

Universidad
Politécnica
de Cartagena



industriales
etsii UPCT

**CARACTERIZACIÓN DEL ORIGEN DE LOS
EVENTOS EXTREMOS DE CONTAMINACIÓN
ATMOSFÉRICA SOBRE LA PENÍNSULA
IBÉRICA MEDIANTE ANÁLISIS
MULTIVARIANTE DE UNA SIMULACION DE
CALIDAD DE AIRE-CLIMA**

Máster: Ingeniería Ambiental y de Procesos Químicos y Biotecnológicos

Especialidad: Profesional

Alumno/a: Rocío Baró Esteban

Director/a/s: José Alberto Egea Larrosa
Pedro Jiménez Guerrero

Cartagena, 25 de julio de 2013

ÍNDICE GENERAL

ÍNDICE	II
LISTA DE FIGURAS	III
LISTA DE TABLAS	IV
RESUMEN	V
ABSTRACT	VI
PUBLICACIONES	VII

1. INTRODUCCIÓN	1
1.1. MOTIVACIÓN	1
1.2. NORMATIVA RELATIVA A LA CALIDAD DEL AIRE Y CONTAMINANTES OBJETO DE ESTUDIO	4
1.3. LA IMPORTANCIA DE LA METEOROLOGÍA DE LA CONTAMINACIÓN ATMOSFÉRICA.	5
1.4. ZONA DE ESTUDIO	6
1.4.1. Orografía	7
1.4.2. Situaciones atmosféricas típicas que afectan la PI	9
1.5. OBJETIVOS	10
2. METODOLOGÍA	11
2.1. DATOS - SISTEMA DE MODELIZACIÓN	11
2.2. DEFINICIÓN DE LOS EVENTOS EXTREMOS- CASO PROBLEMA	13
2.3. ANÁLISIS ESTADÍSTICO	15
2.3.1. Análisis de Componentes Principales	15
2.3.2. Selección del número de componentes	16
2.3.2.1. Matriz de datos ozono	17
2.3.2.2. Matriz de datos PM10	18
2.3.3. Análisis Clúster	18
3. RESULTADOS	22
3.1. OZONO	22
3.1.1. PCs y EOFs	22
3.1.1.1. Primera componente ozono	24
3.1.1.2. Segunda componente ozono	26
3.1.1.3. Tercera componente ozono	28
3.1.1.4. Cuarta componente ozono	30
3.1.2. Análisis clúster para episodios extremos de ozono	32
3.2. PM10	35
3.1.3. PCs y EOFs	35
3.1.3.1. Primera componente PM10	37
3.1.3.2. Segunda componente PM10	39
3.1.3.3. Tercera componente PM10	40
3.1.3.4. Cuarta componente PM10	42
3.1.3.5. Quinta componente PM10	44
3.1.3.6. Sexta componente PM10	46
3.1.4. Análisis clúster para episodios extremos de PM10	49
4. CONCLUSIONES	52
5. BIBLIOGRAFÍA	54

Anexos	57
---------------	-----------

LISTA DE FIGURAS

Figura 1.1. Localización de la Península ibérica	7
Figura 1.2. Orografía de la Península Ibérica	8
Figura 2.1. Configuración espacial del sistema SINQLAIR	12
Figura 2.2. Proceso de obtención de los eventos extremos	14
Figura 2.3. Ejemplo de dendrograma	20
Figura 3.1. Gráfico de sedimentación ozono	22
Figura 3.2. Representación espacial EOF1 ozono	25
Figura 3.3. Tendencia temporal PC1 ozono	26
Figura 3.4. Representación espacial EOF2 ozono	27
Figura 3.5. Tendencia temporal PC2 ozono	28
Figura 3.6. Representación espacial EOF3 ozono	29
Figura 3.7. Representación temporal PC3 ozono	30
Figura 3.8. Representación espacial EOF4 ozono	31
Figura 3.9. Tendencia temporal PC4 ozono	32
Figura 3.10. Clúster de ozono	33
Figura 3.11. Representación espacial Clúster ozono	34
Figura 3.12. Comparativa Clúster EOF1 ozono	35
Figura 3.13. Gráfico de sedimentación PM10	36
Figura 3.14. Representación espacial EOF1 PM10	37
Figura 3.15. Tendencia temporal PC1 PM10	38
Figura 3.16. Representación espacial EOF2 PM10	39
Figura 3.17. Tendencia temporal PC2 PM10	40
Figura 3.18. Representación espacial EOF3 PM10	41
Figura 3.19. Tendencial temporal PC3 PM10	42
Figura 3.20. Representación espacial EOF4 PM10	43
Figura 3.21. Tendencia temporal EOF4 PM10	44
Figura 3.22. Representación espacial EOF5 PM10	45
Figura 3.23. Tendencia temporal PC5 PM10	46
Figura 3.24. Representación espacial EOF6 PM10	47
Figura 3.25. Tendencia temporal PC6 PM10	48
Figura 3.26. Clúster de PM10	49
Figura 3.27. Representación espacial Clúster PM10	50
Figura 3.28. Comparativa Clúster y EOF1 PM10	51

LISTA DE TABLAS

Tabla 3.1. Variabilidad explicada PCs ozono	23
Tabla 3.2. Variabilidad explicada PCs PM10	36

RESUMEN

Caracterización del origen de los eventos extremos de contaminación atmosférica sobre la Península ibérica mediante análisis multivariante de una simulación de calidad de aire-clima

Un gran número de estudios muestra que la Península Ibérica excede algunos de los umbrales de calidad de aire establecidos en la legislación. Los modelos de transporte químico juegan un papel crucial en el pronóstico de los umbrales de las excedencias para la salud humana así como para entender las causas de esos eventos extremos. En este sentido, se ha llevado a cabo la definición de eventos extremos de contaminación atmosférica mediante un proceso de regionalización aplicada a una simulación climatológica de eventos de contaminación atmosférica en la Península Ibérica. Se ha utilizado la técnica estadística del análisis de componentes principales (ACP), que nos dará información sobre el comportamiento temporal de los eventos extremos, junto con las funciones ortogonales empíricas (EOFs), que nos informarán del comportamiento espacial. Para completar el estudio, también se ha realizado un análisis clúster con el fin de valorar la consistencia de los resultados obtenidos tras el ACP. Se ha utilizado datos del sistema de modelización regional MM5-CHIMERE-EMEP (conducido por el reanálisis ERA40) para el periodo 1971-2000. Los contaminantes estudiados han sido ozono y material particulado con diámetro inferior a 10 micras (PM10). Los resultados más destacables vienen dados por la primera componente de ambos contaminantes la cual explica casi el 50% de la variabilidad. Con respecto al PM10, éste muestra un patrón espacial homogéneo en toda la península, el cual puede estar relacionado con episodios de resuspensión, aunque no ha mostrado ninguna tendencia significativa de incremento o descenso de los niveles. Sin embargo, el ozono muestra un gradiente norte-sur con un incremento en los niveles de en el sur de la Península. Los resultados del análisis clúster son muy similares a los obtenidos tras el ACP, por lo que se ha encontrado consistencia en los resultados. En conclusión, la identificación de los patrones espacio-temporales para los principales contaminantes responsables de eventos extremos puede ayudar a una mejor definición de los planes de control y mitigación de la contaminación atmosférica sobre la Península Ibérica.

ABSTRACT

Characterizing the origin of extreme air pollution events over the Iberian Peninsula by clustering air quality-climate simulations

A wide number of studies show that the Iberian Peninsula (IP) exceeds some of the thresholds of air quality established in the legislation. Chemistry transport models play a key role in forecasting the threshold exceedances for human health and ecosystems, and to understand the causes of these extreme air pollution events. In this context, a definition of extreme air pollution events based on a regionalization process has been carried out, applied to a model climatology of air pollution over the Iberian Peninsula. The pollutants studied have been ozone and particulate matter (PM₁₀). In order to identify locations with similar patterns in terms of the studied pollutants, a principal component analysis was carried out, which will give us temporal behaviour information together with the Empirical Orthogonal Functions (EOFs), which will give us the spatial behaviour. To complete the study, a cluster analysis was also performed in order to assess the consistency of our findings. Data from the regional modeling system MM5-CHIMERE-EMEP (driven by ERA40 reanalysis) for the period 1971-2000 is used in this study. The most remarkable results are given by the EOF1 for both pollutants as it accounts almost the 50% of the total variance of the contamination of the exceedances. According to the PM₁₀, it shows an homogenous spatial pattern, it may be related to dust resuspension episodes. Nevertheless, the ozone shows a north-south gradient with an increase of the levels in southern IP. The cluster analysis results are very similar to the principal component results, so there is consistency in the findings. In conclusion, the spatial patterns of simulated air pollution episodes (30 years) over the Iberian Peninsula have been identified, together with an analysis of the temporal trends for O₃ and PM₁₀. In conclusion, the identification of the spatio-temporal patterns of the air pollution episodes can lead to a better definition of the environmental mitigation plans of the air quality over the Iberian Peninsula.

PUBLICACIONES

Este trabajo ha sido presentado como exposición oral en el congreso de la Unión Europea de Geociencia (European Geoscience Union, EGU) el pasado 11 de abril de 2013 en Viena (Austria), y llevó a la siguiente publicación:

Baró, R., Egea, J.A, Lorente-Plazas, R., and Jiménez-Guerrero, P. Characterizing the origin of extreme air pollution events over the Iberian Peninsula by clustering air quality-climate simulations. European Geosciences Union General Assembly Vienna, Austria 04-2013. 15-726.

1. INTRODUCCIÓN

1.1. MOTIVACIÓN

El impacto de la contaminación atmosférica es un tema de vital importancia en el clima y el medio ambiente, el cual continúa suponiendo una traba para la salud mundial. En relación a una evaluación realizada por la Organización Mundial de la Salud (WHO, 2005) sobre los efectos que la contaminación atmosférica tienen sobre la salud humana, se estima que más de dos millones de muertes prematuras anuales pueden ser atribuidas a los efectos de una pobre calidad del aire. Según un reciente estudio de la Agencia Europea del Medio Ambiente (EEA, 2012), la contaminación atmosférica es el factor ambiental con mayor impacto en la salud en Europa y es responsable del mayor número de enfermedades relacionadas con el medio ambiente. A pesar del éxito de la legislación y del éxito en el tratamiento de las emisiones, la población europea está aún respirando aire cargado de sustancias peligrosas. Entender el estado y la influencia de las tendencias de la calidad del aire es un paso crítico en la definición de estrategias de control y mitigación de este problema.

A pesar de las mejoras debido a la legislación europea, el material particulado y el ozono continúan afectando a la salud humana. Además, el pronóstico a corto plazo de la contaminación atmosférica disponible hoy día (generalmente menos de 48 horas) puede obstaculizar la toma de decisiones y el diseño de estrategias de reducción de emisiones para cumplir con los estándares de calidad de aire de la Península Ibérica (PI, de aquí en adelante); de ahí que sea necesario un estudio de las climatologías conducentes a eventos extremos de calidad del aire, cuyo entendimiento nos permitiría una toma de decisiones con un avance mucho mayor que el que los sistemas de pronóstico proporcionan.

Respecto a la PI, un amplio número de estudios (como, por ejemplo, Ziomas et al. 1998; Dueñas et al. 2002; Ribas y Peñuelas, 2004; Palacios et al. 2002;2005, entre otros) muestran que en España se dan numerosas superaciones de los umbrales de calidad del aire establecidos por la legislación, especialmente para ozono troposférico y material particulado y que esta situación se agravará especialmente en el futuro con los impactos climáticos para la región mediterránea (Jiménez-Guerrero et al., 2011;2012).

Según estos estudios realizados para el sur de Europa y la zona mediterránea, los problemas en estas áreas se volverán críticos especialmente en las zonas urbanas para el material particulado de diámetro inferior a 10 micras (PM10) y los óxidos de nitrógeno y en las zonas de fondo a sotavento de las fuentes de emisión para el ozono troposférico.

Adicionalmente, no solo la contaminación de origen antropogénico contribuye de manera decisiva al incumplimiento de las Directivas europeas. El polvo mineral atmosférico proveniente de los desiertos representa una sustancial contribución al contenido global del aerosol troposférico. El Sáhara se revela como la fuente más importante de polvo, siendo África Septentrional es la mayor fuente de polvo mineral en el mundo (Basart et al, 2011; Terradellas et al, 2012). El polvo mineral contribuye de manera importante a los niveles de material particulado suspendido recogidos mediante la red de monitoreo del sur peninsular (Querol et al, 1998; Escudero et al, 2006). Estas últimas contribuyen decisivamente a las superaciones de los valores límite de PM10 de la Directiva de calidad del aire.

