteorológica seleccionadas. ....	18
Figura 2-1 Humedad del combustible vivo de la especie Romero. ....	24
Figura 2-2 Humedad del combustible vivo de la especie Romero por grupo de puntos de muestreo individual. ....	27
Figura 2-2 Humedad del combustible vivo de la especie Romero por grupo de puntos de muestreo individual. ....	28
Figura 2.1-1 Promedio por punto de muestreo en función de la variable humedad de la especie Romero. ....	32
Figura 2.1-2 Mediana por punto de muestreo en función de la variable humedad de la especie Romero. ....	33
Figura 2.1-3 Cuartil inferior (Q25%) de los punto de muestreo en función de la variable humedad de la especie Romero. ....	33
Figura 2.1-4 Cuartil superior (Q75%) de los punto de muestreo en función de la variable humedad de la especie Romero. ....	34
Figura 2.1-5 Diagrama de caja y bigotes de los punto de muestreo en función de la variable humedad de la especie romero. ....	34
Figura 2.2-1 Humedad del combustible vivo de la especie Romero por estaciones del año. ....	36
Figura 2.2-2 Humedad del combustible vivo de la especie Romero por estación del año y por grupos de punto de muestreo. ....	36
Figura 2.2-4 Diagrama de caja y bigotes de los punto de muestreo en función de la variable humedad de la especie romero por estación del año. ....	39
Figura 3-1 Diagrama de los diferentes análisis a realizar y las respectivas variables a utilizar. ....	40
Figura 3.1-1 Diagrama de procesos en el análisis de conglomerados. ....	41
Figura 3.1-2 Dendograma con los datos de HCV de la especie vegetal romero. ....	42
Figura 3.1.1-1 Dendograma presentando 4 y 5 conglomerados. ....	44
Figura 3.1.2.1-1 Dendograma de los análisis realizados con los datos por estación de año. ....	50
Figura 3.1.2.1-2 Dendograma de los análisis realizados con los estadísticos de posición agrupados. ...	50
Figura 3.1.2.1-3 Dendograma de los análisis realizados con los estadísticos de forma individual. ....	51
Figura 3.1.2.2-1 Dendograma de los análisis realizados con los estadísticos de posición agrupados. ...	57
Figura 3.1.2.2-2 Dendograma de los análisis realizados con los estadísticos de forma individual. ....	57

Figura 3.2-1 Ubicación puntos de muestreo nuevos (mayo 2018).....	59
Figura 3.2-2 Análisis de conglomerados por estación del año. ....	61
Figura 3.2-3 Percentiles por clase.....	62
Figura 3.2.1-1 Dendograma por estaciones del año.....	63
Figura 4-1 Comparación entre los tests F y t. ....	66
Figura 4.1-1 Puntos de muestreo con disposición de datos meteorológicos completos. ....	68
Figura 4.2-1 Diagrama de flujo de trabajo.....	69
Figura 4.3.1.1-1 Gráfica de R-cuadrada ajustada por el porcentaje de humedad de la especie vegetal romero (Ro).....	71
Figura 4.3.1.1-2 Gráfica de R-cuadrada ajustada por el porcentaje de humedad de la especie vegetal romero (Ro).....	73
Figura 5-1 Diagrama de flujo de trabajo.....	79
Figura 5.1-1 Histograma .....	81
Figura 5.1-2 Probabilidad Normal.....	81
Figura 5.1-3 Gráfico de caja y bigotes.....	81
Figura 5.1-4 Gráfico de dispersión .....	82
Figura 5.1-5 Mapas de Voronoi con datos de porcentaje de humedad .....	82
Figura 5.2-1 Nube del semivariograma .....	83
Figura 5.2.1-1 Semivariograma omnidireccional .....	84
Figura 5.2.1-2 Semivariograma con diferentes orientaciones.....	85
Figura 5.2.1-3 Semivariograma orientados.....	85
Figura 5.2.2-1 Semivariogramas teóricos.....	86
Figura 5.2.3-1 Tipos de kriging .....	87
Figura 5.2.3-2 Superficie de estimación (método krigeado ordinario) .....	88
Figura 5.2.3-3 Mapa de estimación del porcentaje de humedad del romero (superficie generada con el kriging ordinario) .....	89
Figura 5.2.3-4 Representaciones de los parámetros básicos en un semivariograma. ....	90
Figura 5.2.3.1-1 Gráfico de estimación .....	92
Figura 5.2.3.1-2 Superficie de estimación del porcentaje de humedad del romero.....	92
Figura 5.2.3.1-3 Error estándar de estimación del porcentaje de humedad del romero.....	93
Figura 5.2.3.2-1 Ubicación de los puntos externos.....	94
Figura 5.2.3.2-2. Valor del punto estimado con rango de oscilación respecto a los datos de registros, punto de muestreo Ayora. ....	95
Figura 6.1-1 Dendograma con los datos de humedad de la especie vegetal romero por estación del año. ....	98
Figura 6.1-2 Diagrama lineal de tiempo por puntos de muestreo agrupados en base al promedio por	

estación del año .....	99
Figura 6.1-3 Diagrama lineal de tiempo unión conglomerado 1 y 4.....	101
Figura 6.1-4 Diagrama lineales del conglomerado uno con los datos de humedad calcados por estación del año y meses. ....	102
Figura 6.1-5 Puntos de muestreo distribuidos en los 5 conglomerados. ....	104
Figura 6.2-1 División de conglomerados por datos primavera y puntos de muestreo nuevos.....	105
Figura 6.2-2 Diagrama de dispersión con la categorización de los puntos de muestreo del conglomerado 1. ....	109
Figura 6.2-3 Diagrama de dispersión con la categorización de los puntos de muestreo del conglomerado 2. ....	110
Figura 6.2-4 Diagrama de dispersión con la categorización de los puntos de muestreo del conglomerado 3. ....	111
Figura 6.2-5 Diagrama de dispersión con la categorización de los puntos de muestreo del conglomerado 4. ....	111
Figura 6.2-6 Diagrama de dispersión con la categorización de los puntos de muestreo del conglomerado 5. ....	112
Figura 6.2-7 Especialización de los puntos de muestreos categorizados en la provincia de Valencia. .	113
Figura 6.4-1. Puntos de muestreo con respecto la superficie forestal. ....	126
Figura 6.4-2 Mapa de la superficie de estimación del porcentaje de humedad de la especie vegetal romero en la provincia de Valencia (mayo 2018).....	127
Figura 6.4-3 Mapa de la superficie de estimación del porcentaje de humedad de la especie vegetal romero en la provincia de Valencia de acuerdo a la cobertura forestal (mayo 2018). ....	128
Figura 6.4-4 Mapa de superficie del error estándar de estimación del porcentaje de humedad de la especie vegetal romero.....	129
Figura 6.4-5 Gráfico del valor del punto muestreo estimado con respecto a los datos registrados den 3 años anteriores.....	130

## LISTADO DE TABLAS

Tabla 1.3-1 Clasificación de la cobertura vegetal de la Comunidad Valenciana. ....	5
Tabla 1.3-2 Clasificación de los puntos de muestreo del HCV según el subtipo fitoclimáticos. ....	10
Tabla 1.3-3 Vegetación según los subtipos fitoclimáticos correspondientes a los puntos de muestreo. .	10
Tabla 1.4.3-1 Estaciones meteorológicas seleccionadas. ....	17
Tabla 1.4.3.1-1 Variables meteorológicas calculadas.....	20
Tabla 2-1 Estadísticos humedad del combustible vivo de la especie Romero.....	25
Tabla 2.1-1 Estadísticos humedad del combustible vivo de la especie Romero periodo de 3 años. ....	31
Tabla 2.2-1 Datos del HCV por estaciones del año. ....	35
Tabla 2.2-2 Datos del HCV por estaciones del año en un periodo de 2015 - 2017. ....	38
Tabla 3.1-1 Historial de conglomerados. ....	42
Tabla 3.1.1-1 diferencia entre número de conglomerados. ....	43
Tabla 3.1.2.1-1 Conglomerados de 4 clases de los diferentes análisis con los datos de HCV de la especie vegetal romero por estación del año. ....	46
Tabla 3.1.2.1-2 Conglomerados de 5 clases de los diferentes análisis con los datos de HCV de la especie vegetal romero por estación del año. ....	47
Tabla 3.1.2.1-3 Puntos de muestreo agrupados en los conglomerados de 4 clases. ....	48
Tabla 3.1.2.1-4 Puntos de muestreo agrupados en los conglomerados de 5 clases. ....	49
Tabla 3.1.2.2-1 Conglomerados de 4 clases de los diferentes análisis con los datos de HCV de la especie vegetal romero por mes. ....	54
Tabla 3.1.2.2-2 Conglomerados de 5 clases de los diferentes análisis con los datos de HCV de la especie vegetal romero por mes. ....	54
Tabla 3.1.2.2-3 Puntos de muestreo agrupados en los conglomerados de 4 clases. ....	55
Tabla 3.1.2.2-4 Puntos de muestreo agrupados en los conglomerados de 5 clases. ....	56
Tabla 3.2.1-1 Puntos de muestreo agrupados en los conglomerados de 5 clases por estación del año. .	64
Tabla 3.2.1-2 Percentiles por conglomerado de cada estación del año. ....	65
Tabla 4.1-1 Datos generales puntos de muestreo. ....	67
Tabla 4.3.1.1-1 Selección de modelos ....	71
Tabla 4.3.1.1-2 Modelos destacados con variables del grupo de precipitación. ....	72
Tabla 4.3.1.1-3 Asignación de letra a cada variable dentro de cada grupo. ....	72
Tabla 4.3.1.1-4 Selección de modelos por grupo. ....	72
Tabla 4.3.1.1-5 Modelos destacados con variables de cada grupo. ....	74
Tabla 4.3.1.2-1 Informe del modelo destacado con variables de cada grupo. ....	75
Tabla 4.3.1.3-1 Informe de los modelos por grupo. ....	76
Tabla 4.3.1.3-2 Informe de los modelos por grupo. ....	77

Tabla 4.3.1.3-3 Informe del modelo final.....	77
Tabla 5.1-1 Resumen Estadístico.....	80
Tabla 5.2.3-1 Modelos realizados.....	90
Tabla 5.2.3.1-1 Parámetros del modelo final y los errores de estimación.....	91
Tabla 5.2.3.2-1 Comparativa entre la estimación de HCV obtenida por krigado para mayo de 2018 y los valores observados ese mes en años anteriores.....	94
Tabla 6.2-1 Distribución y clasificación de los puntos de muestreo (mayo 2018).....	106
Tabla 6.2-2 Categorización por conglomerado.....	107
Tabla 6.2-3 Categorización de los puntos de muestreo del conglomerado 1.....	107
Tabla 6.2-4 Categorización de los puntos de muestreo del conglomerado 2.....	108
Tabla 6.2-5 Categorización de los puntos de muestreo del conglomerado 3.....	108
Tabla 6.2-6 Categorización de los puntos de muestreo del conglomerado 4.....	109
Tabla 6.2-7 Categorización de los puntos de muestreo del conglomerado 5.....	109
Tabla 6.3-1 Informe de los modelos con todos los puntos de muestreo.....	114
Tabla 6.3-2 Selección de variables por evaluación.....	115
Tabla 6.3-3 Informe del modelo con las variables seleccionadas por evaluación.....	116
Tabla 6.3-4 Informe del modelo final filtrado por grupo.....	117
Tabla 6.3-5 Informe del modelo final filtrado por grupo.....	118
Tabla 6.3-6 Informe del modelo final filtrado por grupo.....	119
Tabla 6.3-7 Extracto del informe del modelo por evaluación.....	120
Tabla 6.3-8 Selección de variables por evaluación.....	121
Tabla 6.3-9 Extracto del informe del modelo final filtrado por grupo.....	121
Tabla 6.3-10 Selección de variables por evaluación.....	122
Tabla 6.3-11 Informe del modelo final filtrado por grupo.....	122
Tabla 6.3-12 Selección de variables por evaluación.....	123
Tabla 6.3-13 Informe modelo modificado.....	124
Tabla 6.3-14 Extracto del informe del modelo final por conglomerado.....	124

# 1. DATOS GENERALES

## 1.1 Introducción

Según la FAO (2015) anualmente los incendios forestales afectan 350 millones de hectáreas de zonas forestales en todo el mundo, siendo el principal causante el hombre y sus actividades que van relacionadas y ligadas a aspectos socioeconómicos, actividades que en promedio, afectan al 1% de todo el bosque mundial cada año, causando pérdidas humanas, daños a la propiedad, pérdida de biodiversidad, desertificación, entre otros.

Los incendios forestales suponen un problema ambiental de gran trascendencia, tanto a escalas locales como globales, al afectar a la dinámica de la vegetación y las emisiones atmosféricas de muy diversos ecosistemas. Hallazgos arqueológicos y etnológicos sustentan la teoría de que el fuego ha sido utilizado por el hombre para modificar su ambiente geográfico. La quema de vegetación se ha utilizado en algunos sistemas socioeconómicos con fines agrícolas, ganaderos y forestales teniendo ciertas aplicaciones como la eliminación de los restos de cosechas anteriores y preparación del suelo para nuevas siembras esto con respecto a la agricultura; en la ganadería para la regeneración de nuevos pastos y creación de zonas abiertas al ganado y en lo forestal para el mantenimiento de zonas de cultivo forestal (Chuvieco & Pilar, 2004)

Así como existen ciertos ecosistemas que dependen del fuego para su regeneración, también se puede dar incendios intencionados vinculados a conflicto de intereses, acción de pirómanos o venganzas.

El fuego como tal, es una reacción química, que necesita de tres componentes básicos aire para la contribución de oxígeno, combustible para arder y fuente de calor para elevar el combustible a la temperatura de ignición (Denham, 2009). En referencia a los incendios forestales, se definen:

- **Combustible:** toda la materia orgánica muerta y seca que se encuentra acumulada en el suelo natural.
- **Calor:** Fuente de energía calorífica natural o artificial que intervine en este proceso

dando paso a un punto de ignición.

- **Comburente:** Sustancia que interviene en el proceso de combustión, en este caso es el oxígeno.

El proceso de combustión empieza cuando el material combustible tiene las características adecuadas para que una fuente de calor natural o artificial entre en un proceso de combustión, haciendo que el combustible esté en punto de ignición y sea avivado por el oxígeno, mismo que tiene que ser suficiente para dar el inicio a un incendio forestal (Blanco, García, Castellnou, Molina, & Grillo, 2008)

Es importante recalcar que el comportamiento de los incendios forestales está ligado a las condiciones o factores extremos del lugar, mismos que pueden ser físicos y ambientales, es decir meteorológicos, topográficos y al tipo de vegetación predominante de la zona.

Para sofocar un incendio, se debe reducir como mínimo uno de los tres elementos es por eso que en este estudio se enfatiza en la disponibilidad del combustible ya que desacuerdo a sus características contribuye para determinar el proceso de combustión, la inflamabilidad tanto como la combustibilidad del material contribuyen en la facilidad y continuidad de la propagación del fuego. Dentro de la fase de combustión, en un principio, se produce la deshidratación del material evaporando el contenido de agua que puede poseer el mismo, contribuyendo para su propagación o en el caso de que la fuente de calor no sea la suficiente para evaporar el agua contenida se produce la extinción, por lo que el contenido de humedad determina el tiempo de inflamación y si este es mayor que la energía para consumir y evaporar el agua la combustión no puede producirse (Soriano, 2016)

La humedad del combustible se traduce como el contenido de agua que posee la vegetación constituyendo uno de los principales factores para el desarrollo y comportamiento del fuego, sin embargo esta es modificado por factores externos como son las condiciones meteorológicas, morfológicas y propiedades del suelo de la zona, como también las características de la especie y su facilidad de retención de agua.

El combustible se puede presentar de dos formas: el material que yace muerto y se acumula en el suelo, el cual estará condicionada más por los factores externos físicos y ambientales; el material vivo que de acuerdo a la iteración con el ambiente su contenido de humedad varía.

El estudio presente se centra en analizar la humedad del material vivo en la especie

*Rosmarinus officinalis* o también denominado romero en la Comunitat Valenciana ya que al ser especie común de la zona se encuentra en proporciones altas y muy bien distribuidas en toda el área.

La Comunitat Valenciana presenta peligro de incendios en todo el año, sin embargo este se intensifica en el verano ya que presenta altas temperaturas y pocas lluvias. Los incendios forestales constituyen un grave problema en el arco mediterráneo, del que forma parte la Comunidad Valenciana, creando en ocasiones situaciones de grave riesgo y catástrofe con las consecuentes repercusiones negativas sobre el medio ambiente. Según la página oficial de la Generalitat Valenciana se han desarrollado planes de acción especial frente al Riesgo de Incendios Forestales (Decreto 163/1998), que contempla tanto las actuaciones en emergencias producidas por los incendios como las actuaciones en situaciones de preemergencia (Plan de Vigilancia Preventiva). Si a pesar de las medidas preventivas se produce el incendio forestal, se aplica lo contemplado en el Plan Especial frente al Riesgo de Incendios Forestales, movilizándose de inmediato los recursos contemplados en el mismo. (112 Comunitat Valenciana, s.f.)

Además la Conselleria de Agricultura, Medio Ambiente, Cambio Climático y Desarrollo Rural de la Generalitat Valenciana ha lanzado la campaña *Stop al foc bajo el lema 'tu actitud puede marcar la diferencia'* para concienciar a los ciudadanos y ayudar a prevenir posibles incendios que en gran medida se producen por negligencias o descuidos.

## 1.2 Objetivos

### 1.2.1 Objetivo general

Realizar un estudio espacio-temporal de la humedad del combustible vivo de la especie vegetal *Rosmarinus officinalis* dentro de la Comunitat Valenciana, aplicando técnicas de análisis multivariante y geoestadística para la creación de modelos predictivos, para el estudio espacial se aplicará el análisis cluster y el krigado. Para el análisis temporal se hará uso de la regresión múltiple en cada punto de muestreo.

### 1.2.2 Objetivos específicos

- Agrupar los puntos de muestreo por afinidad en función de la variable porcentaje de humedad del combustible vivo, mediante el análisis cluster buscando homogeneidad en el número de puntos por conglomerado.
- Aplicar el análisis de regresión lineal múltiple con la finalidad de encontrar la relación existente entre el porcentaje de humedad del combustible vivo de la especie vegetal *Rosmarinus officinalis* y variables meteorológicas.
- Realizar una clasificación cualitativa basada en los percentiles calculado por agrupaciones de puntos de muestreo afines, en función del porcentaje de humedad del combustible vivo especie vegetal *Rosmarinus officinalis*.
- Clasificar datos actuales de porcentaje de humedad tomados en diferentes puntos de la Comunitat Valenciana, de acuerdo a la clasificación obtenida.
- Generar un mapa de estimación del porcentaje de humedad de la especie vegetal *Rosmarinus officinalis* en la provincia de Valencia, usando datos en un solo mes y con una mayor densificación de puntos de muestreo por medio del método de interpolación kriging y modelos de semivariograma.

### 1.3 Ubicación de la zona de estudio

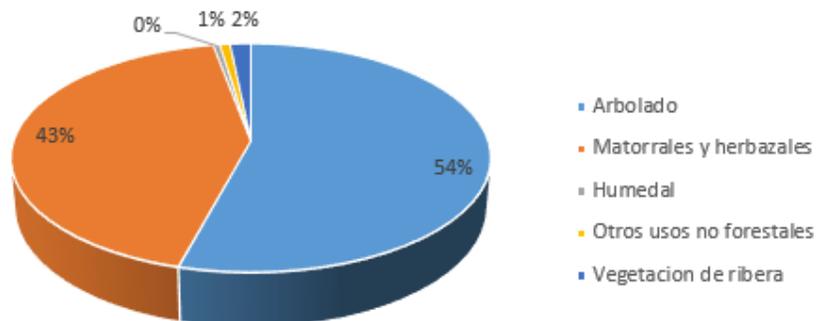
El análisis se lo realizó en la Comunitat Valenciana, localizada al este y sur-este de la Península Ibérica, según los datos obtenidos por el Instituto Cartográfico Valenciano su área total es de 23 266 km<sup>2</sup>, de la cual la cobertura vegetal ocupa 13 195 km<sup>2</sup>, se realizó una clasificación de la cobertura vegetal en base a los ecosistemas determinados en el Plan de Acción Territorial Forestal de la Comunitat Valenciana (PATFOR, 2009), englobándolos en 4 grupos tal como se muestra en la tabla 1.3-1, en donde arbolado ocupa la mayor área con un porcentaje del 54% seguido de matorrales y herbazales con un 43% (ver figura 1.3-1), el resto engloba el 4% del área, dichas clasificaciones a pesar de ser minoritarias por sus características fueron determinadas como una clase por sus característica peculiares .

Tabla 1.3-1 Clasificación de la cobertura vegetal de la Comunidad Valenciana.

Ecosistemas	Área (Km <sup>2</sup> )	Puntos de muestreo
Arbolado	7152.9	Buñol, Castielfabid, Gilet, Villargordo, Biar, Altea, Montanejos, Pobla-Tonesa, Albaida-Muro, Ayora, Villar-Arzobispo
Matorrales y herbazales	5657.1	Chelva, Navarrés, Fredes, Gandia, Llombai
Humedal	67.4	
Otros usos no forestales	105.5	
Vegetación de ribera	212.3	
<b>ÁREA TOTAL</b>	<b>13195.2</b>	

Fuente: PATFOR (2009)

Figura 1.3-1 Porcentaje de área clasificado de la cobertura vegetal de la Comunidad Valenciana.



Fuente: PATFOR (2009)

Cabe destacar que el análisis espacio-temporal de la humedad del combustible vivo (HCV) de la especie vegetal romero en la Comunidad Valenciana se hace a partir de un estudio previo realizado por la empresa Valenciana de Aprovechamiento Energético de Residuos, SA (VAERSA) perteneciente a la Conselleria de Agricultura, Medio Ambiente, Cambio Climático y Desarrollo Rural y el trabajo final de master denominado Análisis de la Humedad del Combustible Vivo (HCV) en la Comunitat Valenciana y la influencia de las variables meteorológicas (Soriano, 2016). De igual manera como Cotutor de este trabajo se cuenta con la ayuda y conocimiento de José Luis Soriano.

Para la obtención de muestras del combustible se realizó la identificación de puntos en diferentes zonas de la Comunidad Valenciana, mismas que deben cumplir condiciones muy específicas como son:

- Áreas de muestreo que contengan variedad de especies vegetales endémicas de la comunidad valenciana en lo que respecta a matorrales, herbazales y arbolado.
- El grupo de puntos de muestreo deben representar a toda la zona de análisis, es decir a la cobertura vegetal existente en la Comunidad Valenciana.
- Se consideraron aspectos climatológicos, topográficos, accesibilidad, entre otros.

Tomando dichas premisas se localizaron 20 puntos de muestreo de los cuales en 16 se toman muestras de la especie vegetal a analizar, estaciones que son mostradas en la figura 1.3-2. Cabe destacar que el apartado 1.4 se describen en un principio 18 puntos de muestreo pero se descartaron 2, ya que la cantidad de datos estimados en dichos puntos eran insuficientes, por lo que a continuación se muestran los puntos finales con los que se trabajó coincidente con los determinados en el trabajo de fin de master nombrado anteriormente.

Figura 1.3-2 Zona de estudio y ubicación de los puntos de muestreo.



Fuente: Cartografía del ICV y VAERSA

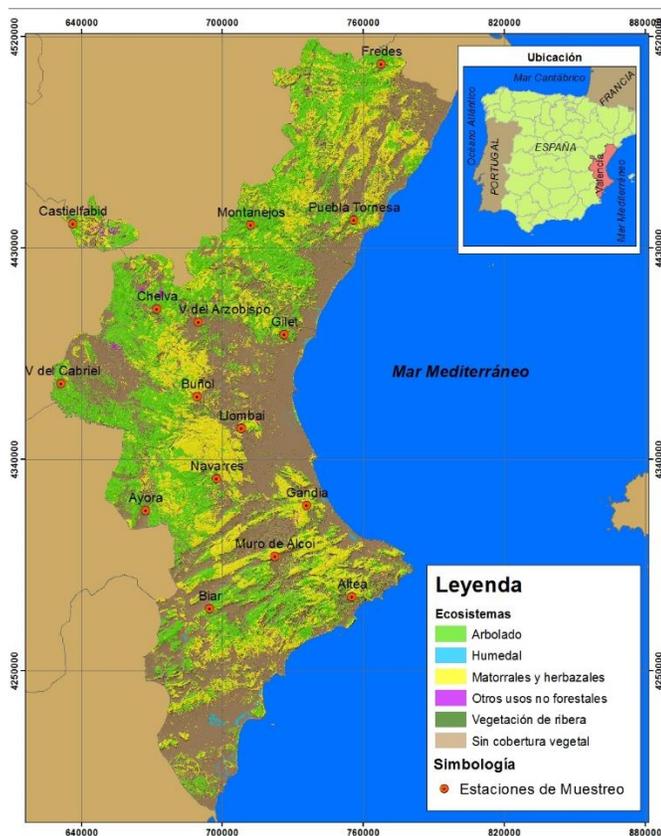
De acuerdo al Plan de Acción Territorial Forestal de la Comunitat Valenciana (PATFOR, 2009), el cual actúa como un instrumento de planificación territorial y forestal, ofreciendo herramientas para su gestión con la finalidad de que todos los estudios o análisis independientes tomen en cuenta el modelo territorial y las actuaciones futuras de la Conselleria de Medio ambiente. Dicho plan se toma en cuenta ya que de acuerdo al tipo de cobertura vegetal se ha tomado las muestras, ubicando a los puntos de muestreo en zonas con cobertura vegetal arbolado, matorral y herbazales siendo el 97% del área de la cobertura vegetal de la Comunidad Valenciana, cabe destacar que la zona forestal en la superficie de la Comunidad Valenciana ocupa un área total del 57% por lo que un poco más del territorio es propensa a desarrollar un incendio ya que cuenta con los elementos suficientes para dicho proceso.

En la Comunidad Valenciana se destaca el bioma perteneciente a zonas templadas que posee un clima mediterráneo presentando precipitaciones no muy abundantes y una estación seca muy marcada, la vegetación predominante son los matorrales, árboles y arbustos del tipo

perennifolio esclerófilo es decir adaptadas a largos períodos de sequía y calor. Como se explica en el apartado 2.3.2 del PATFOR (2009), el clima de la Comunidad Valenciana “Se caracteriza por las temperaturas suaves y la sequía estival, causada por la situación que ocupa en la parte oriental de la Península Ibérica al resguardo de los frentes atlánticos. La cercanía al mar, que regula las temperaturas y aporta humedad, el relieve y la orientación del mismo respecto a la costa, son los factores que, dentro del clima mediterráneo explican las diferencias que aparecen dentro del territorio, con rasgos de aridez.”

En la figura 1.3-3 se muestra la cobertura vegetal perteneciente a la Comunidad Valenciana y la distribución de los puntos de muestreo de los cuales 11 se encuentran en zonas con ecosistemas denominado arbolado y los 5 puntos restantes en un ecosistema de matorrales y herbazales.

Figura 1.3-3 Cobertura vegetal de la Comunidad Valenciana.



Fuente: Cartografía del ICV, VAERSA y PATFOR

Al realizar la clasificación de los puntos de muestreo por clima en relación al tipo de cobertura vegetal existente, se utilizan los subtipos fitoclimáticos determinados en el PATFOR realizado en enero del 2009, mismos que fueron basados en la clasificación de Allue Andrade, 1990, para la Península Ibérica y en las islas Baleares en donde se determinaron 22 subtipos fitoclimáticos de los cuales en la Comunidad Valenciana aparecen 11, en la tabla 1.3-2 se observa los puntos de muestreo y su clasificación según los subtipos fitoclimáticos, donde se puede apreciar 6 subtipos.

Los tipos bioclimáticos donde se realizaron los puntos de muestreo son los denominados Bosques mediterráneos y caducifolios nemorales, mismos que según la caracterización realizada por Allue (1990) son propios de la Península Ibérica.

La determinación de los subtipos fue realizado analizando la vegetación de cada división fitoclimáticos como son los rasgos botánicos, geomorfológicos y edáficos semejantes, además se tuvo en cuenta características del clima en cada zona misma que es representada con un número romano entre paréntesis indicando que hay una relación o rasgos comunes de clima, en este caso se pueden visualizar 2 relaciones con código IV(VI) y VI(IV) lo cual indica que la zona con código IV y VI tienen rasgos del clima VI y IV respectivamente, es decir que la zona pertenecientes a la clasificación como bosque mediterráneo tiene rasgos del clima de la zona perteneciente al bosque caducifolio nemoral y la otra relación ocurre a la inversa, los subíndices indican ser subtipos de un mismo clima.

Tabla 1.3-2 Clasificación de los puntos de muestreo del HCV según el subtipo fitoclimáticos.

Subtipo Fitoclimáticos		Código	Nombre punto de muestreo	
PATFOR (2009)	Allue (1990)			
Mediterráneo extrailicino o ilicino	Mediterráneo genuino	IV2	Llombai	
			Altea	
Mediterráneo ilicino típico más seco		IV3	Ayora	
			Biar	
Mediterráneo ilicino típico menos seco		IV4		Albaida-Muro
				Buñol
				Gilet
			Navarrés	
	Pobla-Tonesa			
Mediterráneo transicional hacia planicaducifolia meseteño	Mediterráneo sub nemoral	IV(VI)1	Villargordo	
Mediterráneo transicional hacia planicaducifolia thetyco		IV(VI)2	Villar-Arzobispo	
			Gandia	
Nemoromediterráneo subesclerófilo	Nemomediterráneo genuino	VI(IV)1	Castielfabid	
			Chelva	
			Montanejos	
			Fredes	

Fuente: PATFOR (2009) y Allue Andrade (1990)

Por último se identifican las formaciones vegetales que corresponden a dichos subtipos tal como se muestra en la tabla 1.3-3

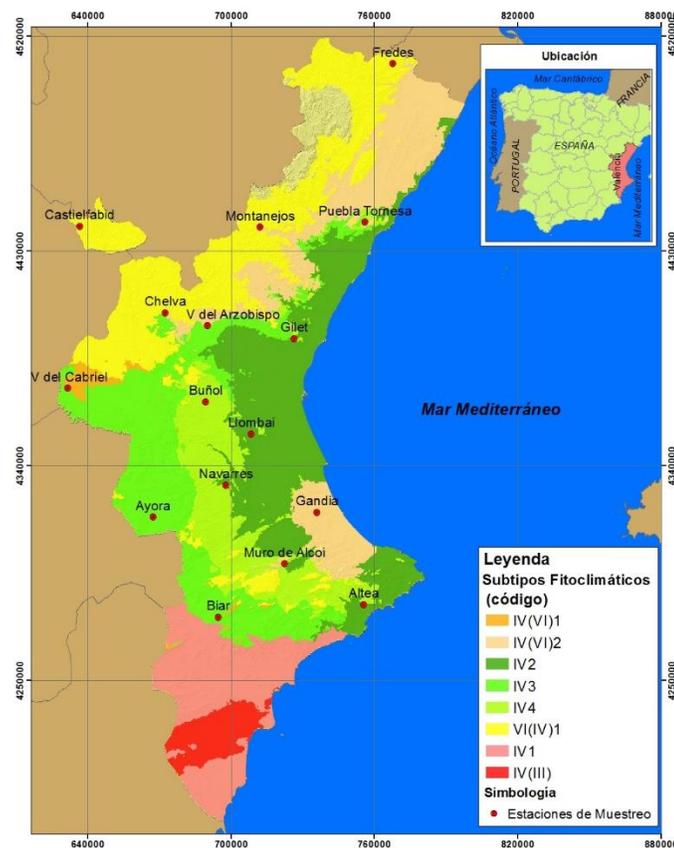
Tabla 1.3-3 Vegetación según los subtipos fitoclimáticos correspondientes a los puntos de muestreo.

VEGETACIÓN	CÓDIGO SUBTIPO FITOCLIMÁTICOS					
	IV2	IV3	IV4	IV(VI)1	IV(VI)2	VI(IV)1
Lentiscales	X	X				
Coscojales	X	X	X	X		X
Acebuchales	X		X			
Encinares	X	X	X	X	X	X
Alsinares	X	X	X	X	X	X
Melojares			X	X	X	X
Quejigares				X	X	X
Robledales pubescentes					X	
Robledales pedunculados					X	

Fuente: Allue Andrade, 1990

En la figura 1.3-4 se muestran los subtipos fitoclimáticos de la comunidad valenciana en relación con los puntos de muestreo, además cabe destacar que en la provincia de Alicante hay dos puntos de muestreo ubicadas al norte de la misma, esto se da ya que el resto de la misma posee un clima semiárido, denominados como: Mediterráneo infraarbóreo estépico y Mediterráneo infraarbóreo subdesértico subtropical (PATFOR, 2009), los cuales son identificados como bosques mediterráneos subáridos con veranos cálidos secos y Moderadamente cálido respectivamente, siendo zonas con poca cobertura forestal y discontinua, “cuya influencia en el comportamiento del fuego es muy puntual” (Soriano, 2016).

Figura 1.3-4 Subtipos Fitoclimáticos de la Comunidad Valenciana.



Fuente: Cartografía del ICV, VAERSA y PATFOR

## 1.4 Datos

### 1.4.1 Humedad del combustible vivo

Como se expresó en el apartado 1.2, el siguiente trabajo se realiza partiendo de una base de datos levantada por la empresa VAERSA, la ubicación y selección de puntos de muestreo tanto como la metodología y análisis de las muestras se los ejecutó en un estudio anterior (Soriano, 2016), donde se enuncia dichos procesos y expresa que la toma de muestras fue efectuada de acuerdo a la identificación de las especies vegetales por punto, haciendo el muestreo por mes y en horas centrales del día entre las 12 y 14 horas.

La metodología usada para la extracción de muestras en campo como el análisis en laboratorio fue la recomendada por el laboratorio del fuego del Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), la cual toma a considerar las siguientes formulas:

$$\text{PesoVerde} = \text{PesoTotal} - \text{PesoBote} \quad (1)$$

El peso verde se obtiene pesando el bote de plástico con la muestra y sin la muestra, el primer valor se denomina peso total y el segundo peso del bote, se restan dichos valores tal como se muestra en la formula (1) y se obtiene el valor del peso de la muestra o peso verde.

Por último se procede a utilizar la fórmula para determinar el porcentaje de humedad del combustible vivo, la cual es expresada de la siguiente forma:

$$HCV = \frac{\text{PesoVerde} - \text{PesoSeco}}{\text{PesoSeco}} * 100 \quad (2)$$

El pesos seco se obtiene eliminando la humedad de las muestras para lo cual son colocadas en una estufa de secado a 100 °C durante 24 horas como mínimo, al finalizar el secado se pesa nuevamente la muestra sin humedad y para finalizar se procede a calcular el porcentaje de humedad utilizando la fórmula (2).

El resultado final es recopilado en una base de datos mismo que consta de información adyacente como: fecha de toma de muestra en campo, especie, punto de muestreo, entre otros.

#### *1.4.2 Humedad del combustible vivo especie vegetal romero*

El análisis de la humedad que contiene la especie vegetal *Rosmarinus officinalis*, o también conocida como Romero es fundamental ya que forma parte de la cobertura vegetal endémica de la comunidad Valenciana y como tal, al ser afectado por factores externos como los fenómenos meteorológicos puede presentar una mayor o menor deshidratación a lo largo de un año, dándose la posibilidad de que dicha especie presente las características adecuadas de un componente combustible que en condiciones adecuadas sea un participante activo de eventos como son los incendios forestales, mismo que pueden generarse de forma natural o antrópica dependiendo del tipo de ignición que se dé a lugar.

Por lo cual a continuación se realiza el estudio de dicha especie con la finalidad de determinar los variantes de humedad en un determinado tiempo y su potencial como elemento primordial en el ciclo del fuego.

Se analiza la variación de humedad de la especie vegetal comúnmente conocida como Romero en la Comunidad Valenciana, contando con datos porcentuales mensuales en un periodo de 4 años, desde abril del 2014 hasta diciembre del 2017, para lo cual se realizó un muestreo en diferentes zonas bien distribuidas con un total de 16 estaciones de muestreo.

Cabe destacar que el contenido de humedad de una especie vegetal varía tomando en cuenta principalmente la fenología sobre la especie, sin embargo no hay que descartar que otros factores inciden en la misma como la retención de humedad del suelo, fenómenos antrópicos, entre otros.

#### *1.4.3 Estaciones meteorológicas*

Las plantas en general tienen una fuerte relación con la cantidad de humedad que se encuentra en el ambiente, de acuerdo a la especie vegetal, la cantidad necesaria le permitirá realizar funciones esenciales como la transpiración, proceso por el cual expulsa agua en forma de vapor a la atmósfera. La transpiración vegetal es de vital importancia ya que así absorbe agua y nutrientes desde el suelo, permitiéndole también regular su temperatura.

Los factores ambientales en los cuales se desarrolla la especie modifican la tasa de transpiración

de las plantas, el grado de humedad en el aire como por ejemplo, si el ambiente es muy seco la planta incrementara su transpiración, absorbiendo más agua del suelo, si el ambiente es húmedo su transpiración será mucho menor. Las corrientes de aire o viento secan el agua de las hojas haciendo que la planta modifique la intensidad de transpiración, de igual forma se relaciona con la temperatura, cantidad de luz, entre otros.

Por lo que se realiza recopilación de datos de variables meteorológicas, con la finalidad de identificar la relación existente con el porcentaje de humedad de la especie vegetal romero y el grado de influencia.

Para lograrlo se hace uso de datos meteorológicos de estaciones que se encuentren dentro de la zona de estudio, es decir la Comunidad Valenciana. (Soriano, 2016)

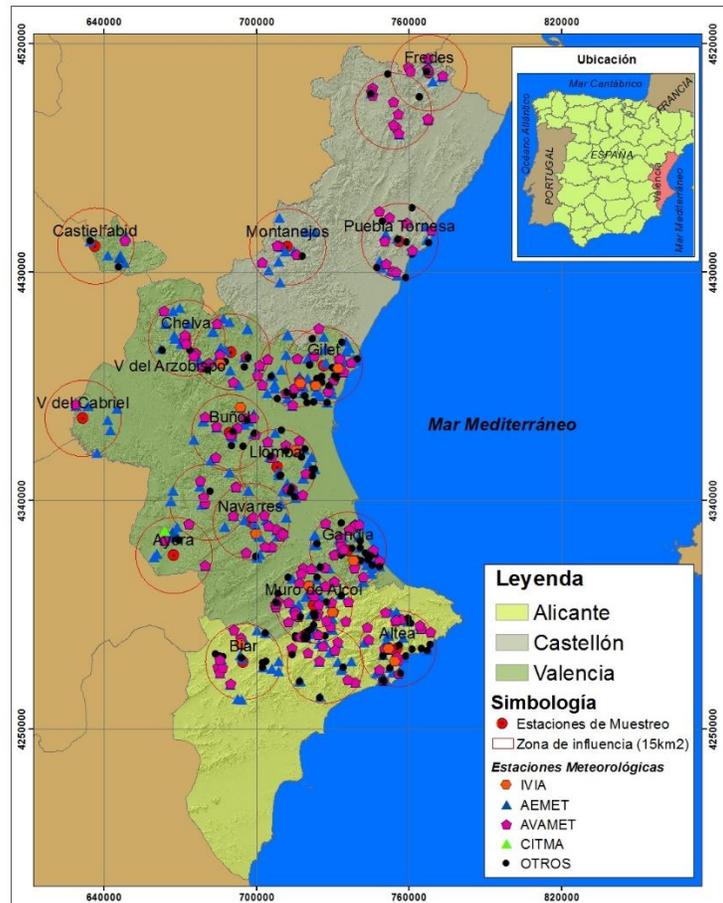
Tal como se muestra en la figura 1.4.3-1 existe una gran cantidad de estaciones por lo que se debe hacer un escudriño para definir las que mejor convengan, los parámetros de selección van de acuerdo a la fuente o entidad gestora, si es privada o pública, la información que levantan, y que estén cerca de las zonas muestreadas, entre otros. Si bien es cierto que resulta imposible obtener datos en los puntos donde se realizó el muestreo de la humedad de la especie vegetal romero, además aplicar métodos como los de interpolación no resulta eficiente ya que la cantidad final de puntos o información necesaria es insuficiente por lo que se procede a trabajar con las estaciones más cercanas.

Las estaciones que se toman principalmente en cuenta son las pertenecientes a entidades gestoras tales como: la Agencia Estatal de Meteorología (AEMET), Asociación Valenciana de Meteorología (AVAMET), Instituto Valenciano de Investigaciones Agrarias (IVIA) y Conselleria de Infraestructura, Territorio y Medio Ambiente (CITMA).



se puede encontrar en el periodo requerido y su obtención se puede realizar vía internet descargando directamente la información desde la página oficial de la institución.

Figura 1.4.3-2 Selección de Estaciones Meteorológica por el área de influencia.



Es importante recalcar que aunque se pudo considera datos de otras fuentes, muchas de ellas eran de pago y de entidades privadas por lo que se descartó estas opciones al presentar dificultad en conseguir la información como también saber si tenían registro de los periodos requeridos. De igual forma la elección de las estaciones se realizó bajo la supervisión de un experto en este tema, quedando a la final con las estaciones mostradas en la tabla 1.4.3-1.

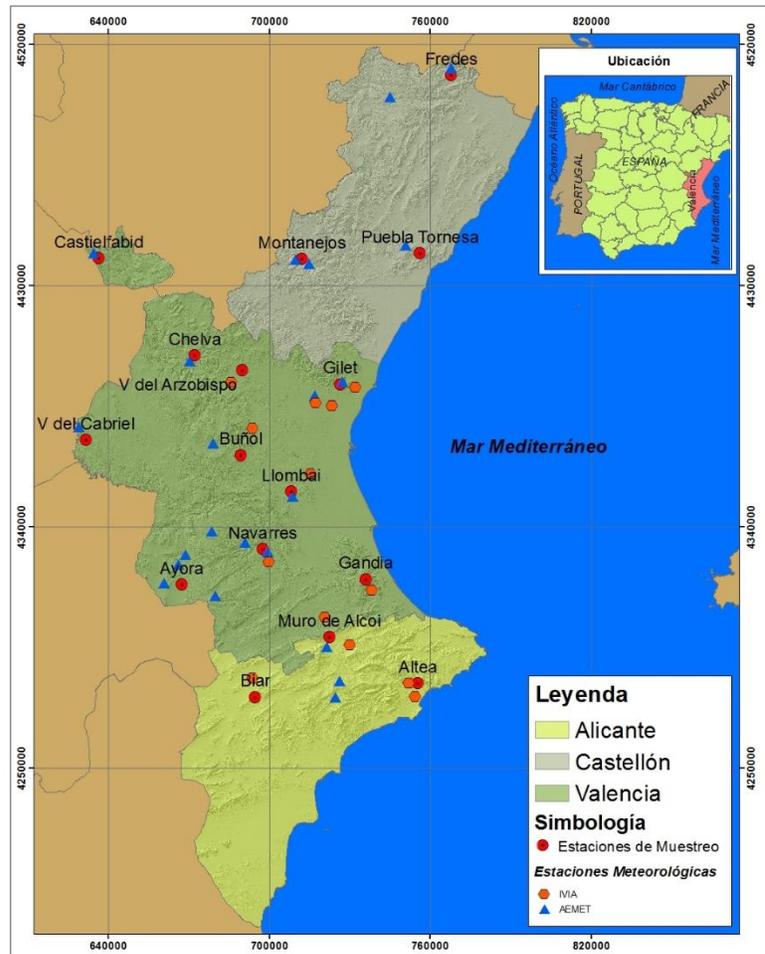
Tabla 1.4.3-1 Estaciones meteorológicas seleccionadas.

Nombre estación	Entidad	Datos
Altea	IVIA	Temperatura, Humedad relativa, Viento y Precipitación
Callosa d'en SarriÓ	IVIA	
Callosa d'en SarriÓ (malla)	IVIA	
BÚtera	IVIA	
Moncada IVIA	IVIA	
Moncada2	IVIA	
Camp de Mirra	IVIA	
Cheste	IVIA	
Gandia Marxuquera	IVIA	
Sagunt	IVIA	
Picassent	IVIA	
BÚlgida	IVIA	
Planes	IVIA	
Bolbaitte	IVIA	
Chulilla	IVIA	

Nombre estación	Entidad	Datos
GILET SANTO ESPIRITU	AEMET	Temperatura, Humedad relativa y Precipitación
LLOMBAY VISTA ALEGRE	AEMET	
ARENOS (C.H. JUCAR)	AEMET	
ARAÑUEL	AEMET	
MORELLA	AEMET	
MURO DE ALCOY P.I.PI	AEMET	
NAVARRRES AYUNTAMIENTO	AEMET	
VILLAFAMES H S	AEMET	
TORREMANZANAS (AYUNTAMIENTO)	AEMET	
BENIFALLIM	AEMET	
PRESA DE CONTRERAS	AEMET	
AYORA CASA MURCIA	AEMET	
AYORA-LA PEDRIZA	AEMET	
TERESA DE COFRENTES	AEMET	
AYORA LA MATEA	AEMET	
BETERA (B.HELICOPTEROS)	AEMET	
SIETE AGUAS	AEMET	
CASTIELFABIB A CEREZO	AEMET	
CHELVA (CONVENTO)	AEMET	
CORTES DE PALLAS (MUELA DE CORTES)	AEMET	
BICORP	AEMET	
POBLA DE BENIFASSA-FREDES	AEMET	
BETERA (B.HELICOPTEROS)	AEMET	

En la figura 1.4.3-3 se muestra de forma espacial las estaciones meteorológicas seleccionadas en referencia con los puntos de muestreo.

Figura 1.4.3-3 Distribución de las estaciones meteorológica seleccionadas.



#### 1.4.3.1 Cálculo de variables meteorológicas.

Los datos de las estaciones metrológicas pedidos son registros diarios y coincidentes con el periodo de tiempo en que se tomaron las muestras de humedad, sin embargo y a pesar de la insistencia no se pudo conseguir los datos en los plazos establecidos por lo que se prosiguió a trabajar con los datos que ya poseía la entidad VAERSA y los que fueron utilizados en el desarrollo del trabajo de fin de master de José Luís Soriano (2016).

Este estudio tiene la finalidad de centrarse en una sola especie vegetal denominada el romero (*Rosmarinus officinalis*) el cual cuenta con un registro de datos del porcentaje de humedad en un periodo de tiempo de 3 años y estudiar la relación que tienen con las variables meteorológicas, tomando en cuenta de que es una especie que aporta gran cantidad de combustible en la zona de estudio ya que es una especie endémica de la región mediterránea y por ende de la Comunidad Valenciana.

La relación entre las variables meteorológicas y el porcentaje de humedad de la especie romero se va a realizar mediante un análisis de regresión. Las variables utilizadas son las 28 indicadas por Soriano (2016, pág.19) donde son clasificadas en grupos, como son: precipitación, temperatura, humedad relativa, viento y punto de rocío.

En dicho trabajo se explica la importancia de cada grupo elegido, es claro que la precipitación es la fuente principal de agua de la planta, la cantidad de lluvia en relación con la permeabilidad del suelo son determinantes en el porcentaje absorción por ende de humedad de la planta a lo largo del año, el análisis realizado es hasta 60 días antes de la toma de muestra. La temperatura influye directamente en la evapotranspiración y de la cual se subdivide en temperatura máxima y media, medirán la pérdida de humedad directamente y son escogidas estas ya que indican la mayor pérdida de humedad. La humedad relativa de igual forma se sub divide en máximas y mínimas, las máximas denotan un mayor intercambio de humedad de las plantas con respecto al ambiente, mientras que con las mínimas se produce la pérdida de humedad por evapotranspiración. Con lo que respecta a los vientos igual se divide en Viento máximo ya que a mayor velocidad de viento mayor evapotranspiración y mayor pérdida de humedad, el levante es un grupo de variables del grupo denominado viento y es propio del viento proveniente del mar el cual contiene y aporta humedad. El último grupo es el denominado Punto de rocío con el cual se desea analizar la influencia que tiene la condensación y la precipitación horizontal sobre la humedad de la especie vegetal romero, se valora el número de días en que se produce condensación, entendiéndose como valor donde la temperatura del punto de rocío alcanza la temperatura mínima del aire.

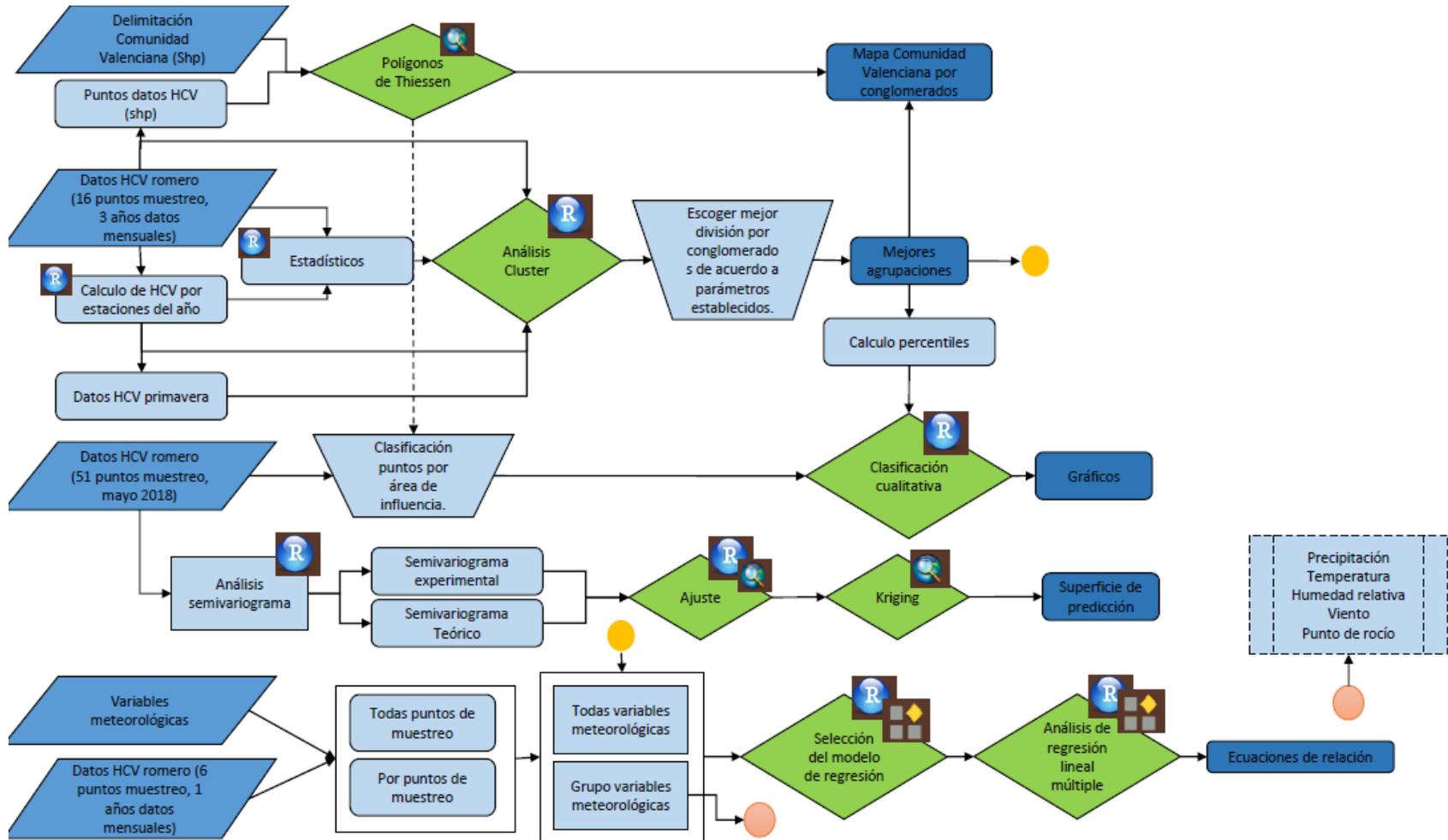
En la tabla 1.4.3.1 se presentan las 28 variables divididas por grupos, una descripción y la medida.

Tabla 1.4.3.1-1 Variables meteorológicas calculadas

Tipo	Variable	Unidad de medida	Descripción
PRECIPITACIÓN	P3	mm	Precipitación acumulada en los últimos 3 días.
	P7	mm	Precipitación acumulada en los últimos 7 días.
	P15	mm	Precipitación acumulada en los últimos 15 días.
	P30	mm	Precipitación acumulada en los últimos 30 días.
	P60	mm	Precipitación acumulada en los últimos 60 días.
TEMPERATURA	Tmax3	°C	Media de las temperaturas máximas en los últimos 3 días.
	Tmax7	°C	Media de las temperaturas máximas en los últimos 7 días.
	Tmax15	°C	Media de las temperaturas máximas en los últimos 15 días.
	Tmax30	°C	Media de las temperaturas máximas en los últimos 30 días.
	Tmedia3	°C	Media de las temperaturas medias en los últimos 3 días.
	Tmedia7	°C	Media de las temperaturas medias en los últimos 7 días.
	Tmedia15	°C	Media de las temperaturas medias en los últimos 15 días.
HUMEDAD RELATIVA	HRmax3	%	Media de la humedad relativa máxima de los últimos 3 días.
	HRmax7	%	Media de la humedad relativa máxima de los últimos 7 días.
	HRmax15	%	Media de la humedad relativa máxima de los últimos 15 días.
	HRmin3	%	Media de la humedad relativa mínima de los últimos 3 días.
	HRmin7	%	Media de la humedad relativa mínima de los últimos 7 días.
	HRmin15	%	Media de la humedad relativa mínima de los últimos 15 días.
VIENTO	Vmax3	km/h	Media de la velocidad máxima media del viento en 10 minutos de los últimos 3 días.
	Vmax7	km/h	Media de la velocidad máxima media del viento en 10 minutos de los últimos 7 días.
	Vmax15	km/h	Media de la velocidad máxima media del viento en 10 minutos de los últimos 15 días.
	Lev3	horas	Media de la suma de frecuencia del viento en los octantes 2 y 3 (vientos de componente E), en los últimos 3 días.
	Lev7	horas	Media de la suma de frecuencia del viento en los octantes 2 y 3 (vientos de componente E), en los últimos 7 días.
	Lev315	horas	Media de la suma de frecuencia del viento en los octantes 2 y 3 (vientos de componente E), en los últimos 15 días.
PUNTO DE ROCÍO	Pto_r3	número de días	Número de días en que se ha producido condensación en los últimos 3 días.
	Pto_r7	número de días	Número de días en que se ha producido condensación en los últimos 7 días.
	Pto_r15	número de días	Número de días en que se ha producido condensación en los últimos 15 días.
	Pto_r30	número de días	Número de días en que se ha producido condensación en los últimos 30 días.

Fuente: Soriano (2016, pág.20)

### 1.5 Diagrama de flujo de trabajo



## 2. ANÁLISIS ESTADÍSTICO DE LA VARIABLE HUMEDAD

Se procede a realizar un análisis descriptivo de la variable humedad del combustible vivo enfocándose en la especie vegetal comúnmente denominada Romero, como primera observación se descarta las estaciones de muestreo Bétera y Muela de Cortes por ausencia de datos continuos presentes en un rango de diciembre 2015 – noviembre 2017 y marzo 2017 - noviembre 2017 respectivamente.

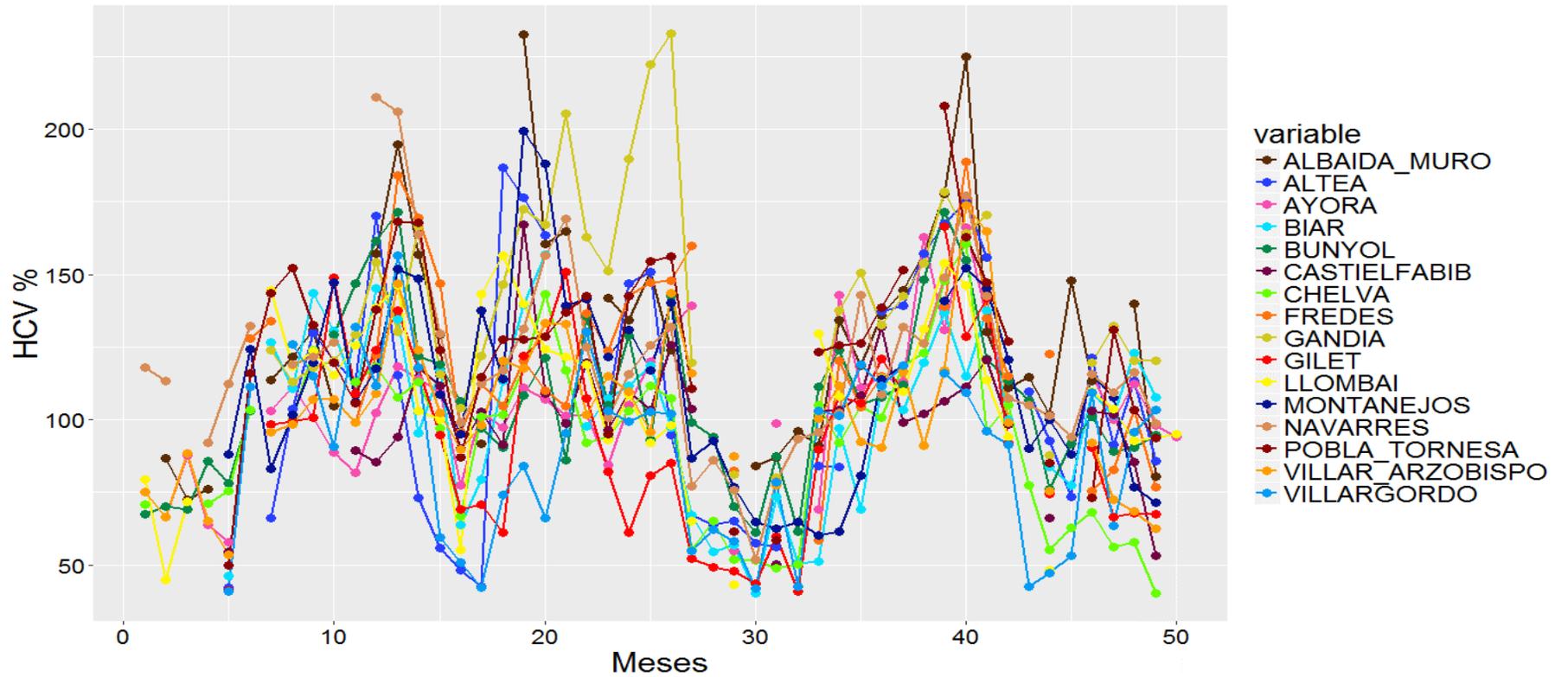
En la figura 2-1 se puede observar los datos mensuales en porcentaje de la humedad de la especie vegetal romero de 16 estaciones muestreadas, con datos tomados por 4 años consecutivos desde abril de 2014, siendo este un año incompleto de muestreo y como se puede observare en la tabla 2-1 en la casilla denominada “Num de EM sin dato”, misma que lleva el conteo de estaciones que no tienen ningún valor en algún mes determinado, los 4 primeros meses del 2014 carecen de valores en la mayoría de las estaciones.