El interés de centrar este estudio en la PI radica en el hecho de que esta región se ha identificado como uno de los puntos calientes respecto a la contaminación atmosférica y el cambio climático (Giorgi, 2006), obteniéndose en distintas proyecciones que es muy probable que sea una de las regiones del planeta que sufrirá mayores variaciones en su climatología, con una modificación en los patrones pluviométricos y un aumento de la desertificación en latitudes bajas más intensos observados en las últimas décadas, con sus correspondientes impactos en la calidad del aire. Desde el punto de vista de la contaminación atmosférica, el Mediterráneo occidental y específicamente la Península Ibérica representan unas de las zonas con mayores problemas derivados de una pobre calidad del aire, especialmente asociadas a contaminantes secundarios como el ozono troposférico o los aerosoles secundarios. Este comportamiento se debe a la complejidad de la zona, donde se presenta una amplia variedad de usos del suelo, compleja orografía, comportamientos diferenciados entre la costa mediterránea y la atlántica etc., (Barros et al., 2003) lo que conduce a una dinámica de transporte y una físico-química de contaminantes muy compleja (Rodríguez et al., 2002):

1. alto índice de insolación en la PI, lo cual favorece los procesos de contaminación fotoquímica y la formación de partículas secundarias respecto al Norte de Europa;
2. la PI y el sur de Europa presentan características meteorológicas que favorecen la acumulación de los contaminantes en el aire. Básicamente estas se traducen en un bajo régimen pluviométrico y en una escasa renovación de las masas de aire, asociada a recirculaciones regionales de las mismas y que son consecuencia de la abrupta orografía que la rodea;
3. importantes aportes de materia mineral en el sur de Europa, los cuales pueden tener lugar mediante procesos de resuspensión local sobre suelos con escasa cobertura vegetal o de transporte a larga distancia desde el norte de África. Las peculiaridades del sur de Europa condicionan los parámetros con los que se debe controlar la contaminación atmosférica por material particulado. Mientras que en países como Alemania o el Reino Unido el PM10 es representativo de la inmisión asociada a las emisiones antropogénicas, no lo es en países Mediterráneos o del sur de Europa (España, Italia y Grecia) a causa de los aportes minerales. De este modo el PM2.5 se presenta como una mejor alternativa para el sur de Europa. Sin embargo, presenta el inconveniente de que gran parte de los nitratos originados a través de los óxidos de nitrógeno y las emisiones de material particulado primario que tienen lugar en determinados procesos industriales quedan fuera de control.

Toda esta complejidad fuerza a disponer de herramientas de modelización que permitan entender el cómo y el porqué de los impactos climáticos sobre los episodios de contaminación atmosférica; así como proporcionar una herramienta de alta resolución que permita abatir estos contaminantes.

Anteriormente se han mencionado estudios episódicos asociados a contaminación atmosférica en la PI. Sin embargo desde una perspectiva de largo plazo, la literatura científica es escasa en relación a climatología de la modelización de la calidad de aire en esta zona de estudio (Jiménez-Guerrero et al., 2012). El mayor problema de la modelización de la calidad del aire para

un periodo climatológicamente significativo radica en que es computacionalmente muy costoso. Sin embargo, genera una útil e inmensa base de datos para estudios de calidad del aire. Además, tenemos que lidiar con la gran complejidad del problema analizado.

Cabría preguntarse, ¿cómo podemos reducir el problema dimensional existente cuando analizamos señales complejas a largo plazo de calidad de aire? Una buena aproximación sería encontrar patrones comunes de los episodios que superan los límites establecidos por las Directivas Europeas. La forma más común para encontrar estos patrones es mediante el empleo de análisis estadísticos como el análisis de componentes principales (PCA) y también el análisis clúster.

1.2. NORMATIVA RELATIVA A LA CALIDAD DEL AIRE Y CONTAMINANTES OBJETO DE ESTUDIO

En primer lugar, cabría plantearse cuáles son los niveles de contaminación atmosférica que se identifican como potencialmente peligrosos para la salud humana u otros ecosistemas. Dichos niveles son recogidos en diferentes normativas nacionales y Europeas, que pasan a ser brevemente descritas a continuación.

La antigua Ley 38/1972 de Protección Ambiental Atmosférico ha sido sustituida por la Ley 34/2007, de calidad del aire y protección de atmósfera, que aporta una nueva base legal para los desarrollos relacionados con la evaluación y gestión de la calidad del aire en España. Esta Ley, cuyo fin último es alcanzar unos niveles óptimos de calidad del aire para evitar, prevenir o reducir riesgos o efectos negativos sobre la salud humana, el medio ambiente y demás bienes de cualquier naturaleza, habilita al gobierno a definir y establecer los objetivos de calidad del aire y los requisitos mínimos de los sistemas de evaluación de la calidad del aire.

La Directiva 2008/50/CE del Parlamento Europeo y del Consejo, de 21 de junio de 2008, relativa a la calidad del aire ambiente y a una atmósfera más limpia en Europa, supone la revisión, a la luz de los últimos avances científicos y sanitarios, y de la experiencia de los Estados miembros, de la normativa europea. Asimismo, ha venido a modificar el anterior marco regulatorio comunitario, sustituyendo la Directiva Marco y las tres primeras Directivas Hijas, e introduciendo regulaciones para nuevos contaminantes, como las partículas

de tamaño inferior a 2,5 micrómetros, y nuevos requisitos en cuanto a la evaluación y la gestión de la calidad del aire ambiente. Esta Directiva, ha sido transpuesta a nuestro ordenamiento jurídico español mediante el Real Decreto 102/2011, que a su vez pretende desarrollar reglamentariamente la Ley 34/2007 en los temas relativos a calidad del aire y simplificar la normativa nacional en dicha materia.

Debido a la importancia de los contaminantes ozono y material particulado (en concreto PM10), éstos han sido los contaminantes objeto de estudio. Según la Directiva Europea 2008/50/CE se definen como:

Ozono: Contaminante transfronterizo que se forma en la atmósfera a partir de la emisión de contaminantes primarios, regulados por la Directiva 2001/81/CE del Parlamento Europeo y del Consejo, de 23 de octubre de 2001, sobre techos nacionales de emisión de determinados contaminantes atmosféricos.

PM10: Partículas que pasan a través del cabezal de tamaño selectivo definido en el método de referencia para el muestreo y la medición de PM10 de la norma EN 12341, para un diámetro aerodinámico de 10 µm con una eficiencia de corte del 50%.

1.3. LA IMPORTANCIA DE LA METEOROLOGÍA DE LA CONTAMINACIÓN ATMOSFÉRICA.

La meteorología, junto con las emisiones y la química atmosférica, es uno de los vectores contribuyentes a los episodios de contaminación atmosférica. Como tal, la definición del ambiente meteorológico ha sido siempre uno de los requisitos principales para la comprensión de la calidad del aire (Seaman, 2003).

Desde que la atmósfera sirve como medio en el cual se liberan los contaminantes atmosféricos, el transporte y la dispersión de estos están significativamente influenciados por parámetros meteorológicos. Para realizar actividades relativas a la planificación de la calidad del aire es imprescindible comprender la meteorología y la climatología de la contaminación atmosférica y su influencia en la dispersión de los contaminantes. Los gestores de la calidad del aire usan este conocimiento con el fin de localizar las estaciones

de monitorización de la contaminación del aire y para desarrollar planes de implementación con el fin de cumplir con los estándares de calidad de aire.

Cuando las condiciones meteorológicas no son propicias para la dispersión de los contaminantes, las agencias gubernamentales de la contaminación atmosférica deben de actuar rápidamente con el fin de asegurar que éstos no se acumulen, y lleguen a niveles inaceptables en el aire que respiramos. Cuando los niveles de contaminación empiezan a ser excesivamente altos, se produce un episodio de contaminación atmosférica.

Alguna de las variables meteorológicas que afectan a la dispersión de los contaminantes, como es el caso de la precipitación, es muy compleja, ya que su descripción incluye un amplio rango de escalas temporales y espaciales. Esto es más notorio en el caso de la PI debido a su relieve orográfico afilado, que intensifica algunos mecanismos de generación de precipitaciones (Serrano et al., 1998).

1.4. ZONA DE ESTUDIO

La zona de estudio que vamos a considerar en este trabajo para definir la climatología de la contaminación atmosférica es la PI (figura 1.1). Ésta limita al norte y oeste con el océano Atlántico y al este con el mar Mediterráneo. Esta característica, junto con su compleja orografía y localización (latitudes medias, ligeramente al sur de la circulación general de las depresiones atlánticas) producen una variedad de condiciones meteorológicas.



Figura 1.1 Localización de la Península ibérica

1.4.1. Orografía

La altitud media de la Península, 650 m, es en gran medida superior a la del conjunto europeo (sólo superada por Suiza). Un 57,7% del territorio español rebasa los 600 m de altitud y cerca de una quinta parte supera los 1000 m (Martín-Vide y Olcina, 2001). Estos valores son claramente descriptivos de la compleja orografía que uno se puede encontrar por la Península (figura 1.2).

La principal estructura que conforma la PI es la Meseta. Ésta se divide en la Meseta (o submeseta) norte y la Meseta sur. El Sistema Central es el accidente geográfico que las divide, orientado de oeste a este en la mitad peninsular. El sector septentrional de la Meseta, más elevado que el meridional, está ocupado por la extensa cuenca del río Duero. Al sur del Sistema Central, la alineación paralela de las Montañas de Toledo separa las cuencas del río Tajo y el Guadiana. El entorno de la Meseta se encuentra limitado por grandes sistemas orográficos que la rodean, a excepción del límite con Portugal, donde la orografía se suaviza sensiblemente (Jorba, 2005).



Figura 1.2 Orografía de la Península Ibérica

Las dos depresiones más importantes de la Península son la formada por el valle del río Ebro, y la del río Guadalquivir. La primera queda aislada de la costa mediterránea por las cordilleras litorales de Cataluña, en cambio, la del Guadalquivir se abre directamente hacia el océano Atlántico. Los Pirineos y el Sistema Ibérico delimitan la depresión del Ebro formando una canalización de masas de aire de gran magnitud. Análogamente, para el Guadalquivir, la Sierra Morena y el Sistema Penibético y Subbético conforman el amplio valle meridional de la Península. Los valles de los ríos Duero, Tago y Guadiana forman depresiones significativas, aunque de menor envergadura que las dos anteriores.

Las dos grandes cordilleras Peninsulares, exteriores a la Meseta son al noreste, los Pirineos, que se elevan en el istmo que une la Península y al sur el Sistema Bético, con una disposición paralela a la costa. Éste último se divide en el Sistema Subbético y el Sistema Penibético. En el litoral del levante peninsular, desde el sur, se alcanzan las cordilleras del Sistema Penibético, el Sistema Subbético, más hacia el norte el límite sureste del Sistema Ibérico, y las cordilleras litorales y prelitorales de Cataluña. Esta

complejidad no es tan acusada en la costa Atlántica, con cordilleras de menor altitud, exceptuando la región de Galicia. Esta complejidad orográfica pone de manifiesto la elevada variabilidad climática que se puede esperar en la Península y en los flujos dominantes en la misma (Jorba, 2005).

1.4.2. Situaciones atmosféricas típicas que afectan a la pi

El viento es una de las variables meteorológicas que más influyen en la dispersión de los contaminantes, sobre todo en el caso del material particulado (Lesniok et al., 2010). El comportamiento de éste está fuertemente asociado con la variabilidad del clima regional. En el caso de la PI, la circulación superficial está controlada por las semi-permanentes altas presiones sub-tropicales localizadas sobre las islas Azores. En invierno, normalmente se centra en latitudes más bajas, por tanto, la PI se ve afectada por la circulación zonal desde el oeste junto con perturbaciones originadas por el frente polar. En verano, el centro de alta presión subtropical se expande hacia latitudes más altas, bloqueando la circulación del oeste sobre la península excepto en regiones del norte. Además, desplazamientos relativos y cambios en la intensidad de esta alta presión subtropical semipermanente permiten una variedad de masas de aire con distintas propiedades físicas que afectan a la circulación sobre la PI a escalas locales y regionales. Las masas de aires polares procedentes de Siberia, continentales tropicales, del Sahara, marítimas polares y las masas de aire árticas llegan a la PI por el norte y noroeste y las masas de aire marítimas tropicales y subtropicales llegan desde el oeste y el suroeste son algunos de los conocidos flujos sinópticos que afectan a la PI. Las diferentes propiedades del aire advectivo pueden introducir polvo debido a flujos del sur procedentes del desierto del Sahara. Además una masa de aire que permanezca lo suficiente sobre el mar Mediterráneo puede aumentar su contenido de agua y moverse hacia la PI penetrando desde el este o sureste. Estas masas de aire del Mediterráneo se ha demostrado que prevalecen durante el verano. La interacción de esta gran variedad de masas de aire que coinciden en la PI y una topografía particularmente compleja producen un amplio rango de climas regionales.

1.5. OBJETIVOS

Considerando lo anterior, el objetivo principal de este trabajo consiste en caracterizar los patrones espacio-temporales comunes de episodios extremos de contaminación (centrándonos en ozono y material particulado, PM10) definidos como aquellos que exceden los límites de calidad de aire establecidos en la legislación anteriormente mencionada en el apartado 1,2, de una climatología de calidad de aire (periodo de estudio 1971-2000) sobre la Península Ibérica. Posteriormente con un análisis clúster se pretende agrupar esos episodios extremos con el fin de resumirlos por áreas geográficas y evolución temporal.

Como objetivos secundarios o derivados, se pretende (1) reducir el problema de dimensionalidad al que nos enfrentamos cuando tratamos con bases de datos climatológicas; (2) buscar aquellos patrones que agrupan el inmenso número de episodios modelizados para la Península Ibérica en el periodo de 30 años de estudio hasta conseguir un número manejable de casos de estudio típicos; (3) testar la aplicación de una meteorología novedosa de clústering en la PI, no aplicada hasta la fecha para el estudio de climatologías de contaminación atmosférica.

Por lo tanto, el objetivo de estudios como el nuestro es obtener un número reducido de grupos de patrones espacio-temporales con una representación importante de la variabilidad observada.

2. METODOLOGÍA

2.1. DATOS - SISTEMA DE MODELIZACIÓN

En este trabajo, se han utilizado los datos de salida de una simulación de contaminación atmosférica para la PI, con una cobertura temporal de 30 años (1971-2000). Dicha simulación ha sido realizada dentro con el sistema de modelización MM5-EMEP-CHIMERE bajo el marco del proyecto SInQIAir..

Dicha base de datos, relativa a los contaminantes ozono troposférico y PM10, ha sido suministrada por el Grupo de Modelización Atmosférica Regional (MAR) de la Universidad de Murcia, grupo de investigación con el que se ha colaborado para la realización de este trabajo. Las simulaciones cubren dos dominios anidados: Europa con una resolución de 90 km y un dominio interior para la PI (utilizado en este trabajo, figura 2.1) con una resolución de 30 km, representativa de los episodios de contaminación a escala regional sobre la zona de estudio. El sistema de modelización se encuentra integrado por tres modelos:

- Modelo meteorológico MM5 (Grell et al., 1994). El modelo mesoescalar MM5 es un modelo de área limitada (Pielke and Segal, 1986) no hidrostático, con coordenadas sigma (que siguen el terreno) especialmente diseñado para reproducir la circulación atmosférica meso escalar (<http://www.mmm.ucar.edu/mm5>). MM5 puede estudiar los fenómenos atmosféricos bajo distintas escalas espaciales y diseñar predicciones de muy alta resolución. También incluye los esquemas más modernos y realistas de parametrización de los procesos físicos (Stensrud, 2007) relacionados con la radiación atmosférica, microfísica de nubes y precipitación, convección por cúmulos, turbulencia, y flujos de energía y momento sobre la superficie terrestre.
- Modelo de emisiones. Se han utilizado las emisiones anuales EMEP correspondientes a cada año simulado, disponibles públicamente (<http://www.ceip.at/emission-data-webdab>). EMEP es el “Co-operative programme for monitoring and evaluation of long range transmission of air pollutants in Europe”. Estas emisiones se basan en la información enviada a EMEP por cada país miembro.

- Modelo de transporte químico CHIMERE (Bessagnet et al., 2004). Es un modelo de código abierto diseñado principalmente para proporcionar predicciones diarias de ozono, aerosoles y otros contaminantes y para realizar simulaciones a largo plazo para diversos escenarios de emisiones. CHIMERE es un modelo multiescala, pudiendo realizar integraciones de calidad del aire desde la mesoescala-escala regional (varios cientos o miles de kilómetros) hasta escalas urbanas con resoluciones espaciales entre 100 y 1-2 km.

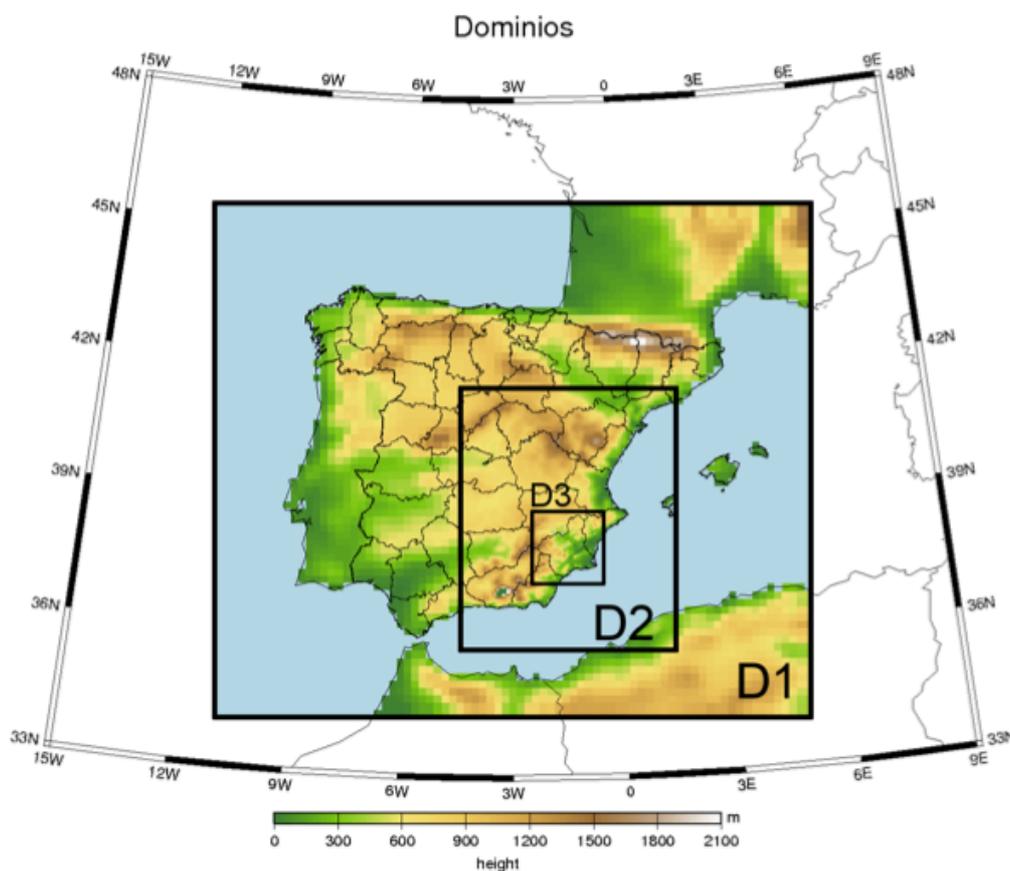


Figura 2.1. Configuración espacial del sistema SINQLAIR

2.2. DEFINICIÓN DE LOS EVENTOS EXTREMOS- CASO PROBLEMA

Como hemos comentado anteriormente, se han utilizado los datos de salida una simulación de climatología de calidad de aire para el periodo comprendido entre 1971-2000.

Para la obtención de los eventos extremos de los contaminantes estudiados (ozono y material particulado, PM10) se han definido como aquellos que, en cualquier punto de la malla de la PI se superen los límites establecidos por la *Directiva 2008/50/CE*. Se ha utilizado como referencia el valor objetivo para la protección de la salud humana, que en el caso del ozono se corresponde con el límite octohorario de **120 $\mu\text{g m}^{-3}$** y para el PM10, el límite medio diario de **50 $\mu\text{g m}^{-3}$** . Sin embargo, tras analizar estos eventos extremos obtenemos una gran cantidad de ellos, 782 para el ozono y 4378 en relación al PM10 (figura 2.2).

Esto supone un inconveniente ya que, no podemos analizar evento por evento con el fin de caracterizar o identificar los episodios de contaminación atmosférica de la Península Ibérica. Si queremos encontrar un patrón común, debemos emplear técnicas estadísticas complejas.

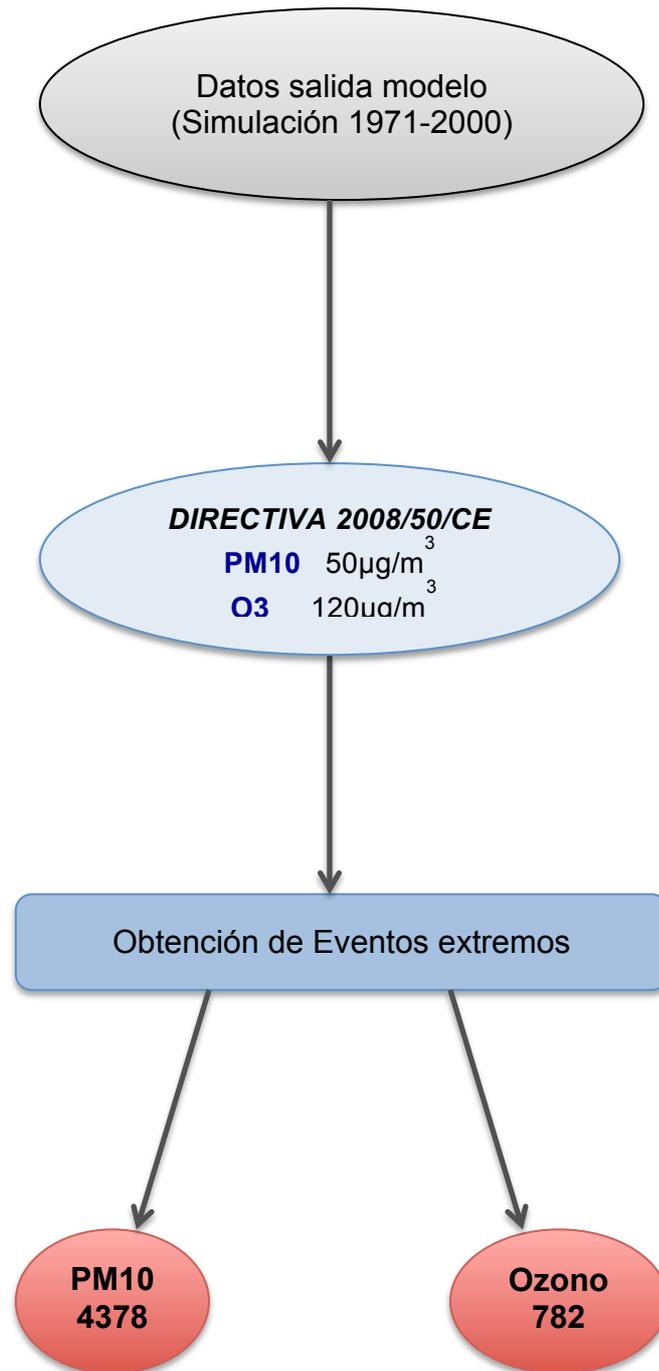


Figura 2.2. Proceso de obtención de los eventos extremos

2.3. ANÁLISIS ESTADÍSTICO

Con el fin de encontrar esos patrones comunes de contaminación atmosférica, es necesario el empleo de técnicas estadísticas complejas como son el análisis de componentes principales (en inglés Principal Component Analysis, PCA) y el análisis clúster. Este último se ha llevado a cabo con el fin de evaluar la consistencia de los resultados obtenidos a partir del análisis de componentes principales. Se ha utilizado el software R para el cálculo de dichos análisis y en el anexo A se muestra el código que se ha ejecutado.

2.3.1. Análisis de Componentes Principales

El análisis de componente principales es posiblemente la técnica estadística multivariante más empleada en las ciencias de la atmósfera. Podemos encontrarla en trabajos como Gómez-Navarro et al. (2010) o Jiménez-Guerrero et al. (2011). Esta técnica se hizo popular gracias al análisis de datos atmosféricos del trabajo de Lorenz (1956), quien denominó a la técnica como análisis de las funciones ortogonales empíricas (en inglés Empirical Orthogonal Function, EOF). Ambos nombres (PCA y EOF) se usan comúnmente, y se refieren al mismo conjunto de procedimientos. A veces, este análisis se refiere incorrectamente como un análisis factorial, que está relacionado pero supone un análisis estadístico multivariante diferente.

La idea central del PCA es reducir la dimensionalidad de un conjunto de datos que contienen un gran número de variables interrelacionadas, mientras se retiene la máxima variabilidad posible de las componentes principales (PCs). Éstas no están correladas y se ordenarán de tal forma que las primeras pocas retengan la mayor parte de la variabilidad presente en todas las variables originales (Jolliffe, 2002).