Al realizar un análisis de la humedad tomando en cuenta la figura 2-1 y la tabla 2-1 se puede observar que los valores de los meses del año 2014 pertenecientes a las estaciones primavera y verano oscilan entre los 44 y 118, cabe destacar que los 4 primeros meses presenta una carencia de datos en la mayoría de las estaciones; mostrando para el mes de septiembre un aumento de humedad notorio, además se puede visualizar que del mes de abril a agosto presentan un valor bajo de humedad en todas las estaciones presentando valores en un rango de 40% a 100%, en septiembre de 2014 se puede notar un incremento considerable en todas las estaciones manteniéndose en un rango de 81 y 153, existiendo un valor atípico en el mes de octubre que presenta un valor mínimo de 66, sin embargo el resto de valores oscilan entre los valores antes mencionado; para los meses de marzo y abril de 2015 se da un incremento brusco, mismo que comienza a decaer en el mes de mayo, el valor más alto en los tres meses mencionados en abril, luego se nota un descenso de humedad hasta el mes de julio de 2016 el cual llegará a un punto en donde sus valores oscilaran en un rango muy pequeño entre estaciones hasta el mes de agosto donde comienza a ver un incremento, corroborando que en dichos meses se hace presente la temporada de verano; en septiembre del 2015 se nota un incremento en la humedad dándose variaciones bajas hasta mayo del 2016, sin embargo en junio se puede observar una disminución considerable la cual se mantiene hasta el mes de

septiembre donde la humedad aumenta manteniéndose en un rango estable hasta llegar a los meses de marzo y abril donde existe un aumento en algunas estaciones más que en otras, tales como en Abaida – Muro con un valor de 225.01, debido en parte a la presencia de un periodo con importantes lluvias en toda la zona de estudio. En el mes de mayo se puede notar un decrecimiento en la humedad de todas las estaciones que se extiende hasta junio, sin embargo de julio de 2017 hasta diciembre del mismo año el comportamiento de la humedad varía enormemente entre las estaciones, oscilando los valores en un rango mucho más amplio con respecto a los demás meses analizados.

El análisis realizado se hizo sin tomar en cuenta la distribución espacial de las estaciones de muestreo a lo largo de la Comunidad Valenciana, tomando solo en cuenta los datos de humedad proporcionados por la entidad correspondiente, es importante destacar ello ya que los puntos de muestreo pueden presentar diferencias significativas, como la capacidad de retención de humedad que puede tener el suelo, la composición de la misma, topografía, entre otros. Parámetros que influyen directamente en la retención de agua de la especie vegetal en estudio.

Figura 2-1 Humedad del combustible vivo de la especie Romero.



Realizado en RStudio

En la figura se muestra la leyenda del eje x codificada por números del 1 al 50 en donde cada uno representa al mes en que fue tomada la muestra de forma ordenada y ascendente, correspondiendo el primer número al mes abril de 2014 y el número final al mes diciembre de 2017, coincidente con el número de muestras totales por punto de muestreo.

Tabla 2-1 Estadísticos humedad del combustible vivo de la especie Romero.

Estadísticos	abril 14	mayo 14	junio 14	julio 14	agosto 14	septiembre 14	octubre 14	noviembre 14	diciembre 14	enero 15
Valor Mínimo	67.6	44.9	69.1	63.8	40.8	103	66	98.43	100.6	88.9
1st cuartil	70.9	66.45	71.8	66.8	42.2	107.3	97.05	101.77	115	107.2
Mediana	75.2	68.35	72.4	73.7	53.4	116.1	113.7	111.31	121.7	120.2
Promedio	82.2	74.67	77.92	75.7	60.1	116.9	112.05	113.48	120.6	119.2
3rd cuartil	79.3	82.55	87.8	83.22	75.5	126.2	130.2	118.93	129.9	129.2
Valor Máximo	118	113.3	88.5	92.1	112.4	132.2	144.6	152.3	143.7	148.7
Num de EM sin dato	11	10	11	10	3	9	5	3	3	3

Estadísticos	febrero 15	marzo 15	abril 15	mayo 15	junio 15	julio 15	agosto 15	septiembre 15	octubre 15	noviembre 15
Valor Mínimo	81.86	85.47	94.01	73.09	55.74	48.13	42.2	61.21	84.05	66.28
1st cuartil	100.65	116.22	127.33	113.75	99.35	65.24	88.69	94.4	117.61	118.6
Mediana	108.99	130.98	146.56	120.07	110.75	87.13	101.78	113.34	129.56	130.97
Promedio	111.31	135.43	147.79	129.48	107.54	79.91	98.15	113.53	141.73	135.31
3rd cuartil	122.57	154.93	169.1	158.72	120.14	95.43	112.97	123.89	168.59	157.75
Valor Máximo	146.77	211.15	206.01	169.45	146.74	106.28	143.26	186.6	232.42	188.13
Num de EM sin dato	2	0	0	0	0	1	0	1	0	0

Estadísticos	diciembre 15	enero 16	febrero 16	marzo 16	abril 16	mayo 16	junio 16	1QJI 16	julio 16	1QAg 16	agosto 16	1QSp 16
Valor Mínimo	86.06	92.04	82.16	61.12	80.78	85	52.35	49.24	43.11	40.28	48.87	40.83
1st cuartil	102.23	104.61	94.81	107.7	100.97	101	66.16	60.27	55.81	43.43	58.56	48.42
Mediana	127.38	124.88	104.08	113.7	117.69	128.8	86.78	64.53	61.67	51.83	75.47	56.07
Promedio	130.32	123.33	107.49	121.27	124.15	128.8	91.63	70.98	64.97	55.13	70.85	62.54
3rd cuartil	147.93	136.53	116.61	136.37	148.19	142.8	113.48	87.82	76.32	61.32	78.99	72.03
Valor Máximo	205.44	162.69	151.11	189.89	222.29	233	159.73	93.94	87.48	83.99	98.68	96
Num de EM sin dato	2	1	0	0	0	0	1	8	1	7	0	8

Estadísticos	sep-16	octubre 16	noviembre 16	diciembre 16	enero 17	febrero 17	marzo 17	abril 17
Valor Mínimo	51.16	61.54	69.31	90.55	98.95	91.01	106.3	109.5
1st cuartil	80.29	100.23	104.48	110.68	110.92	121.29	135.4	141.7
Mediana	98.04	108.03	107	114.46	118.49	131.12	148.5	161.7
Promedio	92.37	109.6	109.95	117.87	121.94	133.81	150.4	156.9
3rd cuartil	103.58	124.47	118.8	132.23	135.54	155.47	168.4	174.2
Valor Máximo	129.65	142.73	150.64	138.61	151.53	162.67	207.9	225
Num de EM sin dato	0	0	2	0	1	5	0	0

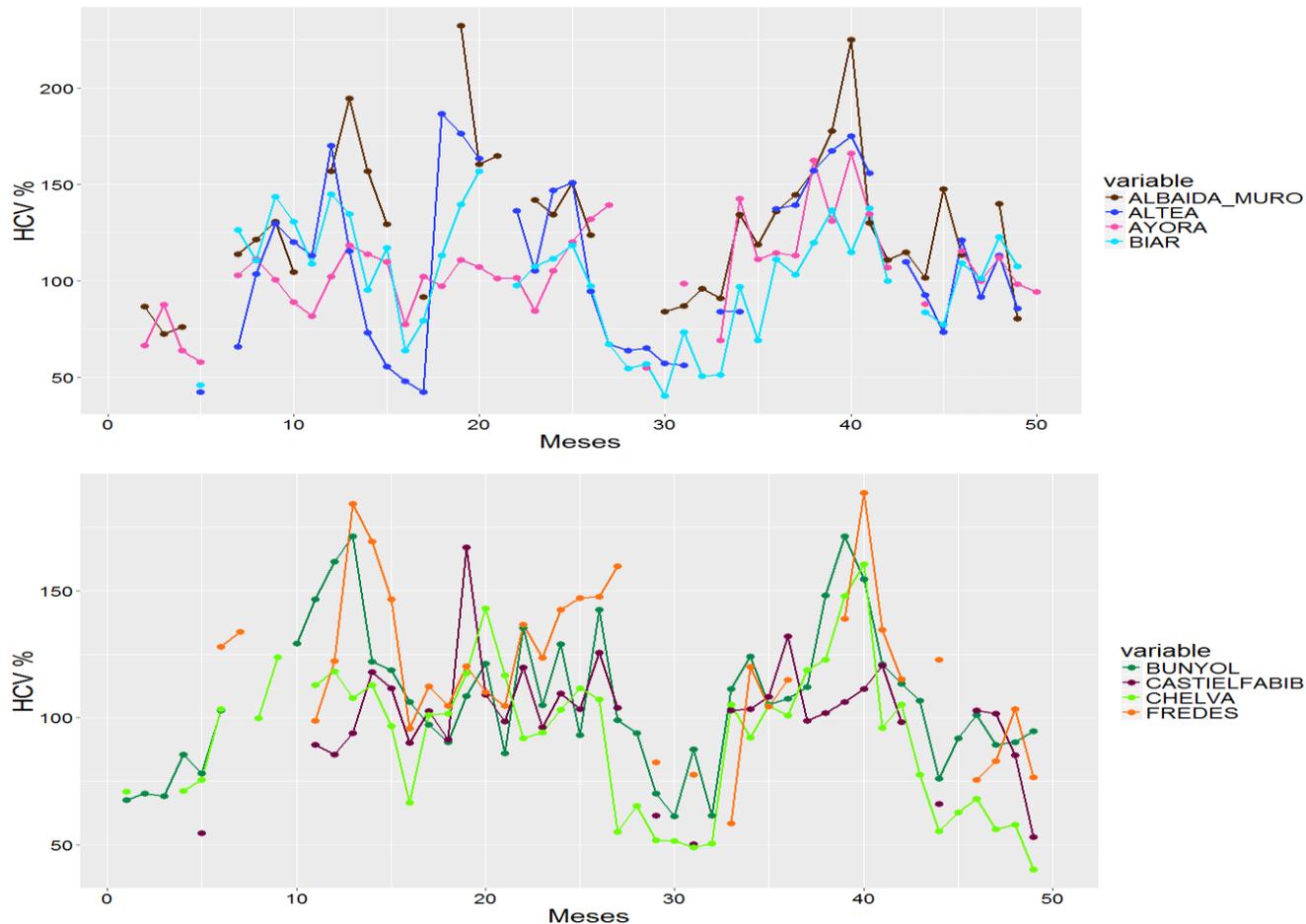
Estadísticos	mayo 17	junio 17	1QJI 17	julio 17	1QAg 17	agosto 17	sep-17	octubre 17
Valor Mínimo	96	91.26	42.47	47.35	53.25	68.04	56.08	57.94
1st cuartil	121	98.5	83.73	72.47	70.77	91.51	77.74	83.2
Mediana	136.2	106.06	104.93	84.31	82.82	109.34	100.1	99.57
Promedio	134.6	105.84	92.32	81.64	86.09	102.03	93.96	97.96
3rd cuartil	145.6	112.86	108.33	94.52	92.39	115.63	105.69	113.98
Valor Máximo	170.4	127.02	114.69	122.79	147.75	121.25	132.34	140.02
Num de EM sin dato	0	2	9	0	8	0	1	0

Estadísticos	noviembre 17	diciembre 17
Valor Mínimo	40.16	94.18
1st cuartil	70.62	94.44
Mediana	89.75	94.69
Promedio	84.28	94.69
3rd cuartil	98.36	94.95
Valor Máximo	120.22	95.21
Num de EM sin dato	0	14

En la figura 2-2 se muestra una gráfica temporal en el periodo de los 4 años por estación de en grupos menores con la finalidad de poder observar la continuidad de los datos, como también mostrar de forma clara el comportamiento de la humedad en cada estación individual contrastando lo mostrado en la tabla 2-1.

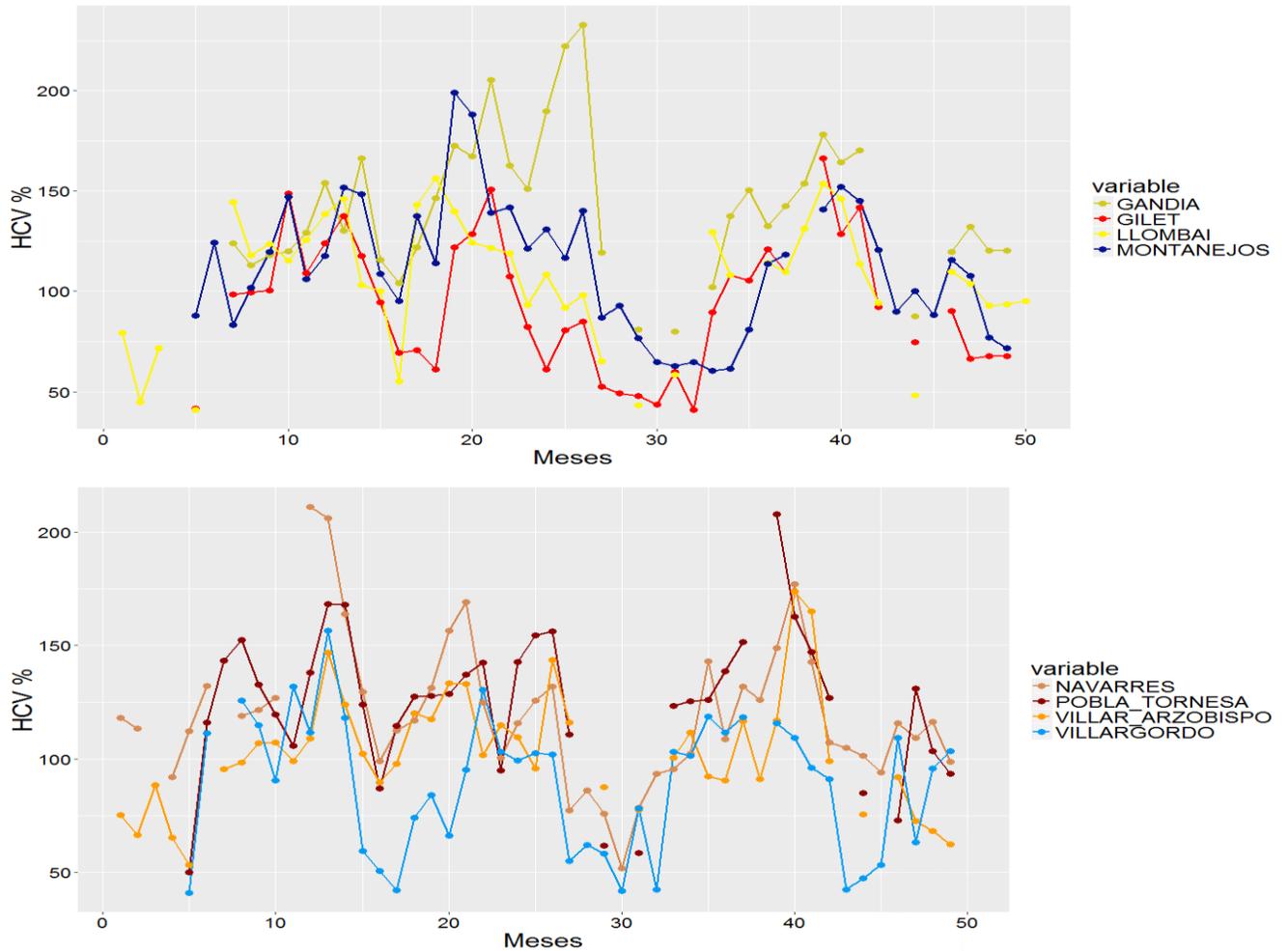
La humedad de la especie vegetal Romero aumenta o disminuye en concordancia con las estaciones de año por lo que se procede a trabajar con datos completos de los últimos 3 años, es decir desde diciembre del 2014 hasta diciembre de 2017 y se realiza un análisis estadístico por punto de muestreo.

Figura 2-2 Humedad del combustible vivo de la especie Romero por grupo de puntos de muestreo individual.



Realizado en RStudio

Figura 2-2 Humedad del combustible vivo de la especie Romero por grupo de puntos de muestreo individual.



Realizado en RStudio

Sabiendo que el objetivo general es conocer la influencia de las variables meteorológicas con respecto a la variable humedad de la especie vegetal romero, al realizar un modelo de regresión lineal con todas ellas su resultado se presenta muy general, tomando en cuenta que el área de estudio es amplio, y las variables a relacionar interactúan de formas diferentes debido a la distribución espacial de los puntos de muestreo dentro de la zona, se debe considerar realizar un análisis diferente. El más indicado sería realizar un análisis por cada punto de muestreo de forma independiente, sin embargo este requeriría mucho tiempo de estudio, al observar la figura 2-2 se puede notar que existe afinidad entre puntos de muestreo con lo cual se podría agruparlos.

El análisis de conglomerados o análisis cluster es una opción acertada, el cual “tiene una importante tradición de aplicación en muchas áreas de investigación. Sin embargo, junto con los beneficios del análisis cluster existen algunos inconvenientes. El análisis cluster es una técnica descriptiva, atórica y no inferencial. El análisis cluster no tiene bases estadísticas sobre las que deducir inferencias estadísticas para una población a partir de una muestra, es un método basado en criterios geométricos y se utiliza fundamentalmente como una técnica exploratoria, descriptiva pero no explicativa.” (Fernández, 2011)

Para realizar el análisis es necesario tener datos completos y para encontrar el mejor modelo de conglomerados que se ajuste a los parámetros requeridos como es la creación de grupos con número de integrantes homogéneos se realiza dos tipos de análisis.

Uno de ellos es por medio de estadísticos que describan de mejor forma a la variable humedad, cabe destacar que los datos directos no pueden ser tomados ya que existen carencia de datos completos de cada punto de muestre y para hacer uso efectivo del análisis cluster como requisito es necesario cumplir dicho parámetro.

Otra línea de estudio es realizada partiendo del agrupamiento por estaciones del año mismo que es basado en las condiciones climáticas dominantes, obteniendo como tal, que los meses de marzo, abril y mayo entran en el grupo correspondiente a primavera; junio, julio y agosto a verano; septiembre, octubre y noviembre a otoño; y diciembre enero y febrero a invierno.

## **2.1 Análisis descriptivo de la variable humedad de la especie romero por mes.**

Se realiza un análisis descriptivo con la variable definida, obteniendo los estadísticos mostrados en la tabla 2.1-1, donde se pueden observar las medidas de posición como: el total de datos o recuento, el promedio o también llamada media aritmética que muestra la suma del valor de todos los datos, dividida para el número de datos; la mediana muestra el valor medio de los datos, en donde una mitad de los valores son menores y la otra son mayores, los datos deben estar ordenados de forma ascendente; la moda el valor que más se repite; el mínimo y máximo el cual informa el rango en que se mueven los datos; Cuartil inferior, valor que indica al 25% de los datos que se encuentran por debajo o son menores y un 75% de datos mayores, el cuartil superior en cambio muestra el valor donde un 75% de datos son menores y solo un 25% mayores, entre el cuartil inferior y cuartil superior se encuentran los datos más cercanos

a la mediana.

En lo que respecta a las medidas de dispersión están la desviación estándar la cual muestra la dispersión de los datos respecto al promedio, en este caso las estaciones presentan valores en un rango de 22 y 42, en donde las estaciones con mayor dispersión es Altea y la de menor Castielfabid; en lo que respecta al coeficiente de variación siendo esta una medida de la dispersión relativa de un conjunto de datos, el cual es expresado en porcentaje, y muestra la dispersión de los datos con respecto a la media se puede decir que la estación con mayor dispersión presentada en este conjunto de datos pertenece a Altea con 37.53% de dispersión.

Dentro de las medidas de forma se muestran los coeficiente de asimetría (sesgo estandarizado) y la curtosis estandarizada mismos que indican si los datos son provenientes de una distribución normal, y el rango (-2 a +2) en el cual esta hipótesis sería cierta, es así que el coeficiente de asimetría al estar dentro del rango se puede decir que los datos provienen de una distribución normal pero hay que tomar en cuenta que mientras más cerca a sus límites existirá menos asimetría alejándose de la idea de que pertenece a una distribución normal, existiendo una desviación significativa, lo esencial sería que el valor se acerque a 0 lo cual mostraría mayor asimetría y por lo tanto una distribución normal, es así que la estación Navarres está fuera de rango con un valor de 2.08 y las estaciones Villar\_Arzobispo y Albaida\_Muro presentan valores muy cercanos a los límites alejándose a la idea de que estos pertenezcan a una distribución normal; el valor de la curtosis estandarizada responden al mismo rango por lo que el sesgo y la curtosis presentan una desviación significativa de la normal, en este caso la estación Castielfabid se encuentra fuera de rango con un valor de 2.85 y las estaciones Navarres y Altea presentan valores muy cercanos a los límites.

Cabe destacar que dichos estadísticos fueron obtenidos de la variable humedad del combustible vivo (HCV) de la especie vegetal Romero de 3 años completos (2015 – 2017), tomando en cuenta que en algunos meses del 2016 y 2017 se realizaron dos muestreos por mes por lo que se debería tener 42 datos por estación, sin embargo ninguna de las estaciones presenta datos temporales completos variando los mismos entre los 32 y 40 datos.

Tabla 2.1-1 Estadísticos humedad del combustible vivo de la especie Romero periodo de 3 años.

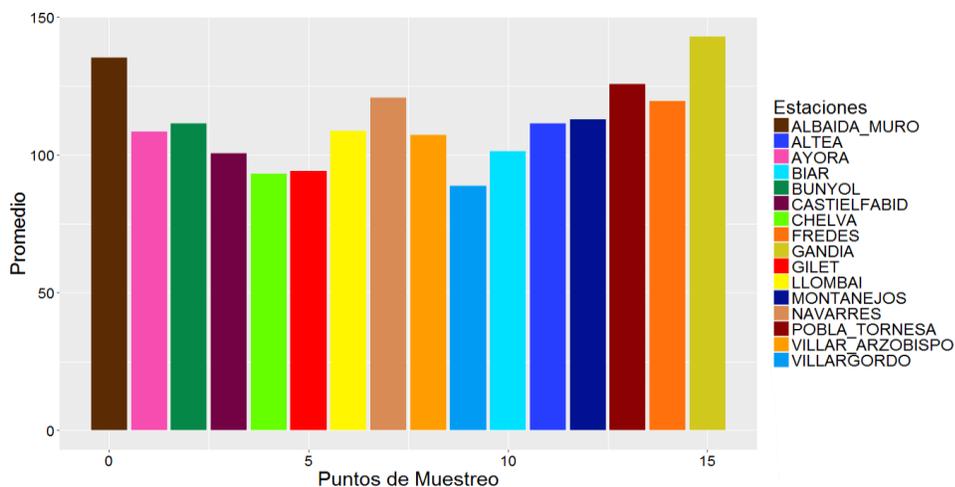
Cod	Estaciones	Recuento	Promedio	Mediana	Desviación Estándar	Coefficiente de Variación	Mínimo
1	AYORA	37.00	108.35	106.95	22.90	21.13	54.78
2	BUNHOL	40.00	111.46	107.23	27.43	24.61	61.32
6	LLOMBAI	36.00	108.74	109.70	28.86	26.54	43.11
8	VILLAR_ARZOBISPO	36.00	107.30	104.63	24.89	23.20	62.54
4	CHELVA	40.00	93.08	100.97	30.13	32.38	40.16
0	ALBAIDA_MURO	33.00	135.31	134.24	36.95	27.31	80.55
5	GILET	38.00	94.07	91.21	33.18	35.27	40.83
9	VILLARGORDO	40.00	88.83	95.88	29.44	33.14	41.78
3	CASTIELFABID	34.00	100.64	102.78	22.05	21.91	50.25
7	NAVARRES	40.00	120.79	116.06	33.23	27.51	51.83
15	GANDIA	35.00	142.99	137.46	36.30	25.39	80.05
10	BIAR	39.00	101.22	107.42	29.77	29.41	40.28
11	ALTEA	37.00	111.53	113.04	41.85	37.53	42.52
12	MONTANEJOS	40.00	112.93	114.93	33.82	29.95	60.32
14	FREDES	32.00	119.51	117.66	31.60	26.44	58.47
13	POBLA_TORNESA	35.00	125.63	127.55	31.39	24.99	58.61

Cod	Estaciones	Máximo	Rango	Cuartil Inferior	Cuartil Superior	Sesgo Estandarizado	Curtosis Estandarizada
1	AYORA	166.15	111.37	98.20	115.61	1.04	1.28
2	BUNHOL	171.52	110.20	92.53	126.53	1.14	-0.13
6	LLOMBAI	156.37	113.26	93.83	127.70	-1.40	0.04
8	VILLAR_ARZOBISPO	173.93	111.39	91.46	117.20	1.74	0.98
4	CHELVA	160.51	120.35	64.08	112.93	0.01	-0.88
0	ALBAIDA_MURO	232.42	151.87	111.04	157.00	1.94	1.00
5	GILET	166.51	125.68	67.62	120.88	0.68	-1.02
9	VILLARGORDO	156.68	114.90	60.91	110.60	-0.22	-1.02
3	CASTIELFABID	167.25	117.00	91.46	109.57	0.14	2.85
7	NAVARRES	211.15	159.32	99.86	131.94	2.08	1.46
15	GANDIA	232.97	152.92	119.61	166.33	1.29	0.34
10	BIAR	156.79	116.51	77.55	119.70	-0.72	-0.87
11	ALTEA	186.60	144.08	73.40	146.96	0.25	-1.47
12	MONTANEJOS	199.28	138.96	87.44	139.64	0.94	-0.07
14	FREDES	188.69	130.22	101.17	140.79	0.71	-0.19
13	POBLA_TORNESA	207.88	149.27	105.75	142.65	-0.15	0.77

Una vez identificado los estadísticos de los datos se procede a analizar de forma detenida 4 estadísticos en concreto, como son promedio, mediana, cuartil inferior (Q25%) y cuartil superior (Q75%) ya que la ser estadísticos de posición podrían ayudar a encontrar similitudes entre los puntos de muestreo en función de la variable de humedad.

En la figura 2.1-1 se presentan el gráfico de barras con los promedios de cada punto de muestreo donde se puede observar que los valores oscilan entre los 88 y 143 correspondientes al punto de muestreo Villagordo y Gandia respectivamente.

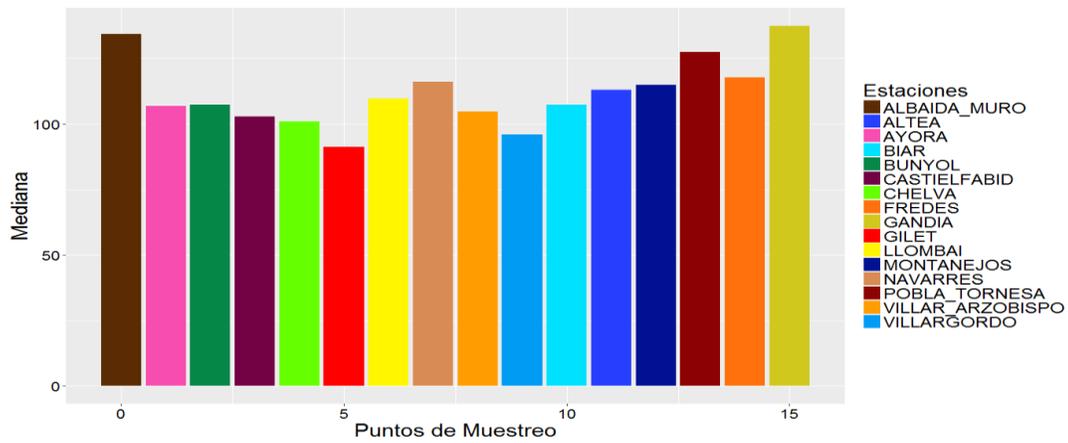
Figura 2.1-1 Promedio por punto de muestreo en función de la variable humedad de la especie Romero



Realizado en RStudio

En la figura 2.1-2 se presentan el gráfico de barras con la mediana de cada punto de muestreo, misma que representa al valor medio en donde la mitad de los datos se encuentra por debajo de este valor y la otra mitad por encima, dando una mejor perspectiva de los datos, claramente se puede observar que varias estaciones oscilan en rangos pequeños por lo cual estadísticamente podrían presentar similitud.

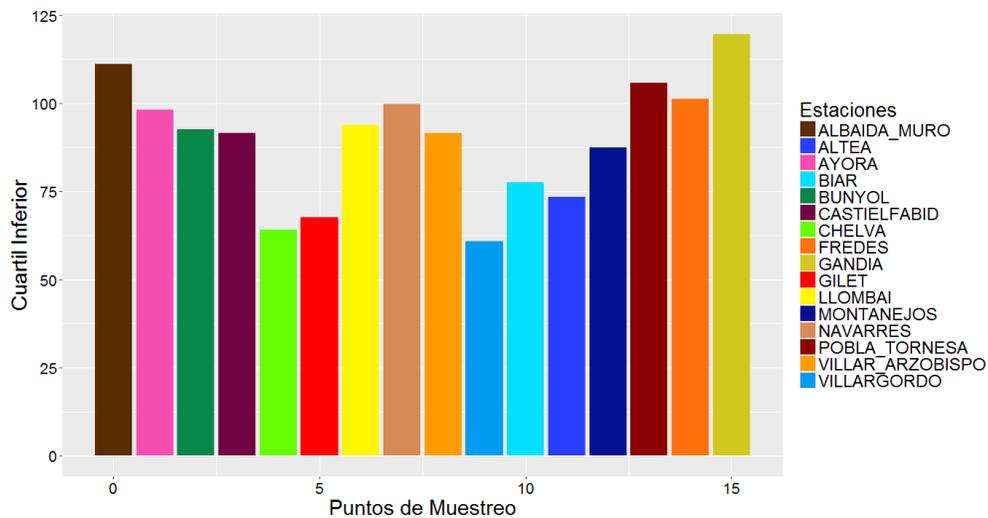
Figura 2.1-2 Mediana por punto de muestreo en función de la variable humedad de la especie Romero.



Realizado en RStudio

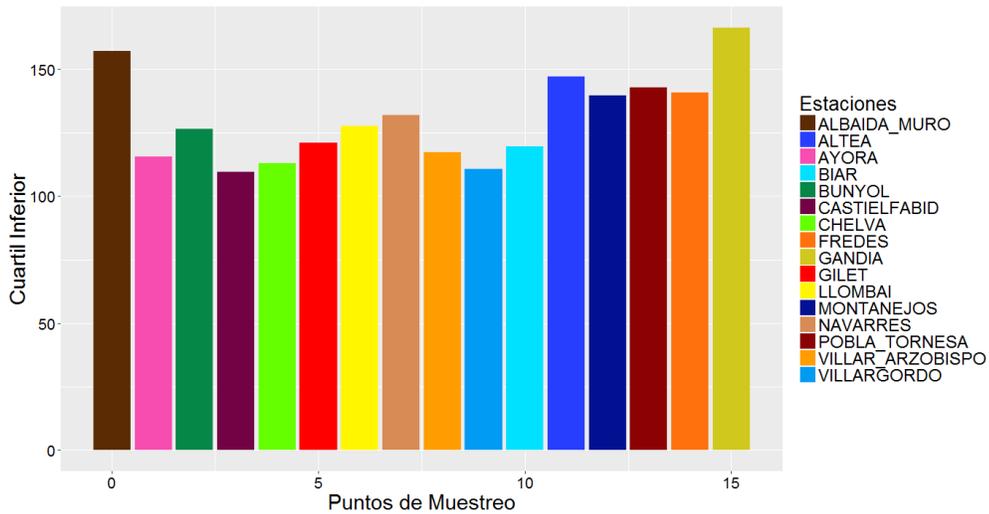
También se ha tomado los Cuartiles inferior (Q25%) y superior (Q75%), con los cuales se puede medir el límite inferior y superior de la dispersión de los datos obteniendo lo que se llama un rango intercuartil mismo que son representados en la figura 2.1-3 y 2.1-4 respectivamente.

Figura 2.1-3 Cuartil inferior (Q25%) de los punto de muestreo en función de la variable humedad de la especie Romero.



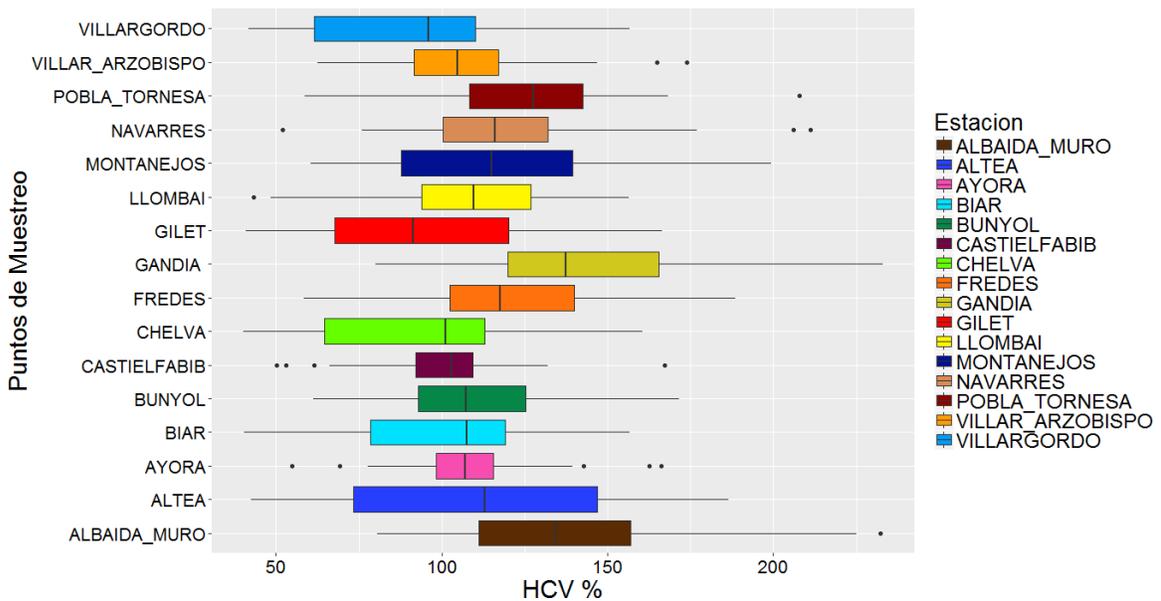
Realizado en RStudio

Figura 2.1-4 Cuartil superior (Q75%) de los punto de muestreo en función de la variable humedad de la especie Romero.



Para una mejor representación de los estadísticos, en la figura 2.1-5 se los representa en un diagrama de caja y bigotes.

Figura 2.1-5 Diagrama de caja y bigotes de los punto de muestreo en función de la variable humedad de la especie romero.



## 2.2 Análisis descriptivo de la variable humedad de la especie romero por estación del año

Se toman los datos de los 3 años establecidos en el apartado anterior y se realiza un promedio entre los meses correspondientes a las estaciones del año, basado en las condiciones climáticas dominantes obtenido como resultado la tabla 2.2-1 con datos completos.

Tabla 2.2-1 Datos del HCV por estaciones del año.

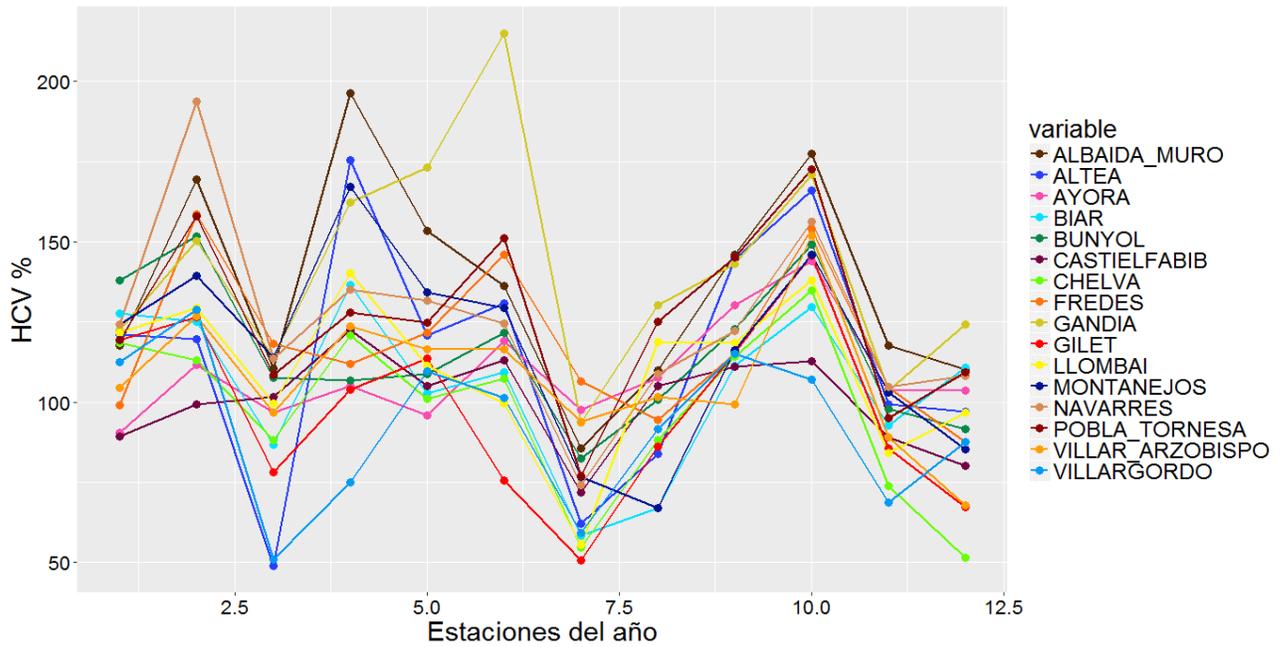
Estación	inv15	prim15	ver15	oto15	inv16	prim16	ver16	oto16	inv17	prim17	ver17	oto17
AYORA	90.463	111.580	96.607	105.117	95.770	119.017	97.613	107.667	130.103	144.003	103.490	103.527
BUNHOL	137.980	151.663	107.460	106.787	108.870	121.570	82.384	100.610	122.653	149.060	97.864	91.467
LLOMBAI	121.543	129.257	99.513	140.260	111.270	99.533	55.553	118.855	118.393	137.893	84.007	96.710
VILLAR_ARZOBISPO	104.360	126.707	96.620	123.693	116.607	116.407	93.707	101.443	99.423	151.910	88.843	67.847
CHELVA	118.420	112.983	88.243	120.797	101.000	107.307	54.484	88.133	114.153	134.873	73.792	51.393
ALBAIDA_MURO	117.695	169.573	110.550	196.520	153.485	136.317	85.520	110.023	145.930	177.620	117.726	110.285
GILET	119.447	126.473	78.233	103.903	113.487	75.633	50.566	86.005	115.280	145.670	85.663	67.293
VILLARGORDO	112.490	128.857	50.843	74.860	109.587	101.333	59.114	91.438	115.125	107.110	68.750	87.547
CASTIELFABID	89.330	99.200	101.483	122.587	104.923	112.917	71.837	104.963	110.997	112.743	89.120	80.017
NAVARRÉS	124.240	193.683	113.723	135.023	131.577	124.470	73.952	108.608	122.223	156.193	104.692	108.217
GANDIA	122.470	150.280	113.953	162.173	173.080	215.050	93.543	130.117	143.040	171.030	103.675	124.303
BIAR	127.763	125.033	86.777	136.553	102.595	109.207	58.454	67.020	111.433	129.790	92.635	110.613
ALTEA	121.040	119.617	48.797	175.493	120.790	130.853	61.942	83.960	144.557	166.130	99.285	96.873
MONTANEJOS	124.333	139.407	113.833	167.160	134.113	129.357	76.746	66.878	116.180	146.053	102.940	85.390
FREDES	98.950	158.703	118.273	111.753	121.693	145.897	106.557	94.347	115.000	154.127	104.453	87.643
POBLA_TORNESA	119.350	158.040	108.527	127.987	124.847	151.160	77.037	125.013	145.070	172.620	95.000	109.350

En la figura 2.2-1 se observa el comportamiento de todos los puntos de muestreo por estaciones de año y como es de esperarse en diferencia con la figura 2-1 muestra una reducción de datos por punto de muestreo, perdiendo información y suavizando la línea con respecto a los valores altos y bajos de humedad en función de las estaciones del año en los 3 años.

De la misma forma en la figura 2.2-2 se muestran los puntos de muestreo en grupos pequeños con la finalidad de observar su comportamiento más detenidamente, claramente dichas agrupaciones no son por afinidad sin embargo se puede observar a detalle el comportamiento de cada una, como por ejemplo el punto Gandía muestra un pico en la primavera del 2016 sobresaliendo por encima por el resto de estaciones, todo lo contrario con Gilet que presenta

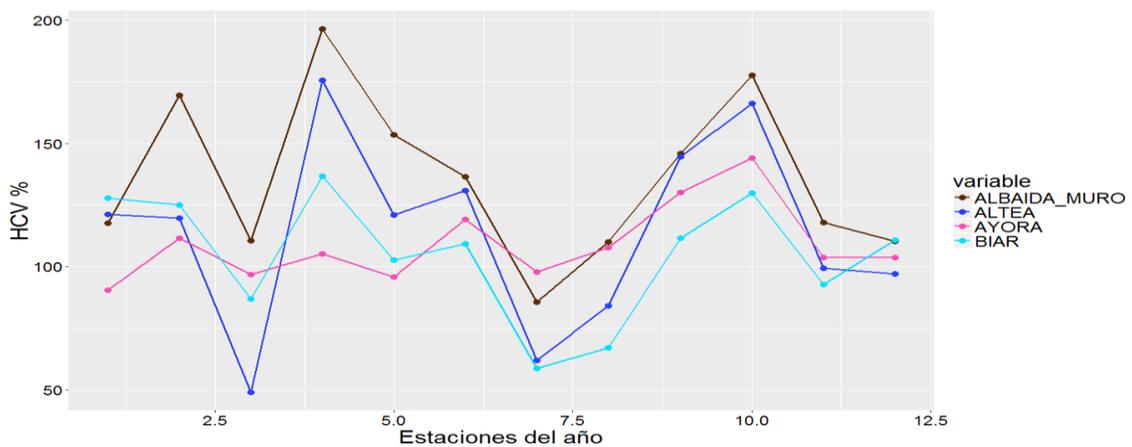
un descenso.

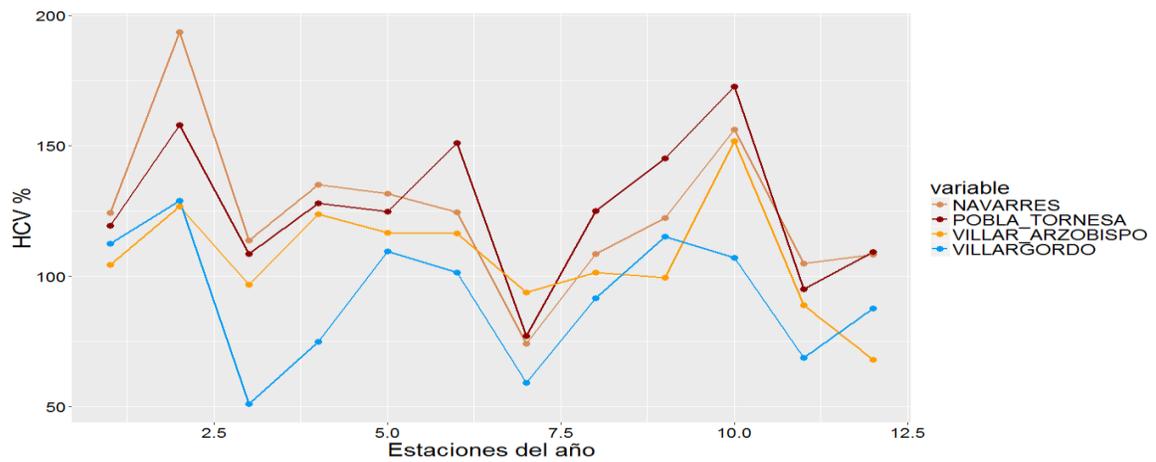
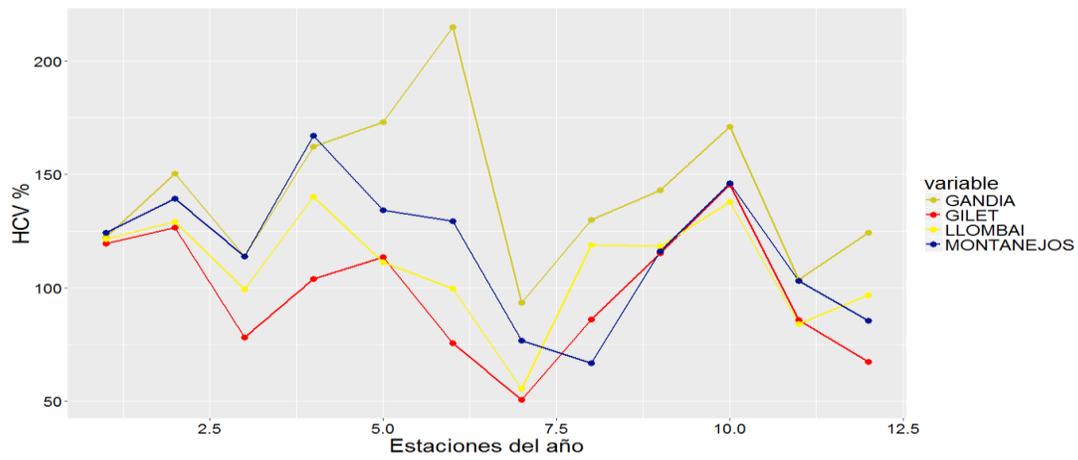
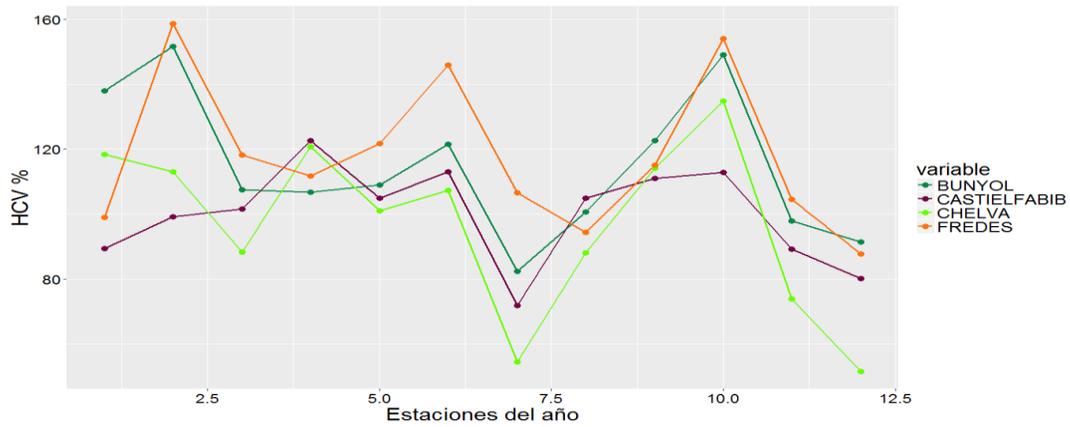
Figura 2.2-1 Humedad del combustible vivo de la especie Romero por estaciones del año.



La leyenda del eje x está codificada por números del 1 al 12, cada uno representa a la estación del año (invierno, primavera, verano y otoño) en ese orden, el primer número corresponde a invierno de 2014 y el final a otoño de 2017.

Figura 2.2-2 Humedad del combustible vivo de la especie Romero por estación del año y por grupos de punto de muestreo.





La leyenda del eje x está codificada por números del 1 al 12, cada uno representa a la estación del año (invierno, primavera, verano y otoño) en ese orden, el primer número corresponde a invierno de 2014 y el final a otoño de 2017.

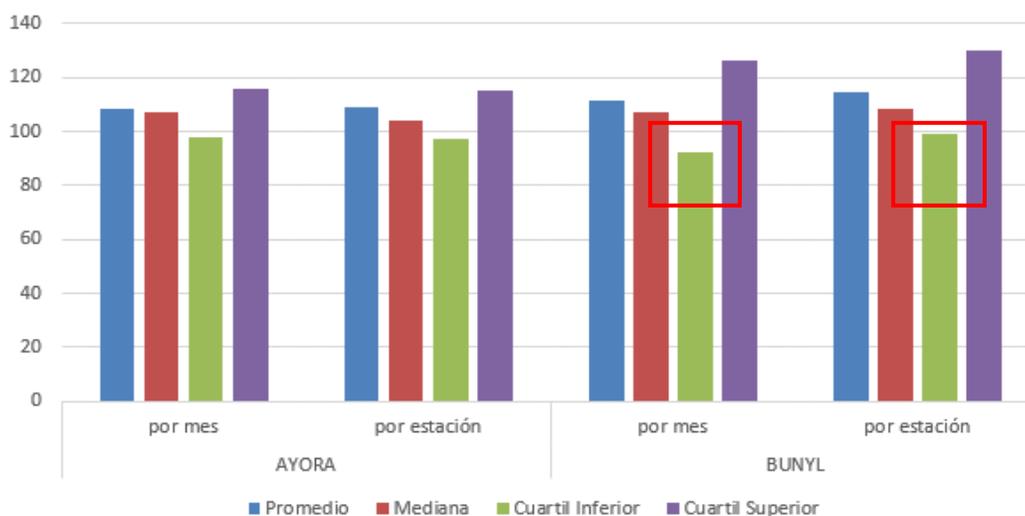
Realizado en RStudio

En la tabla 2.2-2 se muestran los estadísticos determinados en el apartado anterior que se utilizara para las agrupaciones, estos estadísticos fueron tomados en cuenta ya que presentan pequeñas variaciones con respecto a los de la tabla 2.1-1 correspondientes a los datos iniciales de la humedad por mes tal como se puede apreciar en la figura 2.2-3.

Tabla 2.2-2 Datos del HCV por estaciones del año en un periodo de 2015 - 2017.

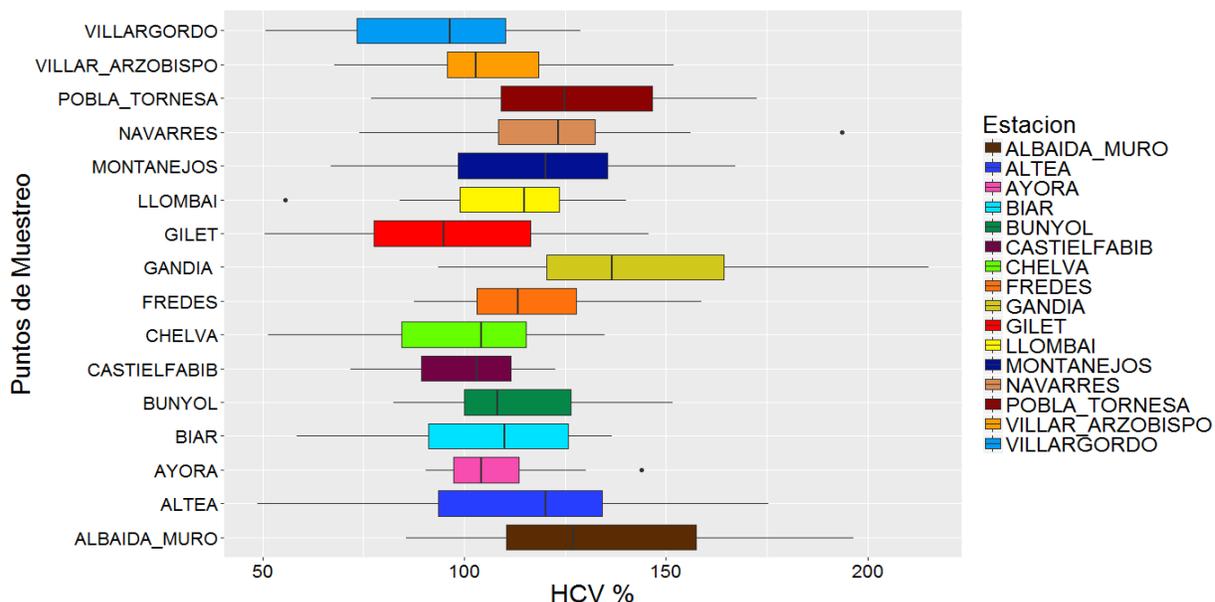
Estaciones	Promedio	Mediana	Cuartil Inferior	Cuartil Superior
AYORA	108.746	104.322	97.110	115.298
BUNYOL	114.864	108.165	99.237	130.317
LLOMBAI	109.399	114.832	98.112	125.400
VILLAR_ARZOBISPO	107.297	102.902	95.163	120.150
CHELVA	97.131	104.153	80.962	116.287
ALBAIDA_MURO	135.937	127.021	110.417	161.529
GILET	97.305	94.954	76.933	117.363
VILLARGORDO	92.254	96.385	71.805	111.038
CASTIELFABIB	100.010	103.203	89.225	111.870
NAVARRRES	124.717	123.232	108.412	133.300
GANDIA	141.893	136.578	118.212	166.602
BIAR	104.823	109.910	89.706	126.398
ALTEA	114.111	120.203	90.417	137.705
MONTANEJOS	116.866	120.257	94.165	136.760
FREDES	118.116	113.377	101.702	133.795
POBLA_TORNESA	126.167	124.930	108.938	148.115

Figura 2.2-3 Comparativa entre estadísticos de dos puntos de muestreo



En la figura 2.2-4 se muestra la gráfica de caja y bigotes de los puntos de muestreo en función de la variable humedad de la especie romero por estación del año que en comparación con la generada utilizando los datos de la variable humedad por mes se puede visualizar una gran disminución de valores atípicos.

Figura 2.2-4 Diagrama de caja y bigotes de los punto de muestreo en función de la variable humedad de la especie romero por estación del año.



### 3. ANÁLISIS DE CONGLOMERADOS

Después que se han definido los datos con los que se va a trabajar, en la figura 3-1 se muestra los diferentes análisis que se van a realizar. De los datos de humedad del combustible vivo de la especie romero por mes, se harán 5 análisis involucrando solo a sus estadísticos, los datos como tal no pudieron ser tomados en cuenta ya que es necesario tener datos completos, tal como se había explicado anteriormente y como dichos datos no cumplen esta cualidad son descartados.

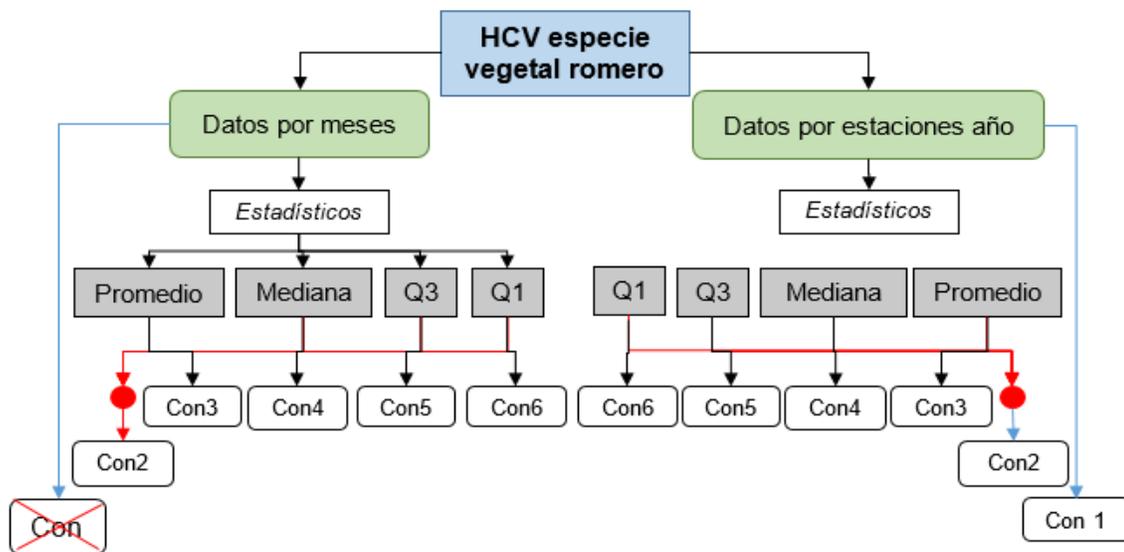
Se tomaran solo los 4 estadísticos de posición determinados en el apartado anterior y se realizara diferentes pruebas con ellos, el primer análisis será tomando todos ellos y los 4 siguientes de forma individual. El análisis de conglomerados individual se realiza con la finalidad de observar la influencia de cada estadístico al realizar las agrupaciones

determinando la ubicación dentro de cada clase de los puntos de muestreo, además se tiene la intención de encontrar la mejor combinación posible, que las agrupaciones presenten homogeneidad dentro de cada clase y grupos con un similar número de integrantes.

Se hace hincapié en controlar el número de integrantes en cada clase o agrupación ya que la tener una clase que abarque el mayor número de individuos generalizaría los datos concluyentes, entonces al tener en cuenta ello, se buscara la cantidad precisa de conglomerados o clases para trabajar.

Del mismo modo se realizan los mismos procesos con los datos calculados por estación del año, a diferencia del anterior grupo de análisis en este se procede incluir los datos generales, ya que cumplen con los parámetros requeridos.

Figura 3-1 Diagrama de los diferentes análisis a realizar y las respectivas variables a utilizar.

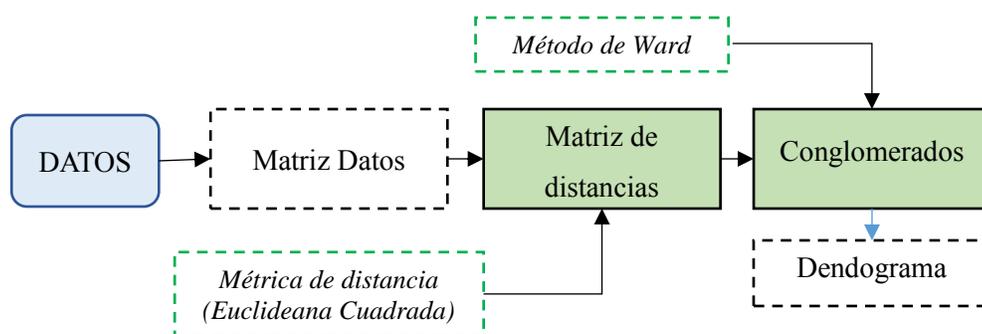


### 3.1 Desarrollo del análisis de conglomerados

El análisis se lo realizó en el programa RStudio, no se usó una librería en específico y a continuación se presenta el procedimiento seguido, cabe destacar que los procesos fueron los mismos en los diferentes análisis por lo cual se tomara como ejemplo explicativo el realizado con los datos por estación del año directamente.

Como se muestra en la figura 3.1-1 con los datos establecidos se crea una matriz, se escoge la métrica de distancia que mejor convenga, es este caso se optó por la Euclídeana Cuadrada, la cual es expresada como la raíz cuadrada de la suma de los cuadrados de las diferencias entre los valores de los elementos, ésta es la medida por defecto que suele usarse para datos de intervalo. Se obtiene la matriz de distancias o también llamada matriz de disimilaridad que viene a ser lo mismo, el método de conglomeración aplicado en el análisis es el de Ward el cual mide la proximidad entre dos grupos calculando las distancias entre sus centroides utilizando las medias de todas las variables. Cabe destacar que el método de Ward es menos sensible a outliers o valores atípicos y tiene tendencias a formar agrupaciones más compactas y de igual tamaño, además dicho método pertenece a los métodos jerárquicos aglomerativos y utiliza diversos criterios para determinar en cada paso del algoritmo a los grupos que se deben unir.

Figura 3.1-1 Diagrama de procesos en el análisis de conglomerados.



En la tabla 3.1-1 se muestra el historial de conglomerados, los diferentes conglomerados o grupos se crean a partir de una matriz de distancias entre los individuos o en este caso puntos de muestreo, por medio de un algoritmo de conglomerado jerárquico el cual agrupa a individuos parecidos en un grupo convirtiendo a dicho grupo en un individuo virtual el cual ira agrupando con otros individuos originales o virtuales hasta generar un único grupo en función a la distancia en que los mismo se han ido uniendo. En este caso cada punto de muestreo es representado por un número, en la primera etapa se agrupan los puntos de muestreo 5 y 7, el signo negativo representa a los puntos de muestreo individuales no pertenecientes a ningún grupo, la primera etapa se puede visualizar hasta la fila 5, en la segunda etapa en la fila 6 se puede apreciar que la estación de muestreo 4 se agrupa al conglomerado 4 este proceso se

realiza hasta conectar todos los puntos en función de la distancia obteniendo como resultado lo mostrado en el dendrograma de la figura 3.1-2.

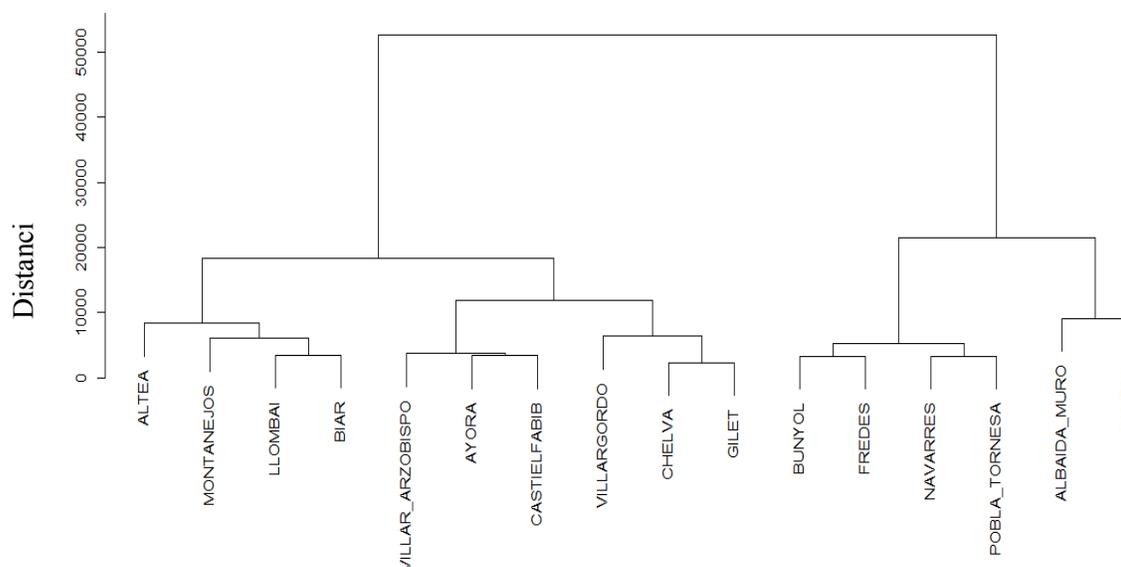
El dendrograma representa la estructura de los datos, teniendo en cuenta la forma en que se van anidando los clusters en función de la distancia, siendo una herramienta útil a la hora de tomar decisiones y elegir las agrupaciones que representen de mejor forma las agrupaciones.

Tabla 3.1-1 Historial de conglomerados.

	[,1]	[,2]		[,1]	[,2]
[1,]	-5	-7	[9,]	-8	1
[2,]	-2	-15	[10,]	-13	8
[3,]	-10	-16	[11,]	-6	-11
[4,]	-1	-9	[12,]	6	9
[5,]	-3	-12	[13,]	10	12
[6,]	-4	4	[14,]	7	11
[7,]	2	3	[15,]	13	14
[8,]	-14	5			

El dendrograma de la figura 3.1-2 muestra la jerarquización, agrupación y anidamiento de los individuos en función de la variable, en este caso de los puntos de muestreo con datos del porcentaje de humedad de la especie vegetal romero, se puede observar que los puntos Navarres y Pobla – Tornesa forman un grupo principal al igual que Buñol y Fredes, mismo que comparten afinidad por lo que en otro nivel se vuelve integrantes de una sola clase.

Figura 3.1-2 Dendrograma con los datos de HCV de la especie vegetal romero.



### 3.1.1 Identificación de número de conglomerados óptimos.

Antes de proceder a identificar los conglomerados que presenten una buena homogeneidad entre los diferentes análisis tal como se muestra en la figura 3.1, es necesario identificar el número de conglomerados con el que se va a trabajar, para lo cual se realizaron varias pruebas, tal como se muestra en la tabla 3.1.1-1.

De manera general se consideró analizar las agrupaciones comprendidas entre 3 y 8, un mayor o menor número de agrupaciones se cree innecesario comparar ya que clases donde el máximo número de puntos este comprendido entre uno y dos miembros o puntos de muestreo no es lo que se está buscando, de igual forma clasificarlos en dos o menos conglomerados se generaliza y abarcarían gran cantidad de puntos.

De todas formas al observar la tabla 3.1.1-1 podemos ver que los conglomerados de 3 agrupan en una sola clase a la mayoría de puntos, por lo que también es descartado, los conglomerados 6, 7 y 8 conservan grupos con un solo integrante lo cual hace notar grupos muy desiguales y lo que se trata es que queden bastante parejos en el número de integrantes de cada conglomerado evitando también que en un grupo exista una sola variable.

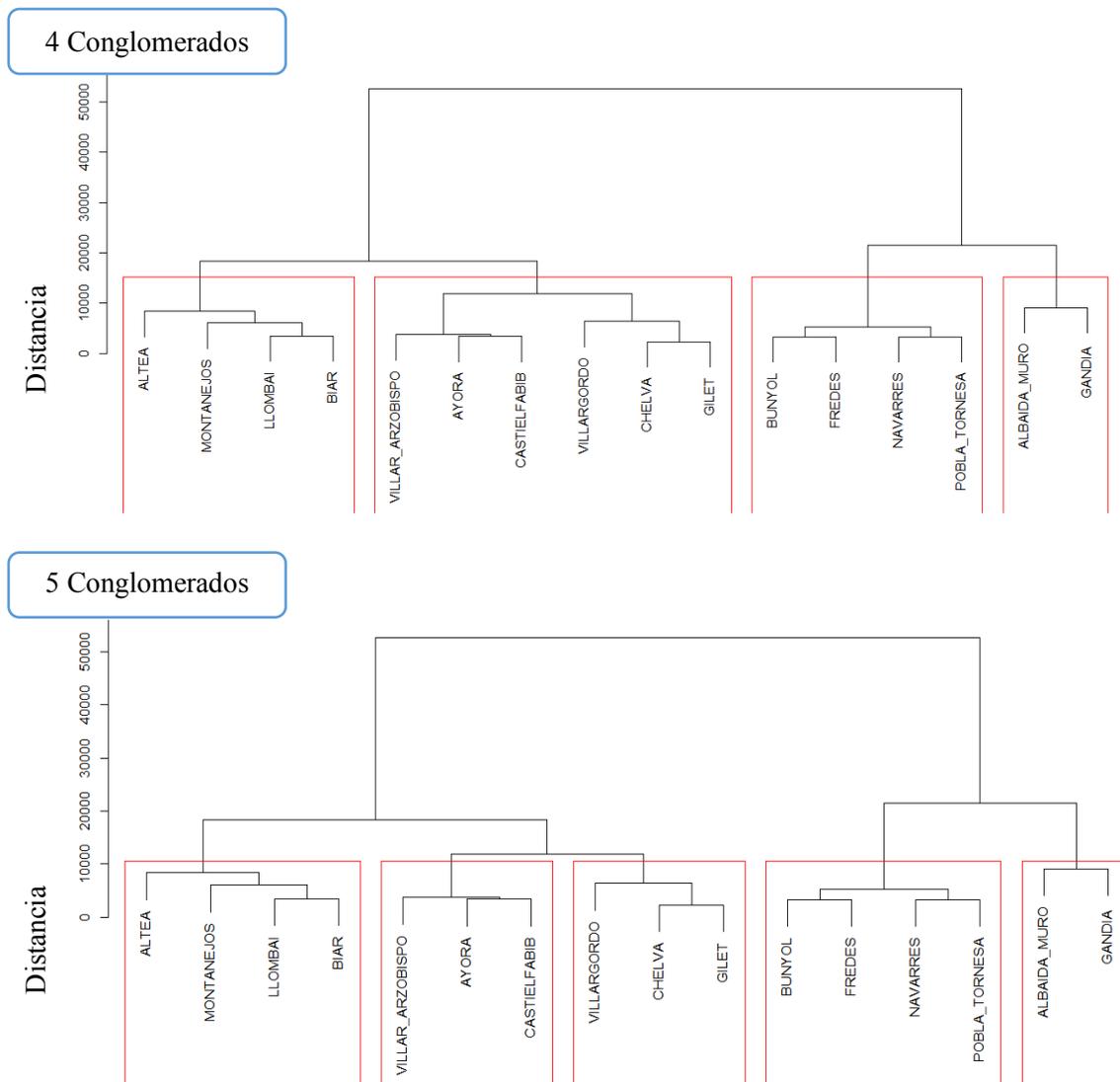
Tabla 3.1.1-1 diferencia entre número de conglomerados.

3 conglomerados		4 conglomerados		5 conglomerados	
Miembros	Porcentaje	Miembros	Porcentaje	Miembros	Porcentaje
10	62,50	6	37,50	3	18,75
4	25,00	4	25,00	4	25,00
2	12,50	4	25,00	4	25,00
		2	12,50	3	18,75
				2	12,50
6 conglomerados		7 conglomerados		8 conglomerados	
Miembros	Porcentaje	Miembros	Porcentaje	Miembros	Porcentaje
3	18,75	3	18,75	3	18,75
4	25,00	4	25,00	4	25,00
4	25,00	3	18,75	3	18,75
3	18,75	3	18,75	2	12,50
1	6,25	1	6,25	1	6,25
1	6,25	1	6,25	1	6,25
		1	6,25	1	6,25
				1	6,25

Como resultado final en la figura 3.1.1-1 se presenta el dendograma obtenido de la variable humedad de la especie romero por estación del año con 4 y 5 conglomerados, se conservan estas dos agrupaciones con la finalidad de hacer un estudio y encontrar el que mejor convenga de acuerdo a las agrupaciones que se van generando.

Cabe desatacar que dicha resolución se tomó observando los demás análisis de conglomerados realizado, tal como se explicó al principio del apartado 3 y correspondiente a la figura 3-1.

Figura 3.1.1-1 Dendograma presentando 4 y 5 conglomerados.



### *3.1.2 Análisis de los conglomerados con las diferentes variables.*

El número de conglomerados escogido cumple con el propósito de agrupar a los puntos de muestreo en un número bastante homogéneo y comportamiento de acuerdo a la humedad de la especie vegetal romero muestreado en cada punto en el periodo de 3 años analizado desde diferentes perspectivas.

El estudio presentado a continuación tiene la finalidad de encontrar diferencias al realizar conglomerados y buscar la que mejor relacione los puntos de muestreo y equilibre la cantidad de puntos de cada agrupación.

Los análisis de conglomerados realizados con los 4 estadísticos se contemplan como una opción al no poder trabajar con los datos generales, y se trata de buscar viabilidad y cuan diferente se presenta en comparación con los demás. El análisis realizado con los estadísticos de forma individual tiene el propósito de estudiar cómo interactúan solos y si se puede identificar como influye dicho estadístico en el análisis donde se los agrupan a todos ellos.

Este análisis se lo realizara en dos partes, el primero será tomando en cuenta los datos por estaciones del año y sus respectivos estadísticos, mientras que el segundo análisis será realizado solo con los estadísticos de los datos por mes, tal como se presentó en el esquema de trabajo figura 3-1.

#### *3.1.2.1 Análisis con los datos de humedad de la especie vegetal romero por estación del año.*

Se realizó el análisis de conglomerados de 4 y 5 agrupaciones siguiendo el procedimiento mostrado en el apartado 3.1.1, como premisa se observa el número de integrantes de cada grupo o clase.

En la tabla 3.1.2.1-1 se puede observar que los grupos de 4 clases formados por los datos directos por estación del año, por agrupación de los 4 estadísticos de posición, estadístico promedio y cuartil inferior presentan grupos con número de individuos semejantes.

El número de puntos de muestreo por cada clase es igual entre el análisis realizado con el grupo de estadístico y promedio.

Los estadísticos mediana y cuartil superior presentan conglomerados con número de integrantes o individuos por clase muy irregulares al punto de tener un solo individuo en una de sus agrupaciones.

Tabla 3.1.2.1-1 Conglomerados de 4 clases de los diferentes análisis con los datos de HCV de la especie vegetal romero por estación del año.

Conglomerado	Estación del año		Estadís de posición		Promedio	
	<i>Num miembros</i>	%	<i>Num miembros</i>	%	<i>Num miembros</i>	%
1	6	37,50	4	25,00	4	25,00
2	4	25,00	6	37,50	6	37,50
3	4	25,00	4	25,00	4	25,00
4	2	12,50	2	12,50	2	12,50
Conglomerado	Mediana		Cuartil Inferior		Cuartil Superior	
	<i>Num miembros</i>	%	<i>Num miembros</i>	%	<i>Num miembros</i>	%
1	6	37,50	6	37,50	6	37,50
2	4	25,00	3	18,75	7	43,75
3	5	31,25	4	25,00	2	12,50
4	1	6,25	3	18,75	1	6,25

La tabla 3.1.2.1-2 muestra el análisis de conglomerados divididas en 5 clases, en donde los resultados del número de integrantes por agrupación se mantiene pero presenta una variación aunque el cuartil inferior tiene un número de integrantes parejo en 3 agrupaciones, mantiene a una con un valor muy alto y otra con uno muy pequeño, es decir la clase con 6 individuos permanece igual que en la división de conglomerados por 4 clases, y crea una clase con un solo individuo.

Los conglomerados realizados con el estadístico, mediana y cuartiles inferiores y superior agrupa en una clase la mayor cantidad de puntos y otra en donde solo existe un integrantes, comparado con los otros 3 que presenta una mejor distribución son descartados.

Tabla 3.1.2.1-2 Conglomerados de 5 clases de los diferentes análisis con los datos de HCV de la especie vegetal romero por estación del año.

Conglomerado	Estación del año		Estadís de posición		Promedio	
	<i>Num miembros</i>	%	<i>Num miembros</i>	%	<i>Num miembros</i>	%
1	3	18,75	4	25,00	4	25,00
2	4	25,00	4	25,00	4	25,00
3	4	25,00	4	25,00	4	25,00
4	3	18,75	2	12,50	2	12,50
5	2	12,50	2	12,50	2	12,50
Conglomerado	Mediana		Cuartil Inferior		Cuartil Superior	
	<i>Num miembros</i>	%	<i>Num miembros</i>	%	<i>Num miembros</i>	%
1	4	25,00	6	37,50	6	37,50
2	4	25,00	3	18,75	5	31,25
3	5	31,25	3	18,75	2	12,50
4	2	12,50	3	18,75	2	12,50
5	1	6,25	1	6,25	1	6,25

En la tabla 3.1.2.1-3 se observa los puntos de muestreo y la agrupación a la que pertenece de acuerdo a los datos con los que se realizó el análisis de conglomerados, el punto de muestreo Ayora y Villar-Arzobispo son las únicas que se mantienen en un mismo grupo en los 6 análisis realizado sin embargo el resto no parece seguir un patrón específico, aunque se puede notar que los puntos se mueven en grupos específicos como por ejemplo Chelva, Gilet, Villargordo y Castielfabib pertenecen a al mismo grupo en los diferentes análisis, Buñol y Fredes presentan el mismo patrón todo lo contrario de Buñol y Fredes clasificados en su mayor parte en el grupos 2, por otro lado el resto de puntos es ubicado en cualquiera de los grupos de forma individual.

En los párrafos anteriores se habló de la similitud del número de clases entre el análisis de conglomerados realizado a partir de la agrupación de los estadísticos y el elaborado con el estadístico individual promedio, se puede notar que los puntos de muestreo además son colocadas en los mismos grupos, se recalca este hecho ya que los mismos comparten similitud con el análisis partiendo de los datos por estación de año directamente sin embargo son solo siete puntos los que son colocados en el mismo grupo.

El número de miembros por clase que destaco también fue el realizado con el cuartil inferior sin embargo en la distribución de los puntos de muestreo es el que más difiere con respecto al resto.

Los análisis realizados con el estadístico individual mediana y cuartil superior, puede ver que

los 9 primeros puntos de muestreo son colocados en las mismas clases, sin embargo al buscar algún patrón ya observando las gráficas lineales y los diagramas de cajas y bigotes no se puede ver ningún patrón específico de relación.

En la tabla 3.1.2.1-4 se observan las estaciones divididas en 5 clases en donde se puede apreciar que las semejanzas presentes en el análisis dividido por 4 clases son en menor número, la distribución de los puntos de muestreo entre el análisis realizado por el agrupamiento de estadísticos y el promedio prevalecen sin embargo la distribución de variables por análisis no muestra ningún tipo de patrón.

Tabla 3.1.2.1-3 Puntos de muestreo agrupados en los conglomerados de 4 clases.

Num	Puntos de muestreo	EstaA	EstaPos	EstaProm	EstaMed	EstaQ1	EstaQ3
1	AYORA	1	1	1	1	1	1
2	BUNYOL	2	2	2	2	1	2
3	LLOMBAI	3	1	1	2	1	2
4	VILLAR_ARZOBISPO	1	1	1	1	1	1
5	CHELVA	1	3	3	1	2	1
6	ALBAIDA_MURO	4	4	4	3	3	3
7	GILET	1	3	3	1	2	1
8	VILLARGORDO	1	3	3	1	2	1
9	CASTIELFABIB	1	3	3	1	4	1
10	NAVARRRES	2	2	2	3	3	2
11	GANDIA	4	4	4	4	3	3
12	BIAR	3	1	1	2	4	2
13	ALTEA	3	2	2	3	4	2
14	MONTANEJOS	3	2	2	3	1	2
15	FREDES	2	2	2	2	1	2
16	POBLA_TORNESA	2	2	2	3	3	4

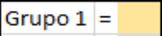
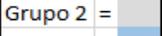
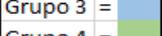
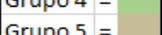
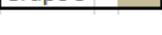
  

EstaA	=	Datos HCV por estación del año	Grupo 1	=	
EstaPos	=	Estadísticos agrupados de posición	Grupo 2	=	
EstaProm	=	Estadístico promedio	Grupo 3	=	
EstaMed	=	Estadístico mediana	Grupo 4	=	
EstaQ1	=	Estadístico cuartil inferior			
EstaQ3	=	Estadístico cuartil superior			

Tabla 3.1.2.1-4 Puntos de muestreo agrupados en los conglomerados de 5 clases.

Num	Puntos de muestreo	EstaA	EstaPos	EstaProm	EstaMed	EstaQ1	EstaQ3
1	AYORA	1	1	1	1	1	1
2	BUNHOL	2	2	2	2	1	2
3	LLOMBAI	3	1	1	2	1	3
4	VILLAR_ARZOBISPO	1	1	1	1	1	1
5	CHELVA	4	3	3	1	2	1
6	ALBAIDA_MURO	5	4	4	3	3	4
7	GILET	4	3	3	4	2	1
8	VILLARGORDO	4	3	3	4	2	1
9	CASTIELFABIB	1	3	3	1	4	1
10	NAVARRRES	2	5	5	3	3	2
11	GANDIA	5	4	4	5	5	4
12	BIAR	3	1	1	2	4	3
13	ALTEA	3	2	2	3	4	2
14	MONTANEJOS	3	2	2	3	1	2
15	FREDES	2	2	2	2	1	2
16	POBLA_TORNESA	2	5	5	3	3	5

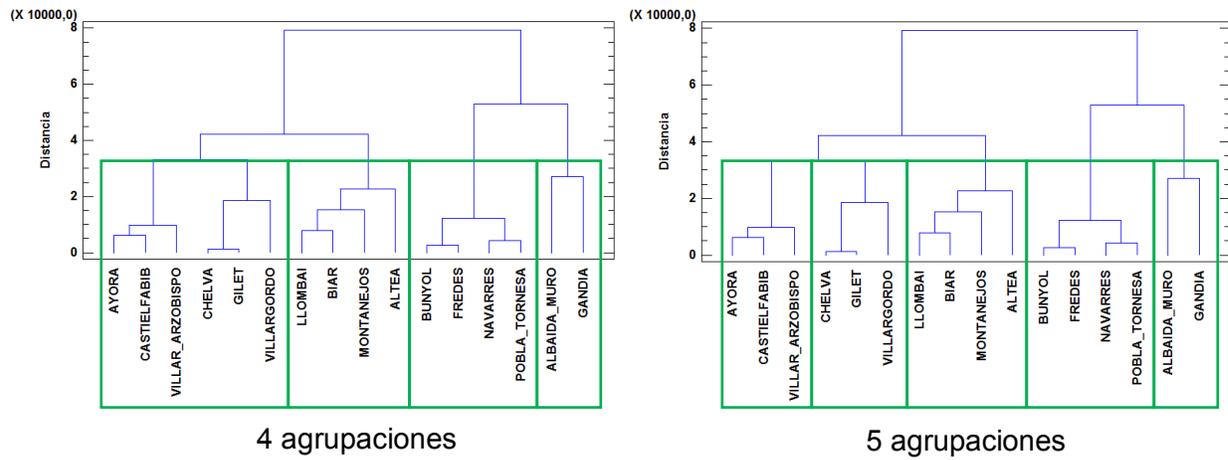
  

EstaA	= Datos HCV por estación del año	Grupo 1	=	
EstaPos	= Estadísticos agrupados de posición	Grupo 2	=	
EstaProm	= Estadístico promedio	Grupo 3	=	
EstaMed	= Estadístico mediana	Grupo 4	=	
EstaQ1	= Estadístico cuartil inferior	Grupo 5	=	
EstaQ3	= Estadístico cuartil superior			

A continuación se presentan los dendrograma de 4 y 5 clases de cada análisis de conglomerados con la finalidad de identificar las diferentes agrupaciones y anidamientos realizadas.

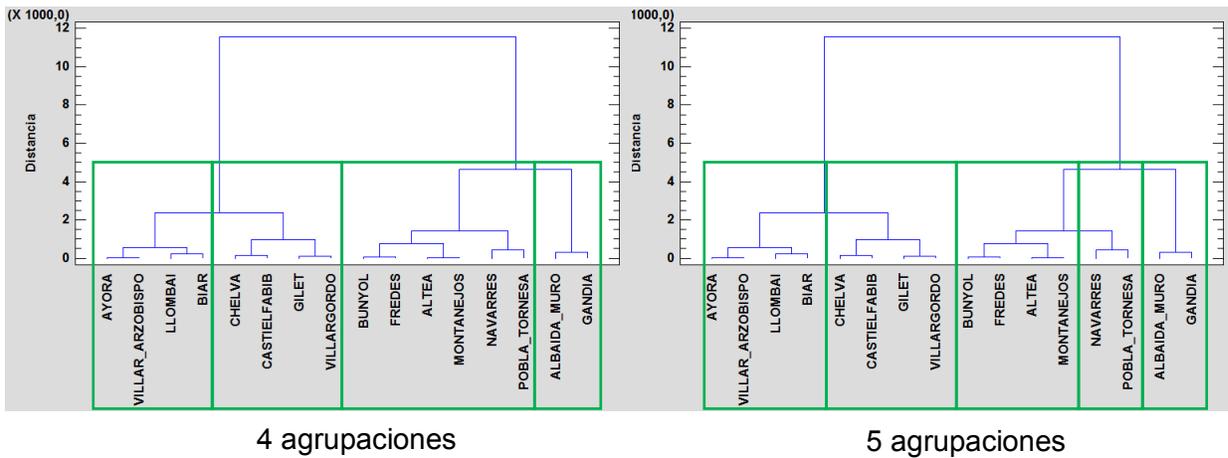
En el figura 3.1.2.1-1 se muestran los diferentes dendrograma obtenidos de los diferentes análisis, los dos primeros gráficos muestran el agrupado jerárquico realizado a partir de los datos de humedad de la especie vegetal romero por estación del año donde se puede observar que las agrupaciones son bastantes simétricas ya que al pasar de 4 a 5 clases, este divide al grupo de mayor número de integrantes en dos, dicho análisis muestra un anidamiento muy bien distribuido de los 16 puntos de muestreo.

Figura 3.1.2.1-1 Dendograma de los análisis realizados con los datos por estación de año.



La figura 3.1.2.1-2 muestra el dendograma obtenido del análisis con los estadísticos de posición agrupados, el cual presenta una distribución de los puntos bastante homogénea.

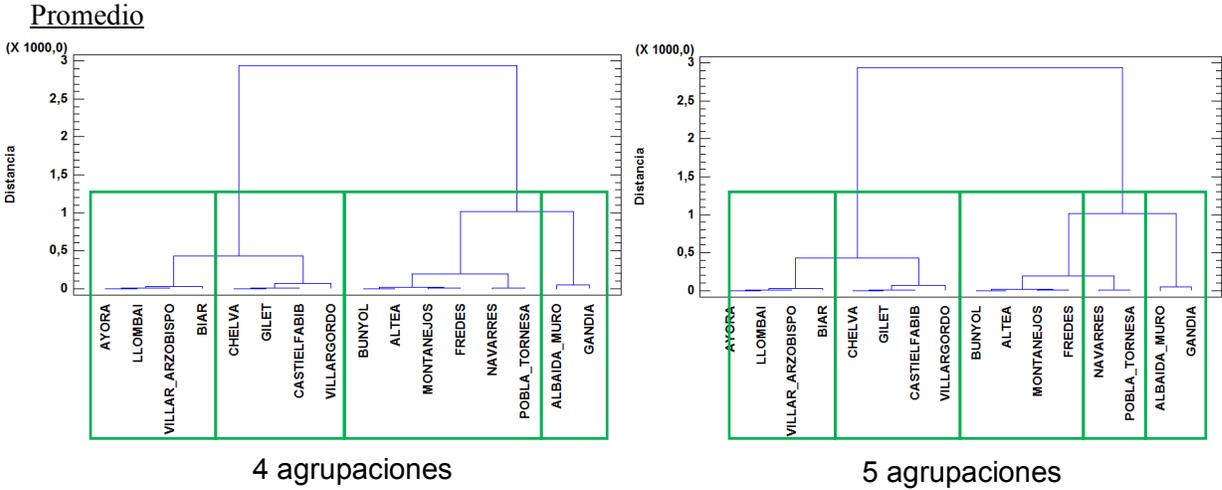
Figura 3.1.2.1-2 Dendograma de los análisis realizados con los estadísticos de posición agrupados.



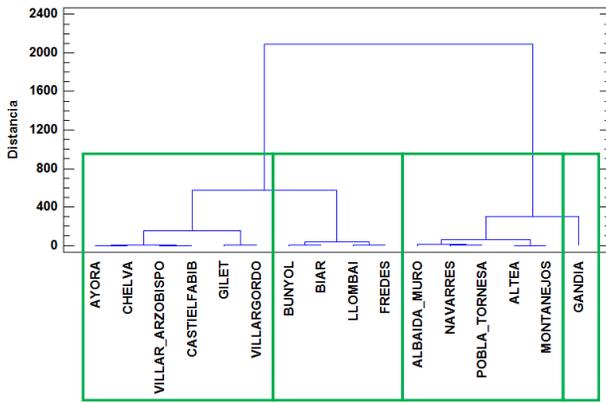
En las figuras 3.1.2.1-3 y 3.1.2.1-4 se muestran a todos los estadísticos con los que se realizó análisis de forma individual y se observa que las agrupaciones presentan variaciones notorias, tomando en cuenta que los estadísticos son obtenidos de una misma variable, la forma en que

las mismas son anidadas presentan cambios claros. Es importante recalcar que estos análisis se hicieron con la finalidad de encontrar una relación con el análisis donde se tomaron todos los estadísticos en grupo. El número de individuos por clase y la distribución de los puntos de muestreo existentes entre el análisis de conglomerados con todos los estadísticos y el promedio, denotaba una relación entre las mismas sin embargo si se comparan los dendograma de los mismos claramente se puede ver que los anidamientos son muy diferentes puede ser que en función de sus matrices de distancias creadas exista una semejanza el cual hace que estas sean parecidas más se puede descartar la idea de que promedio es la variable que más influye de los 4 estadísticos tomados, quedado claro que el análisis de conglomerados depende en su totalidad de la métrica de distancias y el método de conglomerados aplicado.

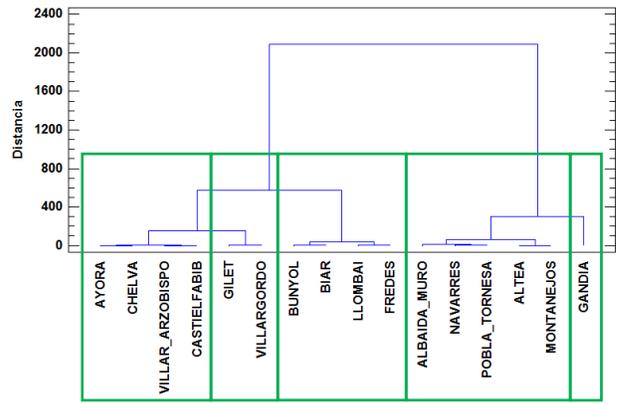
Figura 3.1.2.1-3 Dendograma de los análisis realizados con los estadísticos de forma individual.



Mediana

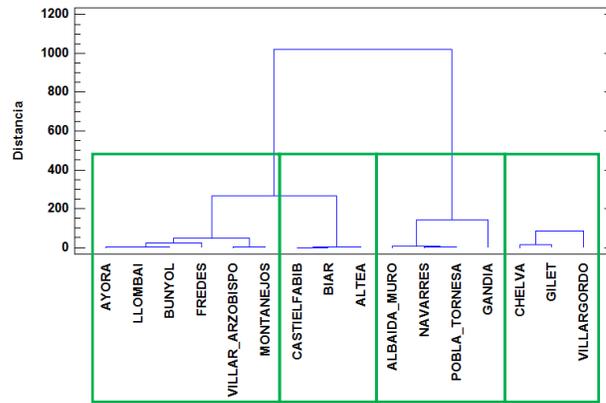


4 agrupaciones

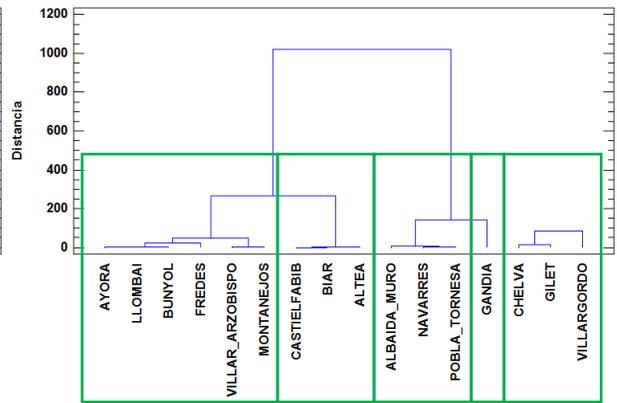


5 agrupaciones

Cuartil Inferior

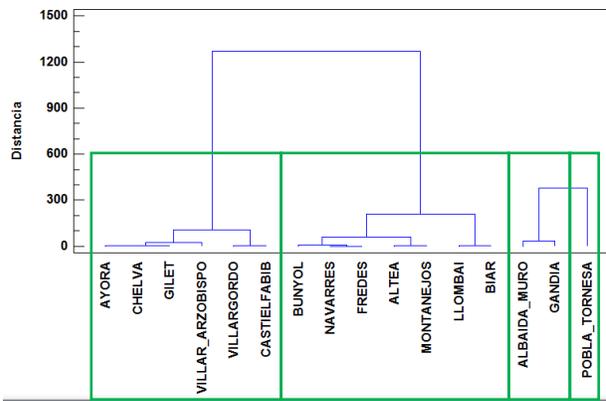


4 agrupaciones

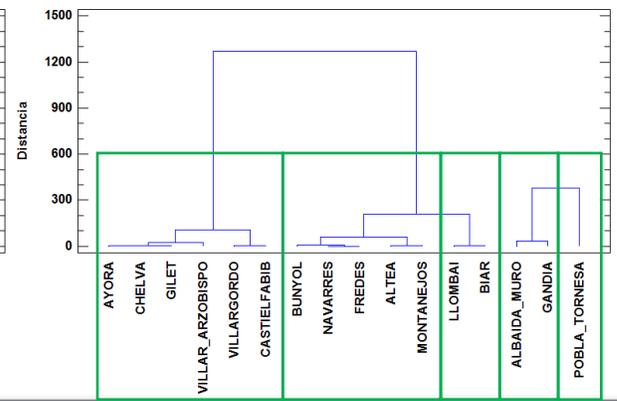


5 agrupaciones

Cuartil Superior



4 agrupaciones



5 agrupaciones

El análisis de conglomerados realizado con los datos por estación del año y sus estadísticos mostro que, los datos por estación de forma directa, el agrupado de los estadísticos y los de estadísticos promedio y cuartil inferior de forma individual, para las agrupaciones de 4 clases presentan agrupaciones aceptables. Al tomar en cuenta la división por 5 clases el análisis con el estadístico cuartil inferior es descartado por presentar dos grupos muy dispares en donde uno agrupaba un mayor número de puntos de muestreo y otro solo uno.

Se notó igualdad en el número de integrantes de las clases formadas por los análisis realizados con los estadísticos agrupados y el promedio, al revisar las tablas 3.1.2.1-3 y 3.1.2.1-4 se visualiza los puntos de muestreo y la clase a la que pertenece, de los 6 análisis realizados donde coinciden los puntos agrupados por clases de estos dos análisis, sin embargo al revisar los dendogramas de los mismos se puede ver que el anidamiento es diferente uno con respecto al otro, coinciden en la formación de clases, mas en el agrupamiento y jerarquización se muestran diferencias que serán más notorias en un número mayor de conglomerados o clases, quedado claro que el análisis de conglomerados depende en su totalidad de la métrica de distancias y el método de conglomerados aplicado.

Al final los análisis de conglomerados de 4 y 5 clases formados por los datos por estación del año, el grupo de estadísticos y promedio presentan un similar número de integrantes en todos los grupos.

### *3.1.2.2 Análisis con los estadísticos de humedad de la especie vegetal romero por mes.*

Como se explicó al principio de este capítulo ahora se expone los análisis de conglomerados realizados con los estadísticos directos de la variable humedad por mes, se realizaron 5 análisis, uno menos que el anterior ya que no se pudo tomar los datos directos por no estar completos.

La tabla 3.1.2.2-1 indica los diferentes análisis con el número de miembros en cada conglomerado, de los cuales ninguno muestra similitud en el número de individuos o miembros dentro de los mismos, sin embargo es importante destacar en todos los análisis la clase o conglomerado 1 agrupa una gran cantidad de puntos de muestreo.

Los que mejor homogeneidad presentan entre ellos son los análisis donde se agrupan los

estadísticos y promedio además muestran número de clases con integrantes en cada una de ellas muy semejantes.

En la tabla 3.1.2.2-2 el cual contiene los resultados de estaciones agrupadas por 5 clases, los que fueron realizados con los estadísticos agrupados, promedio y mediana mantienen al conglomerado o clase 1 con el mismo número de puntos de muestreo mientras que en el resto de clases son distribuidos de mejor forma, los que fueron formados con los cuantiles inferior y superior distribuyen mejor los puntos dentro de las clases. La semejanza entre del número de individuos dentro de las clases formadas con los estadísticos agrupados y el promedio son semejantes.

Tabla 3.1.2.2-1 Conglomerados de 4 clases de los diferentes análisis con los datos de HCV de la especie vegetal romero por mes.

Conglomerado	Estadís de posición		Promedio		Mediana	
	Num miembros	%	Num miembros	%	Num miembros	%
1	6	37,50	6	37,50	7	43,75
2	3	18,75	5	31,25	3	18,75
3	2	12,50	2	12,50	2	12,50
4	5	31,25	3	18,75	4	25,00
Conglomerado	Cuartil Inferior		Cuartil Superior			
	Num miembros	%	Num miembros	%		
1	8	50,00	7	43,75		
2	3	18,75	3	18,75		
3	3	18,75	2	12,50		
4	2	12,50	4	25,00		

Tabla 3.1.2.2-2 Conglomerados de 5 clases de los diferentes análisis con los datos de HCV de la especie vegetal romero por mes.

Conglomerado	Estadís de posición		Promedio		Mediana	
	Num miembros	%	Num miembros	%	Num miembros	%
1	6	37,50	6	37,50	7	43,75
2	3	18,75	3	18,75	2	12,50
3	2	12,50	2	12,50	2	12,50
4	3	18,75	2	12,50	4	25,00
5	2	12,50	3	18,75	1	6,25
Conglomerado	Cuartil Inferior		Cuartil Superior			
	Num miembros	%	Num miembros	%		
1	3	18,75	4	25,00		
2	5	31,25	3	18,75		
3	3	18,75	3	18,75		
4	3	18,75	2	12,50		
5	2	12,50	4	25,00		

En las tablas 3.1.2.2-3 y 3.1.2.2-4 se muestran los análisis de conglomerados con 4 y 5 clases respectivamente con los puntos de muestreos correspondiente a cada uno de ellos, los análisis realizados con los estadísticos de forma individual no guardan alguna relación, si bien es cierto que se puede observar que los 8 primeros puntos de muestreo son clasificadas en los mismos conglomerados en los análisis con los estadísticos agrupados y promedio no se puede saber la influencia de las mismas.

Como se expresó anteriormente los conglomerados de 4 clases tienen un grupo dominante en donde se concentran una gran cantidad de puntos de muestreo, en los de 5 clases se mantienen dichos grupos dominantes en los análisis de estadísticos agrupados, y los estadísticos individuales promedio y mediana, los conglomerados obtenidos de los cuantiles presentan una mejor distribución de puntos entre las clases. Cabe destacar que el análisis de conglomerados por estadísticos separados se lo realizó con el afán de estudiar las diferencias y semejanzas con el análisis en donde se engloban los 4 estadísticos.

De todas formas se procede a observar el anidado de los conglomerados por medio de los dendograma.

Tabla 3.1.2.2-3 Puntos de muestreo agrupados en los conglomerados de 4 clases.

Num	Puntos de muestreo	EstaPos	EstaProm	EstaMed	EstaQ1	EstaQ3
1	AYORA	1	1	1	1	1
2	BUNYOL	1	1	1	1	2
3	LLOMBAI	1	1	1	1	2
4	VILLAR_ARZOBISPO	1	1	1	1	1
5	CHELVA	2	2	1	2	1
6	ALBAIDA_MURO	3	3	2	3	3
7	GILET	2	2	3	2	1
8	VILLARGORDO	2	2	3	2	1
9	CASTIELFABIB	1	2	1	1	1
10	NAVARRRES	4	4	4	1	2
11	GANDIA	3	3	2	3	3
12	BIAR	1	2	1	4	1
13	ALTEA	4	1	4	4	4
14	MONTANEJOS	4	1	4	1	4
15	FREDES	4	4	4	1	4
16	POBLA_TORNESA	4	4	2	3	4

EstaPos	= Estadísticos agrupados de posición	Grupo 1	=
EstaProm	= Estadístico promedio	Grupo 2	=
EstaMed	= Estadístico mediana	Grupo 3	=
EstaQ1	= Estadístico cuartil inferior	Grupo 4	=
EstaQ3	= Estadístico cuartil superior		

Tabla 3.1.2.2-4 Puntos de muestreo agrupados en los conglomerados de 5 clases.

Num	Puntos de muestreo	EstaPos	EstaProm	EstaMed	EstaQ1	EstaQ3
1	AYORA	1	1	1	1	1
2	BUNHOL	1	1	1	2	2
3	LLOMBAI	1	1	1	2	2
4	VILLAR_ARZOBISPO	1	1	1	2	1
5	CHELVA	2	2	1	3	3
6	ALBAIDA_MURO	3	3	2	4	4
7	GILET	2	2	3	3	1
8	VILLARGORDO	2	2	3	3	3
9	CASTIELFABIB	1	4	1	2	3
10	NAVARRES	4	5	4	1	2
11	GANDIA	3	3	2	4	4
12	BIAR	1	4	1	5	1
13	ALTEA	5	1	4	5	5
14	MONTANEJOS	5	1	4	2	5
15	FREDES	4	5	4	1	5
16	POBLA_TORNESA	4	5	5	4	5

EstaPos	= Estadísticos agrupados de posición	Grupo 1	=
EstaProm	= Estadístico promedio	Grupo 2	=
EstaMed	= Estadístico mediana	Grupo 3	=
EstaQ1	= Estadístico cuartil inferior	Grupo 4	=
EstaQ3	= Estadístico cuartil superior	Grupo 5	=

En la figura 3.1.2.2-1 se muestra en dendograma del análisis de mayor interés en esta etapa ya que fue realizado con los estadísticos agrupado, el anidamiento de los puntos de muestreo para conglomerados de 4 y 5 clases mantiene una de ellas con 6 puntos agrupados, teniendo una desigualdad en cantidad por clase incumpliendo uno de los parámetros requeridos. Se puede observar que al realizar conglomerados con 6 clases mejoraría la distribución de los mismos, sin embargo la mitad de las clases serian grupos de dos individuos y lo que se trata de evitar es conglomerados con pocos integrantes.

La figura 3.1.2.2-2 muestra los estadísticos analizados de forma individual se comportan muy diferentes en relación del primero, además los agrupamientos son muy desiguales.

En este grupo de análisis destacan los realizados con los cuartiles ya que el anidamiento presenta una mejor distribución de las variables, sin embargo al ser el primer y el tercer cuartil analizados de forma individual solo destacan clases tomando en cuenta el valor en donde 25% y 75% están por debajo del mismo y el 75% y 25% por encima respectivamente. De forma individual no expresan mucho de la variable humedad analizada por lo que cabe destacar que el realizar análisis por separado tiene la finalidad de observar alguna relación e interacción con

el análisis de conglomerados donde fueron agrupados los mismos.

Figura 3.1.2.2-1 Dendrograma de los análisis realizados con los estadísticos de posición agrupados.

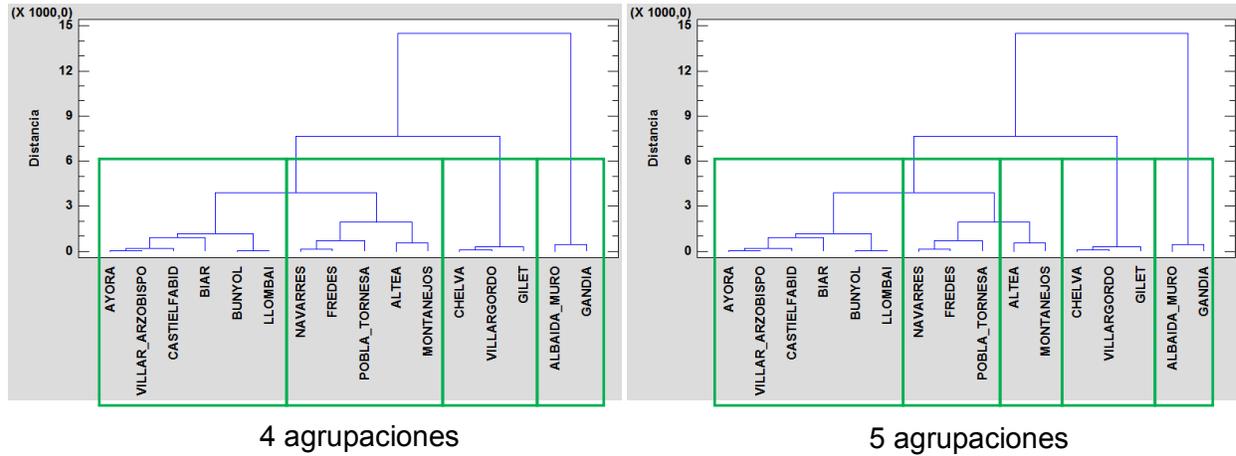
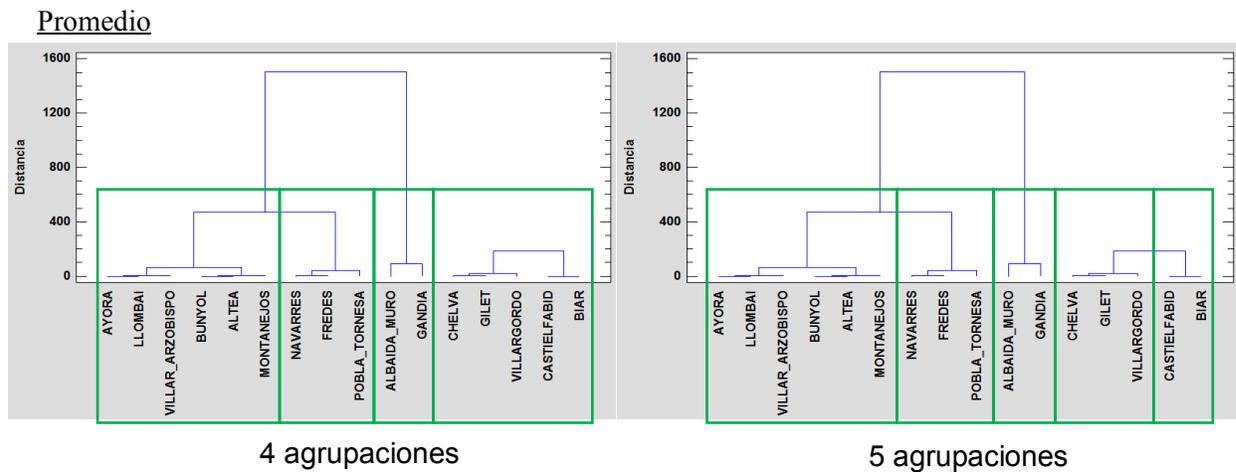
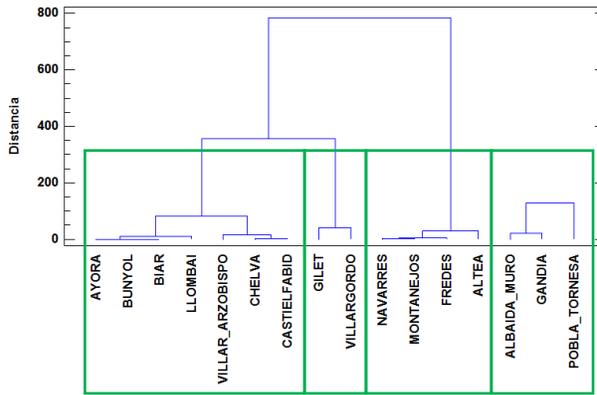


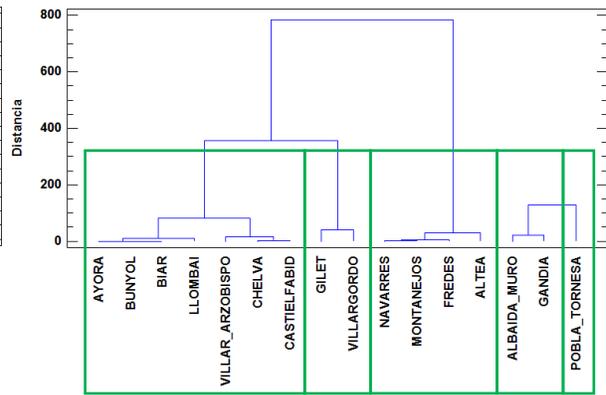
Figura 3.1.2.2-2 Dendrograma de los análisis realizados con los estadísticos de forma individual.



Mediana

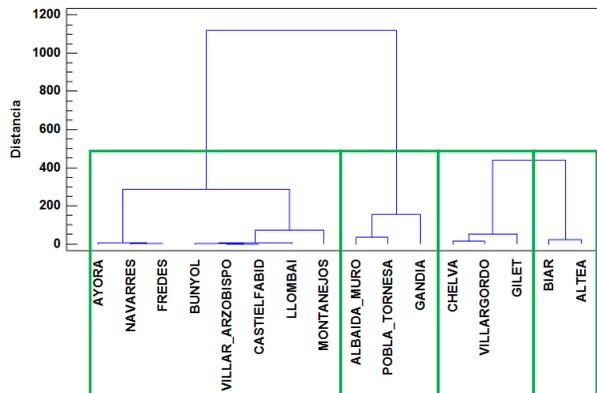


4 agrupaciones

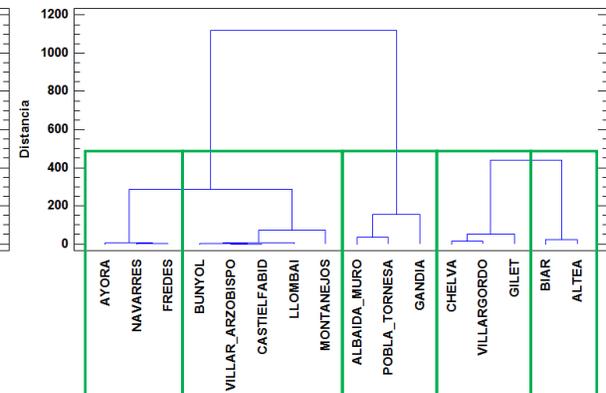


5 agrupaciones

Cuartil Inferior

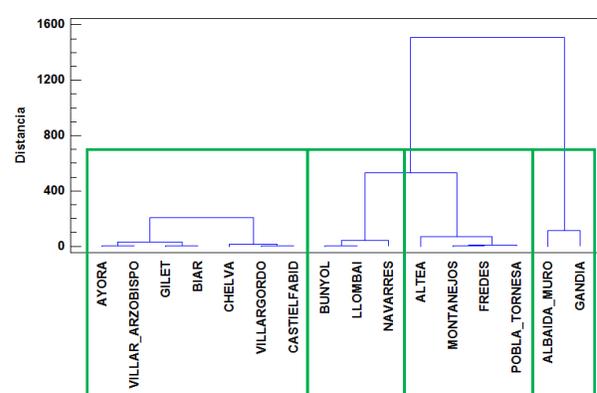


4 agrupaciones

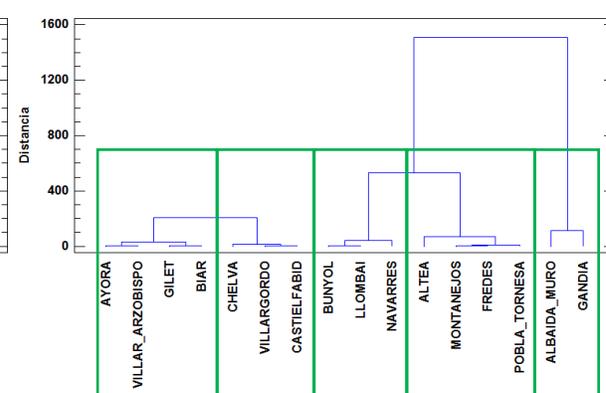


5 agrupaciones

Cuartil Superior



4 agrupaciones



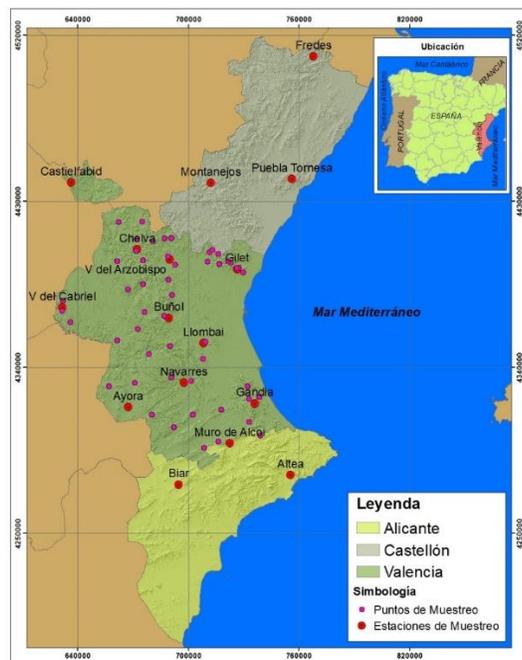
5 agrupaciones

### 3.2 Valoración del porcentaje de humedad en función de los percentiles

Mediante los 16 puntos de muestreo con datos del porcentaje de humedad de la especie vegetal romero en el periodo de 3 años (diciembre del 2014 hasta diciembre de 2017) por estación del año, se procede a realizar una categorización, para lo cual, se toma los datos de cada estación del año (invierno, primavera, verano y otoño) de forma individual y se hace un análisis de conglomerados, las agrupaciones tomadas serán de 5 clases ya que en su mayoría presenta grupos con un número mayor de dos puntos de muestreo y un número de integrantes homogéneo. De cada grupo formado se determinan percentiles, el proceso se realiza de esta forma ya que al identificar puntos de muestreo con un comportamiento similar por estación del año, el cálculo de los percentiles daría una mejor referencia para valoraciones.

Este estudio cuenta con 51 datos tomados en el mes de mayo del 2018 en la provincia de Valencia donde se analizó el porcentaje de humedad de la especie vegetal romero tal como se muestra en la figura 3.2-1, estos puntos serán valorados de acuerdo a los percentiles calculados con el proceso anterior. El análisis utilizado es el realizado con los datos de la primavera.

Figura 3.2-1 Ubicación puntos de muestreo nuevos (mayo 2018).



Fuente: Cartografía del ICV, VAERSA y PATFOR

La categorización se hace partiendo del estadístico percentil, el cual es una medida de posición, en donde se ordenan los datos de la variable de menor a mayor y de acuerdo a su porcentaje se encuentra la posición dentro del conjunto de datos, se determina el valor al que pertenece dicha posición, dividiendo a la variable en dos partes, donde una cantidad de datos estarán por encima y otra por debajo del valor encontrado. Por la naturaleza del estadístico se tomaran varios percentiles ya que cumplen con la función de clasificación requerida.

En estudios anteriores se determinó que la distribución de los puntos de muestreo dentro de la Comunidad Valenciana al ser analizados en conjunto, presentaría resultados muy generales, por lo que se prosiguió a trabajar de la manera presentada a continuación:

El porcentaje de humedad de la especie vegetal romero está ligado a factores externos que lo modifican, como es el caso de variables meteorológicas que al presentarse en periodos de tiempo con características parecidas son catalogados como estaciones del año y se denominan invierno, primavera, verano y otoño. Como se pueden apreciar en gráficas anteriores de líneas temporales donde se observa que el porcentaje de humedad presenta picos muy bajos en verano y altos en primavera se procedió a trabajar de forma independiente con los datos de cada estación del año, es decir invierno, primavera, verano y otoño por separado, esto se realiza tomando en cuenta que existen variaciones muy marcadas y al trabajar con los percentiles ayudara a tener clasificaciones más claras.

Cabe destacar que el agrupar puntos de muestreo con los datos de la variable humedad por estaciones del año pierde información sin embargo ayuda a eliminar valores atípicos mejorando los resultados.