Su finalidad es la representación de los datos originales en un nuevo espacio ortogonal de funciones que representan los modos principales de variación del sistema. Con ello se consigue eliminar el grado de correlación existente en los datos iniciales y que constituye una fuente de ruido significativa para otro tipo de análisis como por ejemplo el de agrupación. Los modos de variación son las funciones ortogonales empíricas (EOFs) se corresponden con los vectores propios de la matriz de correlación o covarianza de las anomalías de los datos originales. Las proyecciones de los

datos originales en estas nuevas funciones reciben el nombre de componentes principales (PCs), siendo su característica principal la inexistencia de correlación entre ellas (Cattell 1966; Wilks 1995).

Esta metodología reduce la gran complejidad y dimensionalidad del fenómeno estudiado descomponiendo un campo espacio-temporal (2.1) en patrones espaciales (EOFs) y sus índices temporales asociados (PCs):

$$C(t,s) = \sum_{i=1}^N PC_i(t) EOF_i(s) \quad (2.1)$$

Donde $C(t,s)$ se correspondería con la variable contaminación, N es el número de modos contenidos en el campo, $EOF_i(s)$ son un conjunto de funciones definidas para el espacio y $PC_i(t)$ sus coeficientes temporales. La diferencia entre el campo original y el primer número del sumatorio nos da el campo residual (2.2).

$$Res(t,s) = C(t,s) - PC_1(t) EOF_1(s) \quad (2.2)$$

Gracias a este PCA se van a agrupar áreas que normalmente presentan el mismo nivel de contaminante durante los eventos extremos estudiados.

2.3.2. Selección del número de componentes

Hay diferentes criterios para elegir finalmente con cuántas componentes nos vamos a quedar:

1. Criterio de la media aritmética

Se seleccionan las Componentes Principales Z_j que verifican (2.3)

$$\lambda_j > \frac{\lambda_1 + \lambda_2 + \dots + \lambda_p}{p} \quad (2.3)$$

Si se utilizan las variables originales tipificadas, equivale al criterio de Kaiser (2.4)

$$\lambda_j > \frac{\lambda_1 + \lambda_2 + \dots + \lambda_p}{p} = 1 \quad (2.4)$$

2. Criterio de la proporción de variabilidad explicada

Consiste en buscar el número de componentes r tal que (2.5):

$$\text{Proporción de variabilidad explicada} = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_r}{\lambda_1 + \lambda_2 + \dots + \lambda_r + \dots + \lambda_p} > P \quad (2.5)$$

Donde P es un valor positivo cercano a 1, normalmente superior a 0.8 ó 0.9

3. Gráfico de sedimentación o diagrama de codo

Consiste en representar los valores propios respecto al número de componentes: las componentes a retener serán todas aquella previas a la “zona de sedimentación”, es decir, la zona donde el gráfico presenta una tendencia horizontal.

2.3.2.1. Matriz de datos ozono

La malla de la PI correspondiente a la simulación empleada consistía en 3417 puntos de malla. Como en el caso del ozono obtuvimos 782 episodios de eventos extremos (días en los que se produjo una superación de los niveles de ozono en cualquier punto de la malla de la PI), hemos trabajado con una matriz de 3417 filas por 782 columnas. De esos 3417 puntos de malla, 1822 eran valores perdidos, es decir, no había datos de contaminación, ya que esa malla también contenía el mar, donde no se registran esos datos. Por lo tanto, esos puntos de grid no se tuvieron en cuenta a la hora de realizar el PCA, por lo que finalmente trabajamos con una matriz de 1595 filas por 782 columnas.

2.3.2.2. Matriz de datos PM10

Al igual que ocurría con el ozono, vamos a tener 3417 puntos de malla que se reducen a 1595. Con respecto al número de eventos, el PM10 presenta 4378 días con superación del umbral del PM10 en algún punto de la malla de la PI. Por lo tanto nuestra matriz de PM10 sería 1595 filas por 4378 columnas. En este caso el número de columnas es superior al número de filas, y al realizar el PCA, no obtendríamos tantas PCs como variables (eventos de contaminación). Por lo tanto antes de llevar a cabo el PCA, previamente realizamos la traspuesta de la matriz, por lo que la nueva matriz era de 4378 filas por 1595 columnas.

2.3.3. Análisis Clúster

El análisis de conglomerados (clúster) tiene por objeto agrupar elementos en grupos homogéneos en función de las similitudes o similitudes entre ellos. El análisis de conglomerados estudia tres tipos de problemas:

Partición de los datos. Disponemos de datos que sospechamos son heterogéneos y se desea dividirlos en un número de grupos prefijado, de manera que:

- (1) cada elemento pertenezca a uno y solo uno de los grupos;
- (2) todo elemento quede clasificado;
- (3) cada grupo sea internamente homogéneo.

Construcción de jerarquías. Deseamos estructurar los elementos de un conjunto de forma jerárquica por su similitud. Una clasificación jerárquica implica que los datos se ordenan en niveles, de manera que los niveles superiores contienen a los inferiores. Estrictamente, estos métodos no definen grupos, sino la estructura de asociación en cadena que pueda existir entre los elementos. Sin embargo, la jerarquía construida permite obtener también una partición de los datos en grupos.

Clasificación de variables. En problemas con muchas variables, es interesante hacer un estudio exploratorio inicial para dividir las variables en grupos. Este estudio puede orientarnos para plantear los modelos formales para reducir la dimensión. Las variables pueden clasificarse en grupos o

estructurarse en una jerarquía.

Los métodos de partición utilizan la matriz de datos, pero los algoritmos jerárquicos utilizan la matriz de distancias o similitudes entre elementos. Para agrupar variables se parte de la matriz de relación entre variables: para variables continuas suele ser la matriz de correlación, y para variables discretas, se construye a partir de la distancia ji-cuadrado (Peña, 2002).

Los métodos de clústering se dividen en jerárquicos y no jerárquicos. Dentro de los métodos no jerárquicos de clústering, el más empleado es el método de las k-medias. La “k” se refiere al número de grupos, y no a la dimensión del vector de datos. Este tipo de métodos tienen como objetivo clasificar las observaciones en un número de grupos prefijado de antemano. Suelen utilizarse cuando hay muchos datos para analizar o bien para refinar una clasificación obtenida mediante un método jerárquico. El método de las k-medias puede resumirse en 4 pasos:

1. Seleccionar k puntos como centros de los grupos iniciales.
2. Asignar cada elemento al grupo de cuyo centro esté más próximo.
3. Definir un criterio de optimalidad y comprobar si reasignando uno a uno cada elemento de un grupo a otro mejora el criterio.
4. Si es posible mejorar moviendo un elemento, hacerlo. Si no es posible mejorar, terminar el proceso.

El criterio de optimalidad que se utiliza en el criterio de las k-medias consiste en minimizar la suma de los cuadrados dentro de los grupos (SCDG), (2.6):

$$SCDG = \sum_{g=1}^k \sum_{i=1}^{n_g} \sum_{j=1}^p (x_{ij:g} - \bar{x}_{j:g})^2 = \sum_{g=1}^k \sum_{i=1}^{n_g} (x_{i:g} - \bar{x}_g)(x_{i:g} - \bar{x}_g) \quad (2.6)$$

donde x_{ijg} es el valor de la variable j en el elemento i del grupo g y \bar{x}_{jg} la media de esta variable en el grupo.

Dentro de los métodos jerárquicos el más empleado es el método de Ward (2.7), que ha sido el utilizado para este estudio. Es un método de clústering muy popular que no funciona con la distancia de las matrices sino que, como método jerárquico, empieza con un único miembro de un grupo, y se van juntando en grupos de dos en cada paso, hasta que todos los datos están

en un único grupo, tras n-1 pasos. El criterio para elegir qué par de grupos se juntan en cada paso es el siguiente: entre todas las formas posibles de juntar en grupos de dos, el par que se junta es aquel que minimiza la suma del cuadrado de las distancias entre los puntos y los centroides de sus respectivos grupos. Esto es, de entre todas las maneras posibles de combinar dos grupos (G)+1 para formar G grupos, esta unión se hace para minimizar la siguiente expresión:

$$W = \sum_{g=1}^G \sum_{i=1}^{n_g} \|x_i - \bar{x}_g\|^2 = \sum_{g=1}^G \sum_{i=1}^{n_g} \sum_{k=1}^K (x_{i,k} - \bar{x}_{g,k})^2 \quad (2.7)$$

Los resultados de un análisis clúster de tipo jerárquico son convencionalmente mostrados mediante el uso de un dendrograma o un diagrama de árbol. Empezando por las “ramas pequeñas” al principio del análisis, cuando cada una de las n observaciones “x” constituyen su propio clúster, un par de “ramas” se juntan en cada paso a medida que los dos clúster más cercanos se unen. Las distancias entre esos clústers antes de ser unidos también se indican en el diagrama (figura 2.3) con la distancia de los puntos de unión desde el estado n-clúster inicial de las ramas (Wilks, 1995).

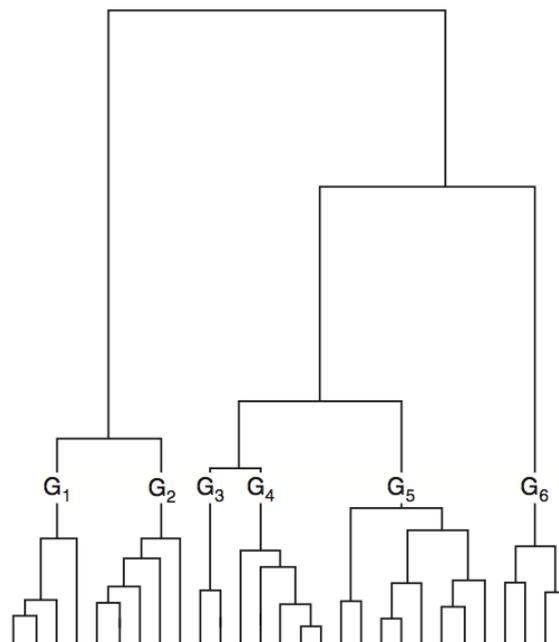


Figura 2.3. Ejemplo de dendrograma

Este análisis clúster ayudará a identificar esos eventos extremos que producen patrones similares de contaminación sobre la PI. Así mismo será útil para agrupar eventos extremos similares y encontrar relaciones con la climatología.

3. RESULTADOS

A continuación se explicarán los resultados más significativos correspondientes a los contaminantes ozono y PM10.

3.1. OZONO

3.1.1. PCs y EOFs

Al realizar el PCA, se obtuvo una componente espacial (1595) que se corresponde con la EOF, y una componente temporal (782) que son las PCs. A la hora de elegir con cuántas componentes nos íbamos a quedar, nos basamos en dos criterios de los comentados con anterioridad: por un lado teniendo en cuenta el diagrama de codos representado en la figura 3.1, y por otro lado la variabilidad explicada representada en la tabla 3.2.

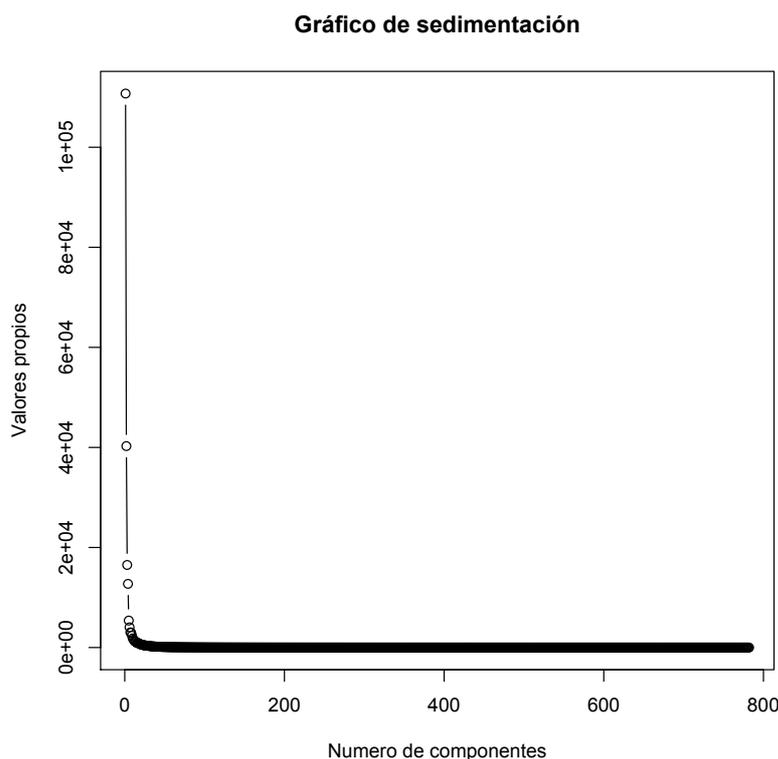


Figura 3.1. Gráfico de sedimentación ozono

A pesar de que se solapan las componentes en el eje x debido a que hay un gran número de éstas (782), podríamos quedarnos a partir de la quinta o sexta componente. Como no queda suficientemente claro, recurrimos al criterio de la variabilidad explicada.