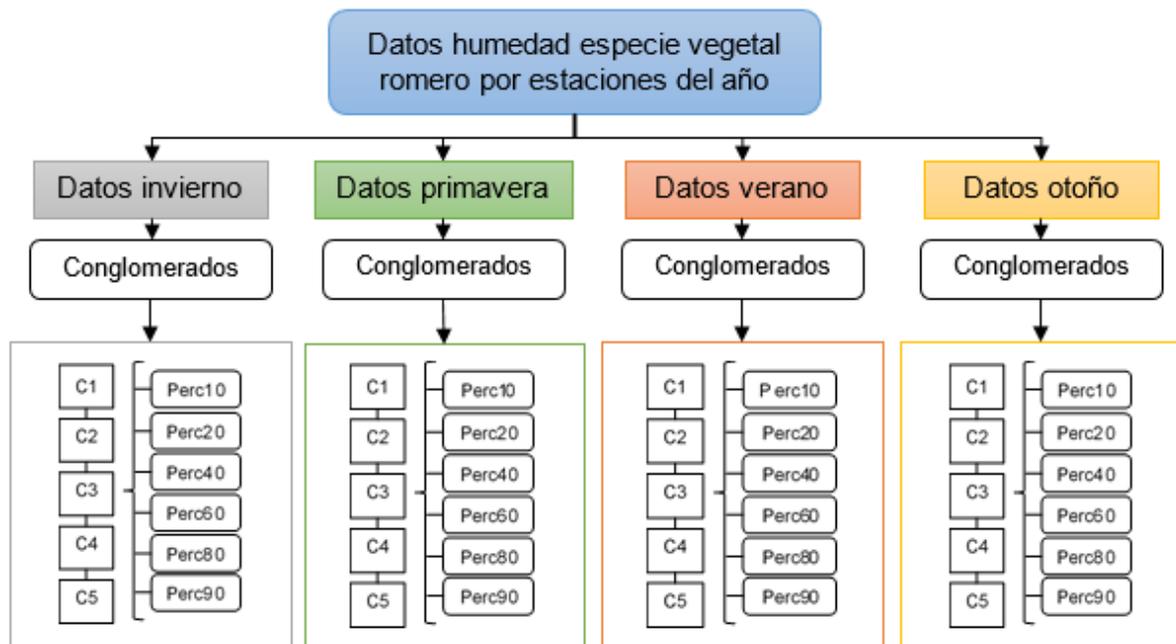
Tal como se presenta en la figura 3.2-2 se procedió a realiza un análisis de conglomerados tomando de forma independiente datos de las estaciones del año y agrupando los puntos de muestreo por semejantes, ya que de igual forma por el área de estudio existen variaciones latentes entre los puntos de muestreo por cada estación del año.

Como resultado se tendrán agrupaciones de puntos de muestreo por cada estación del año, en el apartado 3.1 se realizado un análisis general para la realización de

conglomerados y sus resultados serán tomados en cuenta, es decir el número de clases con las que se trabajara son 5, ya que en su mayoría presenta grupos con un número mayor de dos puntos de muestreo y forma grupos con número de integrantes, lo más homogéneo posible.

Después de tener los puntos de muestreo agrupados en 5 conglomerados por estación del año se determinan los percentiles con los datos de los individuos dentro de cada clase.

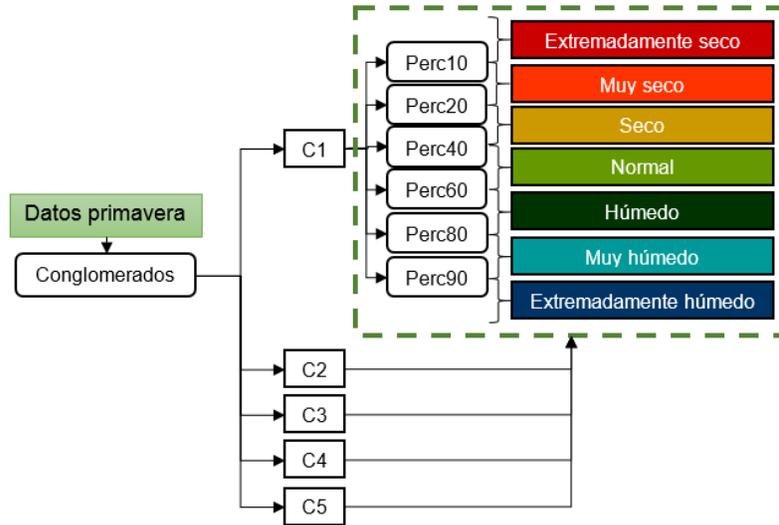
Figura 3.2-2 Análisis de conglomerados por estación del año.



Los percentiles calculados son 10, 20, 40, 60, 80 y 90, y con ello se desean establecer rangos con los que se pueda valorar toma de muestras futuras, partiendo de la naturaleza de cálculo de este estadístico los valores de humedad para la especie vegetal romero que entren por debajo del percentil 10 serán considerados como extremadamente seco, los que se encuentren en el rango de 10 a 20 muy seco, y así sucesivamente hasta llegar al percentil 90 el cual identificara valores extremos de humedad.

Al ser un estudio dentro del ámbito de prevención de incendios, en donde se trata de analizar la humedad de la especie vegetal romero y su subsecuente valoración para ser considerado combustible para un incendio, se considerarán los valores de humedad bajos.

Figura 3.2-3 Percentiles por clase.



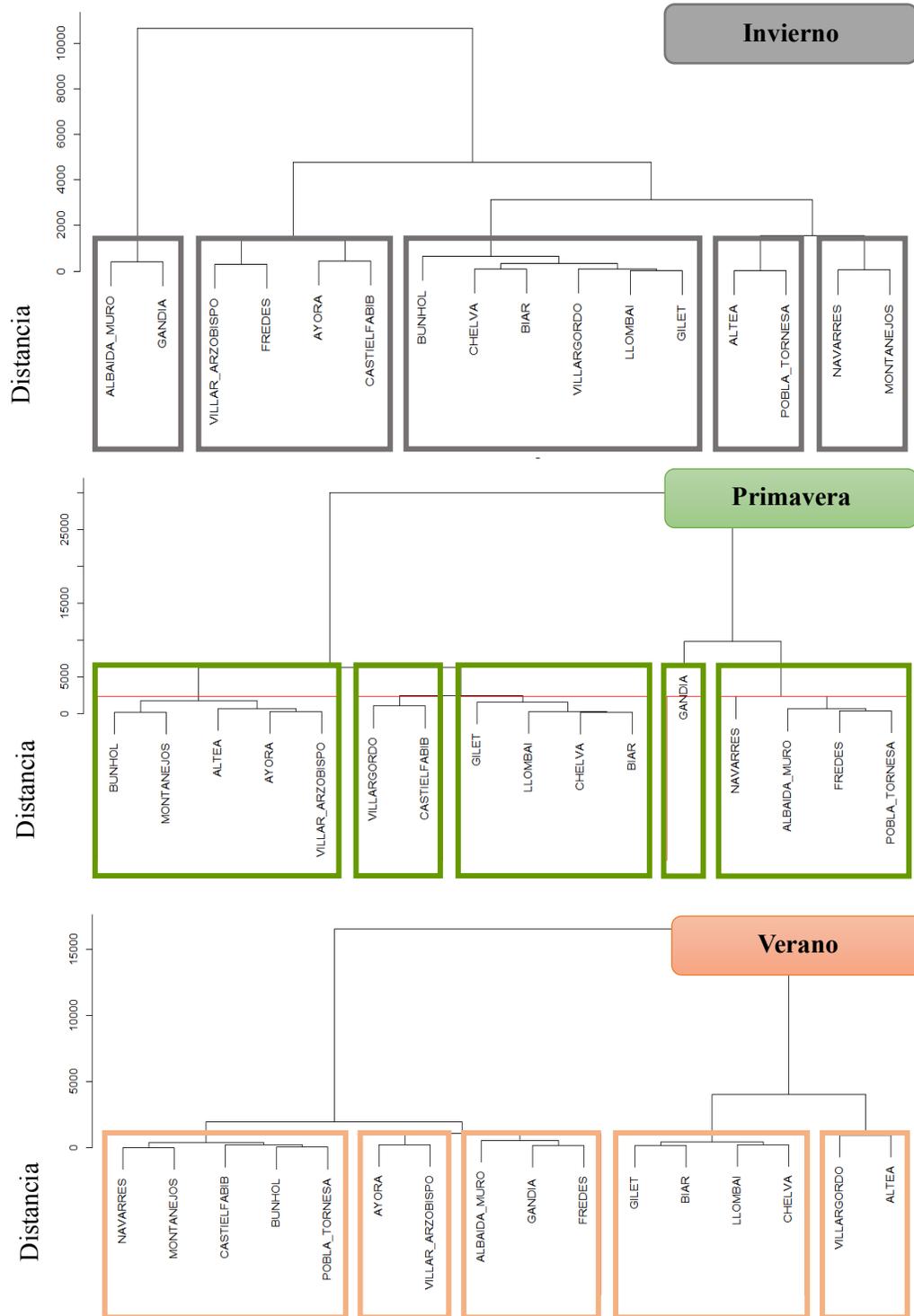
*3.2.1 Análisis de conglomerados por estación del año y cálculo de percentiles.*

En este apartado se procede a determinar por el sistema de conglomerados 5 agrupaciones de los puntos de muestreo con los datos del porcentaje de humedad de la especie vegetal romero por estaciones del año. El procedimiento a seguir es el mismo explicado en el apartado 3.1.1 por lo que solo se procede a presentar resultados.

En la figura 3.2.1-1 se muestra el dendograma con los puntos de muestreos agrupados en 5 clases por estación del año. Cabe destacar que no se toma en cuenta la homogeneidad entre el número de clases. En las tablas 3.2.1-1 se muestran los puntos de muestreo y al conglomerado que pertenece en cada análisis por estación del año.

En la tabla 3.2.1-2 se muestran los percentiles por cada conglomerado y estación del año, esto servirán como clasificadores.

Figura 3.2.1-1 Dendograma por estaciones del año.



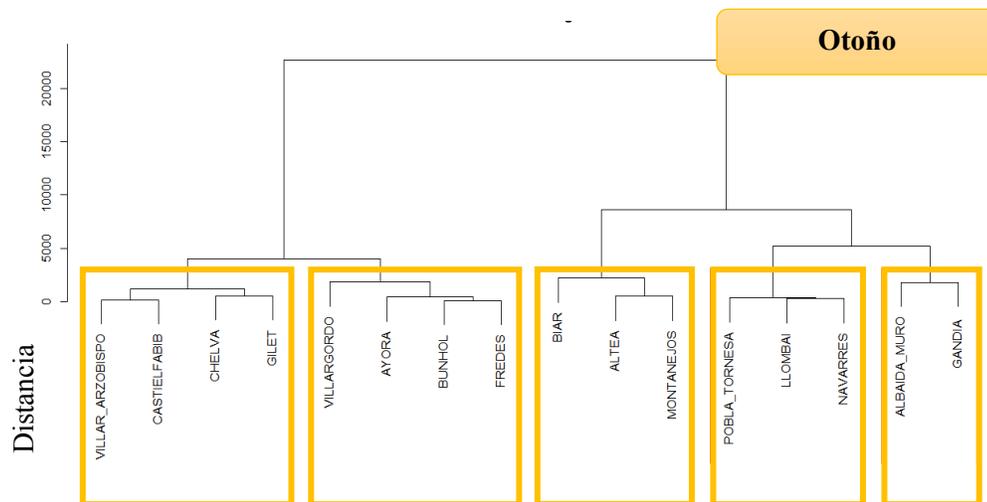


Tabla 3.2.1-1 Puntos de muestreo agrupados en los conglomerados de 5 clases por estación del año.

Num	Puntos de muestreo	Invierno	Primavera	Verano	Otoño
1	AYORA	1	1	1	1
2	BUNHOL	2	1	2	1
3	LLOMBAI	2	2	3	2
4	VILLAR_ARZOBISPO	1	1	1	3
5	CHELVA	2	2	3	3
6	ALBAIDA_MURO	3	3	4	4
7	GILET	2	2	3	3
8	VILLARGORDO	2	4	5	1
9	CASTIELFABIB	1	4	2	3
10	NAVARRES	4	3	2	2
11	GANDIA	3	5	4	4
12	BIAR	2	2	3	5
13	ALTEA	5	1	5	5
14	MONTANEJOS	4	1	2	5
15	FREDES	1	3	4	1
16	POBLA_TORNESA	5	3	2	2

Grupo 1 =	
Grupo 2 =	
Grupo 3 =	
Grupo 4 =	
Grupo 5 =	

Tabla 3.2.1-2 Percentiles por conglomerado de cada estación del año.

Percentil	Invierno					Primavera				
	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
10%	90.99	106.99	120.08	119.20	120.07	117.45	100.31	137.27	100.27	154.43
20%	96.41	110.26	122.47	122.22	120.79	119.50	107.69	146.95	101.33	158.58
40%	101.40	113.29	143.04	124.24	121.04	128.30	117.80	154.95	107.11	166.88
60%	108.57	115.90	145.93	124.33	124.85	141.25	128.14	158.44	112.74	179.83
80%	116.29	120.70	153.49	131.58	144.56	149.58	133.86	172.01	112.92	197.44
90%	121.18	124.19	163.28	132.85	144.81	151.81	137.59	177.12	120.89	206.25
Percentil	Verano					Otoño				
	C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
10%	91.28	75.07	54.59	91.94	49.82	87.56	105.92	67.35	110.15	66.99
20%	93.71	76.98	56.13	99.62	50.84	88.40	108.45	70.28	110.29	77.18
40%	96.61	92.65	75.57	104.87	59.11	92.62	111.25	86.86	124.30	87.69
60%	96.62	102.07	85.00	109.75	61.94	102.36	123.78	102.92	130.12	107.87
80%	97.61	107.67	87.95	115.46	68.75	106.45	130.80	117.63	162.17	148.80
90%	100.55	111.64	92.20	117.84	84.02	107.58	136.07	122.41	179.35	168.83

#### 4. ANÁLISIS DE REGRESIÓN LINEAL MÚLTIPLE

La aplicación del análisis de regresión lineal múltiple tiene la finalidad de encontrar la relación existente entre el porcentaje de humedad de la especie vegetal romero y las 28 variables meteorológicas explicadas en el apartado 1.4.3.1, lo que se busca es que por medio de la selección de variables meteorológicas se encuentre un modelo con el cual se pueda describir la relación existente. Cabe destacar que el análisis de regresión múltiple realiza la descripción de la relación entre las variables independientes y la variable de respuesta por medio de una ecuación.

Para la creación del modelo se especifican dos tipos de variables, las variables independientes o predictoras en este caso son las meteorológicas y con ellas se pretende encontrar las que mejor expliquen y se relacionen con la variable dependiente que en este caso es el porcentaje de humedad de la especie romero.

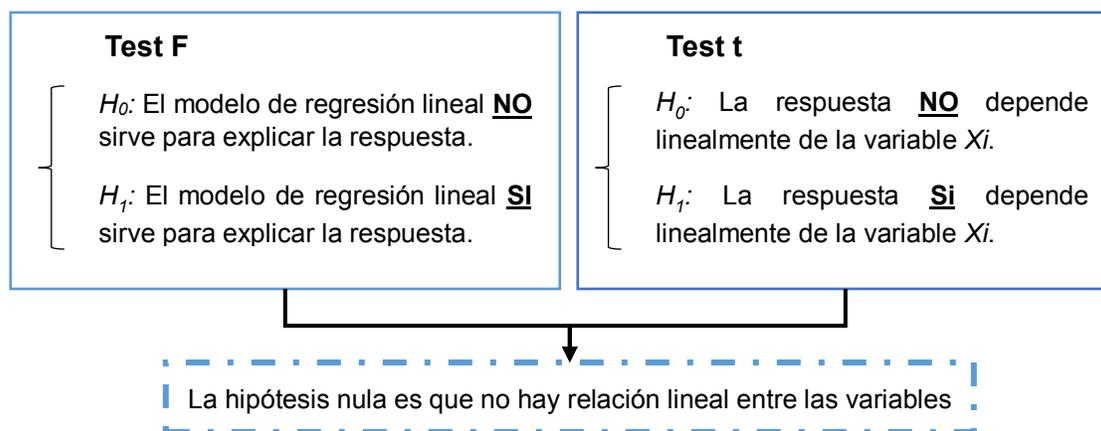
Si bien es cierto que en el modelo se pueden introducir todas las variables, esto puede dar la existencia de redundancias produciendo errores en la estimación de la respuesta, además el modelo se volvería muy complicado y difícil de entender, hay que recordar que con el modelo se trata de simplificar la realidad, no reproducirla.

La evaluación de las variables independientes que entren en el modelo se realizó con el método denominado paso a paso hacia Adelante (Forward Stepwise Regression), en donde las variables serán introducidas secuencialmente de forma automática, el procedimiento termina cuando ya no queden variables que cumplan el criterio de entrada. (Ferrero, s.f.)

El análisis de la regresión lineal múltiple fue realizado por el software Statgraphics Centurion y RStudio, al compararlos se notó diferencia en las selección de variables dentro del modelo, a pesar de que el software RStudio cuenta con varias librerías para realizar análisis de regresión múltiple tales como MASS, leaps, subselect, olsrr, entre otras, de las analizadas se escogió la librería olsrr ya que a diferencia con el resto, el formato de presentación de los resultados es muy amigable para el usuario.

Los criterios de selección con los que trabaja este análisis principalmente es el test de la f de la tabla ANOVA con un nivel de confianza del 95%, el software RStudio también toma en cuenta el teste de la t, valor que puede ser observado en el resumen de los resultados por variable seleccionada, ya que los dos coinciden al determinar la hipótesis nula, tal como se muestra en la figura 4-1.

Figura 4-1 Comparación entre los tests F y t.



Fuente: (Justel, 2015)

Sin embargo se busca trabajar con el valor-p ya que presenta un enfoque alternativo, definiéndose como el más bajo nivel de significación al que puede ser rechazada una hipótesis nula. (Uriel, 2013)

Con RStudio no se pudo conseguir determinar la selección de variables por este parámetro por lo que en este apartado se presentan los procesos realizados con el software Statgraphics Centurion.

#### 4.1 Variables utilizadas en el análisis de regresión lineal múltiple

Como se expuso en el apartado 1.4.3.1 se calculó las variables meteorológicas para el periodo de los 3 años coincidentes con los datos del porcentaje de humedad de la especie vegetal romero, sin embargo en el momento de realizar el estudio de regresión no se pudo contar con todos los datos meteorológicos, solo se tienen datos meteorológicos completos de un año para los 6 puntos de muestreo especificados a continuación.

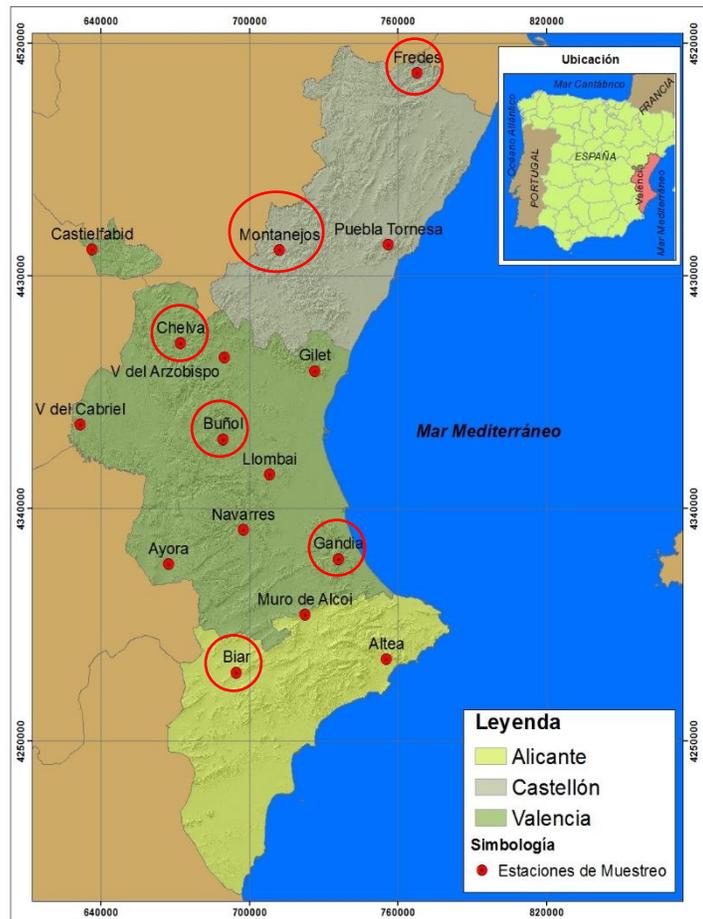
Se cuenta con datos de 6 puntos de muestreo, en donde el período de tiempo más largo del cual se tiene corresponde a Buñol y está comprendido entre abril del 2014 y enero del 2016, el más corto corresponde a Gandía comprendido entre octubre del 2014 y enero del 2016, además Biar, Buñol, Chelva y Fredes no tiene datos completos. Prácticamente se cuenta con un año completo de datos en todas las estaciones, en la tabla 4.1-1 se muestra en número total de datos por mes y su periodo correspondiente, cabe recalcar que se hizo uso de todos los datos disponibles para el presente estudio.

Tabla 4.1-1 Datos generales puntos de muestreo.

Estaciones	Número de datos	Período
Biar	16	sep-2014 / ene-2016
Buñol	19	abr-2014 / ene-2016
Chelva	17	jul-2014 / ene-2016
Fredes	14	sep-2014 / ene-2016
Gandía	16	Oct-2014 / ene-2016
Montanejos	18	sep-2014 / ene-2016

Los puntos de muestreo con los que se trabajó se encuentran bien distribuidos por el área de estudio tal como se muestra en figura 4.1-1.

Figura 4.1-1 Puntos de muestreo con disposición de datos meteorológicos completos.



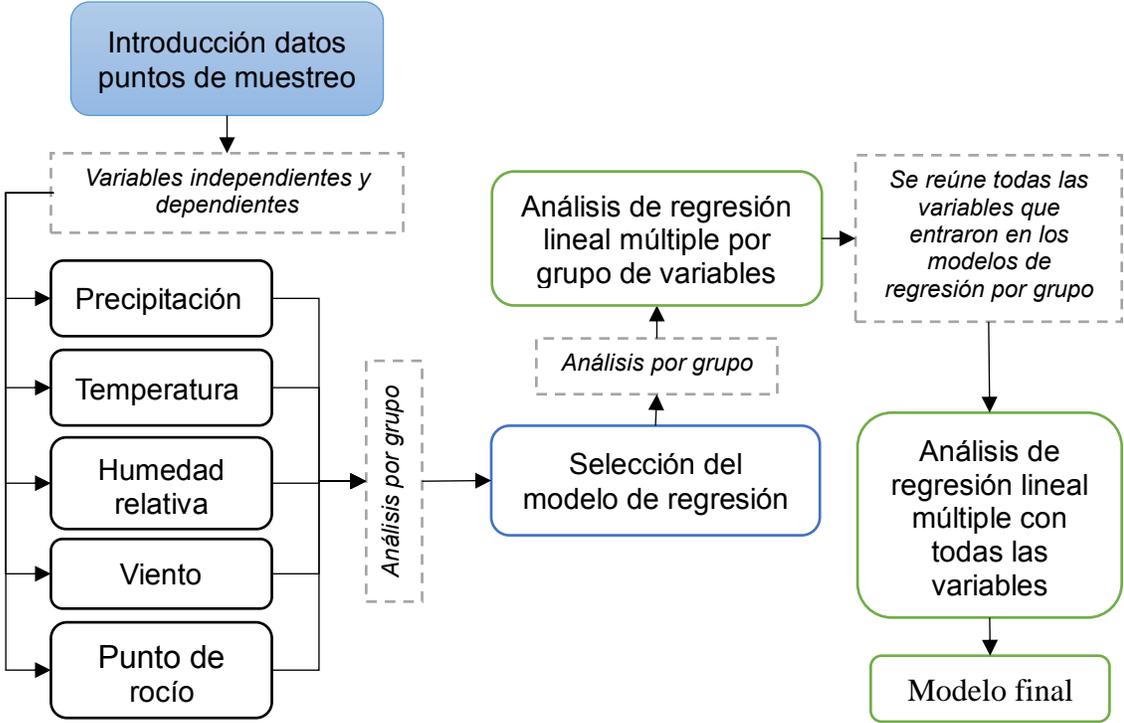
Fuente: Cartografía del ICV y VAERSA

## 4.2 Diagrama de flujo del análisis de regresión lineal múltiple

Realizar el análisis de regresión con las 28 variables meteorológicas por estación no se puede hacer, ya que el número de variables tiene que ser menor que el número de datos menos 1, por ello no se puede calcular la regresión.

Se hacen diferentes análisis con la finalidad de encontrar la mejor relación posible, el primer análisis se hace tomando todas las variables y todos los puntos de muestreo y también de forma individual y se procedió de la forma mostrada en la figura 4.2-1, se tomaron las variables por los 5 grupos mostrados en la tabla 1.4.3.1-1 y se realizó una selección del mejor modelo con las variables meteorológicas en relación de la variable porcentaje de humedad, hay que recordar que de acuerdo al grupo de variables y la forma en que son introducidas en el modelo, varía su influencia o relación con la variable dependiente, es por eso que se realizan dos tipos de pruebas con el fin de encontrar el mejor, el primer análisis se realizó introduciendo todas las variables seleccionadas de los grupos con la finalidad de observar cuales son las que entran en el modelo final; el segundo análisis en cambio se realiza un modelo de regresión lineal por grupo y con las variables que entren en el modelo se realiza un último modelo con todas las variables.

Figura 4.2-1 Diagrama de flujo de trabajo.



### **4.3 Desarrollo del análisis de regresión lineal múltiple**

Los diferentes análisis fueron realizados de la misma forma, por lo cual se tomó como referente el elaborado con los datos del punto de muestreo Biar.

#### *4.3.1 Selección del modelo de regresión lineal múltiple.*

En esta sección se realizó el análisis de diferentes modelos y por ende las variables que mayor relación poseen con respecto a la variable porcentaje de humedad de la especie vegetal romero. Es importante recalcar que estos son posibles modelos, pueda que algunas de las variables no sean estadísticamente significativas en el modelo de regresión.

La evaluación de los modelos es medida por el coeficiente de determinación múltiple  $R^2$  ajustado, el cual denota la bondad de ajuste o el porcentaje de variación de relación que tienen las variables independientes con la variable dependiente en el modelo, dicho de otra forma la descripción de ajuste del conjunto de variables. Si bien es cierto que un modelo puede mejorar al aumentar el número de variables puede ser que la inclusión de una variable no sea significativa para el modelo o no se relacione de forma que explique mejor la variable dependiente, sin embargo para comprobar si el modelo es mejor se utiliza dicho coeficiente.

Mientras mayor sea el valor del coeficiente  $R^2$  ajustado mejor ajuste del modelo se tendrá, de igual forma a menor error cuadrático medio de la estimación variable de estudio (CME) mejor modelo se considerará. También se toma en cuenta la evaluación por los criterios de información AIC (Criterio de información de Akaike), HQC (criterio de información de Hannan-Quinn) y BIC (criterio de información bayesiano) en donde los mejores modelos presentan el menor valor.

A continuación se muestran los resultados de los mejores modelos por los grupos de variables definidas en la figura 4.2-1, realizadas con el software Statgraphics Centurion

##### *4.3.1.1 Selección de mejores modelos por grupos de variables*

El software toma como variable dependiente al porcentaje de humedad de la especie vegetal romero y como variables independientes a las pertenecientes a la agrupación de precipitación

tal como se puede ver en la tabla 1.4.3.1-1, el total de las variables en el grupo son 5 con las que se realizó 32 posibles modelos ajustados, además el programa renombra con una letra mayúscula a cada variable, en donde: A= P3, B= P7, C= P15, D= P30 y E= P60.

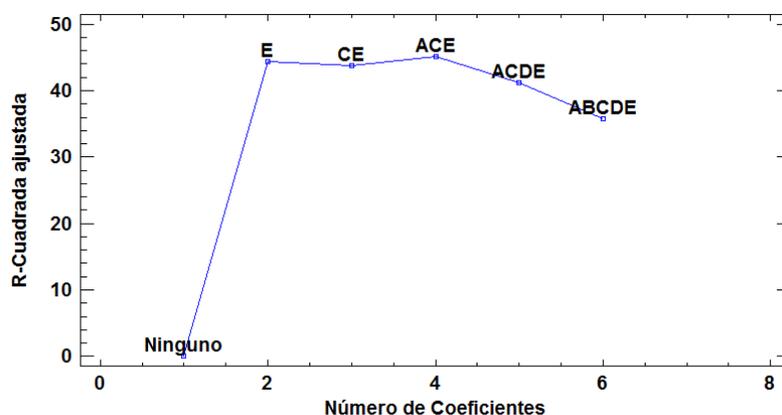
Por ser una cantidad alta de modelos y el interés radica en elegir el mejor solo se muestran los 5 más sobresalientes proporcionados por el valor coeficiente de determinación múltiple  $R^2$  ajustado y los criterio de información tal como se muestra en la tabla 4.3.1-1

Tabla 4.3.1.1-1 Selección de modelos

Num	Coeficiente de determinación múltiple $R^2$ ajustado					Criterio de información					
	CME	R-Cuadrada	R-Cuadrada Ajustada	Cp	Variables incluidas	CME	Coeficientes	AIC	HQC	SBIC	Variables incluidas
1	517,96	56,089	451,113	22,506	ACE	524,925	2	651,325	65,182	660,983	E
2	524,925	480,817	443,732	0,119849	E	530,735	3	664,926	665,668	679,412	CE
3	530,735	512,565	437,575	137,872	CE	551,748	3	668,809	669,551	683,295	AE
4	532,222	54,88	43,6	253,285	BCE	563,039	3	670,835	671,577	685,321	BE
5	550,814	533,037	416,297	290,081	CDE	564,963	3	671,176	671,918	685,662	DE

En la figura 4.3.1.1-1 se muestra los modelos principales en función del número de variables tomadas para su realización en donde claramente destaca en que incluye las variables ACE.

Figura 4.3.1.1-1 Gráfica de R-cuadrada ajustada por el porcentaje de humedad de la especie vegetal romero (Ro)



Realizado en Statgraphics Centurion

Para finalizar se toman en cuenta los modelos que sobresalen de los análisis, el modelo evaluado por los criterios solo posee una variable, misma que está incluida en el mejor modelo

por el coeficiente de determinación, sin embargo se lo selecciona con la finalidad de saber si ingresa o no en el modelo final. (Ver tabla 4.3.1.1-2)

Tabla 4.3.1.1-2 Modelos destacados con variables del grupo de precipitación.

VARIABLES DEL MODELO	VARIABLES
ACE	P3, P15, P60
E	P60

El mismo procedimiento se realiza con el resto de agrupaciones, en la tabla 4.3.1.1-3 se presenta la asignación de cada letra a las variables dentro de los diferentes grupos y en la tabla 4.3.1-14 los 5 más sobresalientes proporcionados por el valor coeficiente de determinación múltiple  $R^2$  ajustado y los criterios de información.

Tabla 4.3.1.1-3 Asignación de letra a cada variable dentro de cada grupo.

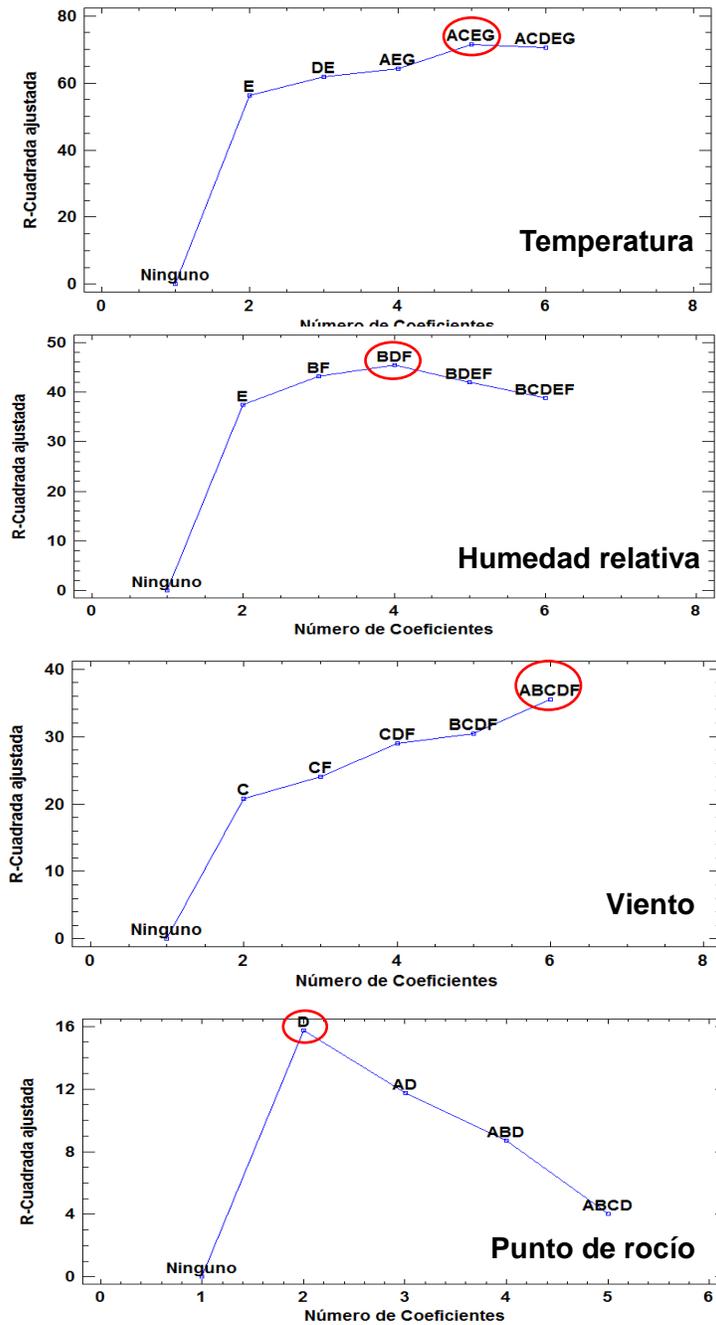
	Temperatura	Humedad relativa	Viento	Punto de rocío
A	Tmax3	HRmax3	Vmax3	pto_r3
B	Tmax7	HRmax7	Vmax7	pto_r7
C	Tmax15	HRmax15	Vmax15	pto_r15
D	Tmax30	HRmin3	lev3	pto_r30
E	Tmedia3	HRmin7	lev7	
F	Tmedia7	HRmin15	lev15	
G	Tmedia15			

Tabla 4.3.1.1-4 Selección de modelos por grupo.

	Num	Coeficiente de determinación múltiple $R^2$ ajustado				Criterio de información						
		CME	R-Cuadrada	R-Cuadrada Ajustada	Cp	VARIABLES INCLUIDAS	CME	Coeficientes	AIC	HQC	SBIC	VARIABLES INCLUIDAS
Temperatura	1	268,198	791,578	715,788	315,147	ACEG	268,198	5	621,673	622,909	645,816	ACEG
	2	277,569	803,905	705,858	46,102	ACDEG	360,118	3	626,143	626,885	640,629	DE
	3	282,458	800,451	700,677	476,185	ABCEG	413,041	2	627,355	627,849	637,012	E
	4	290,448	794,806	692,209	500,971	ACEFG	373,733	3	629,854	630,596	64,434	EG
	5	302,968	764,557	678,941	43,379	ABEG	380,522	3	631,654	632,396	646,141	AE
Humedad relativa	1	514,516	56,381	454,762	167,222	BDF	590,826	2	663,152	663,647	67,281	E
	2	527,609	55,271	440,887	191,835	ADF	535,636	3	665,845	666,587	680,331	BF
	3	535,636	508,064	432,381	0,908341	BF	542,087	3	667,043	667,784	681,529	CF
	4	539,262	542,831	428,539	213,741	ABC	625,581	2	668,868	669,363	678,526	F
	5	541,035	541,328	42,666	217,073	ABF	554,09	3	669,233	669,975	683,719	CE
Viento	1	608,204	57,032	35,548	517,244	ABCDF	747,296	2	686,646	687,141	696,304	C
	2	656,145	490,096	304,676	488,501	BCDF	717,191	3	695,034	695,776	70,952	CF
	3	656,311	489,967	30,45	488,776	ABCF	943,654	1	697,476	697,723	702,305	
	4	661,571	532,618	298,927	597,728	ABCEF	670,55	4	70,081	701,799	720,125	CD
	5	670,55	431,529	289,411	413,525	CD	686,336	4	703,137	704,126	722,452	BCF
Punto de rocío	1	795,193	213,505	157,327	0,289966	D	795,193	2	692,858	693,353	702,516	D
	2	832,856	235,093	117,414	195,263	AD	943,654	1	697,476	697,723	702,305	
	3	834,335	233,734	115,847	197,386	BD	897,381	2	704,948	705,443	714,605	C
	4	852,946	216,641	961,242	224,096	CD	832,856	3	709,986	710,728	724,472	AD
	5	861,43	269,707	871,335	341,174	ABD	834,335	3	710,163	710,905	72,465	BD

En la figura 4.3.1.1-2 se muestra los modelos principales en función del número de variables por grupo y el más destacado está encerrado en un círculo rojo.

Figura 4.3.1.1-2 Gráfica de R-cuadrada ajustada por el porcentaje de humedad de la especie vegetal romero (Ro).



Para finalizar en la tabla 4.3.1.1-5 se presenta los mejores modelos evaluados por grupo, en algunos el mejor modelo es el mismo para los dos evaluaciones como el de temperatura y punto de rocío. Sin embargo serán tomadas en cuenta.

Tabla 4.3.1.1-5 Modelos destacados con variables de cada grupo.

Grupo variables	Código modelo	Nombre variables	Evaluación
Temperatura	ACEG	Tmax3, Tmax15, Tmedia3, Tmedia15	R2 ajustado
	ACEG	Tmax3, Tmax15, Tmedia3, Tmedia15	Criterio de información
Humedad relativa	BDF	HRmax7, HRmin3, HRmin15	R2 ajustado
	E	HRmin7	Criterio de información
Viento	ABCDF	Vmax3, Vmax7, Vmax15, lev3, lev15	R2 ajustado
	C	Vmax15	Criterio de información
Punto de rocío	D	pto_r30	R2 ajustado
	D	pto_r30	Criterio de información

#### 4.3.1.2 Modelos de regresión con todas las variables.

Como se explicó al comienzo de este apartado para el análisis de regresión lineal múltiple se utiliza el método denominado paso a paso hacia Adelante (Forward Stepwise Regression), los criterios de selección con los que se trabaja principalmente es el test de la f de la tabla ANOVA con un nivel de confianza del 95%, Sin embargo se toma en cuenta el valor-p, para la valoración y comparación de modelos se usa el criterio coeficiente R<sup>2</sup> ajustado.

Como se explicó en el diagrama de flujo (Ver figura 4.2-1) se realizan dos tipos análisis el mostrado a continuación se toma todas las variables de los grupos y se analiza en conjunto. Al tener variables escogidas de acuerdo al método de evaluación se procede a escoger el que mejor ajuste tenga de los dos.

La selección de modelos por grupo evaluado por el coeficiente R<sup>2</sup> ajustado excede el número de variables permitidas para realizar el modelo, sin embargo al tomar las variables evaluadas por los criterios de información y juntarlas, se crea el modelo sin problema, las variables que describen mejor a la relación son las mostradas en la tabla 4.3.1.2-1 con un R<sup>2</sup> ajustado del 70,60%.

Tabla 4.3.1.2-1 Informe del modelo destacado con variables de cada grupo.

Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P
CONSTANTE	133,254	17.1397	7.775	0,0000
P60	0,348125	0,124297	2.801	0,0150
Tmedia3	-2.785170	0,758298	-3.673	0,0028

*Regresión por Pasos*

	Variable	R <sup>2</sup>	R <sup>2</sup> ajustado	CME
<b>Paso 0:</b>	CONSTANTE	0	0	943,654
<b>Paso 1:</b>	Tmedia3	59,15%	56,23%	413,041
<b>Paso 2:</b>	P60	74,52%	70,60%	277,419

Error estándar del estimacion	16.6559
Error absoluto medio	13.0166
Estadístico Durbin-Watson	2,30712 (P=0,5915)

De las 8 variables meteorológicas con las cuales se realiza el análisis de regresión lineal múltiple, en el modelo final solo entran 2, la descripción de la relación entre las variables humedad de la especie vegetal romero y las meteorológicas se define por la ecuación mostrada a continuación, siendo que el valor-P en la tabla ANOVA es menor que 0,05, existe una relación estadísticamente significativa entre las variables con un nivel de confianza del 95,0%.

$$R_o = 133,254 + 0,348125 \cdot P60 - 2,78517 \cdot Tmedia3$$

El resumen final realizado por el software indica que:

- El estadístico R<sup>2</sup> enseña que el modelo así ajustado explica 74,52% de la variabilidad en Ro.
- El estadístico R<sup>2</sup> ajustada, que es más apropiada para comparar modelos con diferente número de variables independientes, es 70,60%.
- El error estándar del estimado muestra que la desviación estándar de los residuos es 16,65, este valor puede usarse para construir límites para nuevas observaciones.
- El error absoluto medio (MAE) de 13,0166 es el valor promedio de los residuos.
- El estadístico de Durbin-Watson (DW) examina los residuos para determinar si hay alguna correlación significativa basada en el orden en el que se presentan en el archivo de datos.
- Para determinar si el modelo puede simplificarse, note que el valor-P más alto de las variables independientes es 0,0150, que corresponde a P60. Puesto que el valor-P es menor que 0,05, ese término es estadísticamente significativo con un nivel de confianza del 95,0% por lo que no quisiera eliminar ninguna variable del modelo.

#### 4.3.1.3 Modelos de regresión por grupo de variables.

Se realiza modelos con las variables por grupo, además se hace una comparación de las variables seleccionadas en los modelos destacados de las dos evaluaciones (coeficiente R<sup>2</sup> ajustado y criterios de información) se comparan y se toma el que mejor se ajuste.

Las variables seleccionadas por la evaluación criterio de información en el grupo de precipitación y viento son las que han entrado en el modelo, en el grupo de humedad relativa el modelo que mejor se ajusta integra a la variable seleccionada por la evaluación criterio de información, hasta el momento parece que las variables seleccionadas por dicha evaluación son las que entran directamente en el modelo. Con lo que respecta a temperatura se escoge una variable y el grupo punto de rocío no se relaciona ninguna variable.

Tabla 4.3.1.3-1 Informe de los modelos por grupo.

Grupo	Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P	Evaluación
Precipitación	CONSTANTE	78.0064	11.3024	6.90173	0,0000	R2 ajustado
	P60	0,551314	0,153111	3.60075	0,0029	
Temperatura	CONSTANTE	169,77	13.5758	12.5053	0,0000	R2 ajustado
	Tmedia3	-3.73041	0,828576	-4.50219	0,0005	
Humedad relativa	CONSTANTE	51.0153	20.5888	2.47782	0,0266	Criterio de información
	HRmin7	1.48262	0,46984	3.15558	0,0070	
Viento	CONSTANTE	192.667	36.4445	5.28658	0,0001	R2 ajustado
	Vmax15	-3.50748	1.57787	-2.22292	0,0432	
Punto de rocío	No entra ninguna variable en el modelo					

#### Regresión por Pasos

	Variable	R <sup>2</sup>	R <sup>2</sup> ajustado	CME	Grupo
Paso 0:	CONSTANTE	0	0	943.654	Precipitación
Paso 1:	P60	48,08%	44,37%	524.925	
Paso 0:	CONSTANTE	0	0	943.654	Temperatura
Paso 1:	Tmedia3	59,15%	56,23%	413.041	
Paso 0:	CONSTANTE	0	0	943.654	Humedad relativa
Paso 1:	HRmin7	41,56%	37,39%	590.826	
Paso 0:	CONSTANTE	0	0	943.654	Viento
Paso 1:	Vmax15	26,09%	20,81%	747.296	
No entra ninguna variable en el modelo					Punto de rocío

Tabla 4.3.1.3-2 Informe de los modelos por grupo.

Grupo		
Precipitación	Error estándar del estimacion	22.9112
	Error absoluto medio	18.2596
	Estadístico Durbin-Watson	1,52901 (P=0,1403)
Temperatura	Error estándar del estimacion	20.3234
	Error absoluto medio	16.3909
	Estadístico Durbin-Watson	2,08837 (P=0,4786)
Humedad relativa	Error estándar del estimacion	24.3069
	Error absoluto medio	19.5587
	Estadístico Durbin-Watson	1,83419 (P=0,3362)
Viento	Error estándar del estimacion	27.3367
	Error absoluto medio	19.3723
	Estadístico Durbin-Watson	1,51507 (P=0,1341)
Punto de rocío	No entra ninguna variable en el modelo	

Grupo	Ecuación del modelo ajustado
Precipitación	$Ro = 78,0064 + 0,551314 * P60$
Temperatura	$Ro = 169,77 - 3,73041 * Tmedia3$
Humedad relativa	$Ro = 51,0153 + 1,48262 * HRmin7$
Viento	$Ro = 192,667 - 3,50748 * Vmax15$

Se realiza un modelado final en donde se toman todas las variables que han ingresado en los modelos de cada grupo, el resultado final se presenta en la tabla 4.3.1.2-3, que al comparar con el modelo donde se tomó todos los datos las variables de la evaluación por criterio (ver tabla 4.3.1.2-1) el resultado es el mismo.

Tabla 4.3.1.3-3 Informe del modelo final.

Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P
CONSTANTE	133,254	17.1397	7.775	0,0000
P60	0,348125	0,124297	2.801	0,0150
Tmedia3	-2.785170	0,758298	-3.673	0,0028

Regresión por Pasos	Variable	R <sup>2</sup>	R <sup>2</sup> ajustado	CME
Paso 0:	CONSTANTE	0	0	943,654
Paso 1:	Tmedia3	59,15%	56,23%	413,041
Paso 2:	P60	74,52%	70,60%	277,419

Error estándar del estimacion	16.6559
Error absoluto medio	13.0166
Estadístico Durbin-Watson	2,30712 (P=0,5915)

*Ecuación del modelo ajustado*

$$Ro = 133,254 + 0,348125 * P60 - 2,78517 * Tmedia3$$

## 5. INTERPOLACIÓN ESPACIAL

Se procede a realizar un análisis de la interpolación espacial utilizando los datos de humedad de la especie vegetal romero tomados en mayo del 2018, para ello, se utilizó el método de interpolación kriging, el cual es considerado como un procedimiento geoestadística avanzado, basado en modelos estadísticos que incluyen autocorrelación, es decir, las relaciones estadísticas entre los puntos medidos. Obteniendo como resultado un modelo geoestadística es estimación de la variable de estudio representado en una superficie terrestre.

Para realizar el análisis se contó con 51 datos de humedad, levantados en la provincia de Valencia en el mes de mayo del 2018 tal como se muestra en la figura Ver figura 3.2-1. También se utilizaron los puntos de muestreo los cuales contienen el registro del porcentaje de humedad del 2015 al 2017 del mismo mes, con la finalidad de evaluar el modelo. El número de puntos de muestreo con datos de porcentaje de humedad en un periodo de 3 años es bajo teniendo en cuenta lo requisitos para modelar por el método de interpolación kriging.

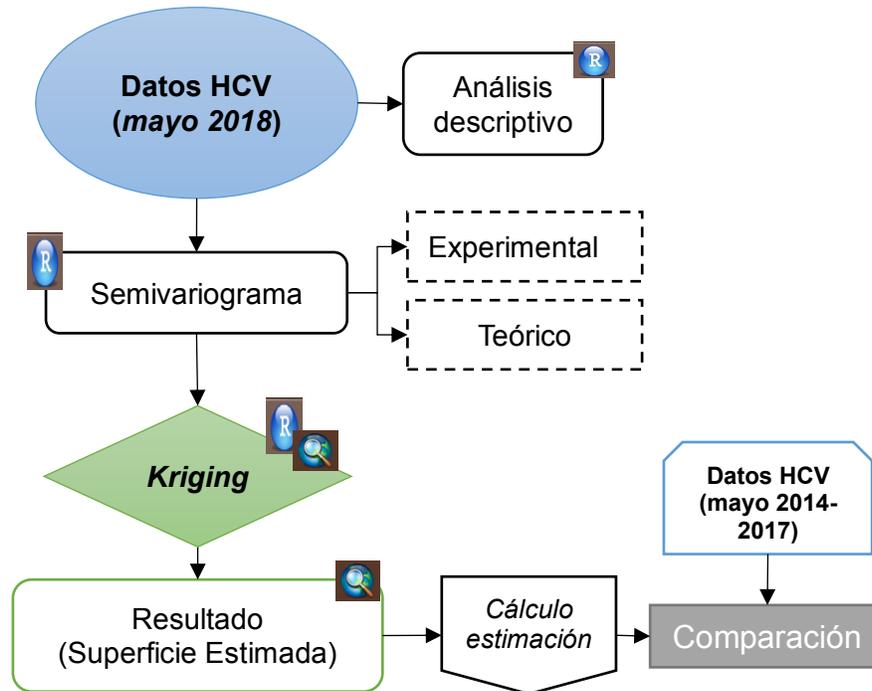
Si bien es cierto que existen varias herramientas de interpolación como IDW (distancia inversa ponderada), el cual es considerado un método determinístico de interpolación, basado directamente en los valores medidos. El método kriging por el contrario genera una superficie estimada partiendo de los puntos medidos sobre el terreno, mismo que refleja la correlación espacial con la que se pueda explicar sus variaciones en la superficie, la diferencia entre estos dos métodos es que en el IDW los pesos están en función de la inversa de la distancia entre las observaciones y en el kriging los pesos dependen de la correlación espacial de la variable de estudio.

El procedimiento a seguir para la obtención de la superficie con los valores estimados es el mostrado en la figura 5-1, para empezar se realiza un análisis descriptivo de la variable, luego se establecen los parámetros del semivariograma experimental y el teórico, por último se procede a aplicar el método de interpolación kriging.

Este análisis de interpolación por el método kriging fue realizado en conjunto mediante el software RStudio y ArcMap, si bien es cierto que el análisis se lo puede realizar por completo utilizando un solo programa, el trabajar en conjunto mejora el producto final. RStudio se utilizó para realizar el análisis preliminar de la variable y la obtención de los parámetros de ajuste

entre los semivariogramas experimenta y teórico, mientras que en ArcMap se realizaron la presentación de los resultados.

Figura 5-1 Diagrama de flujo de trabajo



### 5.1 Análisis descriptivo de los datos de mayo del 2018

El resumen estadístico con respecto al porcentaje de humedad de la especie vegetal romero se muestra en la tabla 5.1-1. Entre las medidas de posición tenemos el número total de datos y el promedio con un valor de 51 y 121.828 respectivamente, el valor de la mediana con un 124.381, la cual indica, que la mitad de datos están por debajo y por encima de dicho valor; el mínimo y máximo el cual informa el rango en que se mueven los datos que son 51,85 y 201,96 respectivamente, entre el cuartil inferior y cuartil superior se encuentran los datos más cercanos a la media. Cabe destacar que los valores del promedio y la mediana son muy próximos.

En lo que respecta a las medidas de dispersión muestra la desviación estándar muestra la dispersión de los datos respecto al promedio y presenta un valor de 34,59; mientras que el coeficiente de variación es una medida de la dispersión relativa de un conjunto de datos, el cual es expresado en porcentaje, y muestra la dispersión de los datos con respecto al promedio, en este caso su valor es 28,39%, al ser menor del 30% se puede decir que los datos medidos son bastante homogéneos.

Dentro de las medidas de forma, se muestra el coeficiente de asimetría (sesgo estandarizado) y la curtosis estandarizada, indicando si los datos son provenientes de una distribución normal, el rango para que se cumple la hipótesis es de -2 a +2, es así que el presentar valores dentro del rango, se puede decir que los datos provienen de una distribución normal.

Tabla 5.1-1 Resumen Estadístico

Recuento	51.00
Promedio	121.83
Mediana	124.38
Desviación Estándar	34.59
Coeficiente de Variación	28.39%
Mínimo	51.85
Máximo	201.96
Rango	150.11
Cuartil Inferior	101.08
Cuartil Superior	146.57
Sesgo Estandarizado	0.13
Curtosis Estandarizada	-0.11

En la figura 5.1-1 se puede observar el histograma donde 30 datos se concentran en el intervalo de 100 y 150 teniendo la mayor acumulación de datos, entre 50 y 100 se tiene 12 datos, y con menor número entre 150 y 225 con 9 datos. En la figura 5.1-2 se presenta el gráfico de probabilidad normal, en donde la concentración de puntos dentro de la recta siguen una distribución normal, los datos aunque no se encuentren encima de la recta están muy próximo lo cual indica que siguen una distribución normal.

Figura 5.1-1 Histograma

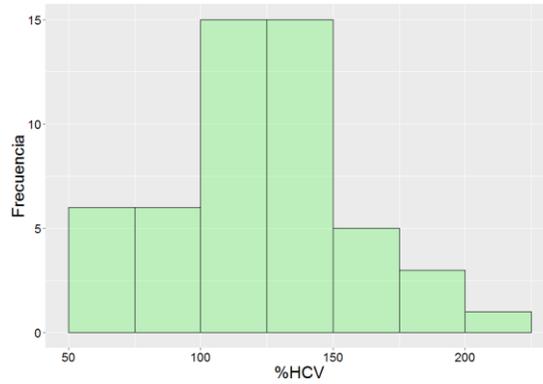
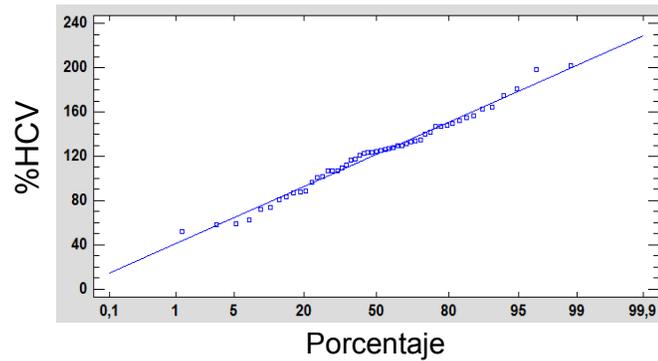
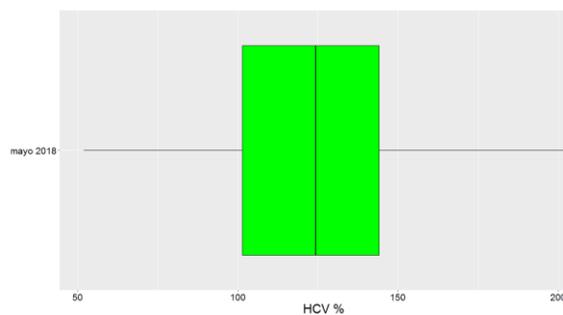


Figura 5.1-2 Probabilidad Normal



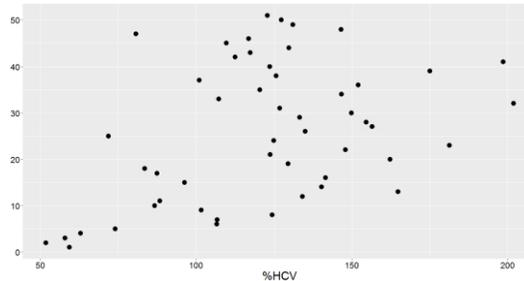
En la figura 5.1-3 Se observa el rango de toda la variable que se extiende desde el 50 hasta los 200, con una concentración de datos desde el rango 101 al 146 (Q1 al Q3), la mayoría de los datos se encuentran desde el Q1 y Q2 (mediana); la mediana y el promedio están cerca y no se visualizan valores atípicos.

Figura 5.1-3 Gráfico de caja y bigotes



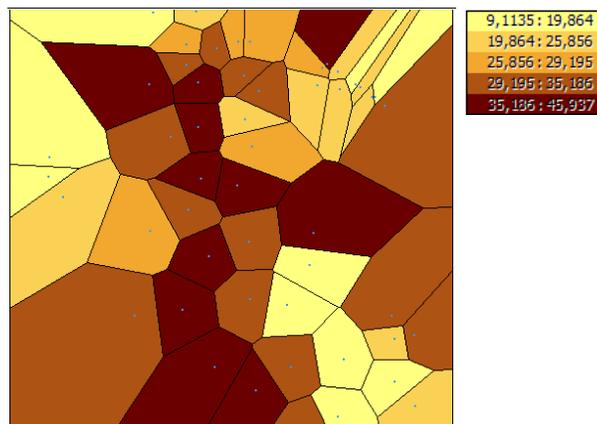
Otra forma de representar los datos es por medio del diagrama de dispersión en donde se puede observar que los valores se concentran entre los 100 y 150 tal como se ha mencionado anteriormente, también se puede notar que hay una agrupación notable entre 50 y 100, sin embargo de 150 en adelante existe pocos puntos y se encuentra mucho más distantes entre ellos. (Ver figura 5.1-4)

Figura 5.1-4 Gráfico de dispersión



Se verifica la estacionariedad de los datos o dicho de otra forma la variación en diferentes áreas de la zona de estudio, es decir, en dos puntos separados a una distancia en diferentes localidades deberían tener diferencias similares en su valor medido ya que kriging es sensible a los cambios bruscos. Esto se puede comparar por medio de un mapa de Voronoi (desviación estándar), en este caso se puede notar cambios bruscos por lo que no denota una adecuada estacionariedad.