Tabla 3.1. Variabilidad explicada PCs ozono

	PC1	PC2	PC3	PC4	PC5	PC6
Desviación estándar	332.765	200.664	128.486	112.834	73.440	63.570
Proporción de varianza	0.496	0.181	0.074	0.057	0.024	0.018
Proporción acumulada	0.496	0.678	0.753	0.810	0.834	0.852

Como podemos observar en la tabla 3.1, la primera PC nos explica casi el 50% de la variabilidad y elegiríamos 4 componentes ya que acumulan un 81% de la variabilidad explicada.

A continuación se van a exponer las representaciones gráficas de las PCs que son las componentes temporales y sus correspondientes EOFs, componentes espaciales, que representan las zonas de igual comportamiento sobre la PI. A continuación describiremos las 4 componentes que hemos elegido

3.1.1.1. Primera componente ozono

En la figura 3.2 se muestra la representación espacial de la concentración de ozono para la primera componente durante el periodo de estudio (30 años). Esta primera componente explica casi el 50% de la variabilidad.

Para la correcta interpretación de las EOFs hay que tener en cuenta que las áreas que estén coloreadas del mismo color presentan el mismo comportamiento en cuanto a la concentración de ozono. Es decir, en todas las áreas de color rojo el comportamiento ha sido igual y éste comportamiento es inverso al que presentan las áreas azules. Las EOFs sólo nos dan información sobre el comportamiento espacial, por lo que si queremos saber cómo ha sido esa tendencia, es decir, si han aumentado o disminuido los niveles de ozono a lo largo de esos 30 años, tenemos que fijarnos en la tendencia temporal (representada en la figura 3.3), obtenida a partir de las PCs. Para ver si la tendencia temporal era significativa, se obtuvo el modelo lineal ajustado por mínimos cuadrados. Esta interpretación se aplica a las restantes EOFs que se expliquen tanto para ozono como PM10.

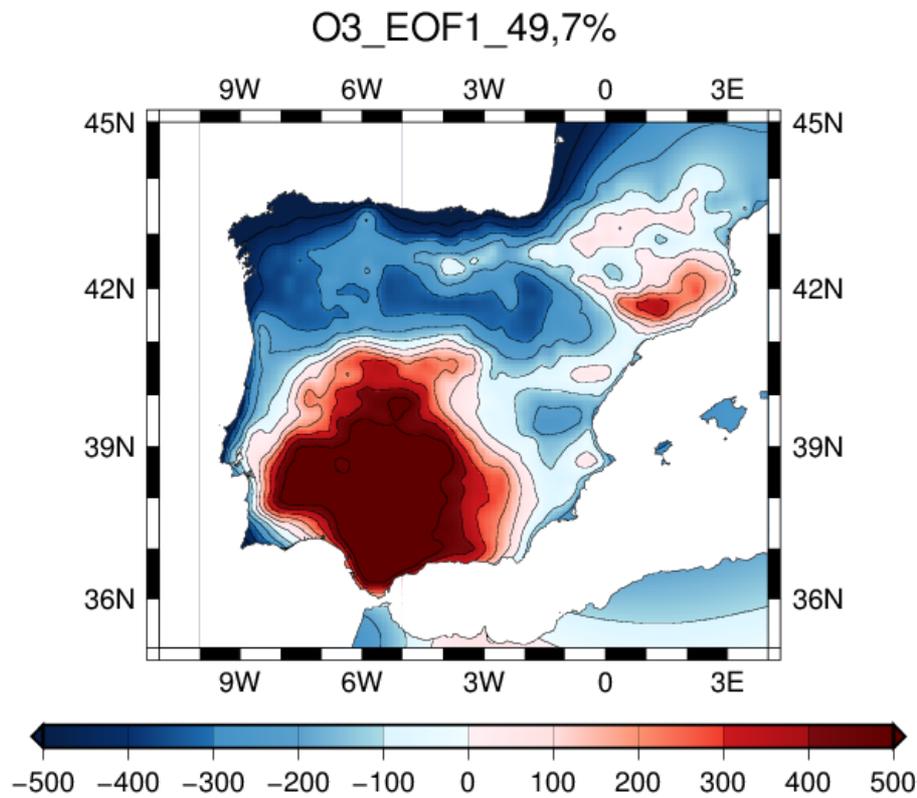


Figura 3.2. Representación espacial EOF1 ozono

La figura 3.3 nos da información sobre cuál ha sido la tendencia temporal. Se realizó el test de significación y el p-valor fue de $1,44e-2$, por lo que es significativo. Como se puede observar, ésta ha sido positiva, es decir ha habido un incremento de la concentración de ozono para los episodios extremos durante los 30 años de estudio, (que estaría localizada en las áreas de color rojo). Esta información combinada con la figura 3.2 nos proporciona los siguientes resultados: la primera componente muestra un destacado dipolo norte-sur con una tendencia positiva en la concentración de ozono en el sur peninsular mientras que en el norte ha habido un descenso de los niveles de episodios extremos durante el periodo de estudio. Cabe destacar que este patrón ha sido el más frecuente, ya que explica casi el 50% de la variabilidad.

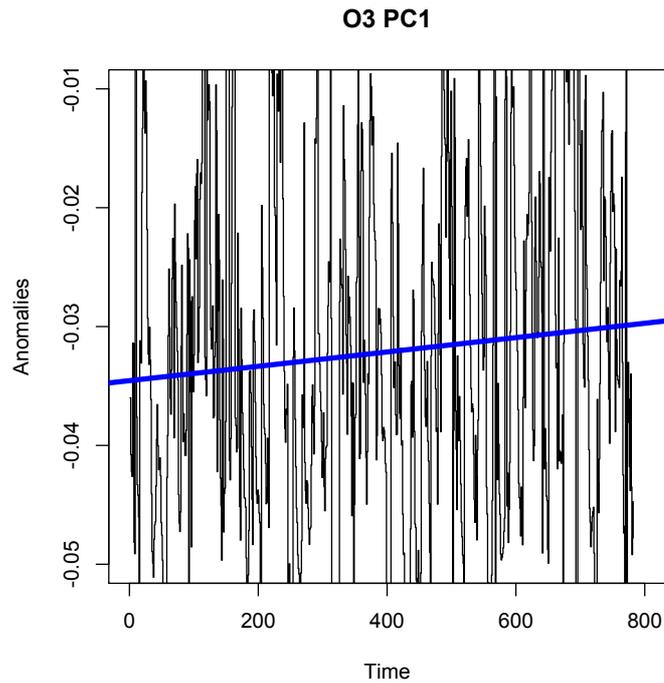


Figura 3.3. Tendencia temporal PC1 ozono

3.1.1.2. Segunda componente ozono

La figura 3.4 representa la EOF2 la cual acumula una variabilidad explicada del 67,8% por lo que esta EOF explica un 18,1%. La tendencia temporal de esta componente viene representada en la figura 3.5.

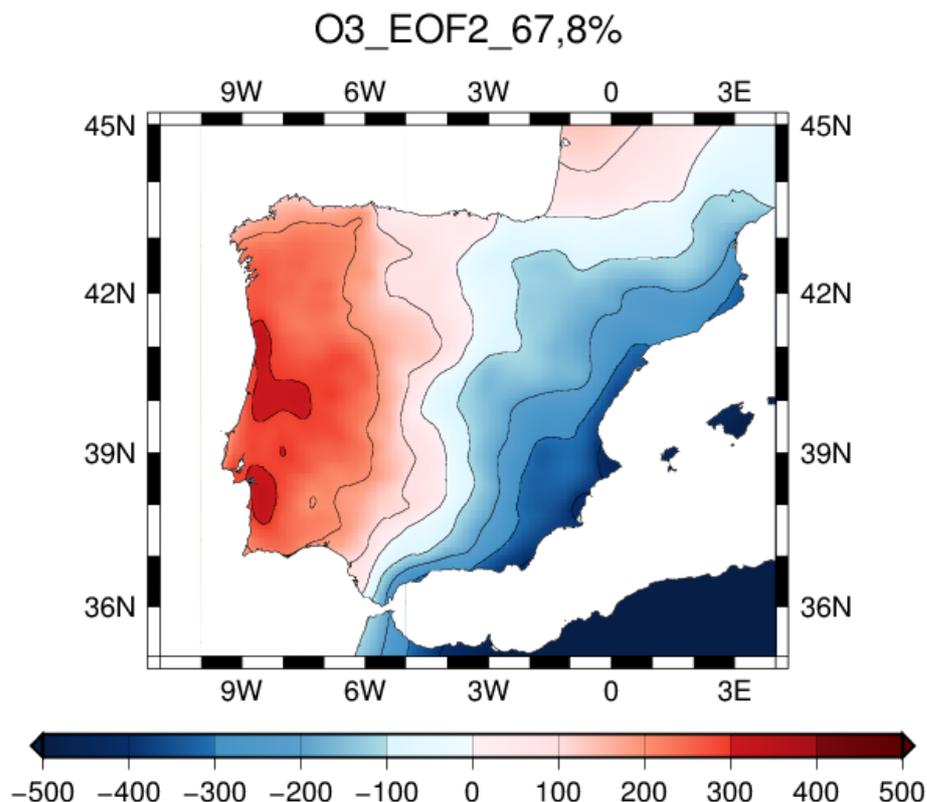


Figura 3.4. Representación espacial EOF2 ozono

En el caso de la segunda componente (figura 3.5), también muestra que ha habido un aumento de la concentración de ozono durante los 30 años de estudio en el oeste peninsular, y un descenso en la concentración de ozono durante episodios extremos en el área Mediterránea. Se realizó el test de significación y el p-valor fue de $2,57e-3$, por lo que es significativo. Teniendo en cuenta la tendencia temporal y espacial, los resultados más destacados de esta componente se pueden resumir en: se observa un gradiente Atlántico-Mediterráneo con una tendencia positiva en las concentraciones de los episodios extremos sobre el oeste de la PI. Sin embargo, podemos observar un descenso de dicha concentración en el levante peninsular.

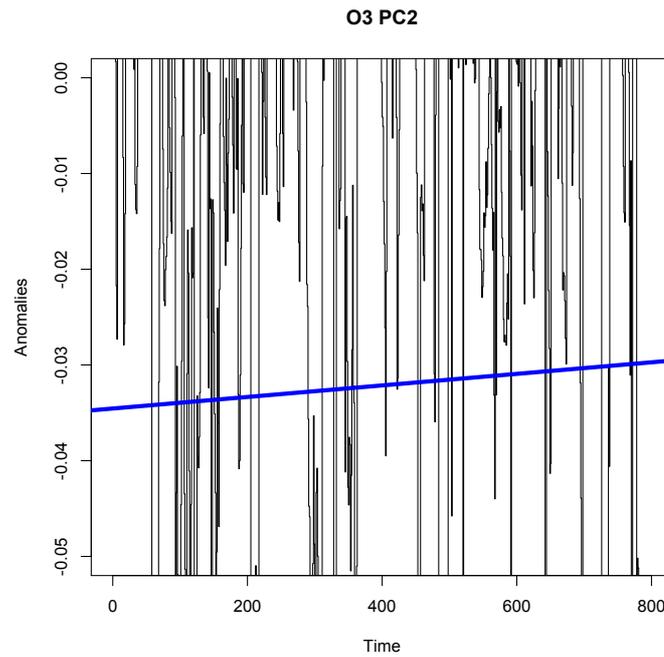


Figura 3.5. Tendencia temporal PC2 ozono

3.1.1.3. Tercera componente ozono

La EOF3 acumula una variabilidad explicada del 75,3% por lo que esta EOF explica un 7,4% (figura 3.6). La tendencia temporal de esta componente viene representada en la figura 3.7.

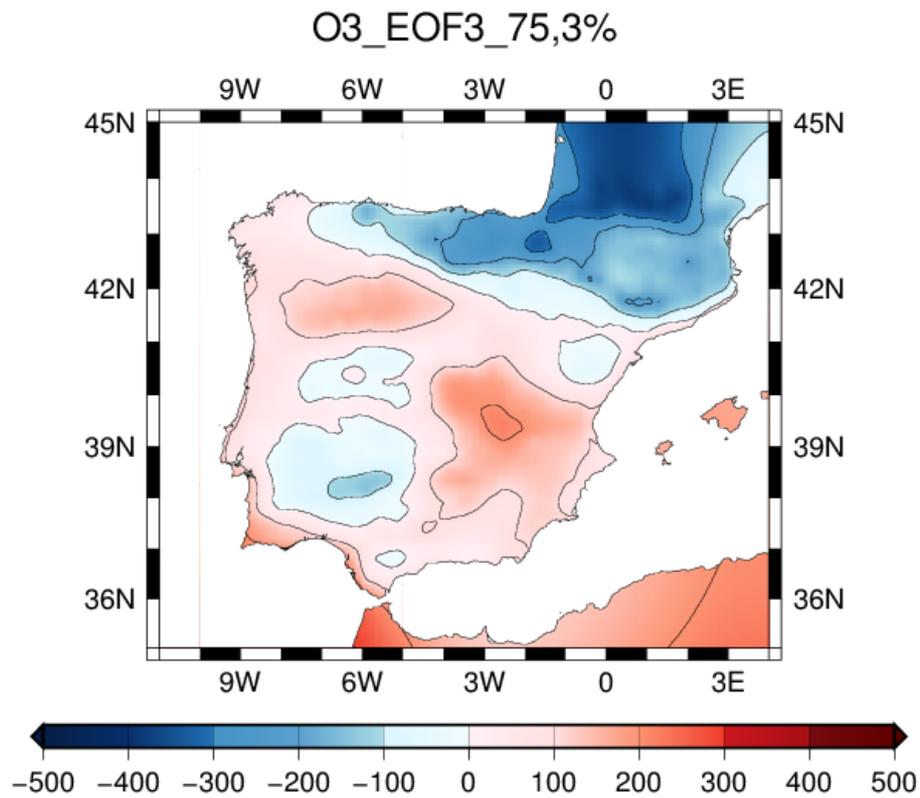


Figura 3.6. Representación espacial EOF3 ozono

En el caso de la tercera componente, al igual que la anterior, muestra que ha habido un aumento de la concentración de ozono durante los 30 años de estudio. El test de significación dio un p-valor de $8,28e-4$, siendo significativo. Teniendo en cuenta la tendencia temporal y espacial, los resultados más destacados son: Se observa un gradiente norte-sur de nuevo, con una tendencia positiva en las concentraciones de los episodios extremos sobre el sur de la PI. Sin embargo, podemos observar un descenso de dicha concentración en el norte peninsular.

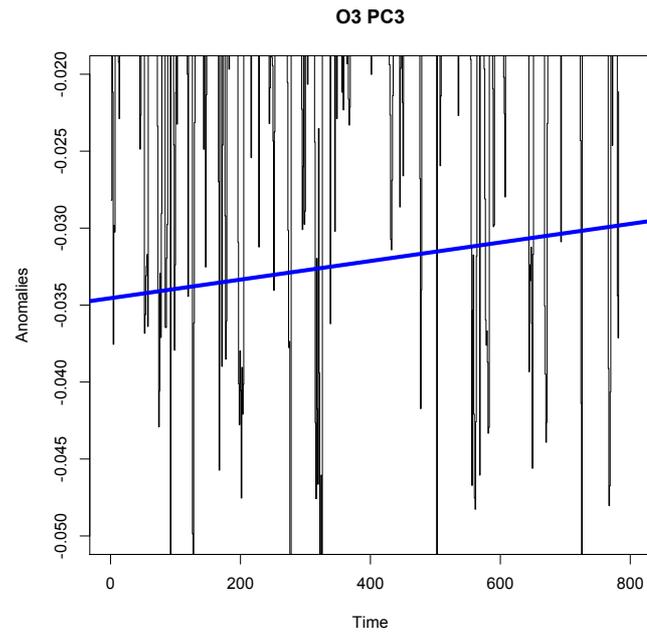


Figura 3.7. Representación temporal PC3 ozono

3.1.1.4. Cuarta componente ozono

La EOF2 (figura 3.8) acumula una variabilidad explicada del 81% por lo que esta EOF explica sólo un 0,05%. La tendencia temporal de esta componente viene representada en la figura 3.9.

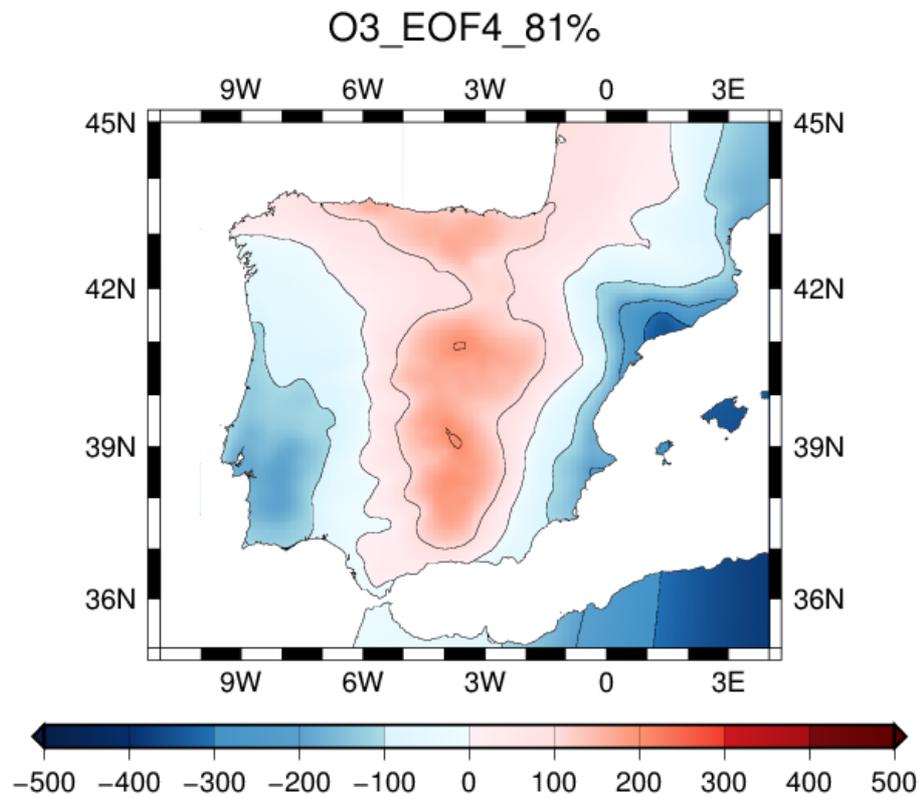


Figura 3.8. Representación espacial EOF4 ozono

En el caso de la última componente (figura 3.9), también ha habido un aumento de la concentración de ozono durante los 30 años de estudio, siendo esta tendencia significativa (p -valor $3,634e-05$). Teniendo en cuenta la tendencia temporal y espacial, los resultados más destacados de esta componente serían que se observa un gradiente oeste-este con una tendencia positiva en las concentraciones de los episodios extremos sobre el centro peninsular. Por otro lado, podemos observar un descenso de dicha concentración en las costas Atlántica y Mediterránea. Cabe destacar que este patrón es el que menos variabilidad explica, por lo que no es muy frecuente en la PI.

Por último cabe destacar que los 782 episodios quedan reducidos a unos pocos patrones que representan la distribución espacial de la concentración de ozono en la PI.

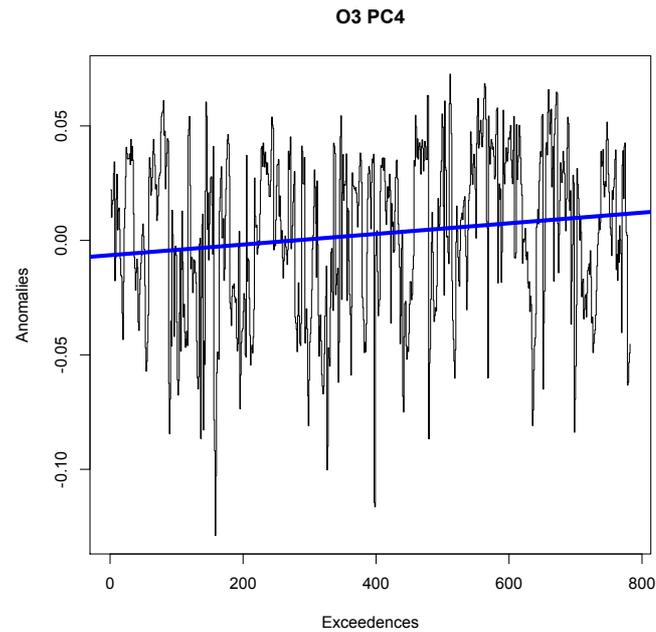


Figura 3.9. Tendencia temporal PC4 ozono

3.1.2. ANÁLISIS CLÚSTER PARA EPISODIOS EXTREMOS DE OZONO

Para la realización del análisis clúster, se ha partido del archivo original de la matriz de 1595 filas por 782 columnas y hemos calculado el dendrograma o diagrama de árbol (figura 3.10) con el objetivo de seleccionar el número de clústeres para el ozono.

Cluster Dendrogram

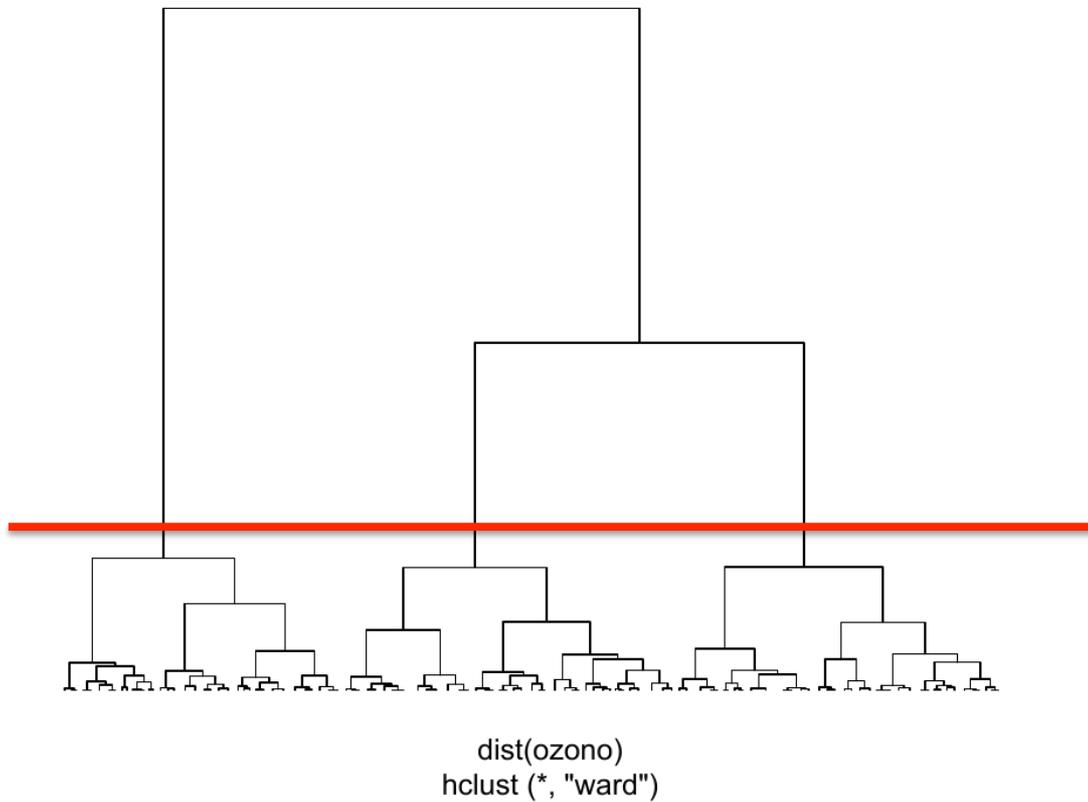


Figura 3.10. Clúster de ozono

A la vista de la figura 3.10, se han seleccionado tres clústeres. Una vez sabíamos el número de clústeres, los representamos gráficamente (figura 3.11).