Figura 5.1-5 Mapas de Voronoi con datos de porcentaje de humedad



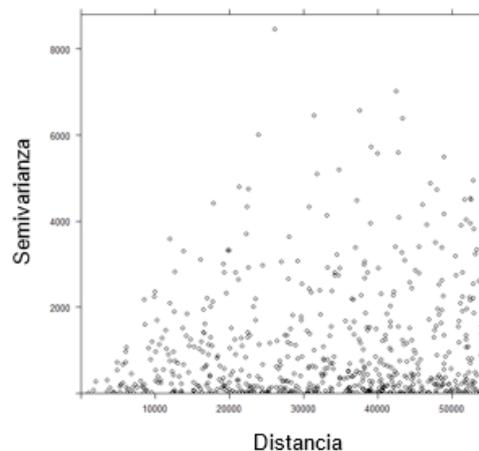
## 5.2 Semivariograma

Para determinar el mapa de superficie de estimación de la variable en estudio se debe realizar un análisis estadístico con la finalidad de determinar la correlación espacial existente entre los datos de humedad de la especie vegetal romero, dicho proceso se realiza a partir de datos muestrales, para ello se puede utilizar dos tipos de funciones: el semivariograma y el correlograma, en el presente estudio usa el semivariograma pues a diferencia del correlograma no requiere el cálculo de la media de la variable.

El variograma expresa la correlación espacial entre los datos, definiéndose como la media de los cuadrados de las diferencias entre pares de muestreo separados por una distancia, la mitad de variograma se conoce como función del semivariograma y caracteriza las propiedades de dependencia espacial. (Gallardo, 2006)

Para empezar en la figura 5.2-1 se muestra la nube del semivariograma, la cual denota el cálculo de la mitad del cuadrado de las diferencias entre los valores de la variable (semivarianza) en cada par de puntos de muestreo en función de la distancia que los separa y tiene la finalidad de poder observar su dispersión como también el detectar valores atípicos. En este caso, se puede observar que la mayor conglomeración se encuentra en los valores de dichas diferencias menores de 2000, entre los 4000 y 6000 existen algunos valores dispersos, sin embargo sobre los 8000 se encuentra un valor muy apartado el cual puede ser considerado como atípico.

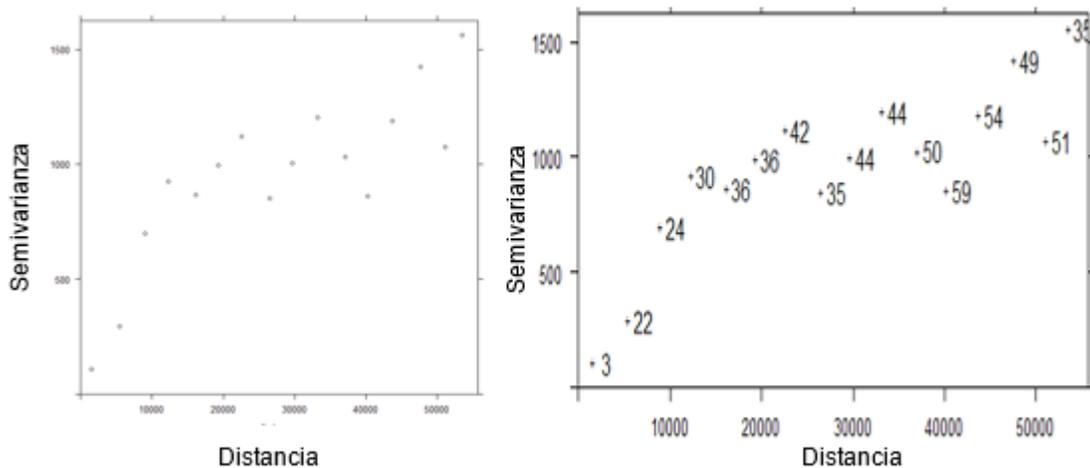
Figura 5.2-1 Nube del semivariograma



### 5.2.1 Semivariograma experimental

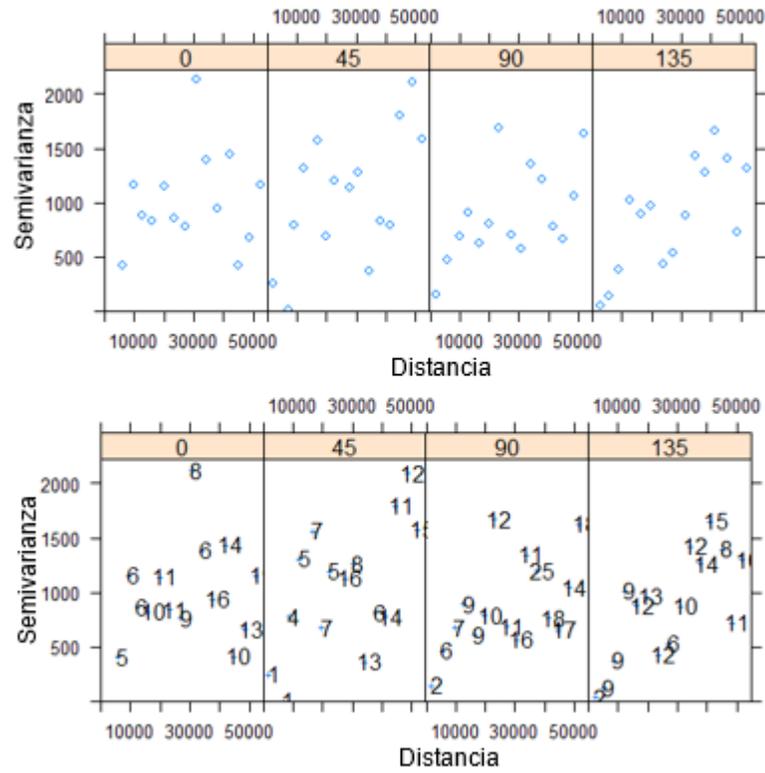
Con el semivariograma experimental se pretende interpretar el comportamiento espacial de la variable porcentaje de humedad sobre la provincia de Valencia reflejando la distancia máxima y la forma en que un punto influye sobre otro punto a diferentes distancias, considerando que la variación del porcentaje de humedad en el espacio es igual en todas las direcciones, se realiza el semivariograma omnidireccional, tal como se muestra en la figura 5.1.1-1, además se muestra el número de pares de datos. En la distancia máxima del semivariograma se presentan valores oscilantes, con una meseta para no definida, ya que los valores se muestran creciente a lo largo de la distancia del semivariograma, esto puede deberse a que el número de datos es bajo o es el mínimo necesario.

Figura 5.2.1-1 Semivariograma omnidireccional



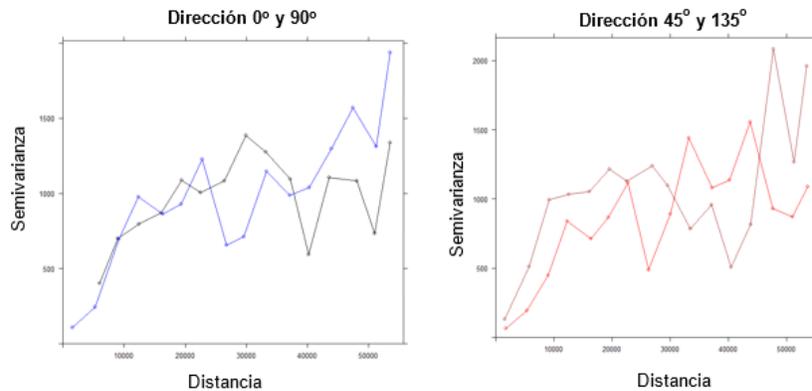
Los semivariograma mostrados en la figura 5.2.1-2, fueron realizados tomando en cuenta las orientaciones principales como son:  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  y  $135^\circ$ , claramente se puede observar que no existe continuidad, además el número de pares de datos no es muy grande al definir la distancia en diferentes direcciones, siendo imposible identificar la existencia de anisotropía. Se dice que existe anisotropía cuando el semivariograma cambia en función de la dirección espacial.

Figura 5.2.1-2 Semivariograma con diferentes orientaciones



En la figura 5.2.1-3 se muestra los semivariogramas con otra presentación corroborando lo anteriormente concluido, además se hace muy notorio las oscilaciones existentes, y al estar en constante crecimiento no se puede identificar el rango ni la meseta, es por ello que se procede a trabajar con el semivariograma omnidireccional.

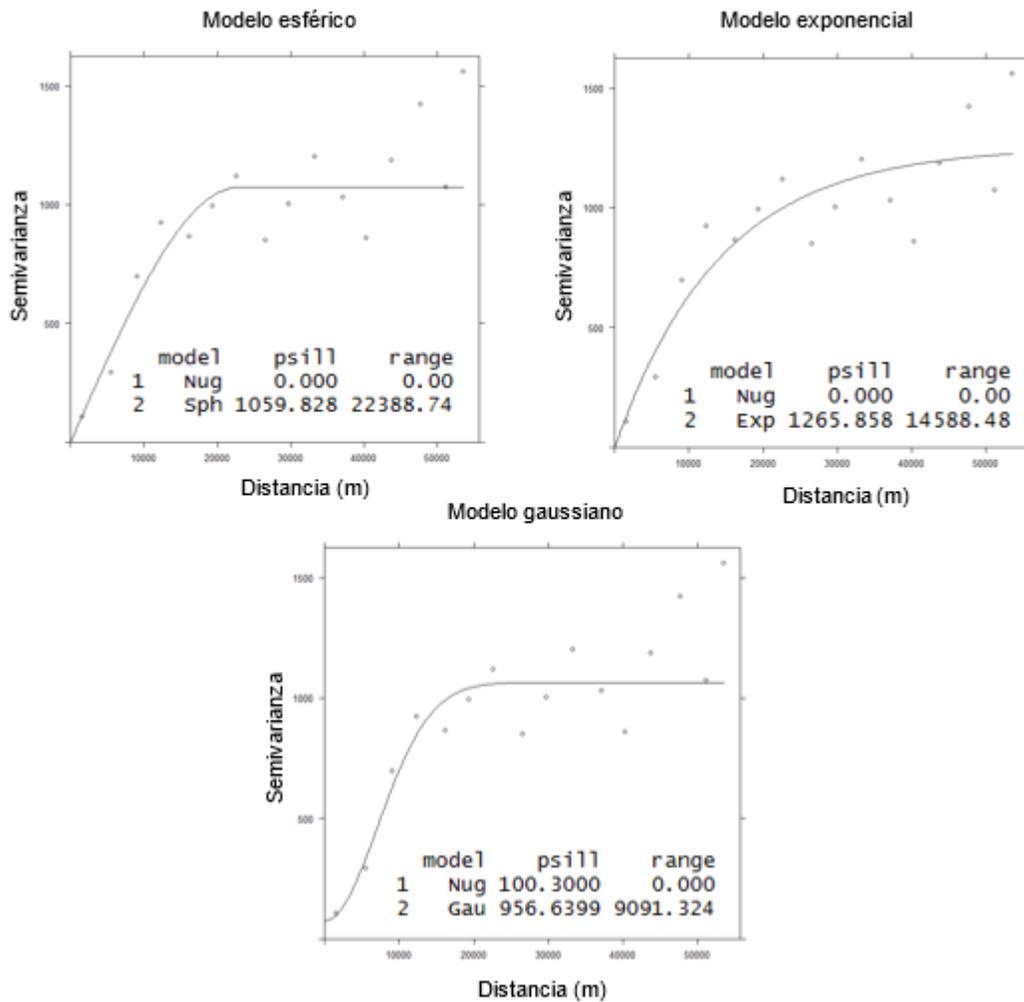
Figura 5.2.1-3 Semivariograma orientados



### 5.2.2 Semivariograma teórico

El semivariograma experimental calcula la autocorrelación para algunas distancias sin embargo es necesario conocer para cualquier posible distancia por lo cual se hace uso de los semivariogramas teóricos, los cuales están divididos en dos grupos los no acotados y los acotados, el segundo grupo son los más utilizados ya que presenta un buen ajuste.

Figura 5.2.2-1 Semivariogramas teóricos



En la figura 5.2.2-1 se muestran los semivariogramas teóricos en referencia con el experimental, de los 3 el que mayor rango presenta es el esférico con un valor de 22388.74 y el menor corresponde al gaussiano con 9091.324, de igual forma hay que tener en cuenta que el valor de la pepita no supere el 50% del valor de la meseta, en este caso el único que

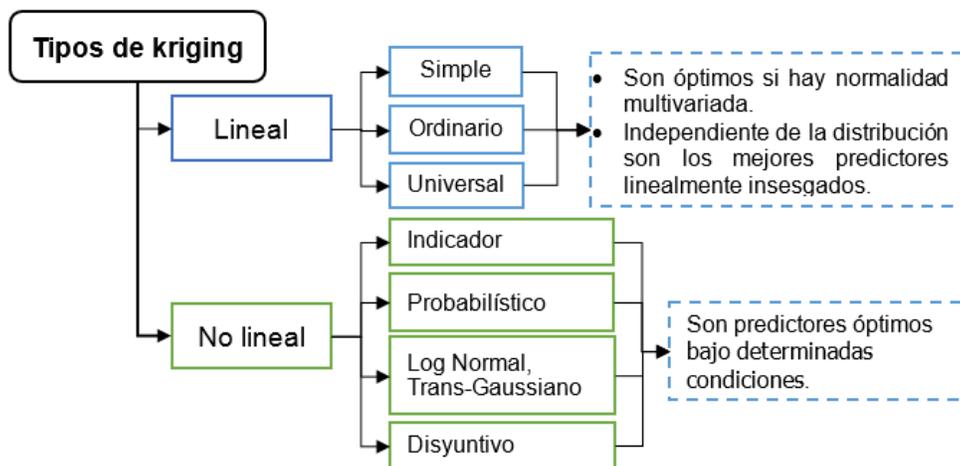
presenta valor diferente de 0 corresponde la modelo gaussiano con valor de 100.3, indicado que si se cumple dicho parámetro.

Para los siguientes procesos se aprecia que el modelo teórico que mejor se ajusta es el modelo gaussiano, por lo cual se utilizara para mostrar los siguientes procesos.

### 5.2.3 Kriging

Kriging es considerado un método de interpolación espacial que se fundamenta en la minimización del error cuadrático y es aplicado con frecuencia con propósitos de estimación de una variable en diferentes localizaciones del espacio. En la figura 5.2.3-1 se nombran los tipos de kriging existentes. (Henao, 2002)

Figura 5.2.3-1 Tipos de kriging

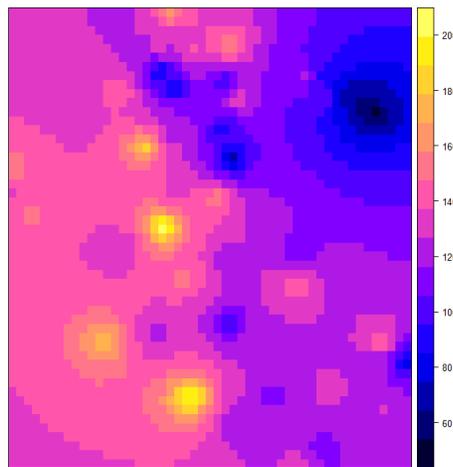


Fuente: (Henao, 2002, pág. 32)

Se usa el krigado ordinario pues a diferencia del krigado simple no requiere una estimación de la media de los datos. El krigado universal se usa cuando existe una deriva externa, la cual podría ser modelizada mediante la utilización de otras variables, tal como se hace en los modelos de regresión vistos en el capítulo 4. Se ha usado el krigado ordinario al no disponer de datos meteorológicos para la fecha de los datos de HCV (mayo 2018).

En la figura 5.2.3-2 se presenta la superficie de estimación de la variable de estudio, obtenido a partir de los 51 datos de humedad de la especie vegetal romero muestreados dentro de la provincia de Valencia, para lo cual se utilizó un modelo de semivariograma gaussiano.

Figura 5.2.3-2 Superficie de estimación (método krigeado ordinario)



*Realizado en: RStudio*

En la figura 5.2.3-3 se muestra la misma superficie de estimación dividida en 7 clases y con información referente al área de estudio. Cabe destacar que ArcMap al ser un programa de procesamientos geoespacial, también posee herramientas enfocadas en el análisis geoestadístico.

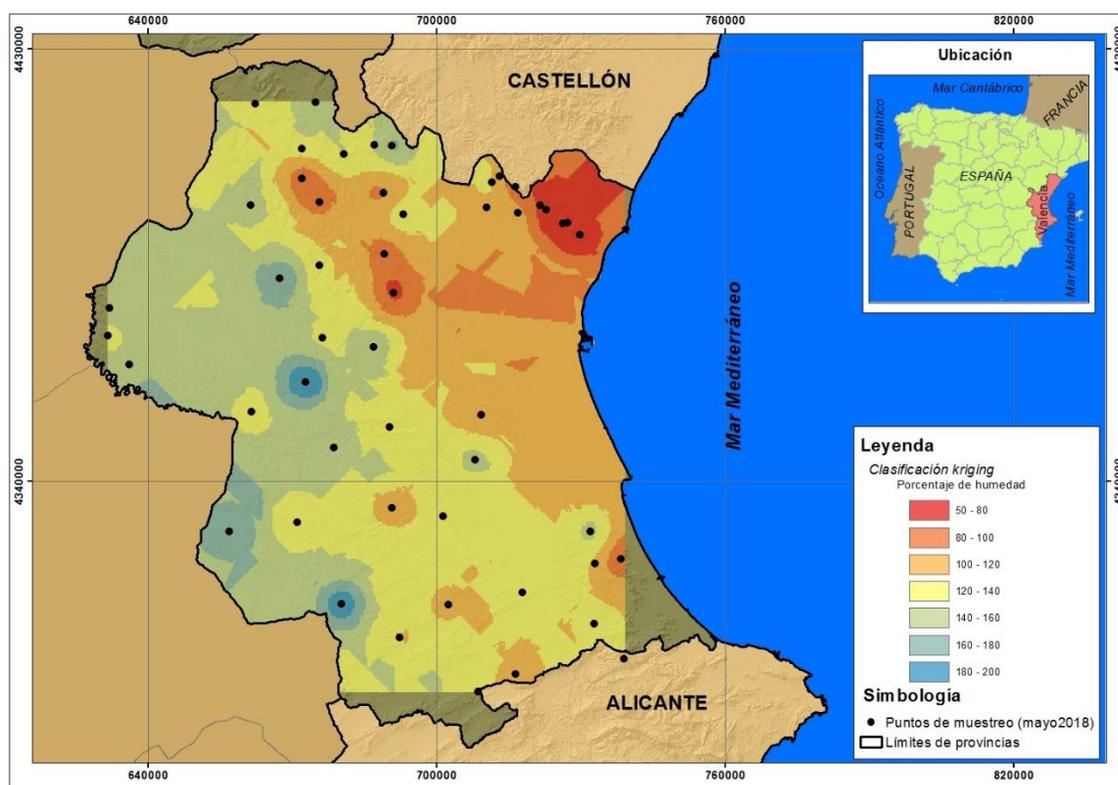
En lo que respecta a la herramienta kriging que proporciona ArcMap, presenta un análisis muy completo en donde se puede realizar un análisis previo, para la exploración de los datos se puede obtener gráficos como son histogramas, NormalQQPlot, inclusive se puede escoger los parámetros del semivariograma experimental. En lo que respecta a la creación de la superficie de estimación de la variable de estudio como tal, el software presenta una alta variedad de parámetros de los cuales se puede modificar y así poder obtener un modelado según las necesidades del usuario, además el software permite realizar modelos partiendo de conocimientos básicos ya que también realiza caculos por defecto, y solo con introducir los puntos o datos, el software tiene la facultad de entregar un producto.

Sin embargo en la figura 5.2.3-3 se muestra el krigeado modelado en el software ArcMap utilizando los parámetros establecidos dados en la figura 5.2.2-1.

Para la validación del Krigeado y escoger el mejor modelo, este puede ser evaluado por la bondad de ajuste del modelo de semivariograma elegido con respecto a los datos muestrales, el más utilizado es la validación cruzada, misma que consiste en excluir la observación de uno

de los n puntos muestrales y con los n-1 valores restantes y el modelo de semivariograma escogido, predecir vía kriging el valor de la variable en estudio en la ubicación del punto que se excluyó. Se piensa que si el modelo de semivarianza elegido describe bien la estructura de autocorrelación espacial, entonces la diferencia entre el valor observado y el valor estimado debe ser pequeña, en el Anexo 9 se presenta la validación cruzada para el modelo final. Este procedimiento se realiza en forma secuencial con cada uno de los puntos muestrales y así se obtiene un conjunto de n “errores de predicción”. (Henaó, 2002)

Figura 5.2.3-3 Mapa de estimación del porcentaje de humedad del romero (superficie generada con el kriging ordinario)



Para obtener el resultado mostrado en la figura anterior, se realizaron diferentes pruebas y para su evaluación se tomó el error cuadrático medio, este proceso se realizó con la finalidad de comparar los modelos con valores por defecto del software ArcMap y los valores de los parámetros obtenidos con RStudio, los parámetros utilizados en cada modelo se puede observar en la tabla 5.2.3-1.

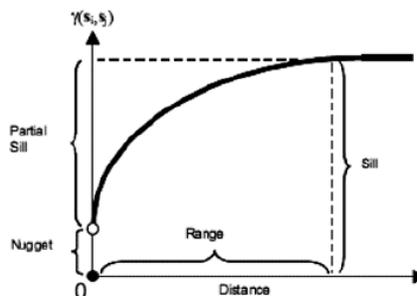
Los modelos generados con los parámetros por defecto calculados por ArcMap dan un error alto, al usar los parámetros obtenidos con el análisis en RStudio para cada modelo teórico, se dan dos situaciones, el modelo esférico empeora mientras con el exponencial y gaussiano bajan, mostrando un menor error medio cuadrático del modelo de estimación utilizando el semivariograma teórico gaussiano.

Tabla 5.2.3-1 Modelos realizados

Kriging ordinario			Parámetros						
Número proceso	Nombre raster	Error Medio cuadrático	Modelo	Mayor range	Partial sill	Nug	Número de puntos	Tamaño de lag	distancia máxima
(np)	(nr)	(rms)	(M)	(mg)	(psill)	(n)	(np)	(tl)	np*tl
1	KriHCVorddef0	27.4811	por defecto	por defecto	por defecto	por defecto	por defecto	por defecto	por defecto
2	KriHCVordEsf1	27.4258	Esférico	por defecto	por defecto	por defecto	por defecto	por defecto	por defecto
3	KriHCVordEsf2	27.7367	Esférico	22388.740	1059.828	0.000	15	3500	52500
4	KriHCVordEsf3	27.7367	Esférico	22388.740	1059.828	0.000	10	5000	50000
5	KriHCVordEsf4	27.7367	Esférico	22388.740	1059.828	0.000	12	4000	48000
6	KriHCVordExp1	27.5945	Exponencial	por defecto	por defecto	por defecto	por defecto	por defecto	por defecto
7	KriHCVordExp2	26.1119	Exponencial	14588.480	1265.858	0.000	15	3500	52500
8	KriHCVordExp3	26.1119	Exponencial	14588.480	1265.858	0.000	10	5000	50000
9	KriHCVordExp4	26.1119	Exponencial	14588.480	1265.858	0.000	12	4000	48000
10	KriHCVordGau1	27.4811	Gaussiano	por defecto	por defecto	por defecto	por defecto	por defecto	por defecto
11	KriHCVordGau2	25.5990	Gaussiano	9091.324	956.640	100.300	15	3500	52500
12	KriHCVordGau3	25.5990	Gaussiano	9091.324	956.640	100.300	10	5000	50000
13	KriHCVordGau4	25.5990	Gaussiano	9091.324	956.640	100.300	12	4000	48000

Los 3 parámetros principales son el denominado mayor rango por ArcMap que hace referencia al rango y corresponde la distancia a partir de la cual dos observaciones son independientes, por ello se llama rango efectivo a la distancia para la cual el semivariograma alcanza el 95% de la meseta.; el Nugget (nug) o Efecto pepita misma que representa la discontinuidad puntual del semivariograma en el origen, cabe destacar que es te no debe superar el 50% de la meseta; Partial sill (psill) hace referencia a la parte de la meseta sin tomar en cuenta el valor del Efecto pepita.

Figura 5.2.3-4 Representaciones de los parámetros básicos en un semivariograma.



Fuente: web

También se tomó en cuenta 3 parámetros con respecto al cálculo del semivariograma experimental: tamaño del lag (paso) la cual hace referencia a la distancia a la que se calcula los puntos del semivariograma y el número de puntos (número de pasos o lag) denotando cabe la redundancia el número de puntos con los que se va a formar el semivariograma experimental.

### 5.2.3.1 Mapa de la superficie de estimación del porcentaje de humedad final

Tal como se explicó anteriormente el modelo teórico utilizado fue el gaussiano, y los valores de los parámetros son los que se encuentran en la tabla 5.2.3.1-1. Al final la estimación del modelo de superficie más óptimo presenta los errores de estimación mostrados en la tabla 5.2.3.1-1.

El error medio cuadrático fue el utilizado para estimar los diferentes modelos, con lo que respecta al error medio su valor debe estar próximo a cero y el error medio cuadrático estandarizado tiene que estar cerca de 1. Como se puede notar las estimaciones con el modelo tengan una precisión media, esto se puede dar ya que el número de puntos con el que se realizó fue pequeño para realizar el modelo.

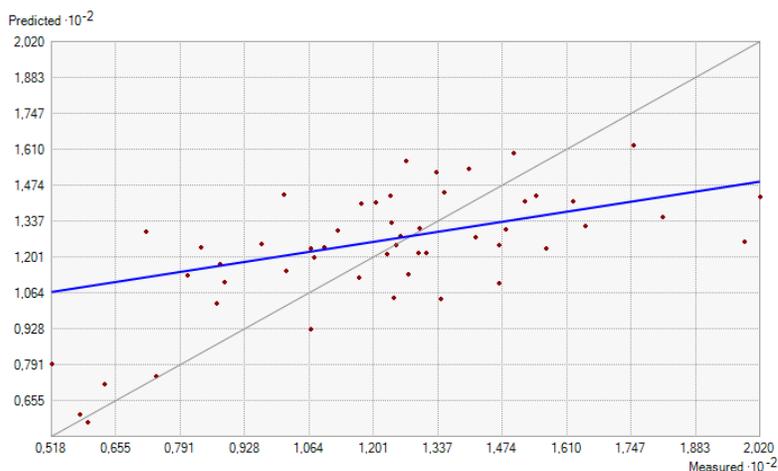
Tabla 5.2.3.1-1 Parámetros del modelo final y los errores de estimación

Parámetros modelo	
Modelo	Gaussiano
Mayor range	9091.324
Partial sill	956.640
Nug	100.300
Número de puntos (pares de datos)	12
Tamaño de lag	4000
distancia máxima	48000

Muestra	51
Media	1.067
Media cuadrática	25.610
Media estandarizada	0.044
Media cuadrática estandarizada	0.782
Error estándar promedio	31.979

En la figura 5.2.3.1-1 se muestra el gráfico de estimación, a medida que los puntos se ajusten más a la línea que pasa por el origen, mejor es el modelo de semivariograma utilizado para realizar el krigeado, pero como se puede ver con respecto a la línea correspondiente a la nube de puntos existe una orientación notable.

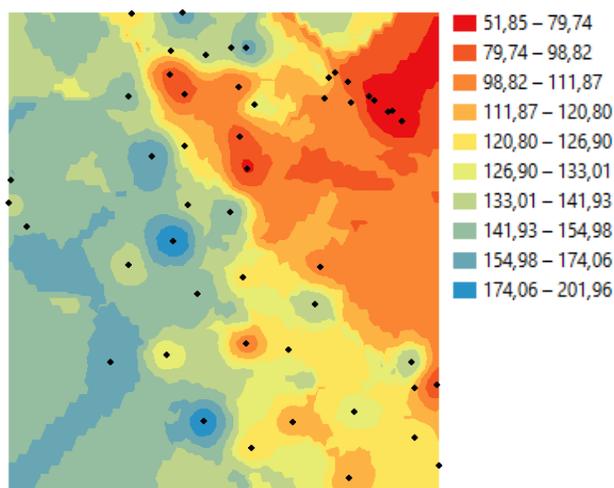
Figura 5.2.3.1-1 Gráfico de estimación



Realizado en: RStudio

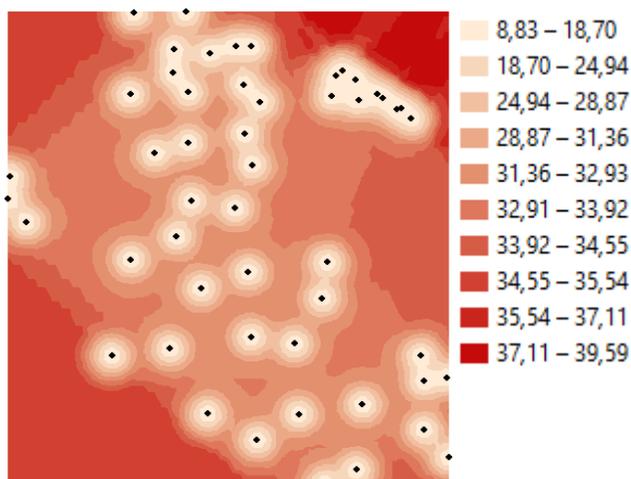
El modelo de superficie final se muestra en la figura 5.2.3.1-2, en donde los valores de porcentaje de humedad se extienden en un rango de 51 a 202. Si se compara con el modelo de krigeado obtenido por RStudio (figura 5.1.3-2), su diferencia radica en la presentación ya que el presentado por ArcMap cada color representa un rango de valores.

Figura 5.2.3.1-2 Superficie de estimación del porcentaje de humedad del romero



En la figura 5.2.3.1-3 se presenta el mapa de error estándar de estimación del porcentaje de humedad del romero, en donde claramente se puede notar que a medida se alejan de los puntos de muestreo con los que se realizó el modelo aumenta el error.

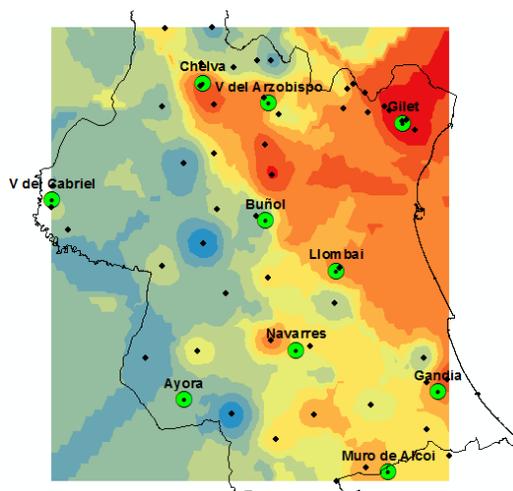
Figura 5.2.3.1-3 Error estándar de estimación del porcentaje de humedad del romero



### 5.2.3.2 Estimaciones con puntos externos

Como se explicó en el principio del apartado 5, se tiene 10 puntos que contienen el registro del porcentaje de humedad de mayo de 2015 al 2017 y se encuentran en el área del modelo de la superficie de estimación, con la finalidad de evaluarlo se procede a realizar la estimación del porcentaje de humedad del romero en dichos puntos. En la figura 5.2.3.2-1 se muestra los puntos externos con los cuales se va a realizar la estimación sobre la superficie de estimación realizada con el método de interpolación kriging y la limitación de la provincia de valencia, con un número total de 10 puntos.

Figura 5.2.3.2-1 Ubicación de los puntos externos



ArcMap cuenta con una herramienta asociada a kriging con la cual se puede extraer el valor de cualquier punto siempre y cuando se encuentre dentro del modelo de superficie, además también asocia al punto el valor del error estándar, el cual se puede considerar como el rango de incertidumbre por la cual el valor del porcentaje en ese punto puede oscilar. El rango de incertidumbre para la mayoría de datos se puede considerar que es alto. (Ver tabla 5.2.3.2-1).

Tabla 5.2.3.2-1 Comparativa entre la estimación de HCV obtenida por kriging para mayo de 2018 y los valores observados ese mes en años anteriores.

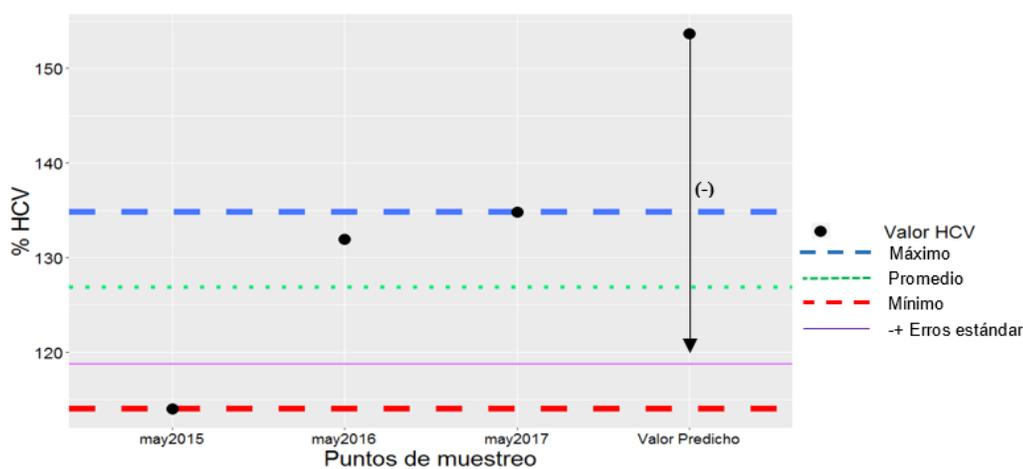
CodFecha	Fecha	AYORA	BUNYOL	LLOMBAI	VILLAR_ARZOBISPO	CHELVA	ALBAIDA_MURO	GILET	VILLARGORDO	NAVARRES	GANDIA
1	may2015	114.00	122.00	103.00	124.00	113.00	157.00	117.80	118.14	163.89	166.33
2	may2016	131.93	142.56	98.08	143.72	107.25	123.86	85.00	102.00	131.98	232.97
3	may2017	134.80	121.08	113.64	164.96	96.10	130.24	141.97	96.00	142.65	170.37
4	Promedio	126.91	128.55	104.91	144.23	105.45	137.03	114.92	105.38	146.17	189.89
5	Valor mínimo	114.00	121.08	98.08	124.00	96.10	123.86	85.00	96.00	131.98	166.33
6	Valor máximo	134.80	142.56	113.64	164.96	113.00	157.00	141.97	118.14	163.89	232.97
7	Estimación kriging	153.674	142.544	108.828	107.298	91.601	121.035	58.314	140.471	123.205	103.028
8	(+ Erros estándar)	34.922	22.038	12.434	14.031	11.833	32.369	8.019	13.454	25.107	23.569
	Rango de oscilación (- Erros estándar)	118.752	120.507	96.394	93.268	79.768	88.666	50.295	127.017	98.098	79.459
	(+ Erros estándar)	188.595	164.582	121.261	121.329	103.434	153.405	66.332	153.925	148.312	126.598

A continuación se realizaron unas gráficas por punto de muestreo con la finalidad de mostrar el valor estimado con los datos de mayo del 2018 y el registro del valor del porcentaje de humedad muestreado de mayo del 2014 al 2017 y así observar la variación de un año a otro.

En las figuras no se mostrará el rango de oscilación del valor estimado, solo se tendrá en cuenta el que se aproxime al registro de datos de los mayos anteriores, para lo cual, se

determinó el promedio, valor máximo y mínimo del registro de porcentaje de humedad de mayo del 2014 al 2017, mismo que son mostrados por líneas entre punteadas que cortan la gráfica paralelas al eje x, el rango de oscilación del punto estimado se muestra de igual forma con una línea constante, con una flecha que parte del valor del punto estimado al rango de oscilación, se indica la distancia existente, además con un signo positivo o negativo se señala si parte del rango presentado es el error estándar que se sumó o restó al valor estimado, como ejemplo se muestra el punto de muestreo denominado Ayora (ver figura 5.2.3.2-2)

Figura 5.2.3.2-2. Valor del punto estimado con rango de oscilación respecto a los datos de registros, punto de muestreo Ayora.



## 6. RESULTADOS

### 6.1 Análisis de conglomerados

El análisis de conglomerados se dividió en dos líneas estudio tal como se expresó al principio del apartado 3, en la primera línea se calculó el porcentaje de humedad de la especie vegetal romero por estaciones del año para cada punto de muestreo, al no tener datos completos por mes y al ser un requisito para el sistema de agrupamiento seleccionado, se presentó como la mejor opción. También se realizó un análisis grupal e individual con los estadísticos promedio, mediana, cuantiles inferior y superior.

El análisis realizado con los estadísticos de forma individual tuvo como objeto, observar la

influencia o las variaciones que presentan con respecto al análisis donde se agrupan todos ellos.

Sabiendo que el sistema de conglomerados utilizado, clasifica elementos de acuerdo a su semejanza y al estar basado en criterios geométricos, las agrupaciones o conglomerados son realizados en base al anidamiento jerárquico en función de la distancia existente entre individuos por lo que no se encontró ninguna relación entre los análisis de los estadísticos de forma individual respecto al grupal.

Los estadísticos obtenidos de los datos por estación del año y por mes presentan variaciones pequeñas si se las compara, sin embargo dentro del análisis de conglomerados, los anidamientos y agrupaciones se muestran diferentes unos con otros, coincidiendo que el análisis de conglomerados es muy sensible y da resultados diferentes al presentar pequeñas variaciones.

Dentro de los análisis con los datos de humedad por estaciones del año, con lo que respecta a los estadísticos, el realizado con el promedio mostro igualdad en el número de integrantes por conglomerados y distribuyo a los puntos de muestreo del mismo modos que el análisis donde se agrupan a todos los estadísticos (ver tablas 3.1.2.1-1,2,3,4), sin embargo al observar los dendogramas donde se presenta el anidamiento y jerarquización se puede notar que las agrupaciones iniciales dentro de cada clase son muy diferentes por lo que se descarta la idea de que el estadístico promedio sea de mayor influencia. Para corroborarlo si se observa el comportamiento de los mismos análisis pero con los obtenidos a partir de los datos de humedad por mes, claramente se puede ver que aunque compartan similares número de individuos en cada agrupación en las clases 4 y 5, no ubica a los puntos de muestreo ni a los conglomerados de la misma forma, es decir el conglomerado 2 del análisis con los estadísticos agrupados tiene 3 integrantes mientras que en el realizado con el promedio presenta 5, además la distribución de los puntos de muestreo son pocos los que coinciden con la misma clase.

Si se analiza los conglomerados con datos por estación el año, teniendo solo en cuenta la distribución y agrupamiento de los puntos de muestreo en las clases 4 y 5, destacan el realizado con el grupo de estadísticos y el promedio de forma individual.

Al contrastar con los resultados del análisis realizado con los datos de humedad por mes, se

aprecia que los divididos en 4 clases tienen un grupo dominante en donde se concentran una gran cantidad de puntos de muestreo, formando a su vez conglomerados con número de integrantes desiguales, en los de 5 clases se mantienen dichos grupos dominantes, en los análisis realizados con los estadísticos agrupados, y los estadísticos individuales promedio y mediana, los conglomerados obtenidos de los cuantiles presentan una mejor distribución de puntos entre las clases. Cabe destacar que el análisis de conglomerados por estadísticos separados se lo realizó con el afán de estudiar las diferencias y semejanzas.

Los análisis con los diferentes estadísticos tiene la finalidad de encontrar una manera fácil de agrupar objetos dentro de una misma variable en caso de que no se tenga datos completos, o ninguna otra forma de rellenar los espacios faltantes. Los estadísticos son una opción muy acertada ya que por medio de ellos se puede realizar una buena descripción de la variable, muy útil para agrupar elementos afines dentro de la misma con el análisis de conglomerados.

Hay que destacar que los estadísticos obtenidos de forma directa de la variable por mes, como se muestra en el diagrama de caja y bigotes (ver figura 2.1-4) tiene más datos atípicos, con respecto a al diagrama de caja y bigotes con datos del porcentaje de humedad por estación del año (ver figura 2.3-1), se le puede atribuir a ello que los conglomerados presenten clases muy desiguales, ya que las agrupaciones son realizadas por el cálculo de distancias en función de los valores de cada punto de muestreo y al ser un sistema sensible a las variaciones tal como se ha venido mostrando.

Al realizar el agrupamiento de los valores de humedad por estaciones del año disminuye la presencia de datos atípicos en la variable, y al obtener los estadísticos de cada uno de ellos se suaviza mucho más los datos, a ello se le puede atribuir que los conglomerados en el anidamiento se vayan relacionando de forma más uniforme, la desventaja que presenta es la pérdida de información.

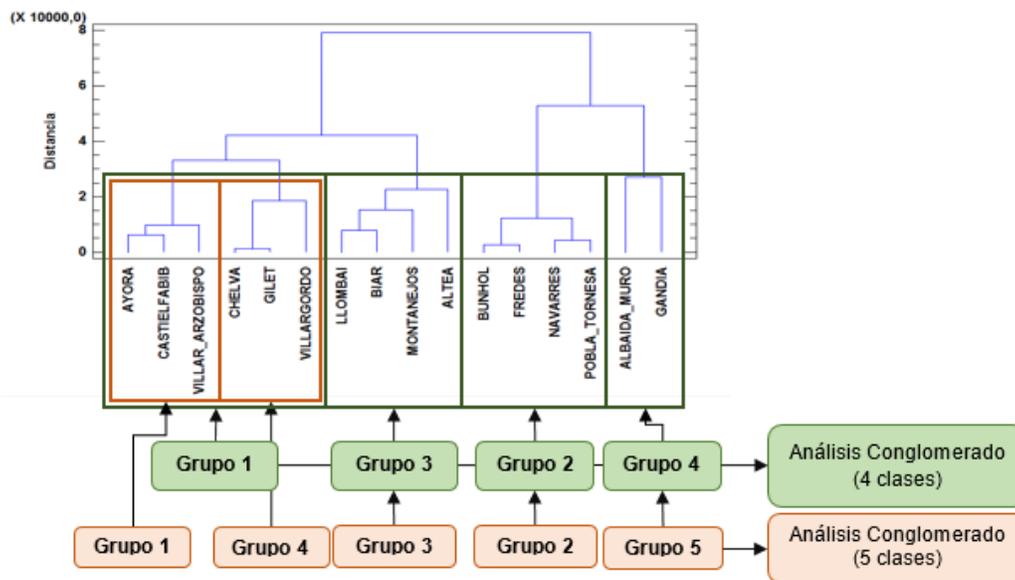
Por lo cual, en este caso la variable con datos del porcentaje de humedad de la especie vegetal romero por estaciones del año es la más apta, ya que la pérdida de información es menor.

La utilización de los estadísticos obtenidos de los datos de humedad por mes de forma directa es descartada ya que presentan anidamientos muy desiguales atribuyendo a que los valores atípicos existentes sean la causa del resultado de los diferentes análisis. Sin embargo no queda descartada la idea de poder utilizarlos bajo un mejor tratamiento, además se debería

probar el introducir más estadísticos dentro del análisis principal.

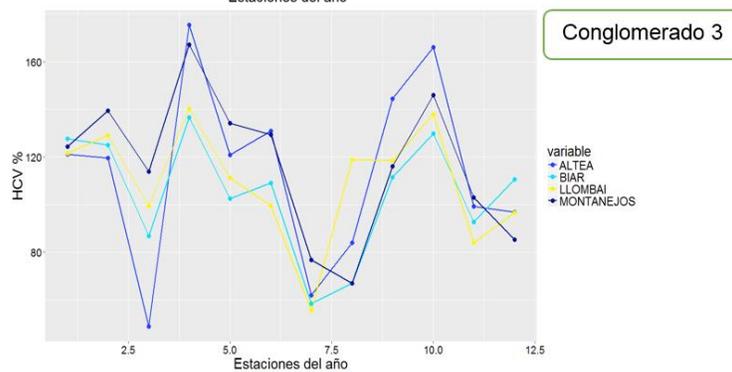
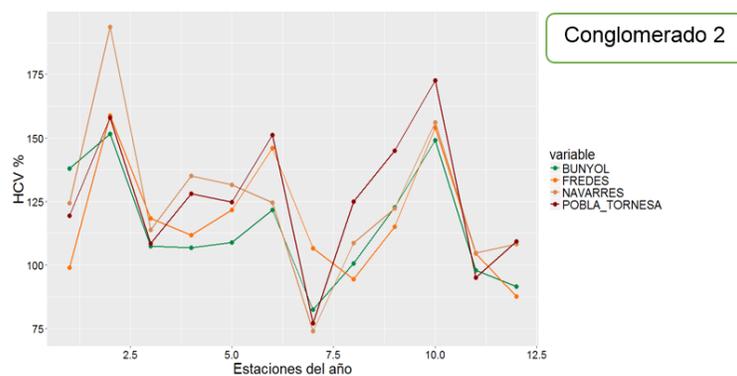
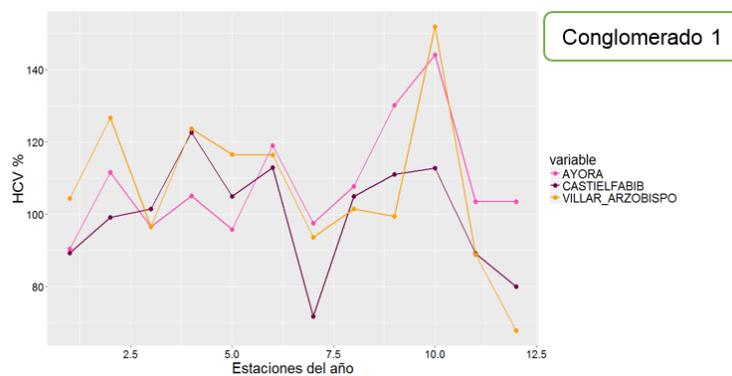
De los diferentes análisis de conglomerados sobresale el realizado con los datos por estación del año, el número de integrantes por grupo son muy parejos entre ellos, la jerarquización generada anida los puntos de muestreo en agrupaciones con número de datos muy similares haciendo posible utilizar cualquiera de las dos divisiones estudiadas. En la figura 6.1-1 se muestra el dendograma con las dos divisiones y sus respectivas clases.

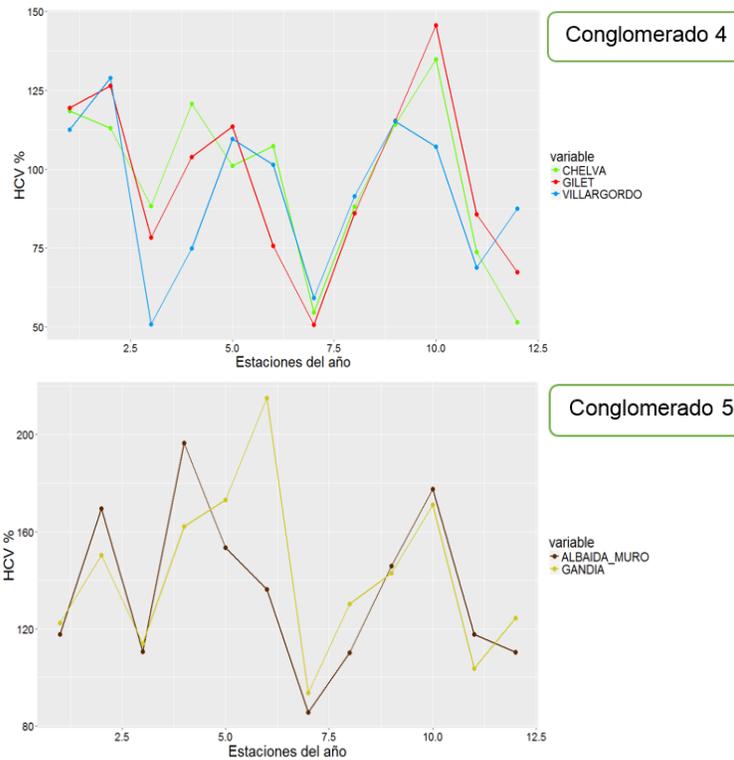
Figura 6.1-1 Dendograma con los datos de humedad de la especie vegetal romero por estación del año.



En la figura 6.1-2 se muestran diagramas lineales de las agrupaciones de puntos de muestreo resultado del análisis de conglomerados en donde claramente se puede observar la existencia de semejanza entre los integrantes de cada punto, el rango del valor de humedad en el que se desplazan los puntos, como la alternación de los picos máximos y mínimos en cada estación estadísticamente muestran similitud.

Figura 6.1-2 Diagrama lineal de tiempo por puntos de muestreo agrupados en base al promedio por estación del año.

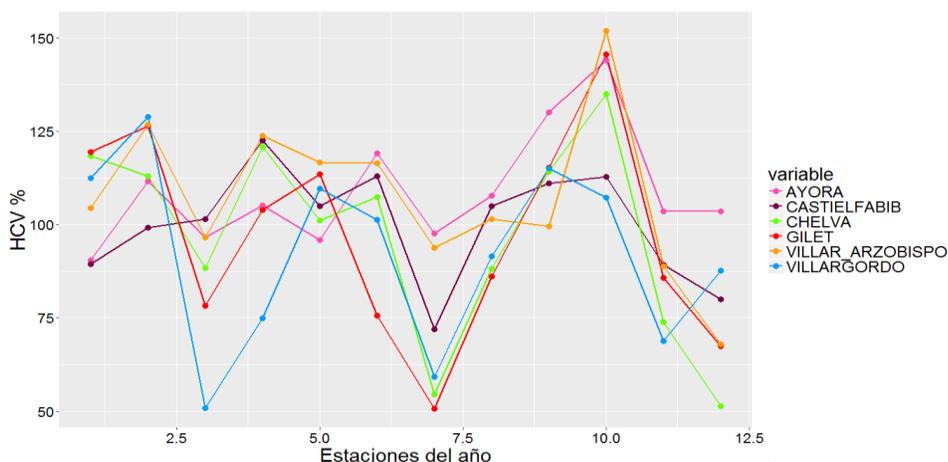




La leyenda del eje x está codificada por números del 1 al 12, cada uno representa a la estación del año (invierno, primavera, verano y otoño) en ese orden, el primer número corresponde a invierno de 2014 y el final a otoño de 2017.

En la figura 6.1-3 se realizó la unión de los grupos 1 y 4 ya que al formar conglomerados de 4 clases son considerados los siguientes a juntarse por afinidad y como se puede apreciar en la gráfica existe coherencia en el anidamiento y jerarquización presentados en el análisis de conglomerados siendo bastante útil al tener opciones en crear grupos partiendo de la afinidad de los elementos dentro de una variable.

Figura 6.1-3 Diagrama lineal de tiempo unión conglomerado 1 y 4.

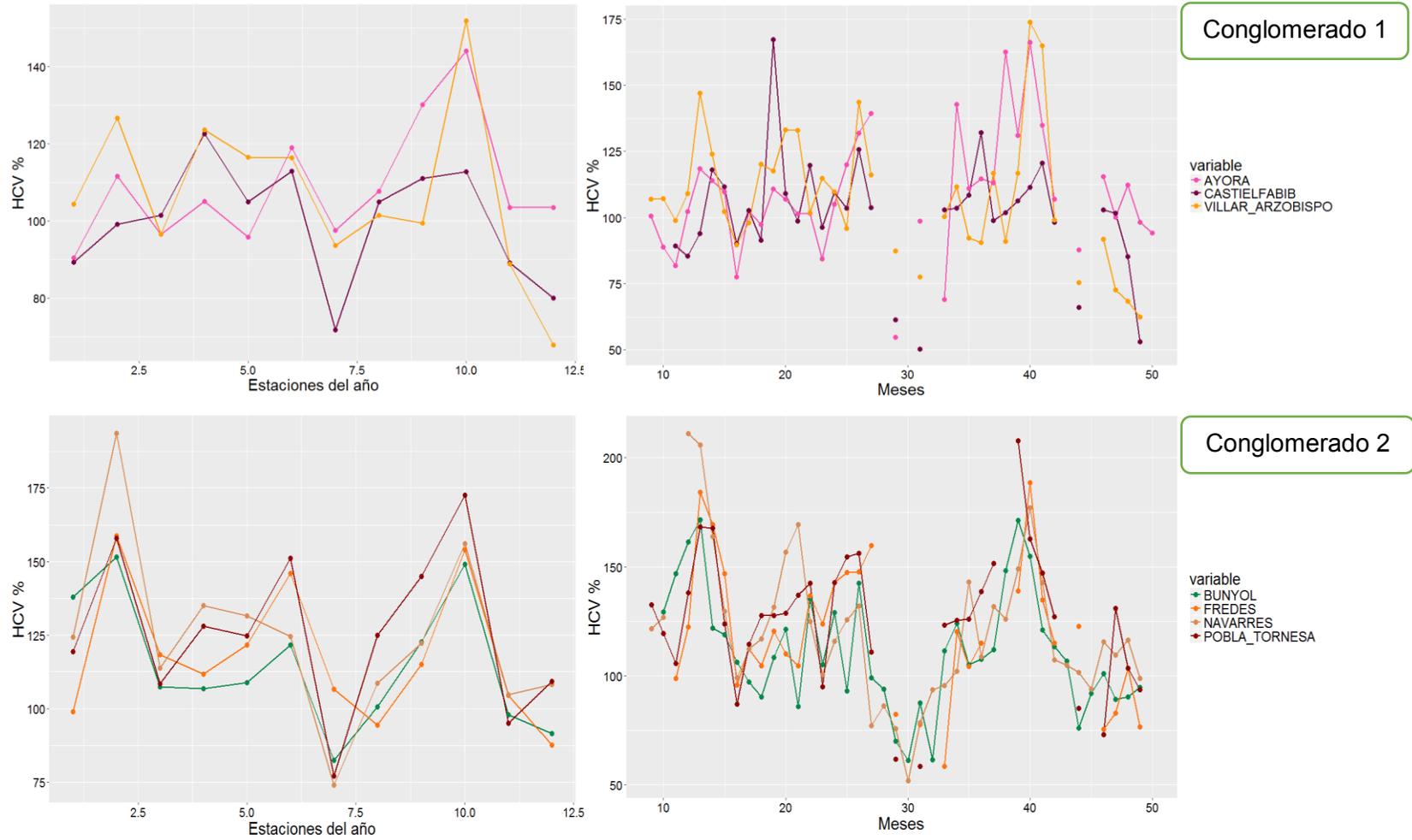


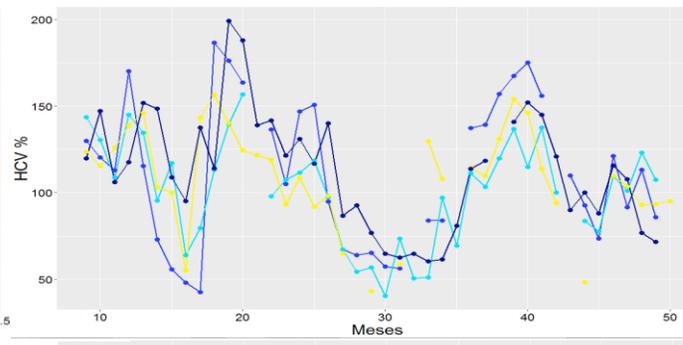
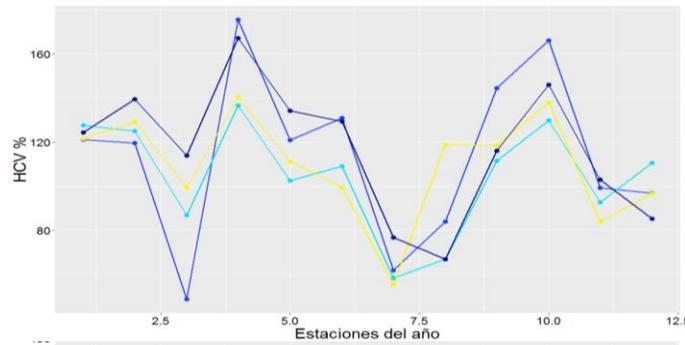
La leyenda del eje x está codificada por números del 1 al 12, cada uno representa a la estación del año (invierno, primavera, verano y otoño) en ese orden, el primer número corresponde a invierno de 2014 y el final a otoño de 2017.

Al comparar los puntos de muestreo de las 5 clases agrupadas entre los datos calculados por estaciones del año y los datos originales por meses, si se observa la figura 6.1-4 es clara la pérdida de información, el punto de muestreo Ayora y Villar-Arzobispo presenta variaciones notorias y si se observa en el resto de conglomerados también se podrá visualizar más casos como el mencionado, sin embargo se presenta como una opción muy coherente.

El clasificar los 16 puntos de muestreo por afinidad en 5 conglomerados se presenta como la mejor opción, ya que son agrupaciones con un número de integrantes muy parejo y no hay clases con un solo integrante, de todas formas si se quiere agrupar en un número menor como 4, podría resultar útil el mismo análisis.

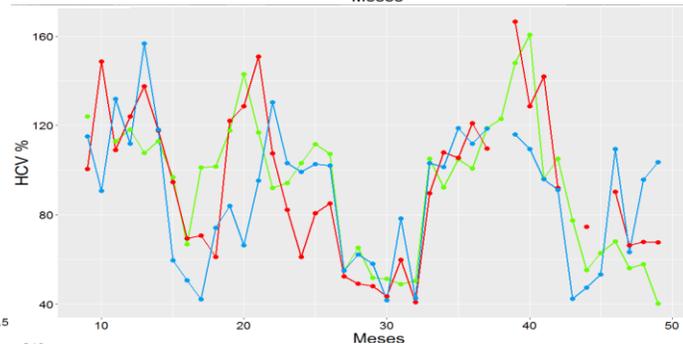
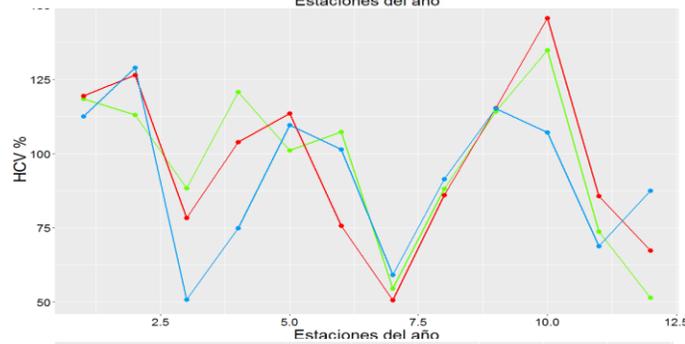
Figura 6.1-4 Diagrama lineales del conglomerado uno con los datos de humedad calcados por estación del año y meses.