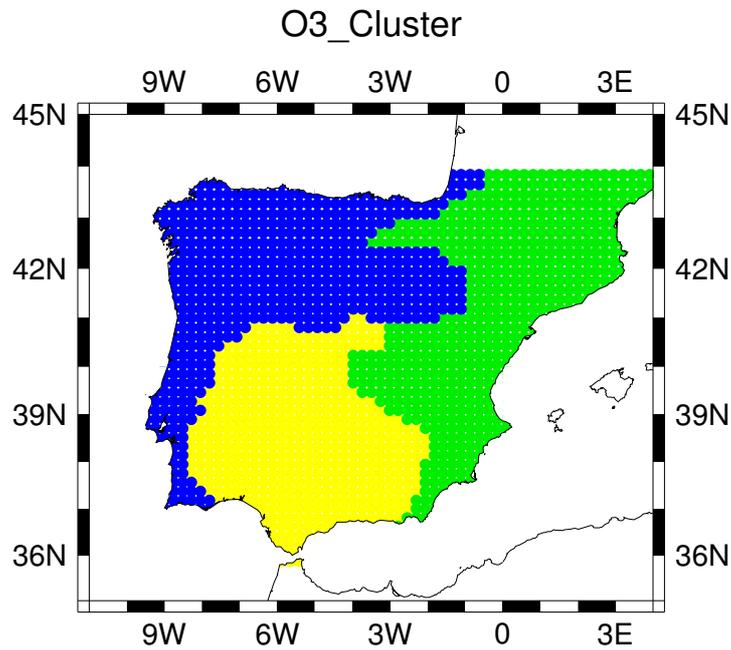


Figura 3.11. Representación espacial Clúster ozono

El resultado más destacable es que la representación espacial de la EOF1 coincide aproximadamente con la representación espacial obtenida en el análisis clúster, tal y como se puede observar en la figura 3.12. Esto nos hace pensar en la alta variabilidad explicada por la primera componente respecto a la señal de espacio-temporal de episodios extremos de ozono troposférico.

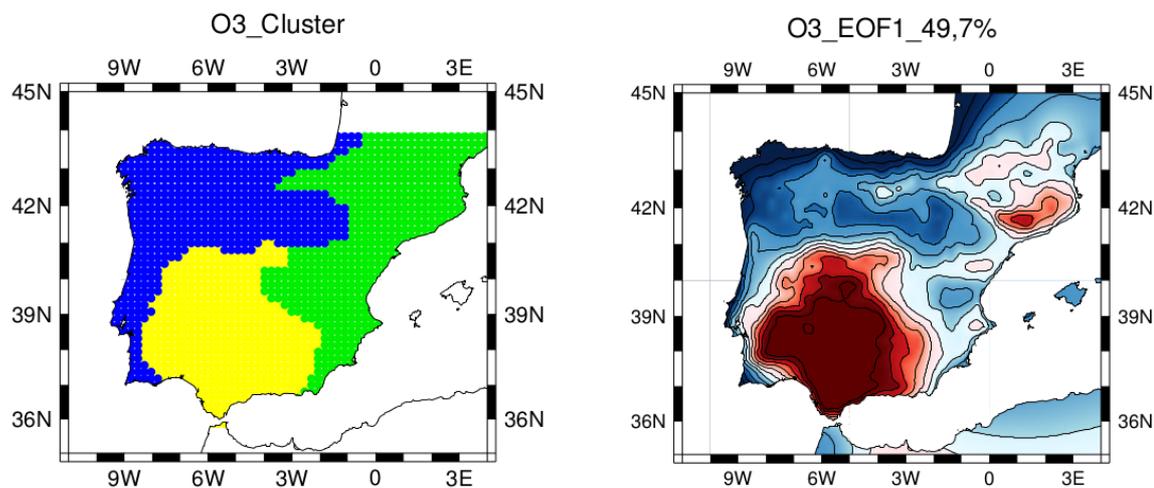


Figura 3.12. Comparativa Clúster EOF1 ozono

3.2. PM10

3.1.3. PCs y EOFs

En el caso del PM10, al realizar el PCA obtuvimos una componente espacial (4378) que se corresponde con la EOF, y una componente temporal (1595) que son las PCs. Por tanto, como para el PM10 hemos realizado la matriz traspuesta, las EOFs se corresponden con las PCs y viceversa. A la hora de elegir con cuántas componentes nos íbamos a quedar, nos basamos en dos criterios de los comentados con anterioridad: por un lado teniendo en cuenta el diagrama de codos representado en la figura 3.13, y por otro lado la variabilidad explicada representada en la tabla 3.2.

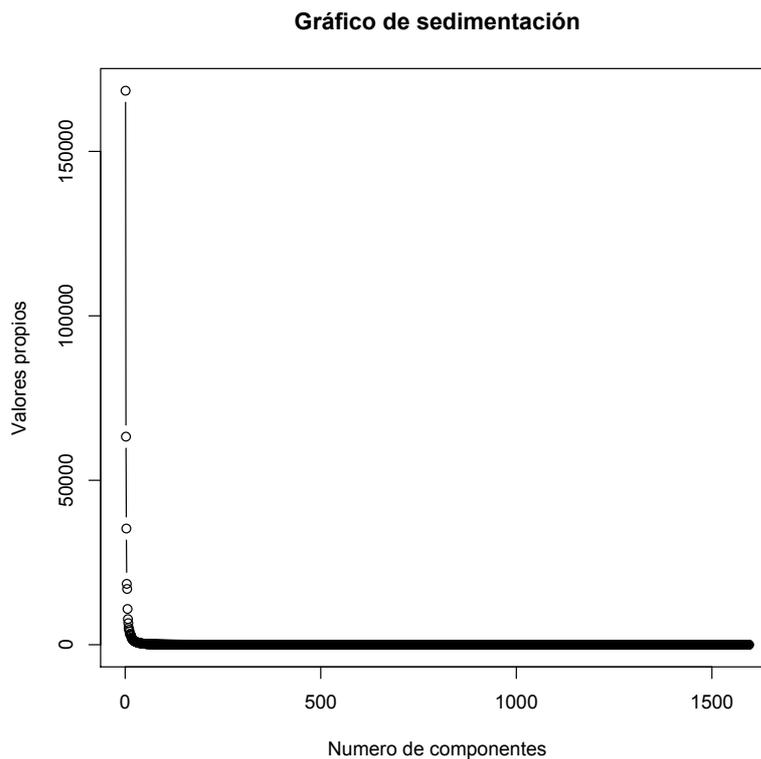


Figura 3.13. Gráfico de sedimentación PM10

En el gráfico de sedimentación del PM10 nos hemos encontrado más solapamiento en el eje x ya que hay muchas más componentes (1595). De nuevo, al no quedar suficientemente claro, recurrimos al criterio de la variabilidad explicada (tabla 3.2)

Tabla 3.2. Variabilidad explicada PCs PM10

	PC1	PC2	PC3	PC4	PC5	PC6
Desviación estándar	410.432	251.541	187.960	136.045	130.221	104.251
Proporción de varianza	0.428	0.160	0.089	0.047	0.043	0.027
Proporción acumulada	0.428	0.589	0.679	0.726	0.769	0.796

Como podemos observar la primera PC nos explica más de un 40% de la variabilidad de los eventos extremos de PM10 sobre la PI. A la vista de la variabilidad explicada por las demás PCs, en este estudio nos hemos quedado con las 6 primeras PCs para el caso del PM10.

Al igual que el ozono, se van a explicar los patrones espaciales para el PM10, y se interpretan de la misma manera, las áreas que presenten el mismo color tienen el mismo comportamiento. A continuación describiremos las 6 EOFs que con anterioridad señalamos que nos quedaríamos.

3.1.3.1. Primera componente PM10

La EOF1 (figura 3.14) explica un 42% de la variabilidad de la señal de los episodios extremos de PM10. La tendencia temporal de esta componente viene representada en la figura 3.15.

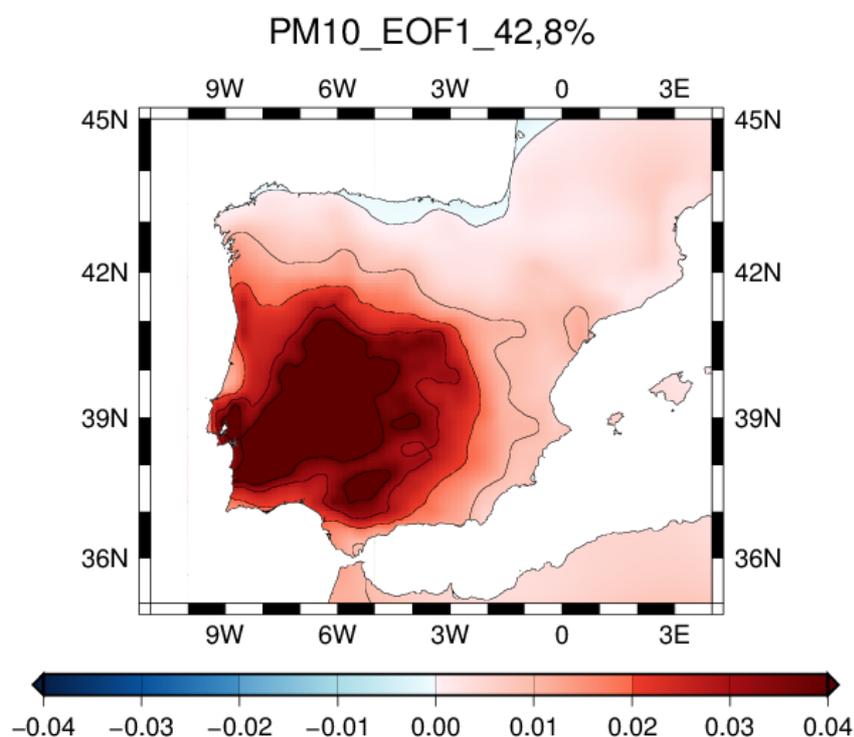


Figura 3.14. Representación espacial EOF1 PM10

La tendencia temporal de la primera componente del PM10 durante los 30 años de estudio ha sido constante (el p-valor fue de $1,562e-1$). Teniendo en cuenta la tendencia temporal y espacial, podemos destacar lo siguiente: hay un patrón homogéneo en toda la PI, siendo más intenso en el suroeste. Este patrón podría estar relacionado con episodios de resuspensión del polvo. Sin embargo, a pesar de que esta componente explica la mayor parte de la variabilidad, no ha mostrado ninguna tendencia positiva en los niveles de PM10 para el periodo climatológicamente significativo estudiado (es decir, no ha mostrado un aumento en los niveles de PM10).

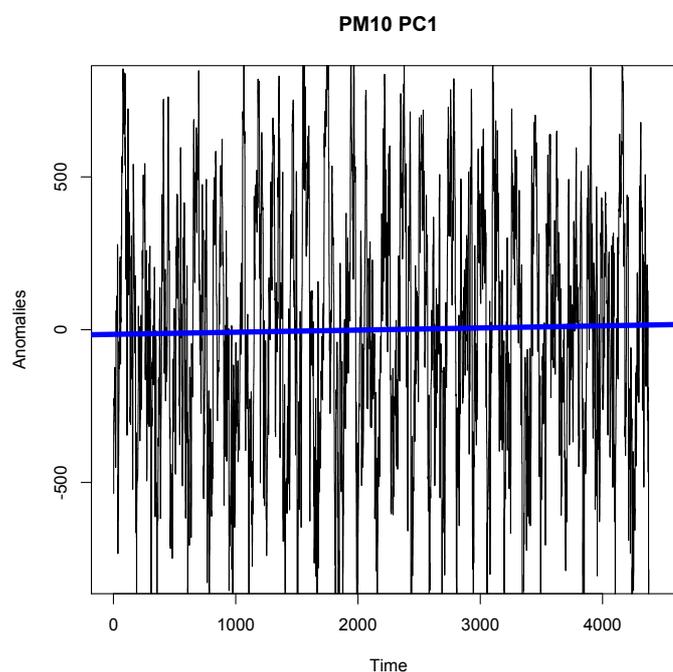


Figura 3.15. Tendencia temporal PC1 PM10

3.1.3.2. Segunda componente PM10

La EOF2 acumula una variabilidad explicada del 58,9%, por lo tanto sólo explica un 16,1% (figura 3.16). La tendencia temporal de esta componente viene representada en la figura 3.17.