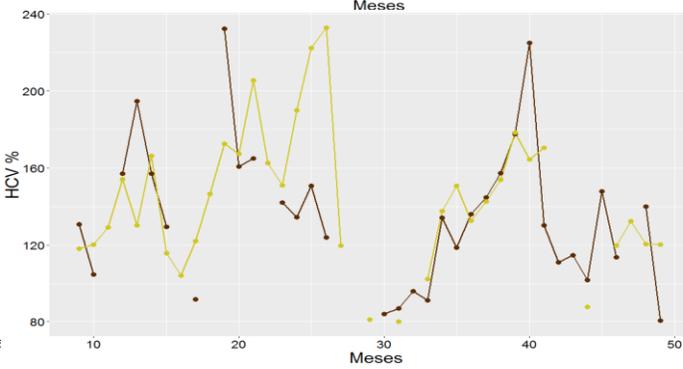
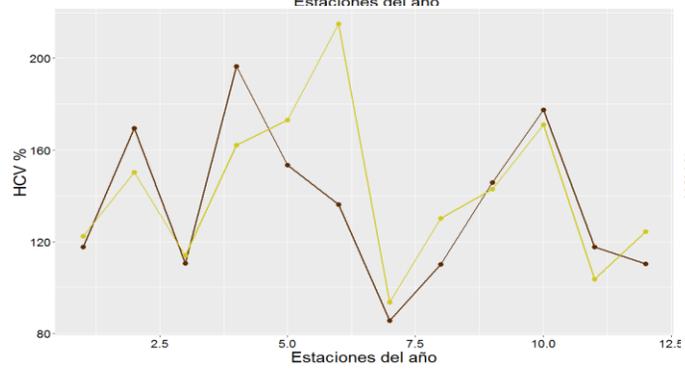
Conglomerado 3

variable  
 ALTEA  
 BIAR  
 LLOMBAI  
 MONTANEJOS



Conglomerado 4

variable  
 CHELVA  
 GILET  
 VILLARGORDO



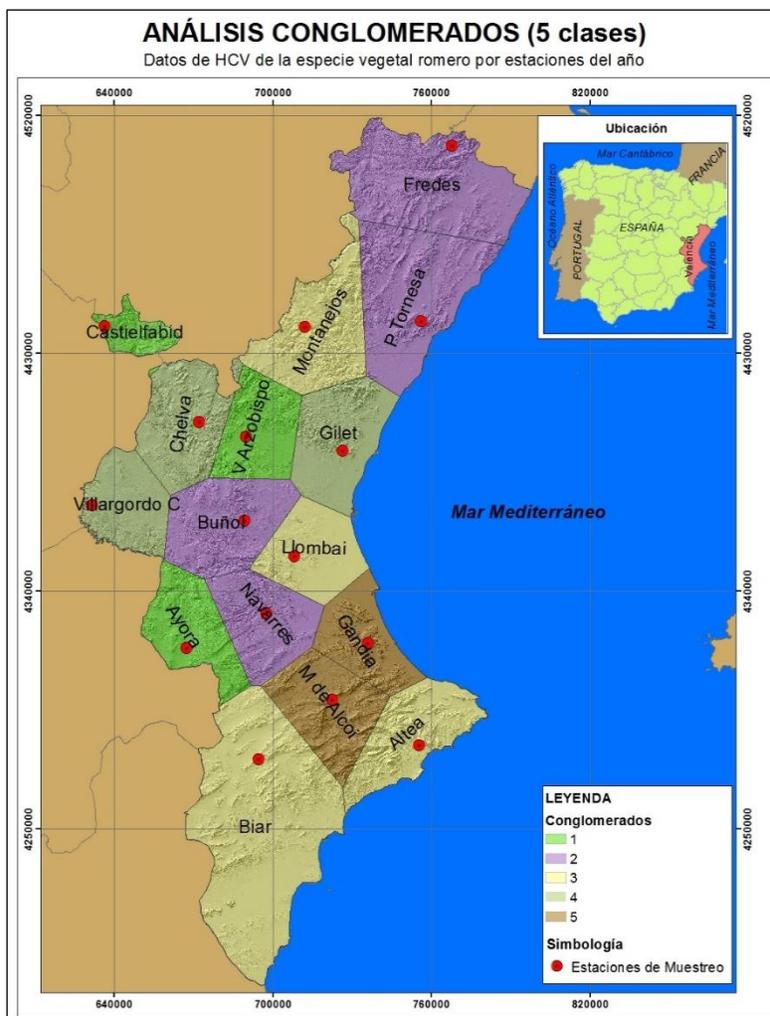
Conglomerado 5

variable  
 ALBAIDA\_MURO  
 GANDIA

La leyenda del eje x está codificada por números del 1 al 12, cada uno representa a la estación del año (invierno, primavera, verano y otoño) en ese orden, el primer número corresponde a invierno de 2014 y el final a otoño de 2017.

A continuación en la figura 6.1-5 se realiza la presentación de las 5 agrupaciones y sus integrantes de manera espacial, la influencia del punto de muestreo y su área de influencia se interpretó mediante polígonos de Thiessen, misma que clasifica el área de la Comunidad Valenciana en zonas determinadas por las distancias más cercanas a los puntos asociados, como son los puntos de muestreo.

Figura 6.1-5 Puntos de muestreo distribuidos en los 5 conglomerados.



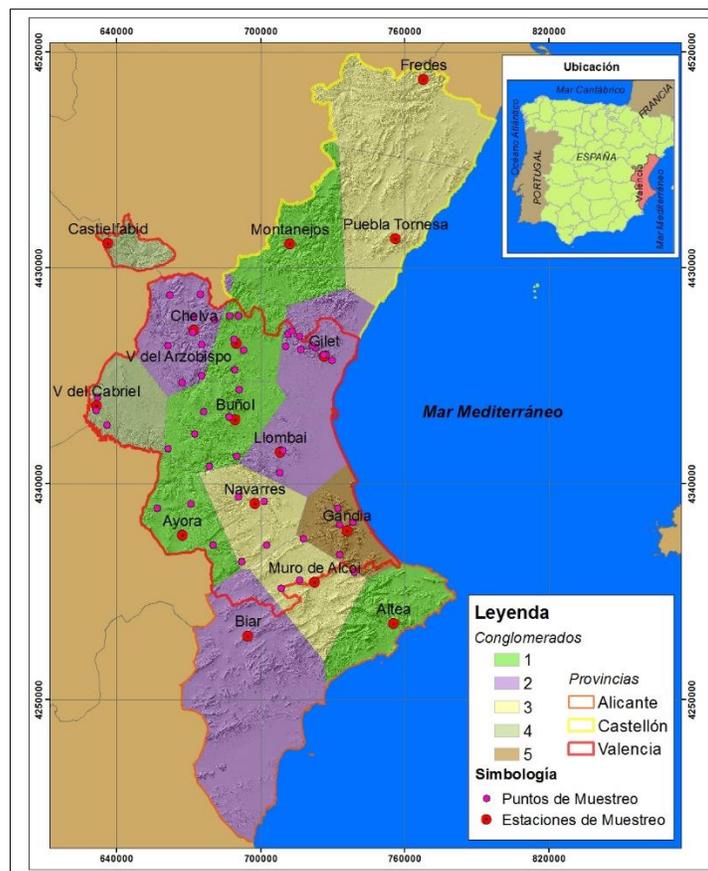
## 6.2 Análisis de percentiles como clasificadores

En el apartado 3.2 Se determinó los percentiles por conglomerado de cada una de las estaciones del año, mismo que serán utilizados como categorizadores para clasificar en base a los datos de la variable humedad de la especie vegetal romero en el periodo de 3 años (diciembre del 2014 hasta diciembre de 2017) la toma de nuevos datos de humedad.

En la figura 3.2-1 se observa la ubicación de 51 puntos de muestreo tomados en mayo del 2018, los cuales son utilizados en esta parte del estudio para clasificarlos de acuerdo a los percentiles establecidos.

Por ser datos tomados en el mes de mayo, el análisis de conglomerados y percentiles obtenidos con los datos de la primavera son los que entraran en la fase de clasificación, los conglomerados y puntos de muestreo a categorizar se muestran en la figura 6.2-1.

Figura 6.2-1 División de conglomerados por datos primavera y puntos de muestreo nuevos.



Para diferencia los puntos de con los que se determinaron los percentiles se los denominará estaciones de muestreo y los datos obtenidos en el mes de mayo serán por punto de muestreo propiamente dichos, aclarado ello, en la figura 6.2-1 los datos de humedad tomados en el mes de mayo de 2018 se encuentran dentro de la provincia de Valencia, dichos puntos para ser categorizados es necesario determinar que estación de muestreo es la que influye sobre ella, ya que de este modo se identificara al conglomerado al que pertenece y así utilizar los percentiles apropiados y clasificarlos. Ya que se ha venido trabajando con los polígonos de Thiessen y si no se tiene alguno otra forma de relacionarlos, esta puede ser una buena opción tomando en cuenta que se crean los polígonos en función a la distancia más cercana a uno de las estaciones.

De esa forma en la tabla 6.2-1 se muestran los puntos de muestreo correspondientes a cada estación, el número de conglomerado al que pertenece y percentiles con los que van a ser catalogados.

Tabla 6.2-1 Distribución y clasificación de los puntos de muestreo (mayo 2018).

Provincia	Estación de muestreo	Num punto muestreo	Num Conglomerado	Percentiles					
				10%	20%	40%	60%	80%	90%
Valencia	AYORA	3	Con 1	117.45	119.50	128.30	141.25	149.58	151.81
	BUNYOL	7							
	VILLAR_ARZOBISPO	5							
	CHELVA	9	Con 2	100.31	107.69	117.80	128.14	133.86	137.59
	GILET	10							
	LLOMBAI	2							
	ALBAIDA_MURO	3	Con 3	137.27	146.95	154.95	158.44	172.01	177.12
	NAVARRES	4							
	VILLARGORDO	3	Con 4	100.27	101.33	107.11	112.74	112.92	120.89
	GANDIA	5	Con 5	154.43	158.58	166.88	179.83	197.44	206.25
<b>Total</b>		<b>51</b>							

Al tomar los percentiles como rangos para catalogar los puntos de muestreo tal como se muestra en la figura 3.2-2 en donde los valores menores al percentil 10 serán tomados como “Extremadamente seco”; entre el percentil 10 y 20 se considera “Muy seco”, entre el entre el percentil 20 y 40 “Seco”; entre el percentil 40 y 60 se considera “Normal”, entre el percentil 60 y 80 se considera “Húmedo”, entre el percentil 80 y 90 se considera “Muy húmedo” y mayores al percentil 90 “Extremadamente húmedo”

Como resultado final se presenta la tabla 6.2-2 donde se muestra el número de puntos de muestreo por categoría en función de los percentiles donde claramente se puede ver que la mayoría de puntos se distribuyen entre los extremos, también es importante recalcar que el conglomerados 1 y 2 son los más número de puntos de muestreo poseen.

Tabla 6.2-2 Categorización por conglomerado.

Conglomerado	Extremadamente seco	Muy seco	Seco	Normal	Húmedo	Muy húmedo	Extremadamente húmedo	Total Puntos por conglomerado
Con1	3	0	3	3	0	0	6	15
Con2	8	4	0	3	1	0	5	21
Con3	7	0	0	0	0	0	0	7
Con4	0	0	0	0	0	0	3	3
Con5	5	0	0	0	0	0	0	5
Total Puntos por categoría	23	4	3	6	1	0	14	51

A continuación en la tablas de la 6.2-3 hasta la 6.2-7 se muestran cada punto de muestreo con su valor del porcentaje de humedad y su categorización, además se muestra la estación de muestreo por la que fue agrupada. En las figuras comprendidas entre la 6.2-2 y la 6.2-6 se presentan los diagramas de dispersión con los puntos de muestreo donde también son dibujados los percentiles denotando la categorización realizada. Es importante destacar que en la gráfica los puntos de muestreo están identificados por un código creado con el cual está de igual forma vinculado en las tablas.

Tabla 6.2-3 Categorización de los puntos de muestreo del conglomerado 1.

Id	PM_ref	Codnum	HCV	Clasificacion_R
39	AYORA	Ayo1	175.06	Extremadamente húmedo
40	AYORA	Ayo2	123.60	Seco
41	AYORA	Ayo3	198.51	Extremadamente húmedo
25	BUNYOL	Bun1	71.87	Extremadamente seco
26	BUNYOL	Bun2	135.07	Normal
27	BUNYOL	Bun3	156.53	Extremadamente húmedo
31	BUNYOL	Bun4	126.91	Seco
32	BUNYOL	Bun5	201.96	Extremadamente húmedo
35	BUNYOL	Bun6	120.45	Seco
36	BUNYOL	Bun7	152.01	Extremadamente húmedo
11	VILLAR_ARZOBISPO	Vill_ar1	88.36	Extremadamente seco
12	VILLAR_ARZOBISPO	Vill_ar2	134.15	Normal
13	VILLAR_ARZOBISPO	Vill_ar3	164.83	Extremadamente húmedo
14	VILLAR_ARZOBISPO	Vill_ar4	140.22	Normal
15	VILLAR_ARZOBISPO	Vill_ar5	96.24	Extremadamente seco

Tabla 6.2-4 Categorización de los puntos de muestreo del conglomerado 2.

<b>Id</b>	<b>PM_ref</b>	<b>Codnum</b>	<b>HCV</b>	<b>Categorización</b>
16	CHELVA	Che1	141.52	Extremadamente húmedo
17	CHELVA	Che2	87.43	Extremadamente seco
18	CHELVA	Che3	83.48	Extremadamente seco
19	CHELVA	Che4	129.58	Húmedo
20	CHELVA	Che5	162.25	Extremadamente húmedo
21	CHELVA	Che6	123.75	Normal
22	CHELVA	Che7	148.00	Extremadamente húmedo
23	CHELVA	Che8	181.37	Extremadamente húmedo
24	CHELVA	Che9	124.87	Normal
1	GILET	Gile1	59.34	Extremadamente seco
2	GILET	Gile2	51.85	Extremadamente seco
3	GILET	Gile3	57.89	Extremadamente seco
4	GILET	Gile4	62.90	Extremadamente seco
5	GILET	Gile5	74.05	Extremadamente seco
6	GILET	Gile6	106.60	Muy seco
7	GILET	Gile7	106.82	Muy seco
8	GILET	Gile8	124.38	Normal
9	GILET	Gile9	101.60	Muy seco
10	GILET	Gile10	86.68	Extremadamente seco
33	LLOMBAI	Llom1	107.32	Muy seco
34	LLOMBAI	Llom2	146.64	Extremadamente húmedo

Tabla 6.2-5 Categorización de los puntos de muestreo del conglomerado 3.

<b>Id</b>	<b>PM_ref</b>	<b>Codnum</b>	<b>HCV</b>	<b>Categorización</b>
44	ALBAIDA_MURO	Alb1	129.75	Extremadamente seco
45	ALBAIDA_MURO	Alb2	109.67	Extremadamente seco
51	ALBAIDA_MURO	Alb3	122.82	Extremadamente seco
37	NAVARRES	Nav1	101.08	Extremadamente seco
38	NAVARRES	Nav2	125.73	Extremadamente seco
42	NAVARRES	Nav3	112.47	Extremadamente seco
43	NAVARRES	Nav4	117.41	Extremadamente seco

Tabla 6.2-6 Categorización de los puntos de muestreo del conglomerado 4.

Id	PM_ref	Codnum	HCV	Categorización
28	VILLARGORDO	Vill1	154.54	Extremadamente húmedo
29	VILLARGORDO	Vill2	133.29	Extremadamente húmedo
30	VILLARGORDO	Vill3	149.82	Extremadamente húmedo

Tabla 6.2-7 Categorización de los puntos de muestreo del conglomerado 5.

Id	PM_ref	Codnum	HCV	Categorización
46	GANDIA	Gan1	116.85	Extremadamente seco
47	GANDIA	Gan2	80.65	Extremadamente seco
48	GANDIA	Gan3	146.57	Extremadamente seco
49	GANDIA	Gan4	131.08	Extremadamente seco
50	GANDIA	Gan5	127.41	Extremadamente seco

Cabe destacar que en los conglomerados 3, 4 y 5 solo se perciben dos tipos de datos Extremadamente seco y extremadamente húmedo, es necesario contar con más puntos de muestreo en la zona para hacer una mejor clasificación espacial por categorías.

Figura 6.2-2 Diagrama de dispersión con la categorización de los puntos de muestreo del conglomerado 1.

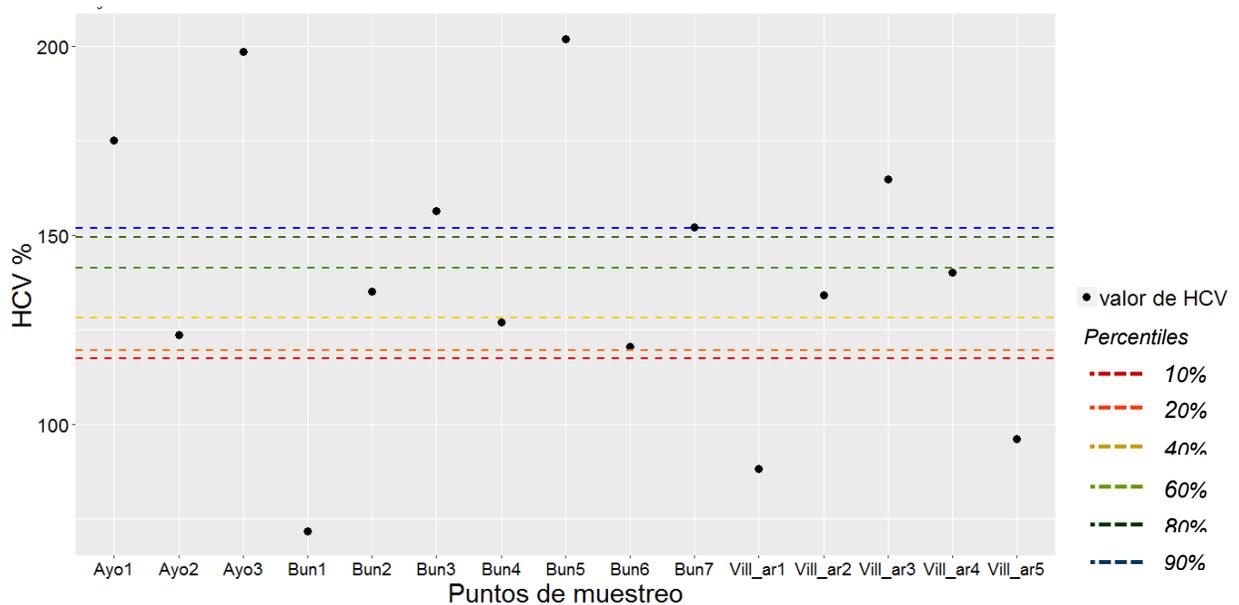


Figura 6.2-3 Diagrama de dispersión con la categorización de los puntos de muestreo del conglomerado 2.

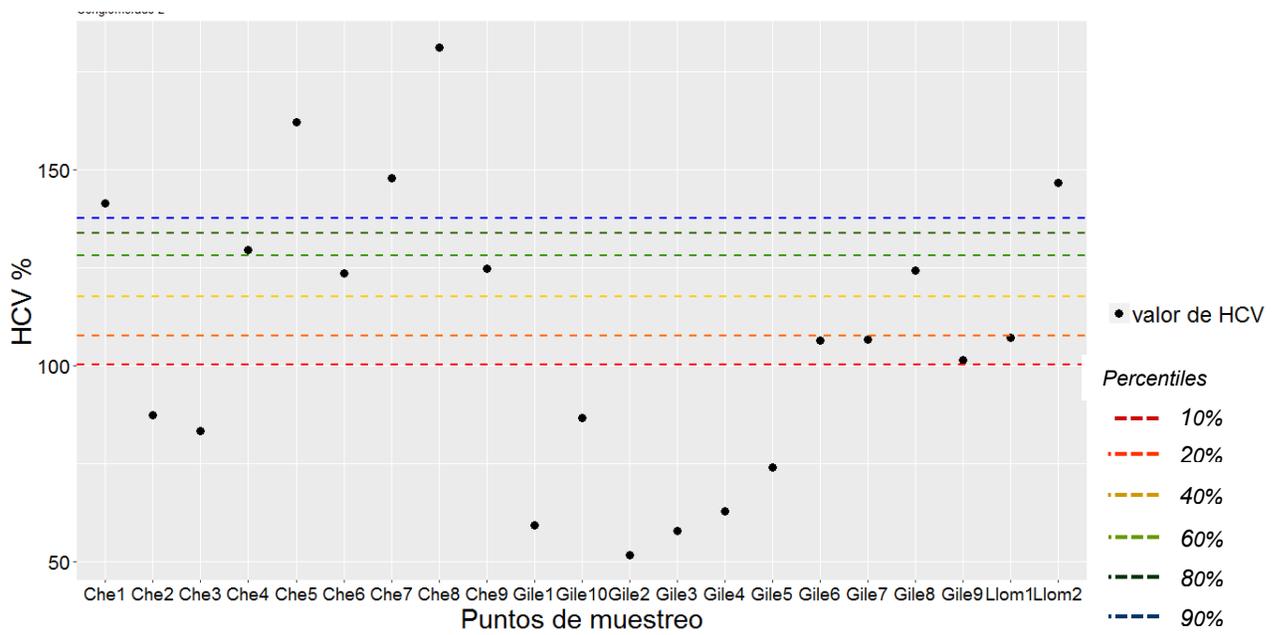


Figura 6.2-4 Diagrama de dispersión con la categorización de los puntos de muestreo del conglomerado 3.

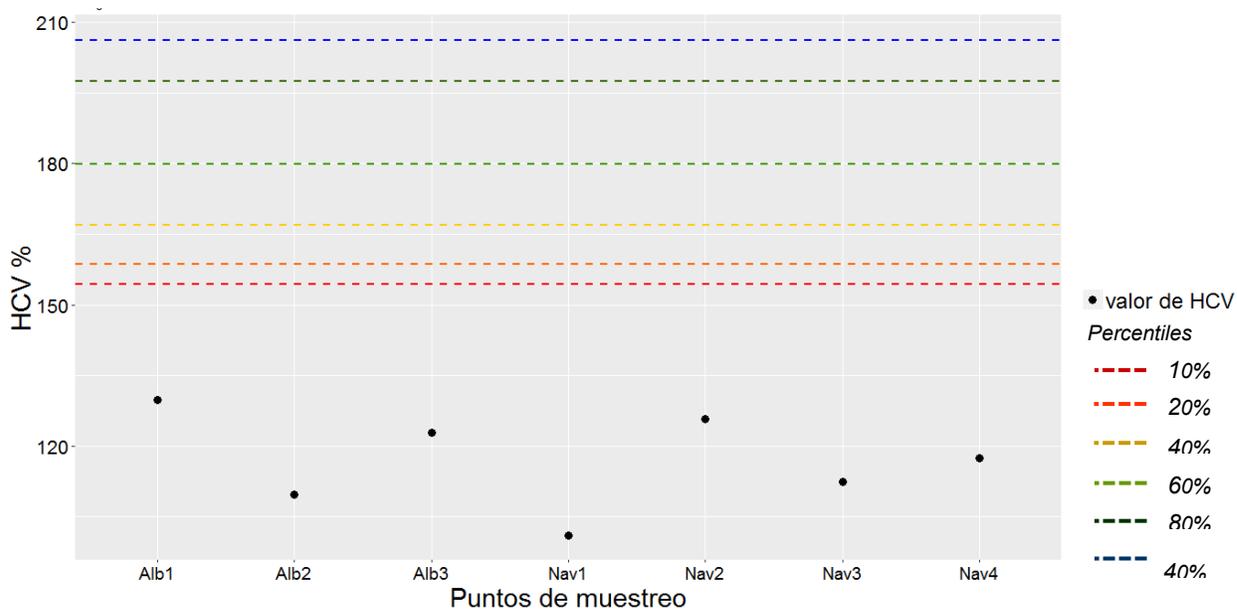


Figura 6.2-5 Diagrama de dispersión con la categorización de los puntos de muestreo del conglomerado 4.

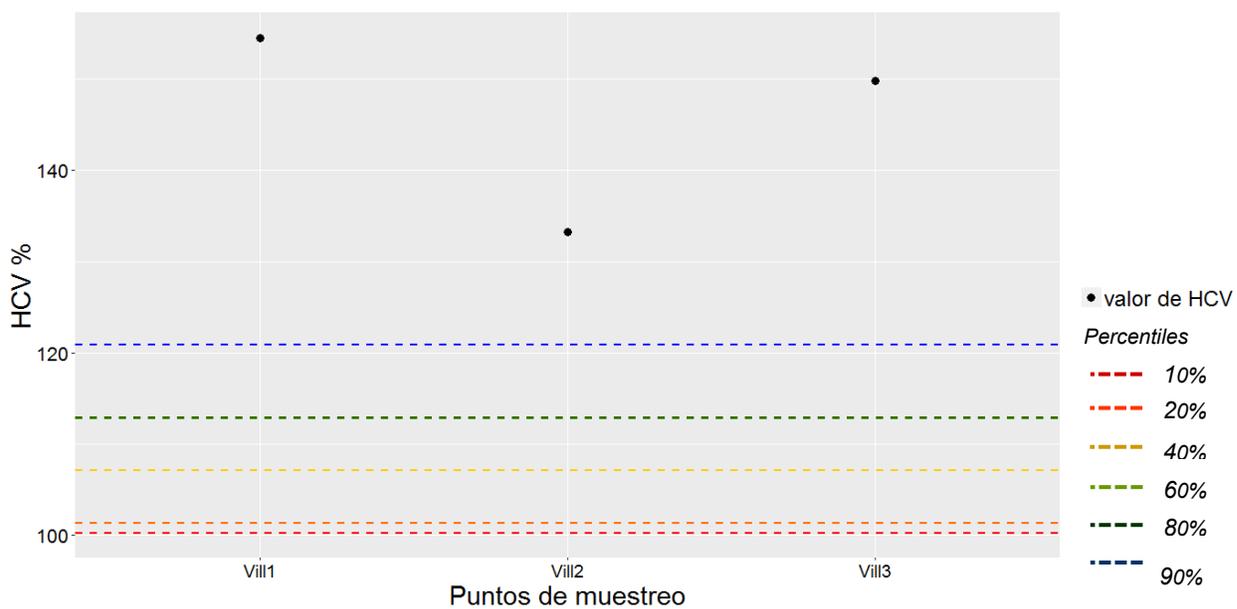
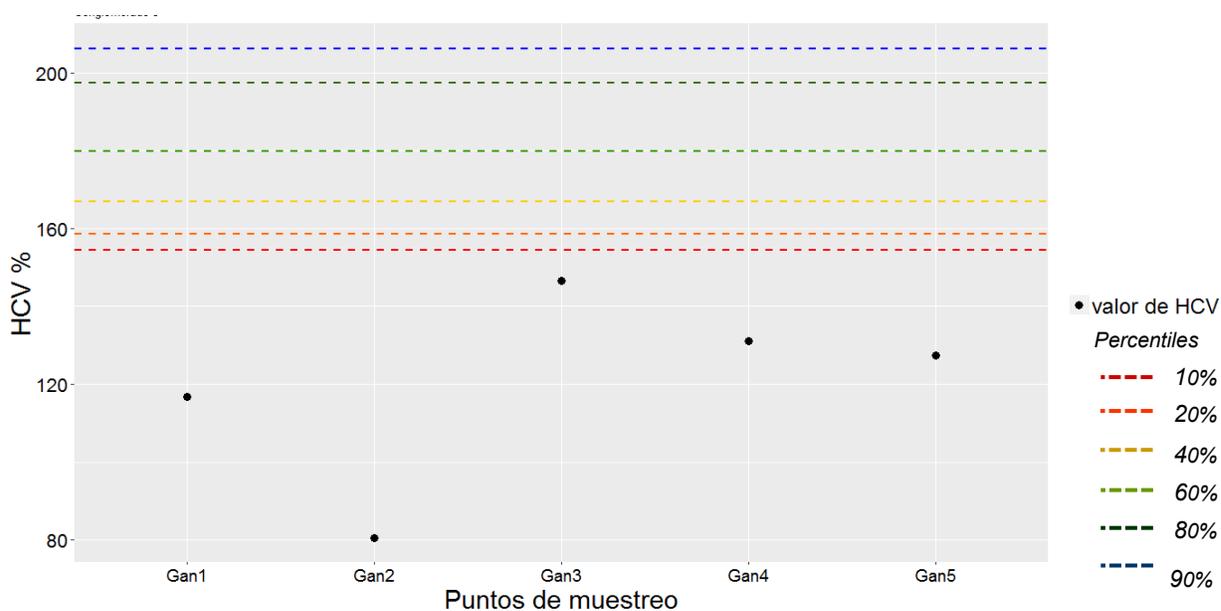
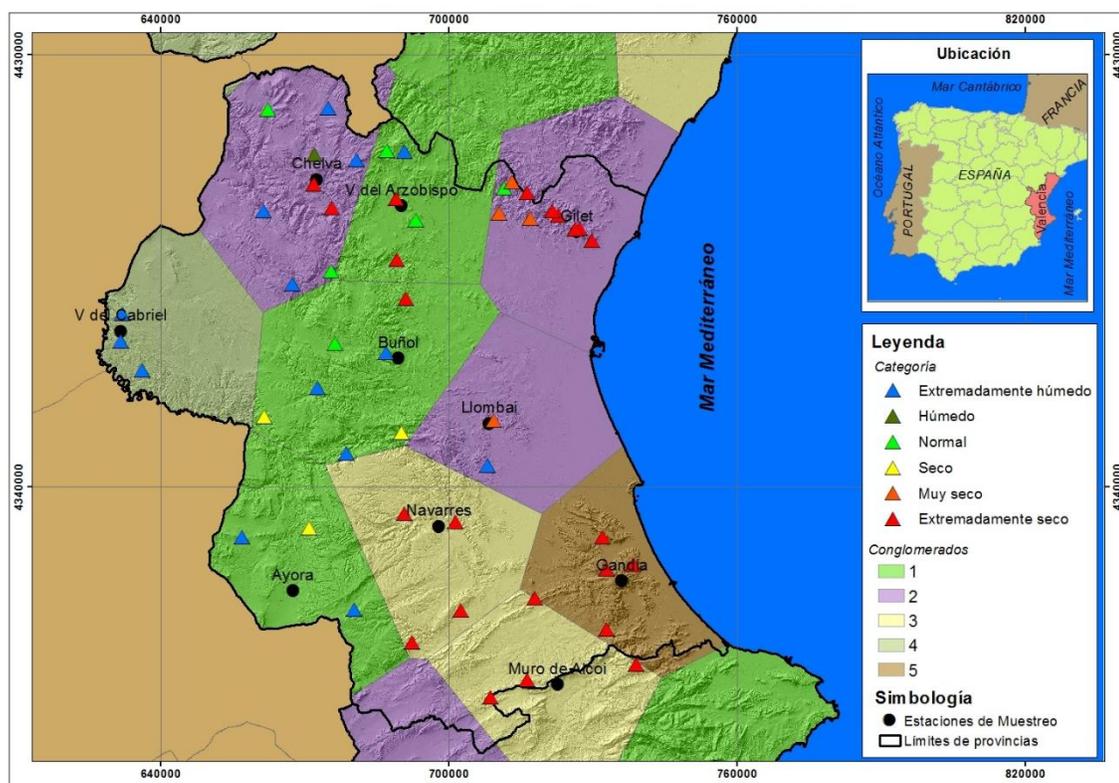


Figura 6.2-6 Diagrama de dispersión con la categorización de los puntos de muestreo del conglomerado 5.



Para finalizar se muestra en la figura 6.2-7 los puntos de muestreo categorizados de manera espacial donde se puede observar que la mayoría de puntos con valores catalogados como extremadamente secos se localizan al sur de la provincia Valenciana, al norte se localizan en la zona denominada Gilet cerca al mar Mediterráneo, en la parte interior también se puede localizar algunos puntos sobre el sistema montañoso Ibérico, los puntos catalogados como extremadamente húmedos se encuentran en el interior provincia.

Figura 6.2-7 Especialización de los puntos de muestreos categorizados en la provincia de Valencia.



### 6.3 Análisis de regresión lineal múltiple

En este estudio se realizaron varios tipos de análisis ya que su finalidad es encontrar los mejores modelos que describan la relación entre las variables meteorológicas y la variable humedad, analizando diferentes resultados.

En un principio se tenía la intención de utilizar datos de 3 años completos de 16 puntos de muestreo, sin embargo no se pudo conseguir la información meteorológica a tiempo para calcular las variables necesarias, por lo que se utilizó la información de la cual se disponía tal como se muestra en el apartado 4.1.

En el primer estudio realizado se evaluó a todos los puntos de muestreo en conjunto tal como se muestra en la tabla 6.3-1, en donde 5 variables entran en el modelo y su ajuste es del 46.06%, ninguna variable del grupo viento entra en el modelo, cabe destacar que todas las

variables tienen un valor-P menor que 0.05, existiendo una relación estadísticamente significativa entre las variables con un nivel de confianza del 95,0%.

Tabla 6.3-1 Informe de los modelos con todos los puntos de muestreo.

Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P
CONSTANTE	95.7945	15.1528	6.3219	0.0000
P3	-0.5132	0.1348	-3.8062	0.0003
P60	0.1237	0.0302	4.0996	0.0001
Tmedia3	-1.8500	0.3711	-4.9848	0.0000
HRmax7	0.4256	0.1740	2.4456	0.0163
pto_r30	0.4658	0.1035	4.4981	0.0000

**Regresión por Pasos**

	Variable	R <sup>2</sup>	R <sup>2</sup> ajustado	CME
<b>Paso 0:</b>	CONSTANTE	0%	0%	924.071
<b>Paso 1:</b>	pto_r30	20.59%	19.78%	741.312
<b>Paso 2:</b>	Tmedia3	33.53%	32.16%	626.929
<b>Paso 3:</b>	P60	39.22%	37.32%	579.242
<b>Paso 4:</b>	P3	45.50%	43.21%	524.797
<b>Paso 5:</b>	HRmax7	48.76%	46.04%	498.651

Error estándar del estimacion	22.3305
Error absoluto medio	16.7171
Estadístico Durbin-Watson	1.049061 (P=0.0028)

**Ecuación de la relación**

$$Ro = 95,7945 - 0,513188*P3 + 0,123686*P60 - 1,84995*Tmedia3 + 0,425625*HRmax7 + 0,465753*pto_r30$$

El ajuste con todas las variables es bajo porque las variables meteorológicas no son capaces de explicar la variabilidad espacial analizada a través del análisis de conglomerados en la sección anterior., hay que considerar que el área de estudio es muy amplio y los factores climáticos afectan de diferentes formas de acuerdo a la zona, sin embargo se realizan 3 modelos más con la finalidad de comparar las variables que intervienen en cada uno de ellos.

Los 3 análisis presentados a continuación se realizaron partiendo de los grupos de variables agrupadas como son precipitación, temperatura, humedad relativa, viento y punto de rocío, en donde se realizaron dos evaluaciones y se seleccionó al mejor modelo de cada una, para las evaluaciones se tomó de coeficiente de determinación múltiple R<sup>2</sup> ajustado y los criterio de información, tal como se mostró en el apartado 4.3.1.1 obteniendo como resultado el presentado en la tabla 6.3-2 donde se muestran las variables seleccionadas por grupo y de

acuerdo a la evaluación.

Tabla 6.3-2 Selección de variables por evaluación.

<b>Grupo variables</b>	<b>Código modelo</b>	<b>Nombre variables</b>	<b>Evaluación</b>
Precipitación	BE	P7, P60	<i>R2 ajustado</i>
	E	P60	<i>Criterio de información</i>
Temperatura	ACEG	Tmax3, Tmax15, Tmedia3, Tmedia15	<i>R2 ajustado</i>
	ACEG	Tmax3, Tmax15, Tmedia3, Tmedia15	<i>Criterio de información</i>
Humedad relativa	BCEF	HRmax7, HRmax15, HRmin7, HRmin15	<i>R2 ajustado</i>
	BCEF	HRmax7, HRmax15, HRmin7, HRmin15	<i>Criterio de información</i>
Viento	BCEF	Vmax7, Vmax15, lev7, lev15	<i>R2 ajustado</i>
		-	<i>Criterio de información</i>
Punto de rocío	AD	pto_r3, pto_r30	<i>R2 ajustado</i>
	AD	pto_r3, pto_r30	<i>Criterio de información</i>

En las tabla 6.3-3 se muestran los informes de los modelos obtenidos por el análisis de regresión lineal múltiple por evaluación, el mejor ajuste presenta el obtenido por las variables seleccionadas de la evaluación con el coeficiente  $r^2$  ajustado y su diferencia radica en ingresa en el modelo la variable P7, de igual forma se puede notar que del grupo de variables viento no entra ninguna.

Tabla 6.3-3 Informe del modelo con las variables seleccionadas por evaluación.

**R<sup>2</sup> ajustado**

Parámetro	Estimación	Error Estándar	EstadísticoT	Valor-P
CONSTANTE	92.6818	15.0562	6.15572	0.0000
P7	-0.158223	0.0670424	-2.36005	0.0204
P60	0.122305	0.031829	3.84258	0.0002
Tmedia3	-1.91066	0.367295	-5.20198	0.0000
HRmax7	0.469311	0.173227	2.70923	0.0080
pto_r3	-1.42986	0.693035	-2.06318	0.0419
pto_r30	0.632572	0.145831	4.33771	0.0

	Variable	R <sup>2</sup>	R <sup>2</sup> ajustado	CME
<b>Paso 0:</b>	CONSTANTE	0%	0%	924.071
<b>Paso 1:</b>	pto_r30	20.59%	19.78%	741.312
<b>Paso 2:</b>	Tmedia3	33.53%	32.16%	626.929
<b>Paso 3:</b>	P60	39.22%	37.32%	579.242
<b>Paso 4:</b>	pto_r3	44.97%	42.65%	529.918
<b>Paso 5:</b>	HRmax7	47.83%	45.06%	507.717
<b>Paso 6:</b>	P7	50.78%	47.60%	484.179
Error estándar del estimación				22.0041
Error absoluto medio				16.1962
Estadístico Durbin-Watson				1.54769 (P=0.0070)

$$Ro = 92,6818 - 0,158223*P7 + 0,122305*P60 - 1,91066*Tmedia3 + 0,469311*HRmax7 - 1,42986*pto_r3 + 0,632572*pto_r30$$

**Criterio de información**

Parámetro	Estimación	Error Estándar	EstadísticoT	Valor-P
CONSTANTE	98.6345	15.2000	6.4891	0.0000
P60	0.0878	0.0289	3.0326	0.0031
Tmedia3	-1.7974	0.3729	-4.8202	0.0000
HRmax7	0.3962	0.1745	2.2703	0.0255
pto_r3	-2.2098	0.6238	-3.5426	0.0006
pto_r30	0.7315	0.1430	5.1146	0.0000

	Variable	R <sup>2</sup>	R <sup>2</sup> ajustado	CME
<b>Paso 0:</b>	CONSTANTE	0%	0%	924.071
<b>Paso 1:</b>	pto_r30	20.59%	19.78%	741.312
<b>Paso 2:</b>	Tmedia3	33.53%	32.16%	626.929
<b>Paso 3:</b>	P60	39.22%	37.32%	579.242
<b>Paso 4:</b>	pto_r3	44.97%	42.65%	529.918
<b>Paso 5:</b>	HRmax7	47.83%	45.06%	507.717
Error estándar del estimación				22.5326
Error absoluto medio				16.5574
Estadístico Durbin-Watson				1.46197 (P=0.0017)

$$Ro = 98,6345 + 0,0877896*P60 - 1,7974*Tmedia3 + 0,396226*HRmax7 - 2,20982*pto_r3 + 0,731532*pto_r30$$

Para finalizar se hace un último modelo, tomando en cuenta que en el grupo de precipitación el incluir una variable en el evaluado por el coeficiente  $R^2$  ajustado hizo la diferencia, se procede a realizar una filtración diferente, se hace un realiza un modelo por grupo y por evaluación seleccionando el que mejor se ajuste, con las variables de los modelos por grupo finales se procede a realizar un último modelo y se compara con los anteriores, el resultado final se muestra en la tabla 6.3-4.

Tabla 6.3-4 Informe del modelo final filtrado por grupo.

Grupo	Pasos	Variable	R <sup>2</sup> ajustado	Ecuación del modelo ajustado	Evaluación
Precipitación	Paso 0:	CONSTANTE	0.00%	$R_o = 103,461 + 0,156503 * P60$	R <sup>2</sup> ajustado o Criterio de
	Paso 1:	P60	18.35%		
Temperatura	Paso 0:	CONSTANTE	0.00%	$R_o = 149,336 - 2,03173 * Tmedia3$	R <sup>2</sup> ajustado o Criterio de
	Paso 1:	Tmedia3	17.43%		
Humedad relativa	Paso 0:	CONSTANTE	0.00%	$R_o = 75,0 + 0,968351 * HRmin15$	Criterio de información
	Paso 1:	HRmin15	12.45%		
Viento					
Punto de rocío	Paso 0:	CONSTANTE	0.00%	$R_o = 113,258 - 1,62183 * pto\_r3 + 0,846418 * pto\_r30$	R <sup>2</sup> ajustado o Criterio de información
	Paso 1:	pto_r30	19.78%		
	Paso 2:	pto_r3	22.93%		

**Modelo final**

Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P	Evaluación
CONSTANTE	129.3250	7.0990	18.2173	0.0000	R <sup>2</sup> ajustado y Criterio de información
P60	0.0933	0.0295	3.1657	0.0021	
Tmedia3	-1.6440	0.3747	-4.3880	0.0000	
pto_r3	-1.9828	0.6290	-3.1521	0.0022	
pto_r30	0.7230	0.1461	4.9497	0.0000	

	Variable	R <sup>2</sup>	R <sup>2</sup> ajustado	CME		
Paso 0:	CONSTANTE	0.00%	0.00%	924.071	Error estándar del estimacion	23.0199
Paso 1:	pto_r30	20.59%	19.78%	741.312	Error absoluto medio	16.843
Paso 2:	Tmedia3	33.53%	32.16%	626.929	Estadístico Durbin-Watson	1.47429 (P=0.0023)
Paso 3:	P60	39.22%	37.32%	579.242		
Paso 4:	pto_r3	44.97%	42.65%	529.918		

$$R_o = 129,325 + 0,0932958 * P60 - 1,64398 * Tmedia3 - 1,98279 * pto\_r3 + 0,723002 * pto\_r30$$

El ajuste del modelo es el más bajo si se compara con los anteriores con tan solo un 42.65%, el mejor modelo es el obtenido de la evaluación con coeficiente  $R^2$  ajustado con un 47.60%.

Las variables pto\_r30, Tmedia3 y P60 se presentan en todos los modelos pero las pertenecientes al grupo viento no se presentó ninguna. Todos los modelos realizados están por debajo del 50% de ajuste, por lo que para mejorarlo y partiendo de la premisa que dicho resultado es por abarcar datos de porcentajes de humedad de diferentes puntos distribuidos

en un área grande y sabiendo que los factores meteorológicos incluyen de forma diferente en cada zona, se procede a realizar el mismo análisis por punto de muestreo.

### Punto de muestreo Biar

Este punto se presentó en el capítulo 4, apartados 4.3.2 y 1.3.3 donde se realizaron 2 modelos, uno el cual incluyó todas las variables y el otro que fue filtrado realizando modelos para cada grupo y variables entrantes se tomó para hacer un modelo final, el ajuste de los dos modelos son exactamente lo mismo 70.60%, y sus variables finales son P60 y Tmedia3, como se puede observar el ajuste es alto si se compara con el realizado donde se incluyeron todos los puntos. Si se hace un análisis por grupo de todos ellos destaca una variable a excepción de punto de rocío que al parecer no guarda una relación como los demás.

### Punto de muestreo Buñol

Los diferentes modelos en el punto de muestreo Buñol de igual forma son iguales por ello se presenta en la tabla 6.3-5 el informe final del modelo realizado filtrando variables por grupos, el ajuste del modelo es del 72.66% e incluye 3 variables, dato curioso es que una de ellas es del grupo de viento, lev3 el cual corresponde al levante o viento proveniente del mar el cual contiene humedad, este punto se encuentra ubicado en medio de la provincia de Valencia por lo que tiene mucho sentido que su influencia determine la humedad en esta zona.

Tabla 6.3-5 Informe del modelo final filtrado por grupo.

Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P	Evaluación
CONSTANTE	127.9320	10.2645	12.4636	0.0000	R <sup>2</sup> ajustado y Criterio de información
Tmedia7	-1.3535	0.6340	-2.1349	0.0497	
lev3 (h)	-2.4098	0.7928	-3.0397	0.0083	
pto_r30	1.3820	0.2962	4.6652	0.0003	

	Variable	R <sup>2</sup>	R <sup>2</sup> ajustado	CME		
Paso 0:	CONSTANTE	0%	0%	944.324	Error estándar del estimación	16.0682
Paso 1:	pto_r30	49,63%	46,67%	503.625	Error absoluto medio	11.47
Paso 2:	lev3	70,29%	66,58%	315.595	Estadístico Durbin-Watson	1.97259 (P=0.2712)
Paso 3:	Tmedia7	77,22%	72,66%	258.186		

$$Ro = 127,932 - 1,35352 * Tmedia7 - 2,40976 * lev3 (h) + 1,38195 * pto\_r30$$

## Punto de muestreo Chelva

En este punto se pudieron hacer 3 análisis, en los modelos finales en donde se tomaron las todas las variables seleccionadas por evaluación entraron P15, Tmax15 y Tmedia15 y su ajuste fue igual con un 69.34% pero al filtrarlo por grupos en el modelo final se elimina la variable P15 alcanzando un 52.79% en el ajuste. Variables de humedad relativa y punto de rocío no entran ni en los moldeos por grupo. Cabe destacar que la ubicación del punto es en el interior de la provincia de Valencia y es un lugar montañoso por lo que la temperatura y la lluvia sean de mayor influencia en la zona.

Tabla 6.3-6 Informe del modelo final filtrado por grupo.

Grupo	Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P	Evaluación
<b>Precipitación</b>						-
<b>Temperatura</b>	CONSTANTE	109.1270	17.9112	6.0927	0.0000	R <sup>2</sup> ajustado o Criterio de información
	Tmax15	8.4697	3.9402	2.1496	0.0496	
	Tmedia15	-12.0740	4.6228	-2.6118	0.0205	
<b>Humedad relativa</b>						-
<b>Viento</b>	CONSTANTE	165.9450	20.2975	8.1756	0.0000	Criterio de información
	Vmax15	-2.4331	0.9896	-2.4587	0.0276	
	lev3 (h)	-3.6509	1.1706	-3.1187	0.0075	
<b>Punto de rocío</b>						-

	Variable	R <sup>2</sup>	R <sup>2</sup> ajustado	CME	Grupo
	-				<b>Precipitación</b>
<b>Paso 0:</b>	CONSTANTE	0%	0%	384.187	<b>Temperatura</b>
<b>Paso 1:</b>	Tmedia15	45,06%	41,39%	225.162	
<b>Paso 3:</b>	Tmax15	58,69%	52,79%	181.380	
	-				<b>Humedad relativa</b>
<b>Paso 0:</b>	CONSTANTE	0%	0%	384.187	<b>Viento</b>
<b>Paso 1:</b>	lev3	33,26%	28,81%	273.516	
<b>Paso 2:</b>	Vmax15	53,39%	46,73%	204.673	
	-				<b>Punto de rocío</b>

### Modelo final

Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P	Evaluación
CONSTANTE	109.1270	17.9112	6.0927	0.0000	R <sup>2</sup> ajustado y Criterio de información
Tmax15	8.4697	3.9402	2.1496	0.0496	
Tmedia15	-12.0740	4.6228	-2.6118	0.0205	

	Variable	R <sup>2</sup>	R <sup>2</sup> ajustado	CME
<b>Paso 0:</b>	CONSTANTE	0%	0%	384.187
<b>Paso 1:</b>	Tmedia15	45.06%	41.39%	225,162
<b>Paso 2:</b>	Tmax15	58,69%	52,79%	181,38

Error estándar del estimación	13.4677
Error absoluto medio	10.5312
Estadístico Durbin-Watson	1.99047 (P=0.3780)

$$Ro = 109,127 + 8,46974 * Tmax15 - 12,074 * Tmedia15$$

Tabla 6.3-7 Extracto del informe del modelo por evaluación.

Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P	Evaluación
CONSTANTE	100.7960	14.7129	6.8509	0.0000	R <sup>2</sup> ajustado o Criterio de información
P15	0.2571	0.0879	2.9251	0.0118	
Tmax15	8.6473	3.1759	2.7228	0.0174	
Tmedia15	-12.1749	3.7257	-3.2678	0.0061	

	Variable	R <sup>2</sup>	R <sup>2</sup> ajustado	CME
<b>Paso 0:</b>	CONSTANTE	0	0	384.187
<b>Paso 1:</b>	Tmedia15	45,06%	41,39%	225.162
<b>Paso 2:</b>	P15	60,88%	55,29%	171.764
<b>Paso 3:</b>	Tmax15	75,09%	69,34%	117.8

Error estándar del estimacion	10.8536
Error absoluto medio	7.66365
Estadístico Durbin-Watson	1.81619 (P=0.2733)

$$R_o = 100,796 + 0,257065 * P15 + 8,64728 * Tmax15 - 12,1749 * Tmedia15$$

El modelo en donde se toma todas las variables no permite ejecutar el software ya que el número de datos es menor al número de variables.

### Punto de muestreo Fredes

En este punto solo se pueden realizar dos modelos, uno en donde se toma todas las variables de la selección por evaluación con los criterios de información y el otro en donde se filtran las variables por grupo de datos meteorológicos, presentan un mismo ajuste y variables entrantes en los modelos por lo cual el mostrado en la tabla 6.3-9 es el filtrado, el ajuste presentado es del 56.50% y las variables Vmax7 y Vmax15, el punto se encuentra al norte de la provincia de Castellón dentro del sistema montañoso.

Los modelos con todas las variables y el seleccionado por la evaluación con el coeficiente r<sup>2</sup> no se pudieron realizar ya que se tiene más variables que datos de humedad.

Tabla 6.3-8 Selección de variables por evaluación.

Grupo variables	Código modelo	Nombre variables	Evaluación
Precipitación	BCDE	P7, P15, P30, P60	R2 ajustado
	BE	P7, P60	Criterio de información
Temperatura	ABCFG	Tmax3, Tmax7, Tmax15, Tmedia7, Tmedia15	R2 ajustado
	BC	Tmax7, Tmax15	Criterio de información
Humedad relativa	ABCDF	HRmax3, HRmax7, HRmax15, HRmin3, HRmin15	R2 ajustado
	BF	HRmax7, HRmin15	Criterio de información
Viento	BCE	Vmax7, Vmax15, lev7	R2 ajustado
	BC	Vmax7, Vmax15	Criterio de información
Punto de rocío	No se puede efectuar el análisis. Los valores son todos iguales.		

Tabla 6.3-9 Extracto del informe del modelo final filtrado por grupo.

Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P
CONSTANTE	133.945	21.576	6.208	0.000
Vmax7	4.577	1.433	3.195	0.009
Vmax15	-5.429	2.262	-2.400	0.035

	Variable	R <sup>2</sup>	R <sup>2</sup> ajustado	CME
<b>Paso 0:</b>	CONSTANTE	0%	0%	686.163
<b>Paso 1:</b>	Vmax7	43,91%	39,24%	416.931
<b>Paso 2:</b>	Vmax15	63,19%	56,50%	298.493

Error estándar del estimacion	17.2769
Error absoluto medio	12.0828
Estadístico Durbin-Watson	2.13292 (P=0.5219)

$$Ro = 133,945 + 4,57706 * Vmax7 - 5,42921 * Vmax15$$

### Punto de muestreo Gandía

Se realizaron 3 modelos de los cuales coincidieron todos con un ajuste del 34.66%, su valor es bastante bajo, como se muestra en la tabla 6.3-10 la evaluación por grupo presenta varias variables por grupo sin embargo al realizar el análisis de regresión, todas son descartadas y solo ingresa la pto\_r30 correspondiente al punto de rocío en 30 días. El modelo resulta extraño ya que al ser un punto cerca de la costa este debería tener influencia de los vientos proveniente del mar, entre otros. El punto quizás no tiene los datos suficientes para analizar la relación existente. Es un punto que se localiza en el interior de la provincia de Alicante.

Tabla 6.3-10 Selección de variables por evaluación.

Grupo variables	Código modelo	Nombre variables	Evaluación
Precipitación	ABC	P3, P7, P15	R2 ajustado
	AB	P3, P7	Criterio de información
Temperatura	ADE	Tmax3, Tmax30, Tmedia3	R2 ajustado
	AE	Tmax3, Tmedia3	Criterio de información
Humedad relativa	ADF	HRmax3, HRmin3, HRmin15	R2 ajustado
	ADF	HRmax3, HRmin3, HRmin15	Criterio de información
Viento	A	Vmax3	R2 ajustado
	A	Vmax3	Criterio de información
Punto de rocío	AC	pto_r3, pto_r30	R2 ajustado
	AC	pto_r3, pto_r30	Criterio de información

Tabla 6.3-11 Informe del modelo final filtrado por grupo.

Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P
CONSTANTE	121.5790	8.5520	14.2165	0.0000
pto_r30	0.4195	0.1402	2.9928	0.0097

	Variable	R <sup>2</sup>	R <sup>2</sup> ajustado	CME
<b>Paso 0:</b>	CONSTANTE	0%	0%	786.702
<b>Paso 1:</b>	pto_r30	39,02%	34,66%	514.026

Error estándar del estimacion	22.6721
Error absoluto medio	18.3085
Estadístico Durbin-Watson	1.00012 (P=0.0128)

$$R_o = 121,579 + 0,419496 * pto\_r30$$

### Punto de muestreo Montanejos

Está en el interior de la provincia de Castellón y presento problemas ya que ninguna variable meteorológica entra en ningún modelado, lo cual denota un problema de datos ya que no corresponden o se relacionan las variables meteorológicas presentadas en un tiempo determinado con el porcentaje de humedad de la especie romero que debería haberse presentado en ese determinado momento.

Tabla 6.3-12 Selección de variables por evaluación.

Grupo variables	Código modelo	Nombre variables	Evaluación
Precipitación	ABC	P3, P7, P15	R2 ajustado
		-	Criterio de información
Temperatura	BDFG	Tmax7, Tmax30, Tmedia7, Tmedia15	R2 ajustado
		-	Criterio de información
Humedad relativa	BCDEF	HRmax7, HRmax15, HRmin3, HRmin7, HRmin15	R2 ajustado
		-	Criterio de información
Viento	CDE	Vmax15, lev3, lev7	R2 ajustado
		-	Criterio de información
Punto de rocío	ABC	pto_r7, pto_r15, pto_r30	R2 ajustado
		-	Criterio de información

Las estaciones al compararlas por el porcentaje de  $r^2$  ajustado tenemos: Biar = 70.60%, Buñol = 72.66, Chelva = 69.34, Fredes = 56.50 y Gandía = 34.66; Montanejos queda descartada ya que claramente se puede notar que tiene algún problema en la relación, Gandía es una estación que también presenta problemas por el porcentaje de ajuste sin embargo, guarda alguna relación.

Se realiza un análisis final en donde se genera un modelo con todos los puntos de muestreo a excepción de Montanejos, en la tabla 6.3-11 se muestra el modelo de mejor ajuste y coinciden los modelos en donde fueron tomadas todas las variables y las seleccionadas por la evaluación con el coeficiente determinante  $r^2$  ajustado, el modelo mejora en un 9%. Las variables entrantes en el modelo son las mismas sin embargo la relación es mejor, de todas formas su valor es bajar por lo que su mejor opción es trabajar de forma individual.

En este caso es posible realizar un análisis por estación ya que es un número manejable, sin embargo al tener una cantidad mayores necesario trabajar con grupos pequeños, se toma en cuenta el análisis de conglomerados realizados en el capítulo 1, cabe recalcar que el análisis se realizó con 16 puntos de muestreo y se usó solo datos de humedad para agruparlos, de todas formas en la clasificación de 5 conglomerados se puede notar que se tiene un punto de muestreo de cada clase a excepción del conglomerado 1, pero si se utiliza la clasificación de 4 conglomerados se resolvería este inconveniente, los conglomerados 2 y 3 son los que tienen dos punto de muestreo por grupo, sin embargo la clase 3 queda descartada para el análisis ya que uno de sus integrantes es el punto Montanejos.

Tabla 6.3-13 Informe modelo modificado.

Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P
CONSTANTE	85.6901	15.6524	5.4746	0.0000
P7	-0.1530	0.0645	-2.3708	0.0203
P60	0.1141	0.0318	3.5895	0.0006
Tmedia3	-1.7317	0.3574	-4.8458	0.0000
HRmax7	0.4923	0.1718	2.8662	0.0054
pto_r3	-1.6915	0.6483	-2.6090	0.0110
pto_r30	0.7317	0.1373	5.3297	0.0000

	Variable	R <sup>2</sup>	R <sup>2</sup> ajustado	CME
<b>Paso 0:</b>	CONSTANTE	0%	0%	897.92
<b>Paso 1:</b>	pto_r30	28.78%	27.89%	647.461
<b>Paso 2:</b>	Tmedia3	41.88%	40.41%	535.113
<b>Paso 3:</b>	pto_r3	48.75%	46.78%	477.853
<b>Paso 4:</b>	P60	53.50%	51.09%	439.191
<b>Paso 5:</b>	HRmax7	56.87%	54.03%	412.782
<b>Paso 6:</b>	P7	59.87%	56.66%	389.123

Error estándar del estimacion	19.7262
Error absoluto medio	14.8331
Estadístico Durbin-Watson	1.59498 (P=0.0202)

$$Ro = 85,6901 - 0,153001*P7 + 0,11411*P60 - 1,73167*Tmedia3 + 0,492301*HRmax7 - 1,69154*pto_r3 + 0,731743*pto_r30$$

En la tabla 6.3-14 se puede observar la creación de un modelo por los dos integrantes con los que se tienen datos meteorológicos del conglomerado 2, los puntos de muestreo Buñol y Fredes, si bien es cierto que el ajuste es bajo con un 42.84%, solo está por debajo del modelo creado con todas las variables disponibles por un 5%, además es importante nombrar que de igual forma se realizaron diferentes modelos y todos coinciden en el mismo ajuste.

En la tabla 6.3-14 se muestra un extracto del resultado final del análisis de regresión lineal por conglomerado.

Tabla 6.3-14 Extracto del informe del modelo final por conglomerado.

Parámetro	Estimación	Error Estándar	Estadístico T	Valor-P
CONSTANTE	94.4182	5.8791	16.0601	0.0000
P7	-0.1773	0.0835	-2.1222	0.0422
P60	0.2096	0.0412	5.0857	0.0000

	Variable	R <sup>2</sup>	R <sup>2</sup> ajustado	CME
<b>Paso 0:</b>	CONSTANTE	0%	0%	886.476
<b>Paso 1:</b>	P60	38.37%	36.38%	563.945
<b>Paso 2:</b>	P7	46.42%	42.84%	506.68

También se realizó el análisis de regresión lineal múltiple con el software RStudio los informes generados del estudio para elegir el modelado que mejor se ajuste con todos los puntos de muestreo se pueden ver en el anexo 7, los ajustes son:

- Modelo generado por selección de variables por grupo:  $R^2$  ajustado del 45.50%
- Modelo generado con todas las variables:  $R^2$  ajustado del 49.90%
- Modelo con las variables seleccionadas por el coeficiente de determinación múltiple  $R^2$  ajustado:  $R^2$  ajustado del 47.80%
- Modelo con las variables seleccionadas por los criterios de información:  $R^2$  ajustado del 48.10%

Aunque los ajustes son similares a los realizados con el software Statgraphics el número de variables dentro de los modelados por RStudio son el doble.

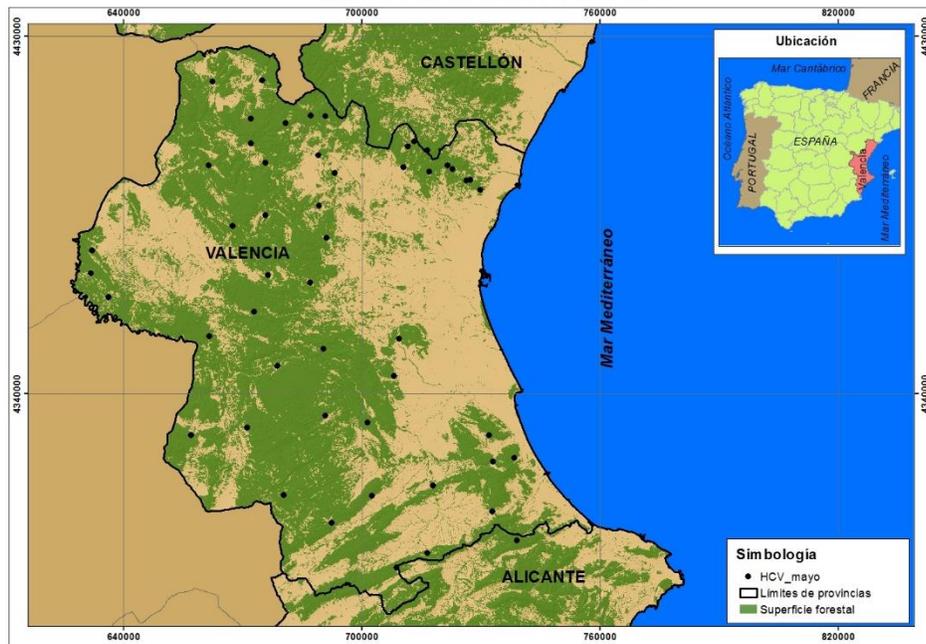
En el anexo 8 se presenta el análisis con el punto de muestreo Biar, aunque el ajuste es bastante bueno, el número de variables introducidas en los modelos sobrepasan los 10, creando una relación bastante compleja.

- Modelo generado por selección de variables por grupo:  $R^2$  ajustado del 96.0%
- Modelo generado con todas las variables:  $R^2$  ajustado del 99.90%
- Modelo con las variables seleccionadas por el coeficiente de determinación múltiple  $R^2$  ajustado:  $R^2$  ajustado del 96.00%
- Modelo con las variables seleccionadas por los criterios de información:  $R^2$  ajustado del 78.10%

#### **6.4 Interpolación espacial**

El resultado final de la interpolación espacial es un modelo de superficie de estimación de la variable humedad de la especie vegetal romero, creada a partir de 51 datos de porcentaje de humedad de la especie vegetal romero levantados en mayo del 2018 en la provincia de Valencia, la ubicación de los puntos muestreados se lo realizó en función de la accesibilidad y de la cobertura forestal tal como se muestra en la figura 6.4-1. Como se puede notar, la costa no cuenta con una cobertura forestal abundante imposibilitando que mejore la estimación del porcentaje de humedad.

Figura 6.4-1. Puntos de muestreo con respecto la superficie forestal.



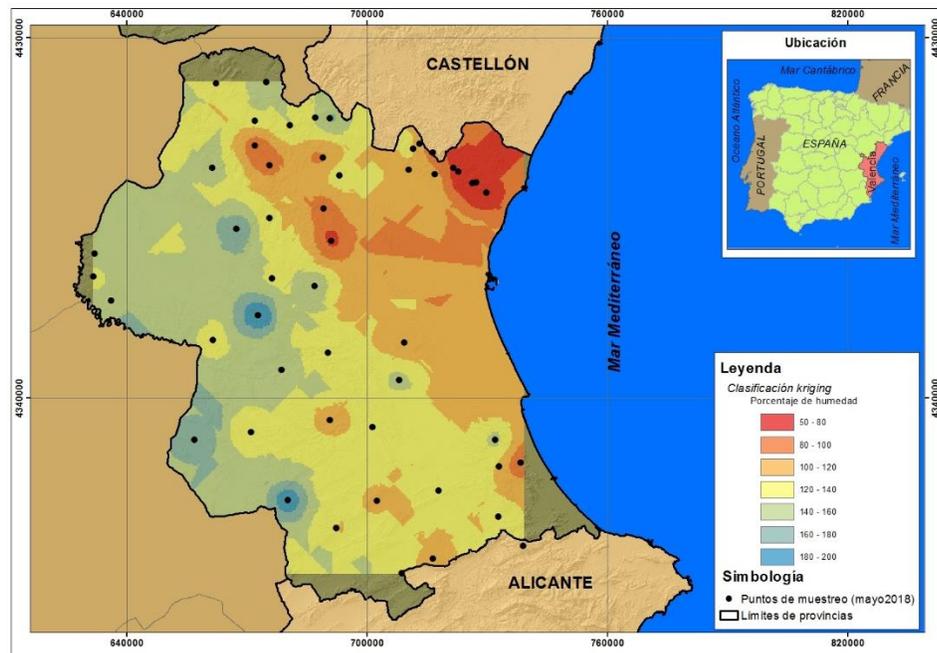
El análisis general de los datos no presentó valores atípicos, también es importante nombrar que los datos provienen de una distribución normal. La creación del kriging o el modelo de superficie de estimación de la variable en estudio con los puntos muestreados se lo analizó y construyó por medio de los semivariogramas, los valores del semivariograma omnidireccional presentan oscilaciones en distancias grandes con valores muy oscilante, dificultando el ajuste con el semivariograma experimental, sin embargo se determinó el modelo Gaussiano como el de mejor ajuste para estos datos, se realizaron varias pruebas en donde los errores de estimación calculados en la validación cruzada, fueron tomados como referente para la elección del mejor modelo, tomando como principal parámetro de validación el error medio cuadrático.

Cabe destacar que al realizar el ajuste de semivariogramas, los puntos levantados son pocos para realizar un modelo por krigeado y por ende su presión en la estimación de los valores no son muy acertados, tal como se denota en la figura 5.1.3.1-1 y el capítulo 5.1.3.

El modelo final se realizó a partir del krigeado ordinario con el ajuste del semivariograma teórico gaussiano y los parámetros mostrados en la tabla 5.2.3.1-1, obteniendo como resultado final el mostrado en la figura 6.4-2, el error medio cuadrático de estimación es de 25.6, el mapa

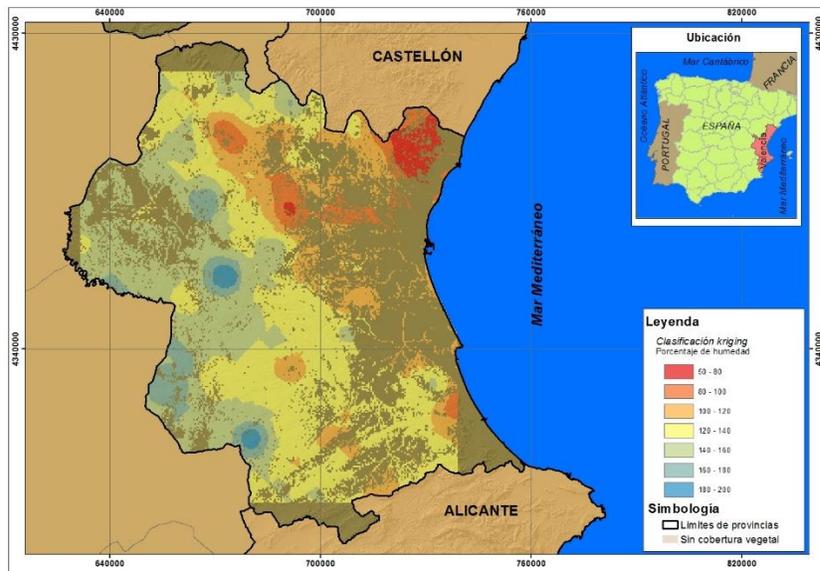
con el modelo de estimación se presenta dividido en 7 clases, además se puede observar que el porcentaje de humedad es menor en la costa, aumentando desde la zona norte-este de la provincia hacia el interior de la provincia donde sus valores son altos.

Figura 6.4-2 Mapa de la superficie de estimación del porcentaje de humedad de la especie vegetal romero en la provincia de Valencia (mayo 2018)



Como se puede observar el modelo de superficie de estimación no cubre por completo a la provincia de Valencia esto se debe a la distribución de los puntos con las que se realizó, sin embargo y tomando en cuenta que los puntos muestreados fueron tomados en función de la cobertura vegetal. En la figura 6.4-3 se muestra el resultado del modelo de estimación del porcentaje de humedad de la especie vegetal romero para el área forestal de la provincia.

Figura 6.4-3 Mapa de la superficie de estimación del porcentaje de humedad de la especie vegetal romero en la provincia de Valencia de acuerdo a la cobertura forestal (mayo 2018).

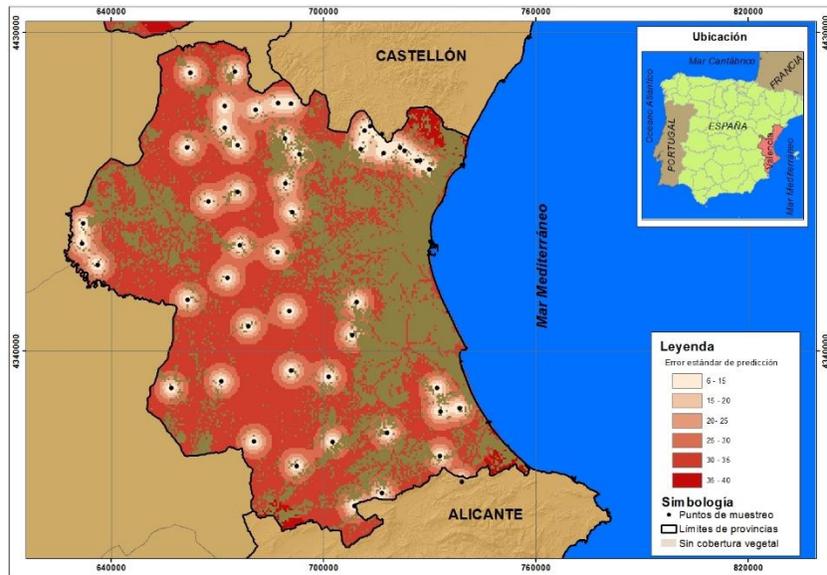


También se genera la superficie en función del error estándar de estimación de la variable en estudio obtenido a partir de la validación cruzada, como es de esperar el error aumenta en cuanto se aleja más de los puntos utilizados para realizar el kriging, el valor mínimo es de 6 considerado no tan bueno, ya que da mucho margen de error.

De igual forma tal como se explicó en el apartado 5.1.3.2 se realizaron gráficas con el valor de los puntos estimados, mismo que tenían un registro anterior de 3 años consecutivos de porcentaje de humedad de la especie vegetal romero, se calculó el promedio, valor máximo y mínimo del registro de porcentaje por puntos de muestreo, los valores son presentados en las gráficas como líneas entre punteadas que cortan al gráfico en el eje Y y es paralela al eje X, con la finalidad de forma límites en función al registro de valores y observar que tan distante está del valor estimado.

En las figuras no se mostrará el rango de oscilación del valor estimado, solo se tendrá en cuenta el límite del rango que se aproxime al registro de datos de los mayos anteriores, para lo cual se dibuja el punto del valor estimado y una flecha indicando la dirección hacia donde está uno de los límites del rango de oscilación, con un signo positivo o negativo, que señala si parte del rango presentado es el error estándar que se sumó o resto al valor estimado.

Figura 6.4-4 Mapa de superficie del error estándar de estimación del porcentaje de humedad de la especie vegetal romero de acuerdo a la cobertura forestal.

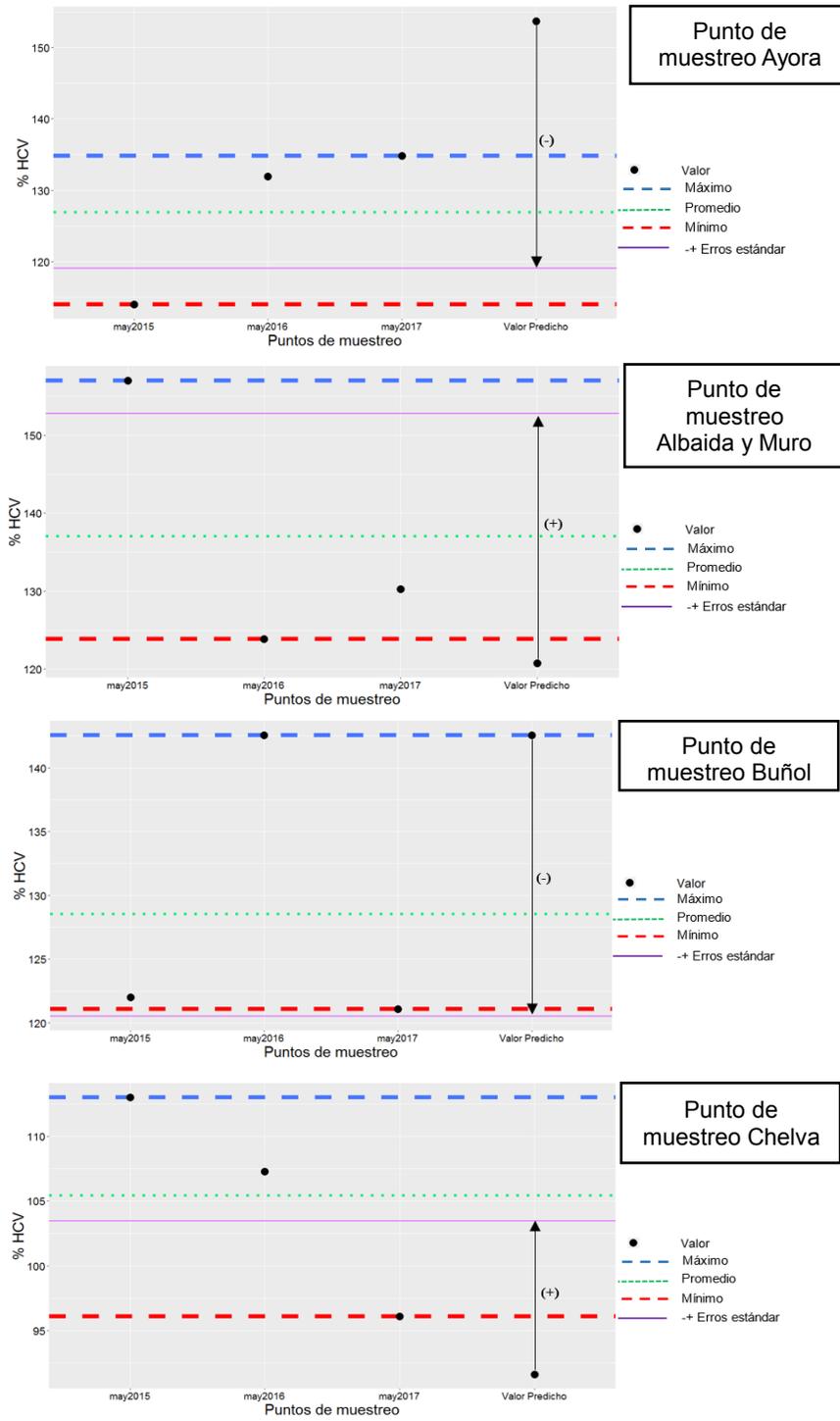


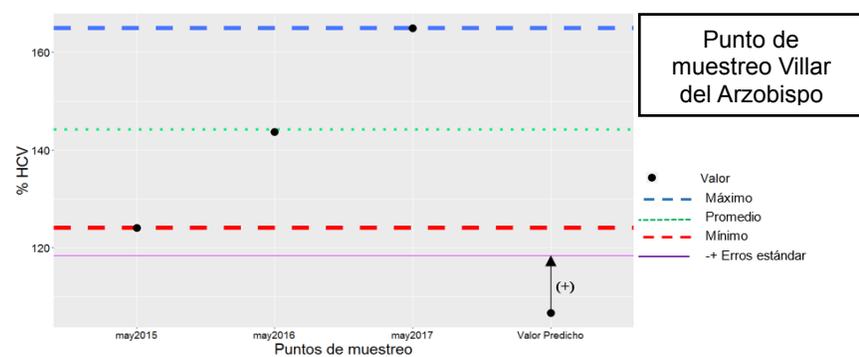
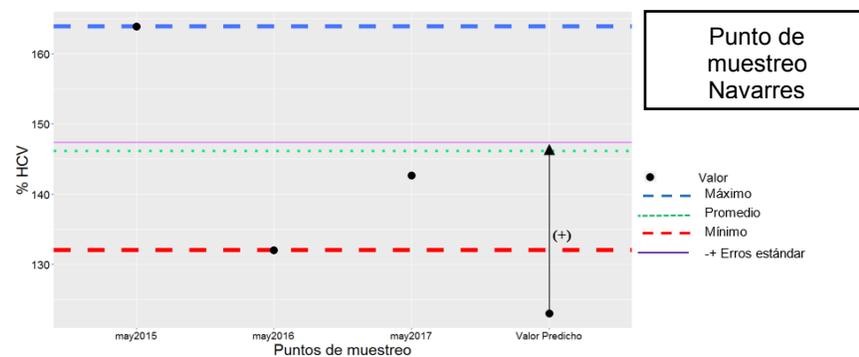
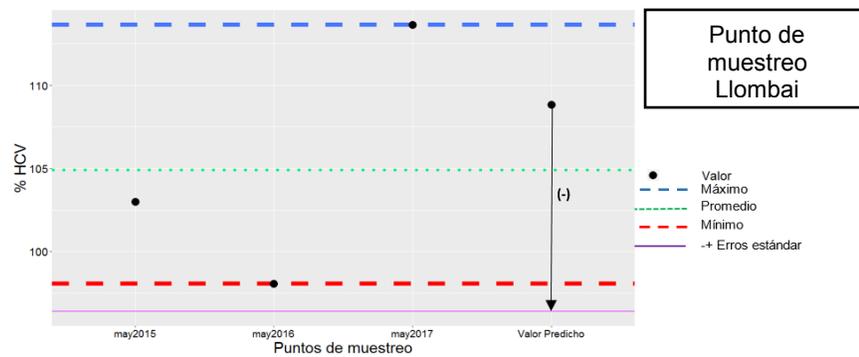
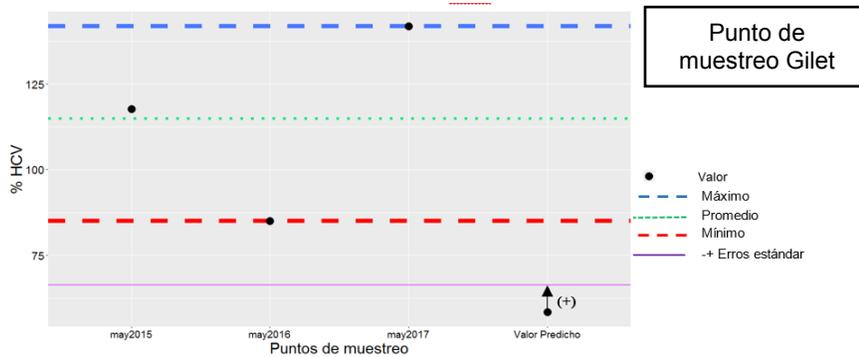
El mapa de superficie de error estándar de estimación presenta valores en un rango de 6 % al 40%.

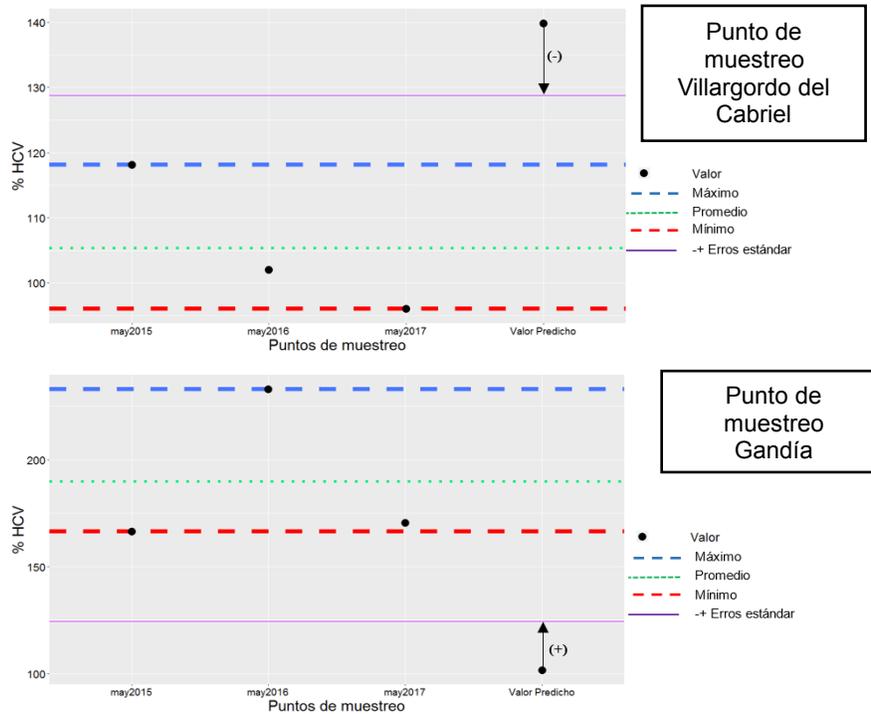
Con lo que respecta a los valores estimados en mayo de 2018 de los puntos de muestreo con datos de los 3 años anteriores para definir el rango en los cuales se mantiene el porcentaje de humedad de la especie vegetal romero, se tiene que el valor estimado de los puntos de muestreo Buñol y Llombai se encuentran dentro de los límites formados por el valor máximo y mínimo, su error estándar de estimación esta entre el 12% y 23%.

Los 4 puntos restantes no se encuentran dentro de los límites pero la oscilación de su valor límite sí. Cabe señalar que el error estándar de dos puntos es bastante alta sobrepasando el valor del 30%.

Figura 6.4-5 Gráfico del valor del punto muestreo estimado con respecto a los datos registrados den 3 años anteriores.







Una validación certera se podría realizar tomando puntos en la misma fecha que no entren en el modelado, sin embargo el número de datos utilizado es ya bastante pequeño y no es aconsejable eliminar un grupo de ellos para calcular los errores, por lo que no se puede descartar ningún punto, Por eso en la validación cruzada solo se ha considera eliminar un punto de muestreo en cada paso y no un conjunto de ellos. Sin embargo el comparar los resultados de los valores estimados, respecto a los datos del registro, da una perspectiva de dirección del valor, es decir en función del comportamiento en los años anteriores se puede verificar si la estimación maximiza o minimiza el valor.

Para que la técnica de interpolación kriging sea óptima, aparte del análisis de los semivariogramas teórico y experimental, los datos deben cumplir con 2 criterios principalmente, primeramente deben tener una distribución normal, en el apartado 5.1 se determinó que los datos cumplen con dicho criterio para lo cual se utilizó las medidas de forma como son el coeficiente de asimetría y la curtosis estandarizada; también se hizo uso del histograma con el cual además se puede comprobar si existen valore atípicos, de igual forma se presentó el gráfico de probabilidad normal (QQ normal) que alinea los datos con la línea de distribución normal, en este caso, los datos siguen una línea recta (ver figura 5.1-2). Si no siguiese una distribución normal se puede aplicar la trasformación de log o un arco. El segundo

criterio tiene que ver con la verificación de la estacionariedad de los datos o dicho de otra forma la variación en diferentes áreas de la zona de estudio, es decir, en dos puntos separados a una distancia en diferentes localidades deberían tener diferencias similares en su valor medido ya que kriging es sensible a los cambios bruscos.

## 7. CONCLUSIONES

Como premisa, es importante nombrar la importancia de realizar un análisis descriptivo de las variables a estudiar, ya que aparte de esclarecer la dimensión de los datos con los que se va a trabajar y la situación que nos presenta las observaciones realizadas, permite identificar valores atípicos que pueden desproporcionar o conducir a una interpretación engañosa de los resultados dentro del estudio.

En el análisis descriptivo de la variable porcentaje de humedad mensual de la especie vegetal romero con datos en un periodo de 3 años completos (diciembre del 2014 - diciembre de 2017), por medio del diagrama de caja y bigotes se determinó la existencia de valores atípicos en los puntos de muestreo Ayora, Villar del Arzobispo, Albaida-Muro, Castielfabib, Navarres y Pobra Tornesa, la idea de que estos valores pueda ser causada por un error sistemático en el proceso de obtención del porcentaje de humedad queda descartada, ya que en las gráficas temporales los valores coinciden con los periodos de año en donde las condiciones climáticas dominantes influyen en la disminución o aumento del valor de la variable estudiada, los valores más bajos y altos registrados son en los meses pertenecientes a verano y primavera respectivamente.

De igual forma, cabe destacar que se realizó el cálculo del porcentaje de humedad de la especie vegetal romero por estaciones del año en el periodo de los 3 años, ya que en el análisis de conglomerados es un requisito fundamental de tener periodos de datos completos y de esta forma también se disminuyó estos tipos de valores, aun al realizar el análisis descriptivo de estos datos, persisten en los puntos de muestreo Navarres, Llombai y Ayora un valor numéricamente distante del resto de los datos. El análisis de conglomerado se utiliza el método de Ward ya que muestra menos sensibilidad a este tipo de valores y forma agrupaciones más compactas y de igual tamaño.

Con lo que respecta a los datos utilizados para el análisis de regresión lineal de los 6 puntos de muestreo, Chelva presenta un valor bajo numéricamente distante del resto de los datos, el valor calculado corresponde al mes de julio de 2015, denotando que en esa fecha se dio una disminución del porcentaje de humedad considerable, el mes está comprendido dentro de la estación del año denominado verano, por lo que se descarta la idea considerarlo como un error sistemático.

## **7.1 Análisis de conglomerados**

El análisis de conglomerados se presenta como una herramienta muy eficaz, si se busca dividir los miembros u objetos semejantes de acuerdo a una o más variables de estudio, la ventaja que presenta, es el de dar seguimiento al anidamiento y formación de los grupos, tanto como a la estructuración de los datos, disponiendo de tablas y gráficos como el dendograma donde se puede notar fácilmente las distancias entre las agrupaciones que se combinan formando las diferentes divisiones y conglomerados. En el presente estudio la división óptima para los puntos de muestreo está entre 4 y 5.

Los datos del porcentaje de humedad de la especie vegetal romero por estación del año es el que mejor describe a la variable, por tanto los conglomerados finales serán de acuerdo a los mismo, con respecto a la división de los grupos, se escoge el de 5 porque la distancia a la que se combinan los grupos a partir de este en adelante son más alejadas y consecuentemente los puntos serán menos semejantes, además el número de integrantes por grupo es muy parejo y no hay clases con un solo integrante. Obtenido en el grupo 1 a los puntos de muestreo Ayora, Castielfabib y Villar del Arzobispo; grupo 2 (Buñol, Fredes, Navarres y Pobla – tornesa); grupo 3 (Llombai, Biar, Montanejos y Altea); grupo 5 (Albaida – Muro y Gandía). Al observar las gráficas temporales se puede notar que sus comportamientos de los puntos en función de la variable de estudio son semejantes, siendo muy evidente en las gráficas lineales temporales mostrados en el apartado 6.1.

El análisis de conglomerados utilizando los estadísticos de forma individual tuvo el objetivo de identificar la influencia de cada uno de ellos con respecto al análisis grupal, sin embargo, al ser un análisis basado en criterios geométricos no se encontró ninguna relación, aunque en algunos casos coinciden los puntos de muestreo en los mismos grupos, hay que tener en

cuenta el anidamiento total presentado en los dendogramas y las distancias entre las agrupaciones que se combinan.

En caso de no tener una variable con datos completos y no se encuentre alguna forma conseguirlo, el uso de los estadísticos puede ser de utilidad, siempre y cuando se consideren estadísticos que describan de la mejor forma la variable, al no tenerlo encuesta esto, las agrupaciones pueden generarse de manera errónea.

El análisis de conglomerado final es el realizado con los datos del porcentaje de humedad por estación del año, en 5 agrupaciones, si se comparan los gráficos temporales con cada punto de muestreo por grupo, se nota la semejanza entre los mismos, para dividir los puntos de muestreo en 4 conglomerados, dos de los grupos de 5 conglomerados se fusionan notando claramente una diferencia, misma diferencia que está relacionada con la distancia a la que estos grupos se unen, denotando así al análisis de conglomerados como una herramienta útil e intuitiva, además el usar datos del porcentaje de humedad de la especie vegetal romero por estaciones del año es la más apta, ya que la pérdida de información es menor.

## **7.2 Análisis de percentiles como clasificadores**

Al especializar los puntos de muestreo con valores categorizados de acuerdo a los percentiles crea un escenario donde se identifican con facilidad zonas de menor porcentaje de humedad de forma rápida, pudiendo así gestionar y realizar tomas de decisiones de forma eficaz.

Determinar los percentiles con un registro de datos anteriores y utilizarlos para categorizar datos actuales presenta potencialidad, sin embargo se podría mejorar, si el registro de datos para determinar los percentiles fuera mayor, de igual forma hay que tomar en cuenta que el área de influencia de los percentiles calculados está determinado por los polígonos de Thiessen y a medida que los puntos a categorizar son más distantes a las estaciones de muestreo, los puntos de muestreo localizados por los límites presentan una confiabilidad de la categorización baja, por lo que también se debería aumentar más estaciones de muestreo dentro de la provincia.

Si se siguen calculando a lo largo de los años los percentiles para cada conglomerado en cada estación del año, se debería tener en cuenta, que esta clasificación podría variar, pues

depende mucho de las variables meteorológicas. Por lo que se recomienda actualizar estos estudios con los datos nuevos de cada año, para así ajustar más el resultado.

### **7.3 Análisis de regresión lineal múltiple**

Con lo que respecta al análisis de regresión lineal múltiple, el modelar la relación existente entre las variables meteorológicas y la humedad de la especie vegetal romero de los 6 puntos de muestreo unidos, generan un modelo con un ajuste bajo, esto se puede atribuir a que el área de estudio es demasiado grande para el número de puntos con los que se trabajó en donde los factores meteorológicos interactúan de diferentes formas y más si se relaciona con la humedad de la especie vegetal del terreno.

El relación entre la variación temporal del porcentaje de humedad de la especie vegetal romero y las variables meteorológicas al analizarlas de forma independiente mejoran, sin embargo en estaciones como Montanejos se determina poca relación ya que el registro temporal es pequeño, mientras que los puntos de muestreo Biar y Buñol tienen  $r^2$  ajustados notables, con un 70.60% y 72.66% respectivamente, cabe destacar que la variable meteorológica temperatura media se presenta en los dos modelos.

El punto de muestreo Buñol como primera variable meteorológica ingresada en el modelo corresponde al grupo de punto de rocío con un  $r^2$  ajustado del 46.7%, es así, que la condensación en los últimos 30 días tiene mayor influencia en esta zona con respecto al porcentaje de humedad para la especie romero, la variable del grupo de viento, lev3 tiene un influencia notable correspondiendo al levante o viento proveniente del mar el cual contiene humedad, este punto de muestreo se encuentra ubicado en medio de la provincia de Valencia por lo que tiene mucho sentido que ejerza influencia el mar mediterráneo, aumentando un 19.91% en el  $r^2$  de ajustado, la variable temperatura media influye en menor proporción con un 6.08% en el aumento del  $r^2$  ajustado final. Con lo que respecta al punto de muestreo Biar, la variable temperatura media de 3 días describe la relación de forma notable con un  $r^2$  ajustado del 56.23%, complementado por la variable precipitación en 60 días que aumenta el  $r^2$  ajustado del modelo final en un 14.37%.

El análisis de regresión realizado con los datos en conjunto indica una baja relación, debido a la variabilidad espacial existente de la que se ha venido comentando en apartados anteriores,

tanto el período como el número de datos para el análisis de regresión por punto de muestreo es corto, los modelos podrían mejorar si aumenta el periodo de datos para el análisis, en especial para el punto de muestreo Montanejos ya que muestra poca relación con las variables meteorológicas.

El análisis de conglomerados y de regresión nos permite concluir que la variable porcentaje de humedad de la especie vegetal romero posee una variabilidad espacial que no se puede predecir con un único modelo de regresión para toda la comunidad Valenciana. Por esto se tiene que calcular un modelo de regresión para las zonas con una evolución temporal semejante.

#### **7.4 Interpolación espacial**

El modelado de superficie generado a partir del método kriging es un complemento de gran utilidad, sin embargo presenta varias desventajas en lo que respecta al análisis del porcentaje de humedad, como se observó la cantidad de puntos necesarios debe superar las 51 muestras, para poder generar un modelo de semivariograma con un buen ajuste y así obtener errores de estimación aceptables o bajos.

Tomando en cuenta la metodología utilizada, se puede decir que presenta muchas ventajas, entre ellas se podría nombrar el hecho de la existencia de zonas pocos accesibles de las cuales se podría obtener un valor coherente de estimación en un momento o tiempo específico, sin embargo hay que considerar que a mayor número de puntos muestreados, mejor es el resultado.

Cabe destacar que la toma de muestras no debe solo ser dentro de la zona de estudio ya que si se desea obtener una cobertura completa del área, se debería planificar tomar puntos externos creando una zona de amortiguamiento para mejorar la estimación en los límites.

Los puntos de muestreo estimados que mayor problema presentan son los ubicados en el límite de la superficie generada por kriging, presentando dos casos en particular Ayora y Albaida – Muro con un error estándar de estimación mayor de 32, la razón principal es que los puntos se encuentran al límite de la superficie de estimación, donde el error es más alto, Gandía y Villargordo a pesar de tener un error estándar de estimación del 24 y 13

respectivamente, quedan fuera de los límites establecido por el registro de datos a pesar de tener puntos cercanos con los que se realizó el modelado, influyendo su ubicación cerca del límite del modelo, por ello es recomendable fijar un área de amortiguamiento. Gilet a pesar de tener el error de estimación bajo queda fuera de los límites establecido, esto puede ser causado por la mala distribución de los puntos de muestreo con los que se realizó el modelo alrededor del mismo, la uniformidad de los puntos con los que se realizó el modelo influye en la estimación, esto ocurre de igual forma con el puntos de muestreo Villar del Arzobispo. Los puntos de muestreo restantes: Buñol, Llombai, Chelva y Navarres, aunque su error de estimación oscile entre los 12 y 25, dos de los puntos se ubican dentro de los límites establecidos (Buñol y Llombai), los otros dos aunque no entren en el rango, al restar o sumar su error estándar cabe la posibilidad de que su valor se encuentre entre el valor máximo y mínimo del registro de datos en los 3 años, además estos puntos presentan una particularidad, están ubicados de tal forma que alrededor de los mismos tienen una mejor distribución de los puntos con los que se realizó el modelo, por lo que es recomendable no solo tener en cuenta el número de puntos considerable para la realización del modelo, sino también deben presentar uniformidad en su distribución, se podría tener en cuenta planificar la ubicación de los puntos a estimar, en base a dichos parámetros.

## 8. BIBLIOGRAFIA

- 112 Comunitat Valenciana. (s.f.). *Incendios forestales*. Obtenido de Incendios forestales: <http://www.112cv.gva.es/>
- Andrés, A. D. (2011). *Biogeografía*. Recuperado el febrero de 2018, de <http://biogeografia.net.au.net/bioclima7a.html>
- Blanco, J., García, D., Castellnou, M., Molina, D., & Grillo, F. (2008). *Curso básico de incendios forestales*.
- Chuvieco, E., & Martín, M. d. (2004). *Nuevas tecnologías para la estimación del riesgo de incendios forestales*. Madrid: Bouncopy S.A.
- Chuvieco, E., & Pilar, M. d. (2004). *Nuevas tecnologías para la estimacion del riesgo de incendios forestales*. Madrid: Bouncopy S.A.
- Conselleria de Agricultura, Medio Ambiente, Cambio Climático y Desarrollo Rural. (2009). *Plan de Acción Territorial Forestal de la Comunitat Valenciana (PATFOR)*. Obtenido de <http://www.agroambient.gva.es/es/web/medio-natural/patfor>
- Conselleria de Vivienda, Obras Públicas y Vertebración del Territorio. (05 de 2018). *Instituto Cartográfico Valenciano*. Obtenido de <http://www.icv.gva.es/es>

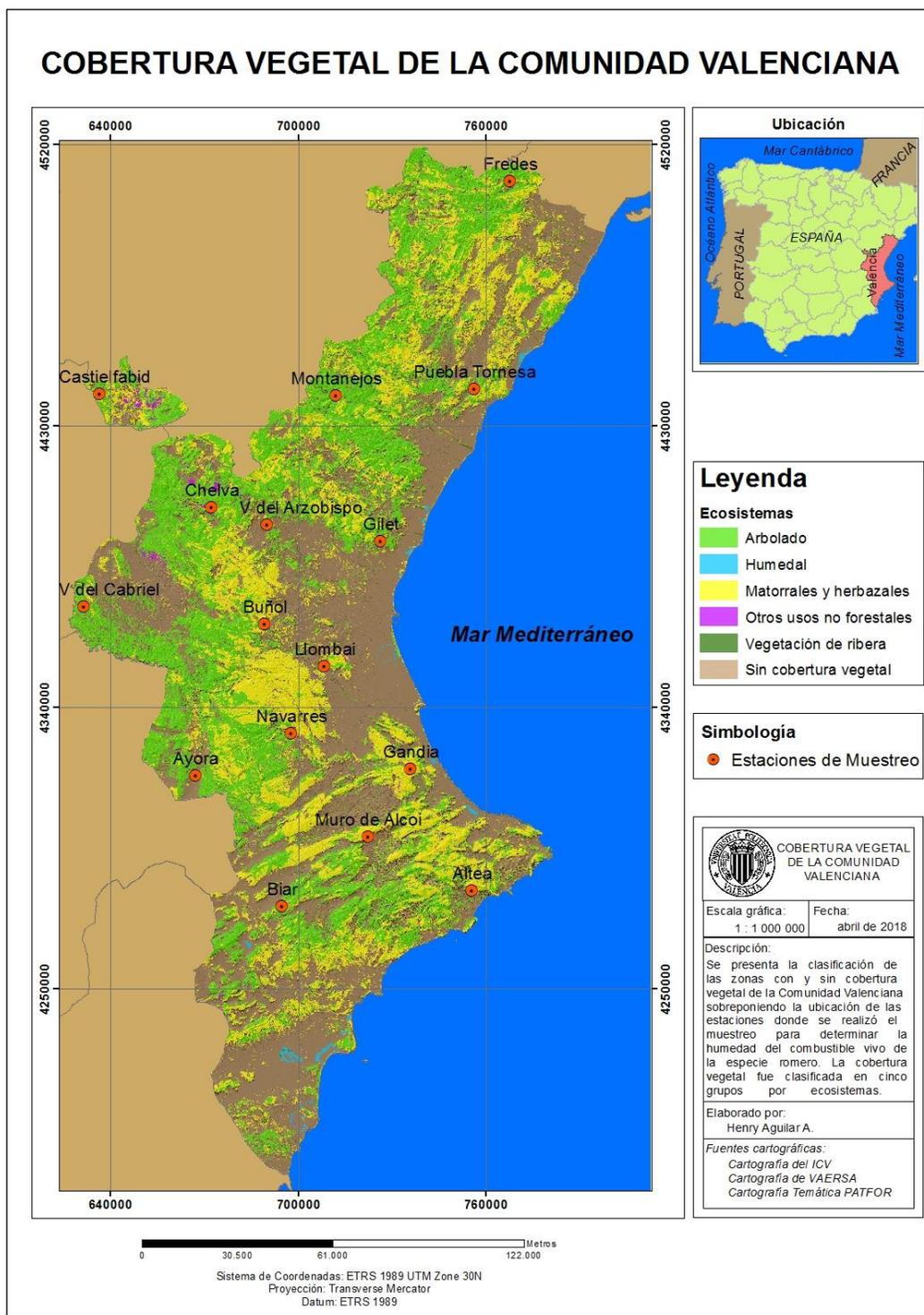
- Denham, M. (2009). *Predicción de la evolución de los incendios forestales guiada dinámicamente por los datos*. Tesis doctoral, Universitat Autònoma de Barcelona, Escola d'Arquitectura de Computadors i Sistemes Operatius, Barcelona.
- Erik Cobo, J. C. (2014). *Prueba de significación y contraste de hipótesis*. Barcelona. Obtenido de [https://ocw.upc.edu/sites/all/modules/ocw/estadistiques/download.php?file=715001/2013/1/54990/09\\_ps-5331.pdf](https://ocw.upc.edu/sites/all/modules/ocw/estadistiques/download.php?file=715001/2013/1/54990/09_ps-5331.pdf)
- ESRI. (s.f.). *ArcGIS Pro*. (ESRI, Editor) Recuperado el 10 de 08 de 2018, de ArcGIS Pro: <https://pro.arcgis.com/es/pro-app/tool-reference/3d-analyst/how-kriging-works.htm>
- FAO. (2015). *Evaluación de los recursos forstales munidales 2015*. Roma: Organización de las Naciones Unidas para la Agricultura y la Alimentación.
- Fernández, S. d. (2011). *Análisis Conglomerados*. Ciencias Economicas y Empresariales. Madrid: Universidad Autonoma de Madrid.
- Ferrero, R. (s.f.). *maximaformacion*. Recuperado el 30 de 07 de 2018, de maximaformacion: <https://www.maximaformacion.es/data-science-r/seleccion-paso-a-paso-e-importancia-de-los-predictores/>
- Gallardo. (3 de 2006). *Revista Científica y Técnica de Ecología y Medio Ambiente*. Obtenido de Revista Científica y Técnica de Ecología y Medio Ambiente: <https://fjferre.webs.ull.es/Bibliog/Biblio/Geoestadistica.pdf>
- Henaó, R. G. (2002). *Introducción a la geoestadística*. Bogotá: Universidad Nacional de Colombia.
- Justel, A. (2015). *Modelo de regresión lineal múltiple*. Madrid: Universidad Autónoma de Madrid. Obtenido de [http://www.uam.es/personal\\_pdi/ciencias/ajustel/docencia/gradoCCAA/2015AMBIENTALES-4-RegresionMultiple.pdf](http://www.uam.es/personal_pdi/ciencias/ajustel/docencia/gradoCCAA/2015AMBIENTALES-4-RegresionMultiple.pdf)
- Lázaro, E. S. (s.f.). *Especies forestales de España. Caracteres culturales*. Obtenido de <http://especiesforestales.com/Fitoclimaticos.htm>
- Soriano, J. L. (2016). *Análisis de la Humedad del Combustible Vivo (HCV) en la Comunitat Valenciana y la influencia de las variables meteorológicas*. Valencia.
- The Comprehensive R Archive Network. (s.f.). *The Comprehensive R Archive Network*. Recuperado el 18 de 12 de 2017, de [https://cran.r-project.org/web/packages/olsrr/vignettes/variable\\_selection.html](https://cran.r-project.org/web/packages/olsrr/vignettes/variable_selection.html)
- Uriel, E. (2013). *Contraste de hipótesis en el modelo de regresión múltiple*. Valencia. Obtenido de <https://www.uv.es/=uriel/4%20Contraste%20de%20hipotesis%20en%20el%20modelo%20de%20regresion%20multiple.pdf>

## 9. ANEXOS

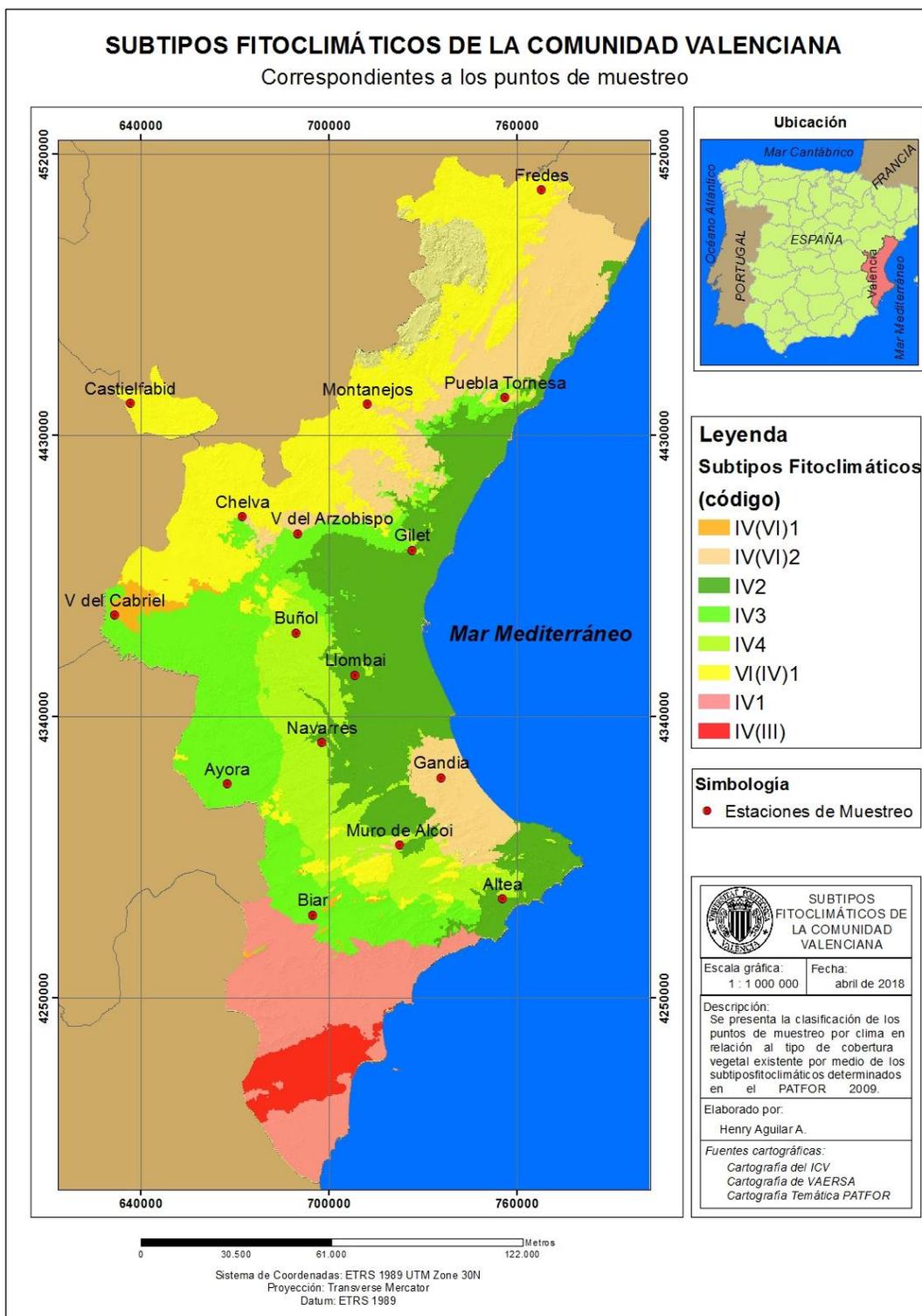
### ANEXO 1. Zona de estudio y ubicación de los puntos de muestreo.



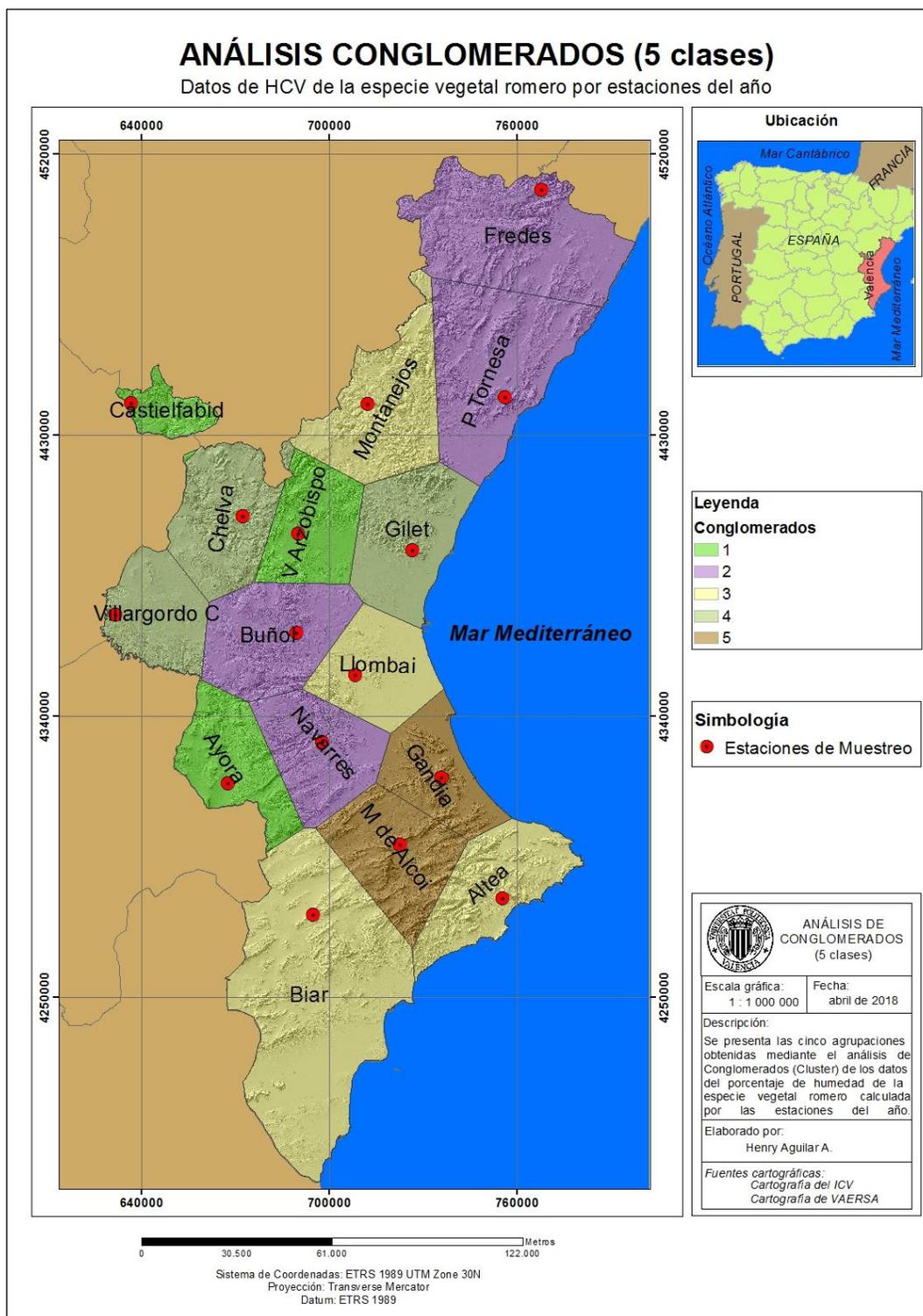
ANEXO 2. Mapa de la cobertura vegetal de la Comunidad Valenciana.



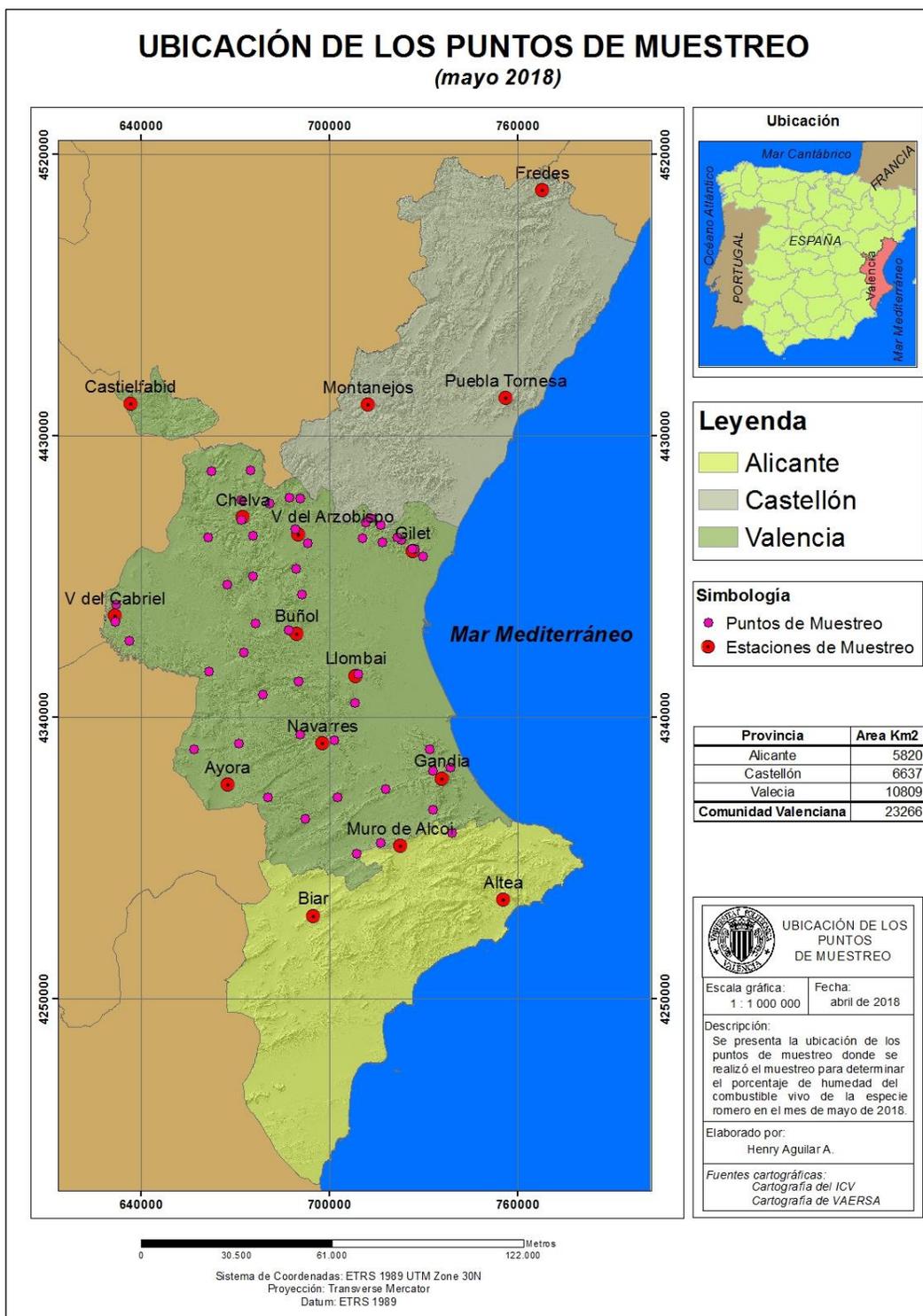
ANEXO 3. Mapa de subtipos fitoclimáticos de la Comunidad Valenciana.



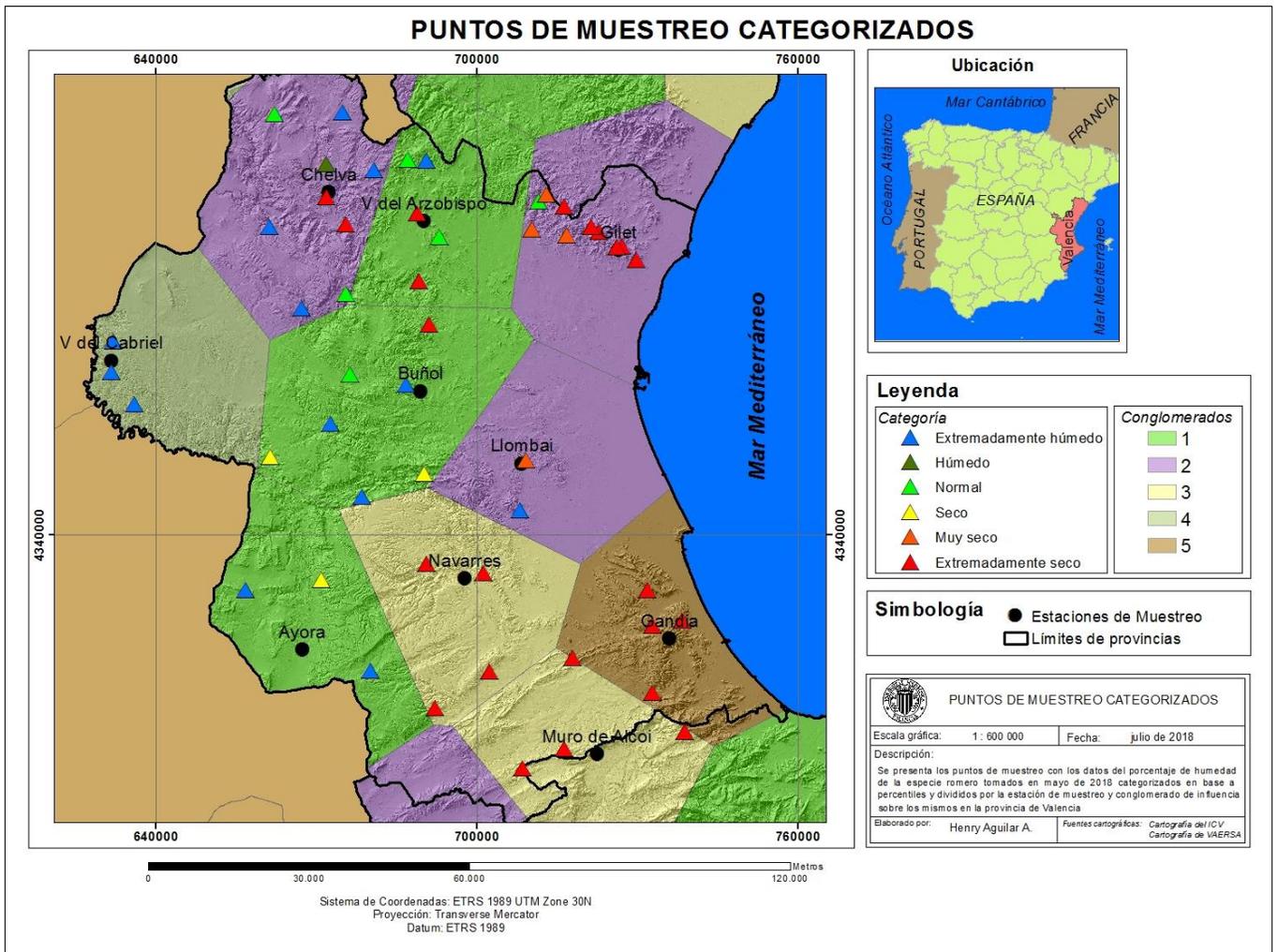
ANEXO 4. Mapa división por conglomerados de los puntos de muestreo.



ANEXO 5. Mapa división por conglomerados de los puntos de muestreo.



ANEXO 6. Mapa de los puntos de muestreo categorizados en la provincia de Valencia.



ANEXO 7. Modelos generados con RStudio.

**Selección de variables por grupo**

Grupo	Pasos	Variable	R <sup>2</sup> ajustado	Beta	Evaluación
Precipitación	Paso 0:	CONSTANTE	0.00%	101.999	R <sup>2</sup> ajustado o Criterio de información
	Paso 1:	P60	18.40%	0.187	
	Paso 2:	P7	19.80%	-0.115	
Temperatura	Paso 0:	CONSTANTE	0.00%	141.992	R <sup>2</sup> ajustado o Criterio de información
	Paso 1:	Tmedia3	17.40%	-4.283	
	Paso 2:	Tmax3	18.20%	1.961	
Humedad relativa	Paso 0:	CONSTANTE	0.00%	72.914	R <sup>2</sup> ajustado o Criterio de información
	Paso 1:	HRmin15	12.50%	1.617	
	Paso 2:	HRmin7	14.00%	-0.605	
Viento	Paso 0:	CONSTANTE	0.00%	120.253	R <sup>2</sup> ajustado
	Paso 1:	lev7	-0.50%	-3.094	
	Paso 2:	lev15	1.00%	2.986	
Punto de rocío	Paso 0:	CONSTANTE	0.00%	113.258	R <sup>2</sup> ajustado o Criterio de información
	Paso 1:	pto_r30	0.198	0.846	
	Paso 2:	pto_r3	0.229	-1.622	

**Modelo generado**

Pasos	Variable	R <sup>2</sup> ajustado	Beta
Paso 0:	CONSTANTE	0.00%	98.937
Paso 1:	pto_r30	19.80%	0.589
Paso 2:	Tmedia3	32.20%	-3.547
Paso 3:	P60	37.30%	0.103
Paso 4:	pto_r3	42.70%	-1.206
Paso 5:	P7	44.10%	-0.122
Paso 7:	HRmin15	45.00%	0.485
Paso 8:	Tmax3	45.50%	1.794

**Modelado con todas las variables**

Pasos	Variable	R <sup>2</sup> ajustado	Beta
Paso 0:	CONSTANTE	0.00%	96.382
Paso 1:	pto_r30	19.80%	0.655
Paso 2:	Tmedia3	32.20%	-3.586
Paso 3:	P60	37.30%	0.095
Paso 4:	P3	43.20%	-0.147
Paso 5:	HRmax7	46.00%	0.392
Paso 6:	pto_r7	46.70%	-1.114
Paso 7:	P15	47.20%	0.137
Paso 8:	P7	48.60%	-0.228
Paso 9:	Tmedia15	49.80%	1.954
Paso 10:	lev3	49.90%	-0.628

**Modelado con las variables seleccionadas por el coeficiente de determinación múltiple R<sup>2</sup> ajustado.**

Pasos	Variable	R <sup>2</sup> ajustado	Beta
Paso 0:	CONSTANTE	0.00%	92.12
Paso 1:	pto_r30	19.80%	0.633
Paso 2:	Tmedia3	32.20%	-2.897
Paso 3:	P60	37.30%	0.135
Paso 4:	pto_r3	42.70%	-1.391
Paso 5:	HRmax7	45.10%	0.432
Paso 6:	P7	47.60%	-0.172
Paso 7:	Tmedia15	47.80%	1.14

**Modelado con las variables seleccionadas por los criterio de información.**

Pasos	Variable	R <sup>2</sup> ajustado	Beta
Paso 0:	CONSTANTE	0.00%	91.13
Paso 1:	pto_r30	19.80%	0.656
Paso 2:	Tmedia3	32.20%	-2.932
Paso 3:	P60	37.30%	0.131
Paso 4:	pto_r3	42.70%	-1.456
Paso 5:	HRmax7	45.10%	0.452
Paso 6:	P7	47.60%	-0.163
Paso 7:	Tmedia15	47.80%	1.294
Paso 8:	lev3	48.10%	-0.725

ANEXO 8. Modelos generados con RStudio punto de muestreo Biar

**Selección de variables por grupo**

Grupo	Pasos	Variable	R <sup>2</sup> ajustado	Beta	Evaluación
Precipitación	Paso 0:	CONSTANTE	0.00%	78.01	R <sup>2</sup> ajustado o Criterio de información
	Paso 1:	P60	44.40%	0.55	
Temperatura	Paso 0:	CONSTANTE	0.00%	192.075	R <sup>2</sup> ajustado o Criterio de información
	Paso 1:	Tmedia3	56.20%	-19.438	
	Paso 2:	Tmedia15	60.40%	16.397	
	Paso 3:	Tmax3	64.30%	8.11	
	Paso 4:	Tmax15	71.60%	-9.486	
Humedad relativa	Paso 0:	CONSTANTE	0.00%	-107.877	R <sup>2</sup> ajustado
	Paso 1:	HRmin15	33.70%	2.378	
	Paso 2:	HRmax7	43.20%	1.762	
	Paso 3:	HRmin3	45.50%	-0.937	
Viento	Paso 0:	CONSTANTE	0.00%	249.398	R <sup>2</sup> ajustado
	Paso 1:	Vmax15	20.80%	-7.982	
	Paso 2:	lev15	24.00%	-38.916	
	Paso 3:	lev3	28.90%	10.045	
	Paso 4:	Vmax7	30.50%	6.629	
	Paso 5:	Vmax3	35.50%	-1.975	
Punto de rocío	Paso 0:	CONSTANTE	0.00%	103.80	R <sup>2</sup> ajustado o Criterio de información
	Paso 1:	pto_r30	15.70%	1.11	

**Modelo generado**

Pasos	Variable	R <sup>2</sup> ajustado	Beta
Paso 0:	CONSTANTE	0.00%	90.012
Paso 1:	Tmedia3	56.20%	-18.75
Paso 2:	HRmax7	78.00%	2.372
Paso 3:	Tmedia15	85.40%	21.042
Paso 4:	Vmax3	85.80%	-1.591
Paso 5:	Tmax15	86.00%	-10.88
Paso 6:	HRmin3	88.80%	0.116
Paso 7:	lev15	90.50%	-17.43
Paso 8:	P60	92.00%	-0.244
Paso 9:	pto_r30	93.80%	-0.654
Paso 10:	Tmax3	96.00%	4.873

**Modelado con todas las variables**

Pasos	Variable	R <sup>2</sup> ajustado	Beta
Paso 0:	CONSTANTE	0.00%	-324.497
Paso 1:	Tmedia3	56.20%	-20.729
Paso 2:	HRmax7	78.00%	9.25
Paso 3:	Tmedia15	85.40%	0.843
Paso 4:	lev7	88.00%	-24.133
Paso 5:	Vmax3	89.00%	-0.639
Paso 6:	P60	90.60%	-0.753
Paso 7:	P30	92.70%	0.801
Paso 8:	Tmedia7	93.90%	15.248
Paso 9:	lev3	94.70%	19.067
Paso 10:	HRmax3	95.30%	-3.401
Paso 11:	pto_r15	97.50%	-0.849
Paso 12:	pto_r7	99.20%	0.531
Paso 13:	P3	99.60%	-0.506
Paso 14:	Vmax15	99.90%	0.524

Modelado con las variables seleccionadas por el coeficiente de determinación múltiple R<sup>2</sup> ajustado.

Pasos	Variable	R <sup>2</sup> ajustado	Beta
Paso 0:	CONSTANTE	0.00%	90.012
Paso 1:	Tmedia3	56.20%	-18.747
Paso 2:	HRmax7	78.00%	2.372
Paso 3:	Tmedia15	85.40%	21.042
Paso 4:	Vmax3	85.80%	-1.591
Paso 5:	Tmax15	86.00%	-10.877
Paso 6:	HRmin3	88.80%	0.116
Paso 7:	lev15	90.50%	-17.431
Paso 8:	P60	92.00%	-0.244
Paso 9:	pto_r30	93.80%	-0.654
Paso 10:	Tmax3	96.00%	4.873

Modelado con las variables seleccionadas por los criterio de información.

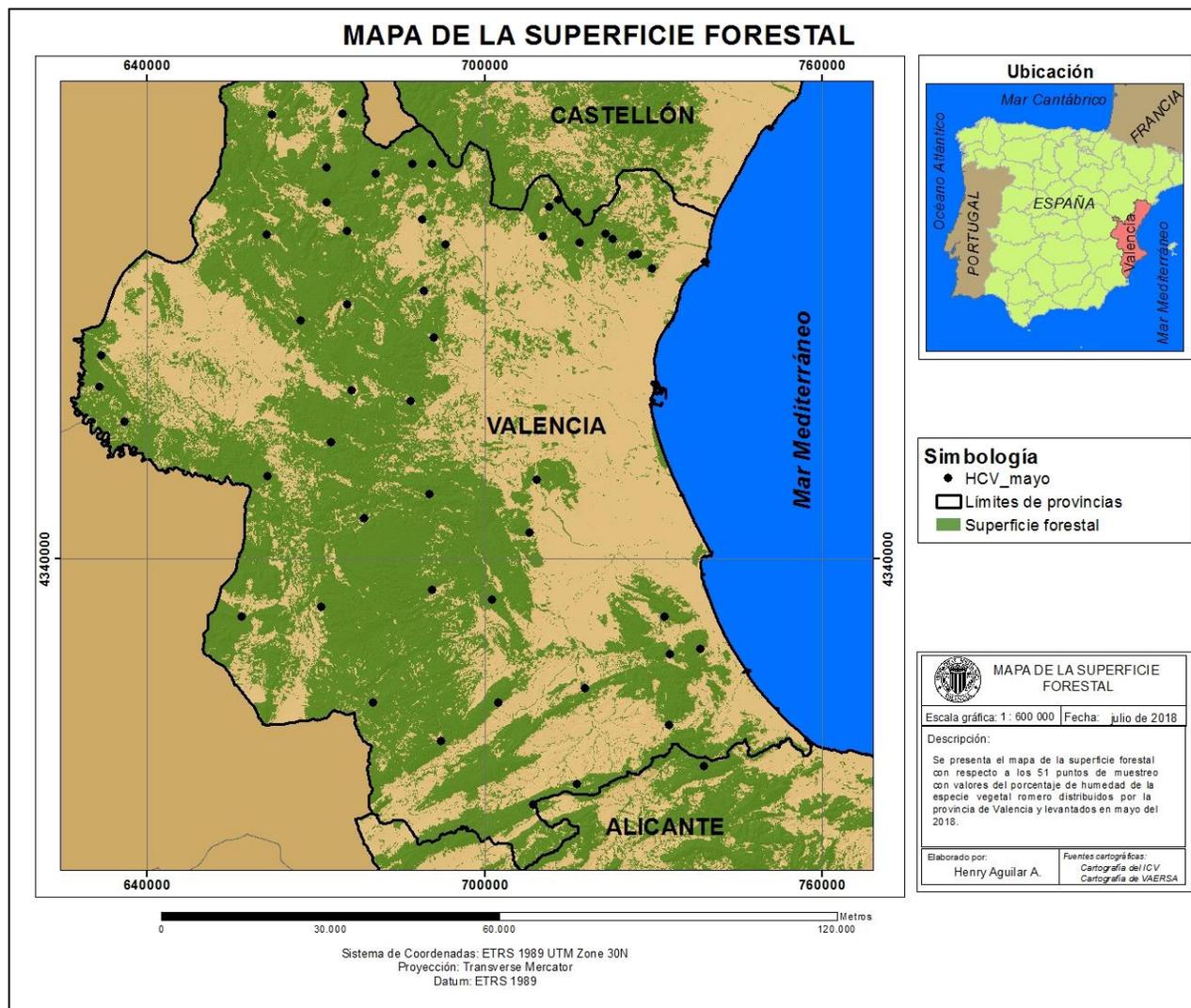
Pasos	Variable	R <sup>2</sup> ajustado	Beta
Paso 0:	CONSTANTE	0.00%	91.377
Paso 1:	Tmedia3	56.20%	-11.982
Paso 2:	P60	70.60%	0.107
Paso 3:	Tmax3	72.60%	3.572
Paso 4:	HRmin7	73.90%	0.824
Paso 5:	Tmedia15	78.10%	5.42

#### ANEXO 9. Resultado de la validación cruzada

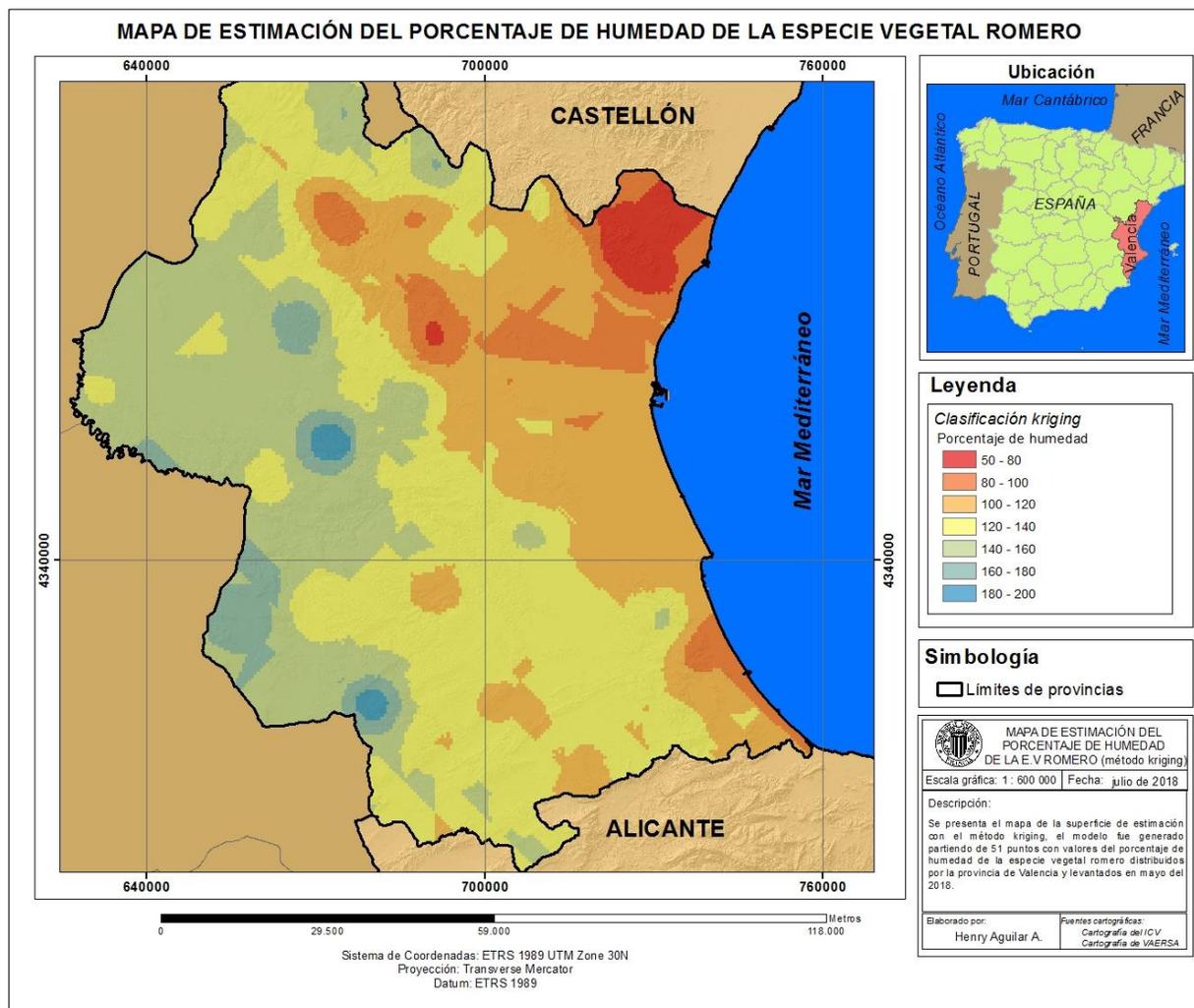
Medidas	Predicho	Error	StdError	Stdd_Error	NormValue	ID	Incluido
59.343	57.416	-1.927	14.664	-0.131	-0.248	0	Yes
51.849	79.580	27.731	27.917	0.993	1.352	1	Yes
57.885	60.528	2.642	14.955	0.177	0.049	2	Yes
62.905	71.685	8.780	16.321	0.538	0.513	3	Yes
74.051	74.669	0.619	16.943	0.037	-0.098	4	Yes
106.602	92.968	-13.635	28.700	-0.475	-0.754	5	Yes
106.820	123.371	16.551	30.808	0.537	0.458	6	Yes
124.381	104.599	-19.781	19.162	-1.032	-1.238	7	Yes
101.596	115.002	13.406	18.725	0.716	0.967	8	Yes
86.678	102.665	15.986	26.076	0.613	0.754	9	Yes
88.364	110.762	22.398	34.162	0.656	0.821	10	Yes
134.149	104.130	-30.019	32.744	-0.917	-1.049	11	Yes
164.834	131.850	-32.984	27.685	-1.191	-1.486	12	Yes
140.221	153.590	13.369	26.398	0.506	0.351	13	Yes
96.244	125.317	29.073	32.949	0.882	1.139	14	Yes
141.515	127.634	-13.881	33.591	-0.413	-0.690	15	Yes
87.428	117.376	29.949	32.944	0.909	1.238	16	Yes
83.480	123.723	40.243	31.509	1.277	1.890	17	Yes
129.578	121.746	-7.832	33.025	-0.237	-0.351	18	Yes
162.248	141.333	-20.916	35.459	-0.590	-0.892	19	Yes
123.749	133.258	9.509	36.737	0.259	0.098	20	Yes
148.003	130.700	-17.303	34.775	-0.498	-0.821	21	Yes

Medidas	Predicho	Error	StdError	Stdd_Error	NormValue	ID	Incluido
181.368	135.400	-45.967	34.217	-1.343	-1.654	22	Yes
124.872	124.785	-0.087	34.317	-0.003	-0.148	23	Yes
71.868	129.752	57.884	34.242	1.690	2.334	24	Yes
135.072	144.938	9.867	34.398	0.287	0.198	25	Yes
156.527	123.284	-33.243	34.456	-0.965	-1.139	26	Yes
154.543	143.347	-11.195	33.057	-0.339	-0.513	27	Yes
133.290	152.415	19.125	31.974	0.598	0.690	28	Yes
149.818	159.599	9.781	34.621	0.283	0.148	29	Yes
126.905	156.960	30.055	34.537	0.870	1.049	30	Yes
201.956	142.966	-58.990	34.374	-1.716	-1.890	31	Yes
107.323	120.122	12.799	34.974	0.366	0.248	32	Yes
146.635	124.855	-21.780	34.396	-0.633	-0.967	33	Yes
120.454	140.854	20.400	34.511	0.591	0.629	34	Yes
152.011	141.431	-10.579	34.475	-0.307	-0.458	35	Yes
101.082	144.072	42.990	34.432	1.249	1.654	36	Yes
125.728	128.129	2.402	34.429	0.070	0.000	37	Yes
175.063	162.718	-12.344	36.348	-0.340	-0.570	38	Yes
123.600	143.429	19.829	34.487	0.575	0.570	39	Yes
198.513	125.928	-72.585	35.614	-2.038	-2.334	40	Yes
112.469	130.240	17.771	34.503	0.515	0.404	41	Yes
117.406	140.491	23.085	35.102	0.658	0.892	42	Yes
129.755	131.279	1.524	34.740	0.044	-0.049	43	Yes
109.667	123.889	14.222	34.764	0.409	0.299	44	Yes
116.845	112.315	-4.531	31.542	-0.144	-0.299	45	Yes
80.650	113.116	32.465	32.601	0.996	1.486	46	Yes
146.573	110.143	-36.431	33.685	-1.082	-1.352	47	Yes
131.084	121.963	-9.121	34.531	-0.264	-0.404	48	Yes
127.408	113.878	-13.529	35.648	-0.380	-0.629	49	Yes
122.820	121.463	-1.357	36.089	-0.038	-0.198	50	Yes

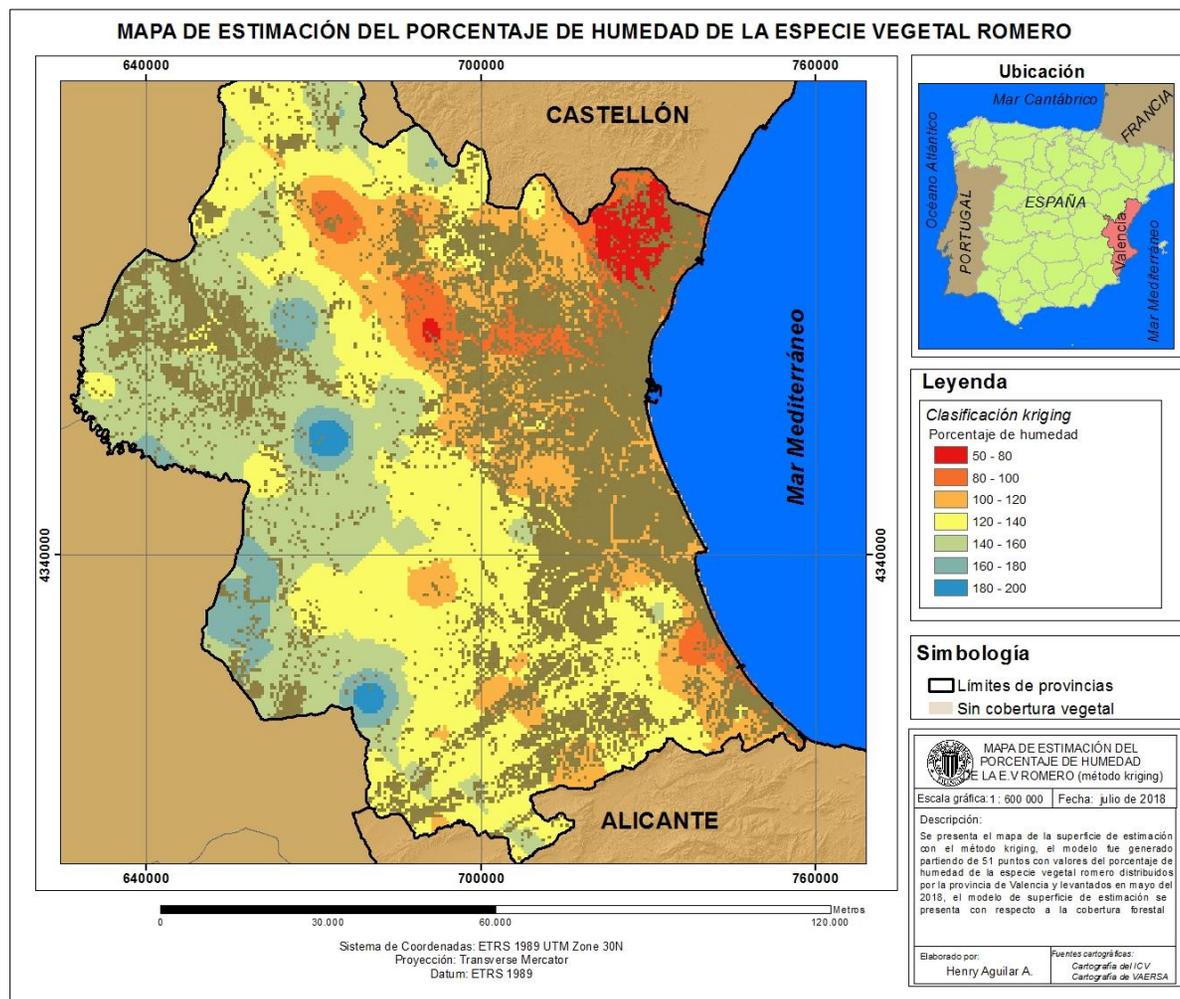
ANEXO 10. Mapa de los puntos muestreados en mayo 2018 con respecto a la superficie forestal.



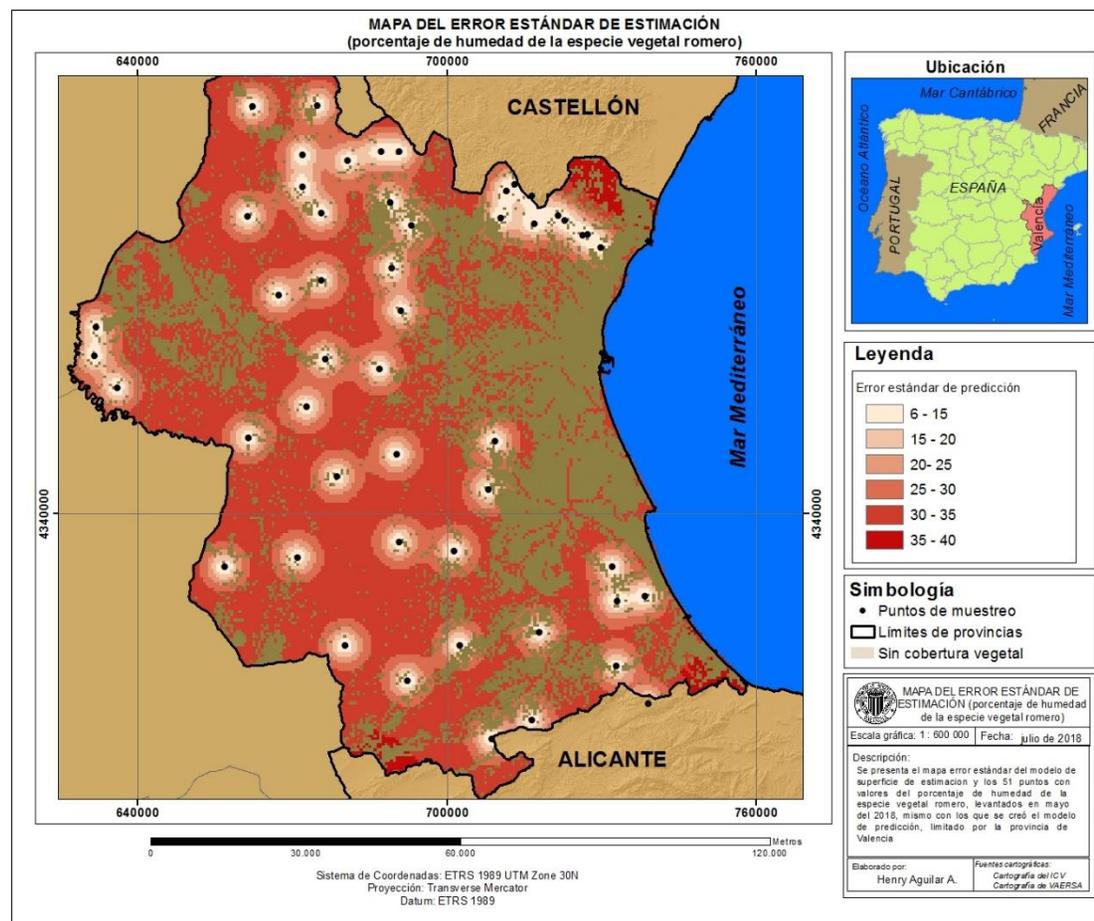
ANEXO 11. Mapa de superficie de estimación del porcentaje de humedad de la especie vegetal romero en la provincia de Valencia (mayo 2018).



ANEXO 12. Mapa de superficie de estimación del porcentaje de humedad de la especie vegetal romero extendido en la provincia de Valencia en función de la superficie forestal (mayo 2018).



ANEXO 13. Mapa de superficie del error estándar de estimación en función de la superficie forestal.



ANEXO 14. Código realizado en RStudio utilizado en los diferentes análisis

```

#-----
#INSTALACION LIBRERIAS
install.packages("RODBC")
install.packages("XLConnect")
#-----
#ACTIVACION LIBRERIAS
#libreria para importacion de libro de excel y coneccion hoja
library(RODBC)
#libreria para conectar, cargar, entre otros. Realizar procesos con excel
library(XLConnect)
#libreria para realizar graficos
library(ggplot2)
library(reshape2)
#-----
#CONECCION
#Coneccion con el libro de excel
canalestas <- odbcConnectExcel2007("E:/TFM/Tomos/2.AnalisisEstadistico/EstadisticosRo.xls")
canalhcv <- odbcConnectExcel2007("E:/TFM/Tomos/2.AnalisisEstadistico/Romero_HCV_2.1.xls")
#Visualizacion de las hojas dentro del libro de excel
sqlTables(canalestas)
sqlTables(canalhcv)
#Coneccion con la hoja de excel
sqlFetch(canalestas, "STATG_pmt")
sqlFetch(canalhcv, "hcvRo2")
sqlFetch(canalhcv, "hcvRot")
sqlFetch(canalhcv, "hcvRot2")
#debido a que no se puede manipular estadisticamente los datos se crea una nueva hoja en R para manipular
estaT <- sqlFetch(canalestas, "STATG_pmt")
hcv1 <- sqlFetch(canalhcv, "hcvRo2")
hcvT <- sqlFetch(canalhcv, "hcvRot")
hcvT2 <- sqlFetch(canalhcv, "hcvRot2")
#Se cierra el canal Excel
odbcClose(canalestas)
odbcClose(canalhcv)
#-----
#-----
#introducir el color por estacion
CorAyora<-c("#f74cb0")
CorBunhol<-c("#058748")
CorLlombai<-c("#fff600")
CorVillar<-c("#ff9d00")
CorChelva<-c("#65ff00")
CorAlbaida<-c("#5b2b03")
CorGilet<-c("#ff0000")
CorVillargordo<-c("#029bf4")
CorCastielfabib<-c("#720244")
CorNavarres<-c("#d88b54")
CorGandia<-c("#dlc81d")
CorBiar<-c("#00e1ff")
CorAltea<-c("#283eff")
CorMontanejos<-c("#021191")
CorFredes<-c("#ff710c")
CorPobla<-c("#8c0202")
#-----
#-----
#introducir base de datos
datos <- estaT
#introducir valores en x
x1 <- datos$cod
#introducir valores en y
y1 <- datos$Promedio
#leyenda
x2 <- "Puntos de Muestreo"
y2 <- "Promedio"

#GRAFICOS BARRAS
Graf_barras <- ggplot(data = datos, aes(x=x1, y=y1, fill = Estaciones)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = c(CorAlbaida,CorAltea,CorAyora,CorBiar,CorBunhol,CorCastielfabib,
                              CorChelva,CorFredes,CorGandia,CorGilet,CorLlombai,CorMontanejos,
                              CorNavarres,CorPobla,CorVillar,CorVillargordo)) +
  labs(x=x2,y=y2) +
  theme(axis.title.x = element_text(size = 22, colour = "black"),
        axis.title.y = element_text(size = 22, colour = "black"),
        axis.text.x = element_text(size = 16, colour = "black"),
        axis.text.y = element_text(size = 16, colour = "black"),
        legend.text = element_text(size = 18),
        legend.title = element_text(size = 20))
#-----
#-----
#GRAFICO LINEA TEMPORAL POR MES TODOS LOS PUNTOS MUESTREO
#Introducir base de datos
datos <- hcvT
#Introducir leyenda

```

```

x2 <- "Puntos de Muestreo"
y2 <- "%HCV"

#Introducir datos para crear nueva variable con datos
datos1 <- melt(datos,id.vars="CodFecha", measure.vars=c("ALBAIDA_MURO", "ALTEA", "AYORA", "BIAR", "BUNYOL",
"CASTIELFABIB",
"CHELVA", "FREDES", "GANDIA", "GILET", "LLOMBAI",
"MONTANEJOS",
"NAVARRES", "POBLA_TORNESA", "VILLAR_ARZOBISPO",
"VILLARGORDO"))

Graf_LT <- ggplot(datos1, aes(x=CodFecha, y=value, group=variable )) +
  geom_line(aes(colour=variable),size=1) +
  geom_point(aes(colour=variable),size=3) +
  scale_colour_manual(values = c(CorAlbaida,CorAltea,CorAyora,CorBiar,CorBunhol,CorCastielfabib,
CorChelva,CorFredes,CorGandia,CorGilet,CorLlombai,CorMontanejos,
CorNavarres,CorPobla,CorVillar,CorVillargordo)) +

  labs(x=x2,y=y2) +
  theme(axis.title.x = element_text(size = 22, colour = "black"),
axis.title.y = element_text(size = 22, colour = "black"),
axis.text.x = element_text(size = 16, colour = "black"),
axis.text.y = element_text(size = 16, colour = "black"),
legend.text = element_text(size = 18),
legend.title = element_text(size = 20))

#-----
#-----
#Introducir base de datos
datos <- hcvT
#Introducir leyenda
x2 <- "Puntos de Muestreo"
y2 <- "%HCV"
#Introducir datos para crear nueva variable con datos
datos2 <- melt(hcv1,id.vars='Estacion',
measure.vars=c("f9","f10","f11","f12","f13","f14","f15","f16","f17","f18","f19","f20",
"f21","f22","f23","f24","f25","f26","f27","f28","f29","f30",
"f31","f32","f33","f34","f35","f36","f37","f38","f39","f40",
"f41","f42","f43","f44","f45","f46","f47","f48","f49","f50"))

grafic_CB <- ggplot(data = datos2) +
  geom_boxplot(aes(x=Estacion, y=value, fill = Estacion)) +
  labs(x=x2,y=2) + coord_flip() +
  theme(axis.title.x = element_text(size = 22, colour = "black"),
axis.title.y = element_text(size = 22, colour = "black"),
axis.text.x = element_text(size = 16, colour = "black"),
axis.text.y = element_text(size = 16, colour = "black"),
legend.text = element_text(size = 18),
legend.title = element_text(size = 20)) +
  scale_fill_manual(values = c(CorAlbaida,CorAltea,CorAyora,CorBiar,CorBunhol,CorCastielfabib,
CorChelva,CorFredes,CorGandia,CorGilet,CorLlombai,CorMontanejos,
CorNavarres,CorPobla,CorVillar,CorVillargordo))

#-----
#-----
#ANALISIS INICIAL POR MES ESTADISTICOS
#Estadisticos datos completos
estaHCVT<-summary(hcvT)
#Estadisticos datos 3 anhos completos
estaHCVT2<-summary(hcvT2)
#-----
#-----
#CREAR LIBRO EXCEL Y EXPORTAR LOS DATOS GENERADOS EN r
wb <- loadWorkbook("E:/TFM/Tomos/2.AnalisisEstadistico/Estadisticos_r.xlsx", create = TRUE)
createSheet(wb, name="pm")
createName(wb, name="pm", formula = "pm!$A$1")
writeNamedRegion(wb, estaHCVT, name="pm")
saveWorkbook(wb)

```

```

#-----
#INSTALACION LIBRERIAS
install.packages("RODBC")
install.packages("XLConnect")
install.packages("dendextend")
#-----
#ACTIVACION LIBRERIAS
#libreria para importacion de libro de excel y coneccion hoja
library(RODBC)
#libreria para conectar, cargar, entre otros. Realizar procesos con excel
library(XLConnect)
#-----
#-----
#CONECCION
#Coneccion con el libro de excel
canalEstaAn <- odbcConnectExcel2007("E:/TFM/Tomos/2.AnalisisEstadistico/Ro_hcv_modificado_estacionesAnho2.3.xls")
canalhcv <- odbcConnectExcel2007("E:/TFM/Tomos/2.AnalisisEstadistico/EstadisticosRo.xls")
#Visualizacion de las hojas dentro del libro de excel
sqlTables(canalEstaAn)
sqlTables(canalhcv)
#Coneccion con la hoja de excel
sqlFetch(canalEstaAn, "EM2")
sqlFetch(canalEstaAn, "EstadisticosEAT")
sqlFetch(canalhcv, "STATG_pmt")
#debido a que no se puede manipular estadisticamente los datos se crea una nueva hoja en R para manipular
em2 <- sqlFetch(canalEstaAn, "EM2")
EstaEAT <- sqlFetch(canalEstaAn, "EstadisticosEAT")
hcv <- sqlFetch(canalhcv, "STATG_pmt")
#Se cierra el canal Excel
odbcClose(canalEstaAn)
odbcClose(canalhcv)
#-----
#####MÉTODO CLUSTER#####
#-----
#CONGLOMERADOS CON DATOS DE ESTACIONES DEL ANHO DE HCV
#-----
#MÉTODO JERARQUICO
#Se crea la matriz x con la finalidad de descartar el nombre de las estaciones, se coloca el num de filas
#en este caso es (1:16)
xEA = em2[1:16,-1:-2]
#Se toman las columnas con los datos a utilizar, se tipifica y estandarizar los datos (opcional)
yEA = scale(xEA[,2:12])
#Se coloca el nombre a cada fila
row.names(xEA) = as.character(em2[1:16,2])
#Se identifica el nombre de cada fila
em2.tipif=xEA
#Se define la matriz de disimilaridad, llamada también la matriz de distancias
matriz_disimilaridadEA=dist(em2.tipif)^2
#la medida de disimilaridad utilizada es la distancia euclidean. Otra forma de especificar la matriz de
#distancias y aprovechar en definir la forma de aglomeracion es mediante la siguiente linea
#La medida de distancia (euclidean) y el método de agrupamiento del análisis cluster no jerarquico,
#en este caso el método de ward
matriz_distEA=hclust(dist(em2.tipif)^2,method="ward.D")
#Historial de conglomeración
matriz_distEA$merge
#identificar los conglomerados
row.names(em2.tipif)[c(5,7)]
#dendograma
plot(matriz_distEA)
#dendograma colocando el numero de grupos manualmente
rect.hclust(matriz_distEA, 4)
#Asignacion del grupo mostrado en la tabla principal(base de datos em2)
em2$clusternumberEA<-cutree(matriz_distEA, 4)
#lista que permite identificar los componentes o estaciones de muestreo de cada clúster
print(rect.hclust(matriz_distEA,k=4))

```

```

#-----
----
#INSTALACION LIBRERIAS
install.packages("RODBC")
install.packages("XLConnect")
install.packages("dendextend")
#-----
----
#ACTIVACION LIBRERIAS
#libreria para importacion de libro de excel y coneccion hoja
library(RODBC)
#libreria para conectar, cargar, entre otros. Realizar procesos con excel
library(XLConnect)
#libreria para realizar graficos
library(ggplot2)
library(reshape2)
#-----
----
#-----
----
#CONECCION
#Coneccion con el libro de excel
canalhcv_datos <- odbcConnectExcel2007("E:/TFM/Tomos/AnalsisConglomeEA/HCVmayo2018/Categor_PMmayo2018.xlsx")
#Visualizacion de las hojas dentro del libro de excel
sqlTables(canalhcv_datos)
#Coneccion con la hoja de excel
sqlFetch(canalhcv_datos, "PM_rstudio")
#debido a que no se puede manipular estadisticamente los datos se crea una nueva hoja en R para manipular
hcvPerc <- sqlFetch(canalhcv_datos, "PM_rstudio")
#Se cierra el canal Excel
odbcClose(canalhcv_datos)
#attach(Comhcv)
#-----
----
##Introduccion archivo para clasificar
ArchPrin<-hcvPerc

##Introducir codigo Color
Colp1 = "#ff0000"
Colp2 = "#ff6600"
Colp3 = "#ffcc00"
Colp4 = "#339900"
Colp5 = "#336600"
Colp6 = "#0000ff"

#-----
----
#extraccion de puntos hcv actuales segun su ubicacion y el cuadrante o poligono del PM que lo influye
albaida<-ArchPrin[ArchPrin$PM_ref=="ALBAIDA_MURO",]
ayora<-ArchPrin[ArchPrin$PM_ref=="AYORA",]
bunhol<-ArchPrin[ArchPrin$PM_ref=="BUNYOL",]
chelva<-ArchPrin[ArchPrin$PM_ref=="CHELVA",]
gandia<-ArchPrin[ArchPrin$PM_ref=="GANDIA",]
gilet<-ArchPrin[ArchPrin$PM_ref=="GILET",]
llombai<-ArchPrin[ArchPrin$PM_ref=="LLOMBAI",]
navarres<-ArchPrin[ArchPrin$PM_ref=="NAVARRES",]
villar<-ArchPrin[ArchPrin$PM_ref=="VILLAR_ARZOBISPO",]
villargordo<-ArchPrin[ArchPrin$PM_ref=="VILLARGORDO",]

##Introducir agrupacion por conglomerados
Con1_pHCV<-rbind(ayora,bunhol,villar)
Con2_pHCV<-rbind(chelva,gilet,llombai)
Con3_pHCV<-rbind(albaida,navarres)
Con4_pHCV<-rbind(villargordo)
Con5_pHCV<-rbind(gandia)

#-----
-----
#-----
-----
#-----CLASIFICAION valores DE HCV PUNTOS DE MUESTREO-----
-----
##ANALISIS por CONGLOMERADO #####
##Introducir conglomerado
Con <- Con1_pHCV
##Introducir leyenda grafico
#nombre eje x y y
jx <- "Puntos de muestreo"
jy <- "HCV %"
#Subtitulo
sb <- "Conglomerado 2"
#nombre leyenda
lg <- "valor de HCV"

```

```

#dimensiones del fichero
D1<-dim(Con)
#percentiles de clasificacion
p10<-Con$P10[1]
p20<-Con$P20[1]
p40<-Con$P40[1]
p60<-Con$P60[1]
p80<-Con$P80[1]
p90<-Con$P90[1]
#-----
#Clasificacion

for(i in 1:D1[1]){
  if(Con$HCV[i]<p10){
    Con$Clasificacion_R[i]<-"Extremadamente seco"
  }
  else if(Con$HCV[i]>p10 & Con$HCV[i]<p20){
    Con$Clasificacion_R[i]<-"Muy seco"
  }
  else if(Con$HCV[i]>p20 & Con$HCV[i]<p40){
    Con$Clasificacion_R[i]<-"Seco"
  }
  else if(Con$HCV[i]>p40 & Con$HCV[i]<p60){
    Con$Clasificacion_R[i]<-"Normal"
  }

  else if(Con$HCV[i]>p60 & Con$HCV[i]<p80){
    Con$Clasificacion_R[i]<-"Húmedo"
  }
  else if(Con$HCV[i]>p80 & Con$HCV[i]<p90){
    Con$Clasificacion_R[i]<-"Muy Húmedo"
  }
  else{
    Con$Clasificacion_R[i]<-"Extremadamente húmedo"
  }
}
#-----
#Grafica Conglomerado con limites

grafCon1 <- melt(Con,id.vars='Codnum', measure.vars="HCV")

graf_Lim <- ggplot(grafCon1, aes(x = Codnum, y = value, colour = variable)) +
  geom_point(size=3) +
  geom_hline(yintercept = p10, color = Colp1, linetype = "dashed", size=0.8) +
  geom_hline(yintercept = p20, color = Colp2, linetype = "dashed", size=0.8) +
  geom_hline(yintercept = p40, color = Colp3, linetype = "dashed", size=0.8) +
  geom_hline(yintercept = p60, color = Colp4, linetype = "dashed", size=0.8) +
  geom_hline(yintercept = p80, color = Colp5, linetype = "dashed", size=0.8) +
  geom_hline(yintercept = p90, color = Colp6, linetype = "dashed", size=0.8) +
  labs(subtitle = sb,
       y = jY,
       x = jx,
       caption = sb) +
  scale_color_manual(labels = c(lg),
                    values = c("black")) +
  theme(axis.title.x = element_text(size = 22, colour = "black"),
        axis.title.y = element_text(size = 22, colour = "black"),
        axis.text.x = element_text(size = 14, colour = "black"),
        axis.text.y = element_text(size = 16, colour = "black"),
        legend.text = element_text(size = 18),
        legend.title = element_text(size = 20)) +
  labs(color = "")

plot(graf_Lim)
#-----
#CREAR LIBRO EXCEL Y EXPORTAR LOS DATOS GENERADOS EN r
#Insertar directorio de salida y hoja de calculo
wb <- loadWorkbook("E:/TFM/Tomos/AnalsisConglomeEA/HCVmayo2018/ClasificacionHCV_percR.xlsx", create = TRUE)
#crear hoja
createSheet(wb, name="ConUno")
createName(wb, name="ConUno", formula = "ConUno!$A$1")
writeNamedRegion(wb, Con, name="ConUno")
saveWorkbook(wb)

```

```

#-----
----
#INSTALACION LIBRERIAS
install.packages("RODBC")
install.packages("XLConnect")
install.packages("dendextend")
#-----
----
#ACTIVACION LIBRERIAS
#libreria para importacion de libro de excel y coneccion hoja
library(RODBC)
#libreria para conectar, cargar, entre otros. Realizar procesos con excel
library(XLConnect)
#libreria para realizar graficos
library(ggplot2)
library(reshape2)
library(readxl)
library(dplyr)
library(olsrr)
#-----
----
#-----
----
#CONEXION
#Coneccion con el libro de excel
canalhcv <- odbcConnectExcel2007("E:/TFM/Por_estacion/Ro_modificado.xlsx")
#Visualizacion de las hojas dentro del libro de excel
sqlTables(canalhcv)
#Coneccion con la hoja de excel
sqlFetch(canalhcv, "Ro_modificado")
#debido a que no se puede manipular estadisticamente los datos se crea una nueva hoja en R para manipular
Ro <- sqlFetch(canalhcv, "Ro_modificado")
#Se cierra el canal Excel
odbcClose(canalhcv)
attach(Ro)

#-----
----
#Estacion de muestreo
biar<-Ro[Ro$pto_muestreo=="BIAR",]
bunhol<-Ro[Ro$pto_muestreo=="Buñol",]
chelva<-Ro[Ro$pto_muestreo=="Chelva",]
fredes<-Ro[Ro$pto_muestreo=="Fredes",]
gandia<-Ro[Ro$pto_muestreo=="Gandia",]
montanejos<-Ro[Ro$pto_muestreo=="montanejos",]

#Variables metereologicas
#precipitacion = P3 + P7 + P15 + P30 + P60
#temperatura = `Tmax3 ` + Tmax7 + Tmax15 + Tmax30 + Tmedia3 + Tmedia7 + Tmedia15
#Humedad relativa = HRmax3 + HRmax7 + HRmax15 + HRmin3 + HRmin7 + HRmin15
#Viento = Vmax3 + Vmax7 + Vmax15 + lev3 + lev7 + lev15
#Punto de rocío = pto_r3 + pto_r7 + pto_r15 + pto_r30

##Introduccion punto de muestreo
VarMet <- biar

#SELECCION DEL MODELO DE REGRESION -----
model1.1 <- lm(Ro ~ P3 + P7 + P15 + P30 + P60, VarMet)
modelo_meteo= ols_all_subset(model1.1)
#plot(modelo_meteo)

#MEJOR MODELO DE REGRESION -----
##Precipitacion
mejor_modelo_meteo <- ols_best_subset(model1.1)
#plot(modelo_modelo_meteo)

#REGRESION HACIA ADELANTE PASO A PASO (RHAPP) -----
##Precipitacion (salida detallada)
model2.1 <- lm(Ro ~ P3 + P7 + P15, VarMet)
RHAPP_meteo = ols_step_forward(model2.1, details = TRUE)
#plot(RHAPP_meteo)

##Temperatura (salida detallada)
model2.2 <- lm(Ro ~ Tmax15 + Tmax30 + Tmedia15, VarMet)
RHAPP_temperatura= ols_step_forward(model2.2, details = TRUE)
#plot(RHAPP_temperatura)

#REGRESION HACIA ADELANTE PASO A PASO FINAL (RHAPP) -----
model3 <- lm(Ro ~ P15 + P3 + P7 + Tmedia15 + Tmax15 + Tmax30 + HRmax15 + HRmin15 + lev7 +
            Vmax15 + Vmax3, VarMet)
RHAPP_modelo_final = ols_step_forward(model3, details = TRUE)
#plot(RHAPP_modelo_final)

```

```

#Instalar librerias
install.packages("rgdal")
install.packages("maptools")
install.packages("shapefiles")
install.packages("rgl")
install.packages("sp")
install.packages("geoR")
install.packages("raster")
install.packages("gstat")

#Cargar librerias
library(rgdal)
library(maptools)
library(shapefiles)
library(rgl)
library(sp)
library(geoR)
library(raster)
library(gstat)
#library(help="gstat")
library(lattice)
library(spatial)

#El paquete gstat se tiene que cargar con la opcion Install Packages de la ventana de la derecha#
# Leer capa shape de puntos que contiene los datos de los puntos de observacion.
# El shape de puntos se lee con la funcion 'readShapePoints' perteneciente
# a la libreria 'sp', y devuelve un objeto tipo SpatialPoints de R, mas concretamente
# de la libreria 'sp' de R. Un objeto SpatialPoints es simplemente una capa espacial
# de puntos en R.

#Lectura shapefile
datos_hcv <- readShapePoints("E:/TFM/Tomos/shp/HCV_mayo.shp")
shp <- datos_hcv
hcv <- shp$hcv
View(shp@data)

#-----
#SEMIVARIOGRAMAS EXPERIMENTAL
#datos de humedad del combustible vivo
#-----
#Nube del semivariograma: Para examinar si existen valores atipicos#
vario0<-gstat::variogram(shp@data$hcv~1,shp,cloud=TRUE)
plot(vario0,main="Nube del semivariograma",col="black")
#-----
#variograma omnidireccional
gstat::variogram(shp@data$hcv~1, shp)
vario1<-gstat::variogram(shp@data$hcv~1, shp)
#width=3500,cutoff=55000
plot(vario1,main="Semivariograma omnidireccional",plot.numbers=FALSE,col="black")
#-----
# variograma en diferentes direcciones dos horizontales y dos verticales (0,45,90 y 135)
vario2.1<-gstat::variogram(shp@data$hcv~1, shp,alpha=c(0,45,90,135))
vario2.2<-gstat::variogram(shp@data$hcv~1, shp,alpha=c(0,20,40,60,80,120,150,180))
plot(vario2.1,col="black")
plot(vario2.2,col="black")

#Modelos de variograma, direccion=0 grados (apunta al norte)#
vario5<-gstat::variogram(shp@data$hcv~1, shp,alpha=0)
plot(vario5)
#Dibuja el numero de pares de puntos usados en cada punto#
plot(vario5, plot.numbers = TRUE, pch = "+")

# variograma en diferentes direcciones
v = gstat::variogram(shp@data$hcv~1, shp,alpha=c(0,90))
plot(v, group.id = FALSE) # en la misma grafica
plot(v, group.id = TRUE) # en distintos paneles
plot(v, multipanel=FALSE) # en distintos colores

v = gstat::variogram(shp@data$hcv~1, shp,alpha=c(45,135))
plot(v, group.id = FALSE) # en la misma grafica
plot(v, group.id = TRUE) # en distintos paneles
plot(v, multipanel=FALSE) # en distintos colores

#-----
#SEMIVARIOGRAMAS TEORICO
#datos de humedad del combustible vivo
#-----
#Calculamos la varianza de los datos
var(shp@data$hcv)

#modelo esferico
model.1<-fit.variogram(vario1,vgm(1197,"Sph",10000,nugget=1))
model.1
plot(vario1, model=model.1,main="Modelo esférico",col="black")

```

```

#modelo exponencial
model.1<-fit.variogram(variol,vgm(1197,"Exp",10000,nugget=0))
model.1
plot(variol, model=model.1,main="Modelo exponencial",col="black")

#modelo gaussiano
model.1<-fit.variogram(variol,vgm(1197,"Gau",10000,nugget=0))
model.1
plot(variol, model=model.1,main="Modelo gaussiano",col="black")
#-----
#-----
#Crear una malla regular para impresion de resultados#
s<-cbind(shp@data$X,shp@data$Y)
s

#Create grid of prediction points:
sp1<-seq(min(s[,1]),max(s[,1]),length=51)
sp2<-seq(min(s[,2]),max(s[,2]),length=51)
sp<-expand.grid(sp1,sp2)
plot(sp)

shp.grid<-sp
names(shp.grid)<-c("X","Y")
shp.grid[1,]

#Si no le decimos cuales son las coordenadas no funciona como objeto#
coordinates(shp.grid) = ~X+Y
class(shp.grid)
gridded(shp.grid) = TRUE
class(shp.grid)
#-----
#krigeado ordinario
#Como no hemos indicado ningun modelo de semivariograma usa el metodo de la inversa de la distancia#
hcv.ok<-krige(hcv~1,shp,shp.grid)
spplot(hcv.ok["var1.pred"], main = "Metodo krigeado ordinario")

#Con modelo de semivariograma
hcv.ok<-krige(hcv~1,shp,shp,model=model.1)
hcv.ok
spplot(hcv.ok[1], main = "HCV: estimacion con el krigeado ordinario")
#Dibujamos la varianza#
spplot(hcv.ok["var1.var"], main = "HCV: varianza del krigeado ordinario")
#Podemos calcular y dibujar el error de prediccion#
hcv.ok$se2 = sqrt(hcv.ok$var1.var)
spplot(hcv.ok["se2"], main = "error en el krigeado ordinario")
View(hcv.ok@data)
#-----
#Otros graficos#
g <- gstat(formula=shp@data$hcv~1,data=shp,model=model.1)
x <- predict(g, shp.grid)
image(x, 1, main="Estimacion krigeado con datos")
points(shp@data$X, shp@data$Y,bg="black",pch=21)
image(x, 2, main="varianza krigeado con datos")
points(shp@data$X, shp@data$Y,pch = "+")
#help(image)
#-----
#La estimacion por krigeado en las coordenadas de los datos del problema tiene un error casi nulo#
x1 <- predict(g, newdata = shp)
x1
View(x1@data)
x1$x1.res <- shp$hcv - x1$var1.pred
quantile(x1$x1.res,seq(0,1,0.25))
#-----
#Validacion cruzada, vecindario de 40 puntos#
#Quitando solo el punto a estimar#
x<-krige.cv(hcv~1,shp,model=model.1,nmax=40)
x[1,]
#-----
#Pintamos la variable "residual" de objeto x#
bubble(x, "residual", main = "Validacion cruzada para hcv: residuos")
#-----
#validacion cruzada:
#Validacion cruzada hace una partición del conjunto de datos en N partes.
#Para toda observación en una parte,las predicciones se basan en las partes N-1 restantes;
#esto se repite para cada una de las N partes.#
x<-krige.cv(hcv~1,shp,model=model.1,nmax=40,nfold=5)
x[1,]
bubble(x, "residual", main = "Residuos para factorR con 5 particiones")
#-----
#Especificamos el fold de cada dato#
length(hcv)
c(rep(1,25), rep(2,26))
x<-krige.cv(hcv~1,shp,model=model.1,nmax=40,nfold = c(rep(1,25), rep(2,26)))
summary(x)

```

```
bubble(x, "residual", main = "hcv: residuos con dos particiones")
#-----
#Comparamos con la eleccion de las particiones aleatoria#
x<-krige.cv(hcv~1,shp,model=model.1,nmax=40,nfold=2)
bubble(x, "residual", main = "Residuos para factorR con 2 particiones")
#-----
# Error medio: tiene que estar cerca de 0
mean(x$residual)
# media cuadrática normalizada, tiene que estar cerca de 1
mean(x$zscore^2)
```