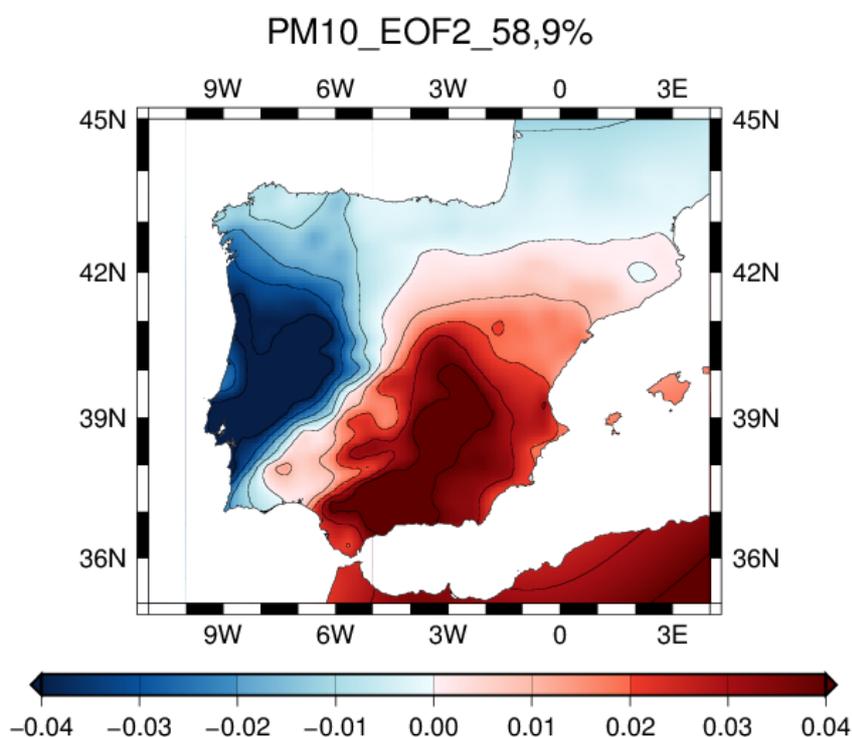


Figura 3.16. Representación espacial EOF2 PM10

La segunda componente del PM10 muestra que ha habido un aumento de los niveles de dicho contaminante durante los 30 años de estudio (el p-valor: $1,404e-14$, es significativo). Teniendo en cuenta la tendencia temporal y espacial, los resultados más destacados de esta componente son: se observa un gradiente noroeste-sureste con una tendencia positiva en las concentraciones de los episodios extremos sobre la costa mediterránea y el sureste peninsular. Por otro lado, podemos observar un descenso de dicha concentración de episodios extremos en la costa Atlántica. A pesar de que

esta componente sí muestra una tendencia positiva en los niveles de PM10, explica sólo un 16,1% de la variabilidad explicada.

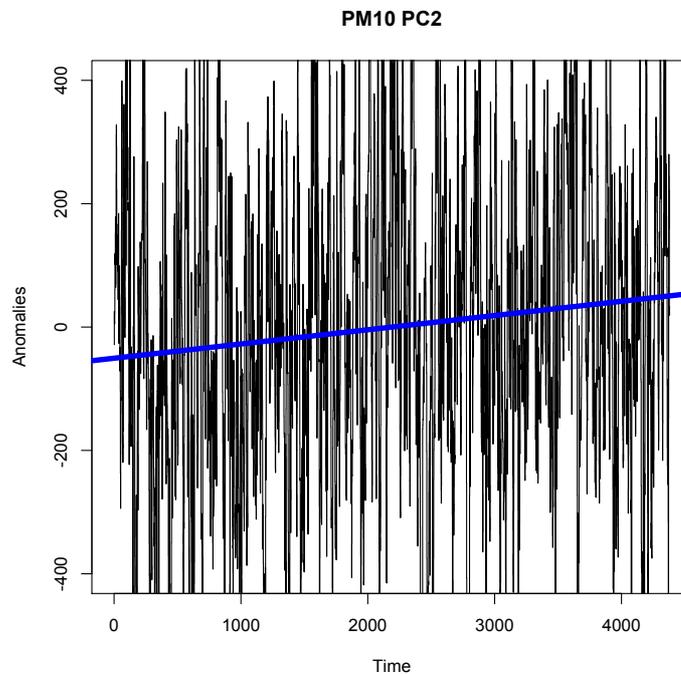


Figura 3.17. Tendencia temporal PC2 PM10

3.1.3.3. Tercera componente PM10

La EOF3 acumula una variabilidad explicada del 67,9%, por lo tanto sólo explica un 8,9% (figura 3.18). La tendencia temporal de esta componente viene representada en la figura 3.19.

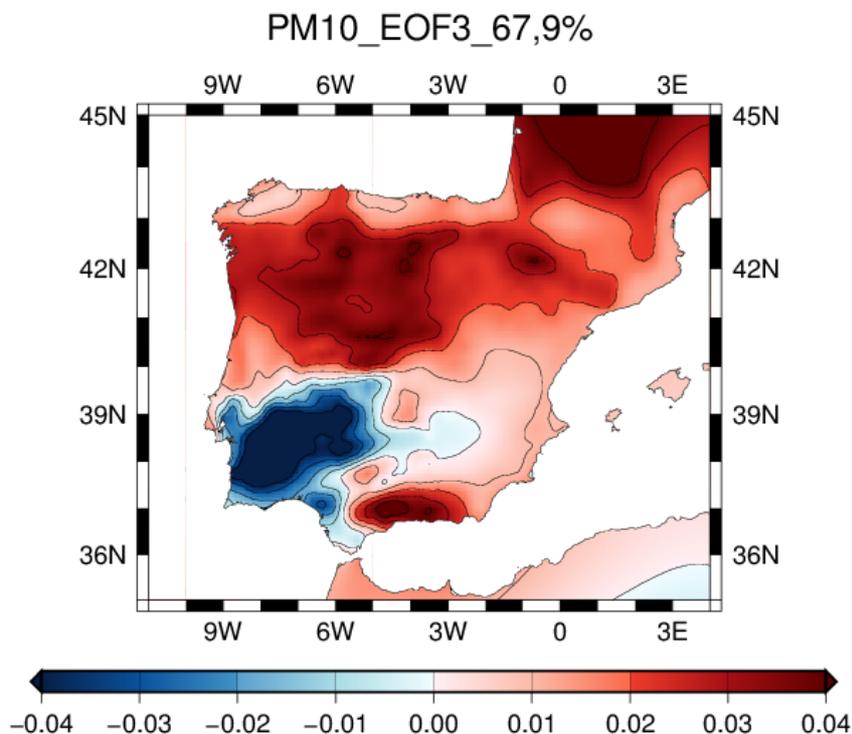


Figura 3.18. Representación espacial EOF3 PM10

En el caso de la tercera componente, también muestra que se ha producido un aumento de la concentración de PM10 durante los 30 años de estudio (con un p-valor de $2,751e-05$, significativo). Teniendo en cuenta la tendencia temporal y espacial, se observa un gradiente norte-sur, con una tendencia positiva en las concentraciones de los episodios extremos sobre el norte de la PI, especialmente sobre la meseta norte. Sin embargo, podemos observar un descenso de dicha concentración en el suroeste peninsular.

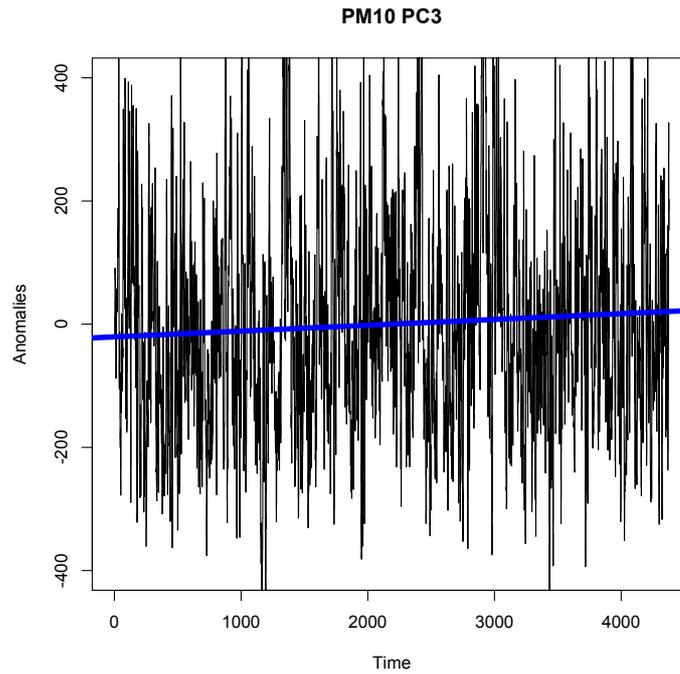


Figura 3.19. Tendencial temporal PC3 PM10

3.1.3.4. Cuarta componente PM10

La EOF4 acumula una variabilidad explicada del 72,6%, por lo tanto solo explica un 4,7% (figura 3.20). La tendencia temporal de esta componente viene representada en la figura 3.21.

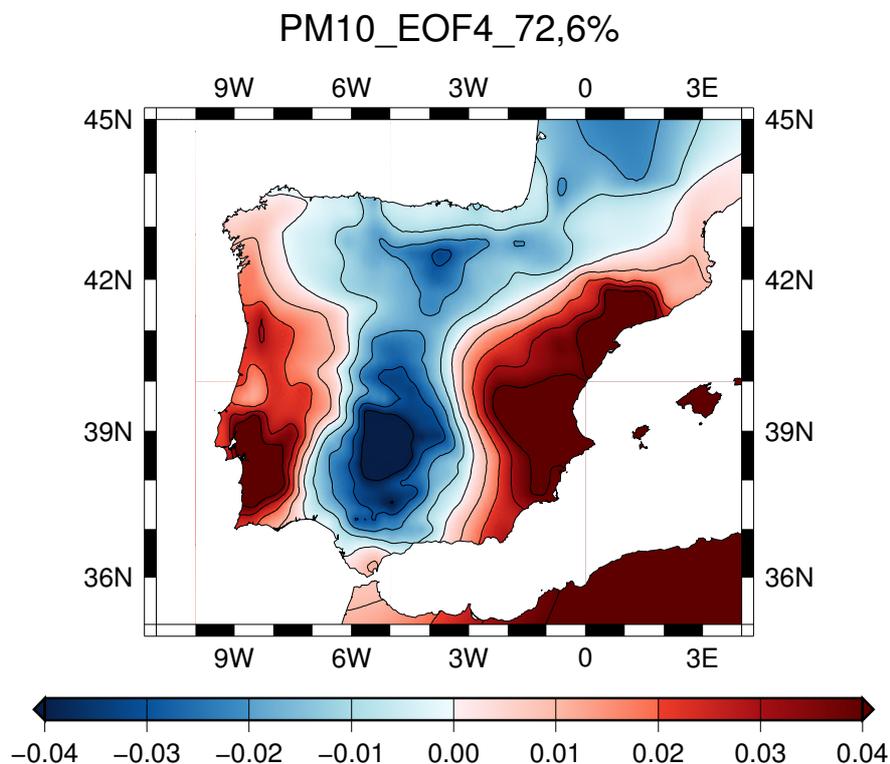


Figura 3.20. Representación espacial EOF4 PM10

La cuarta componente muestra un p-valor de 0,1811, por lo que la tendencia de la concentración de PM10 durante los 30 años de estudio asociada a este patrón ha sido constante. Al no mostrar una tendencia positiva, sólo cabría destacar los aspectos relacionados con la distribución espacial, los cuales muestran similar comportamiento en la costa del levante y la costa oeste peninsular, en contra posición a la zona central.

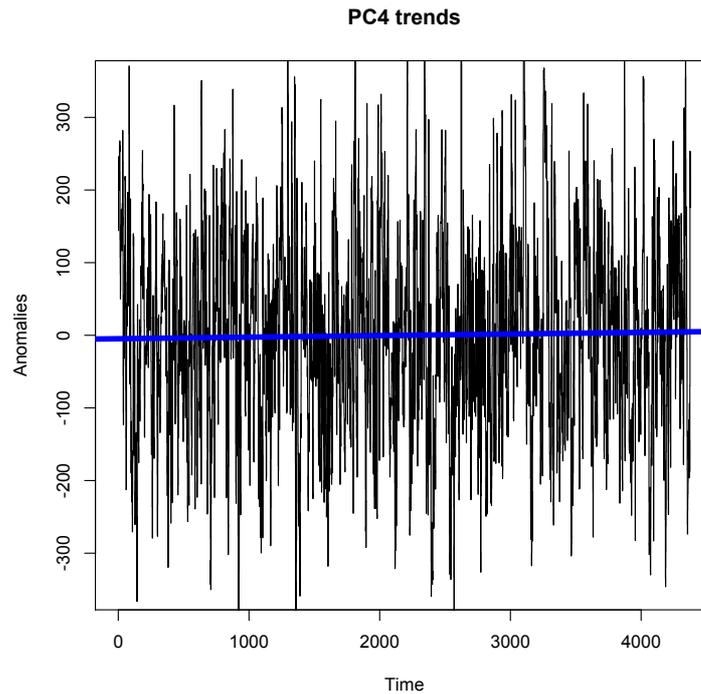


Figura 3.21. Tendencia temporal EOF4 PM10

3.1.3.5. Quinta componente PM10

La EOF5 acumula una variabilidad explicada del 76,9%, por lo tanto solo explica un 4,3% (figura 3.22). La tendencia temporal de esta componente viene representada en la figura 3.23.

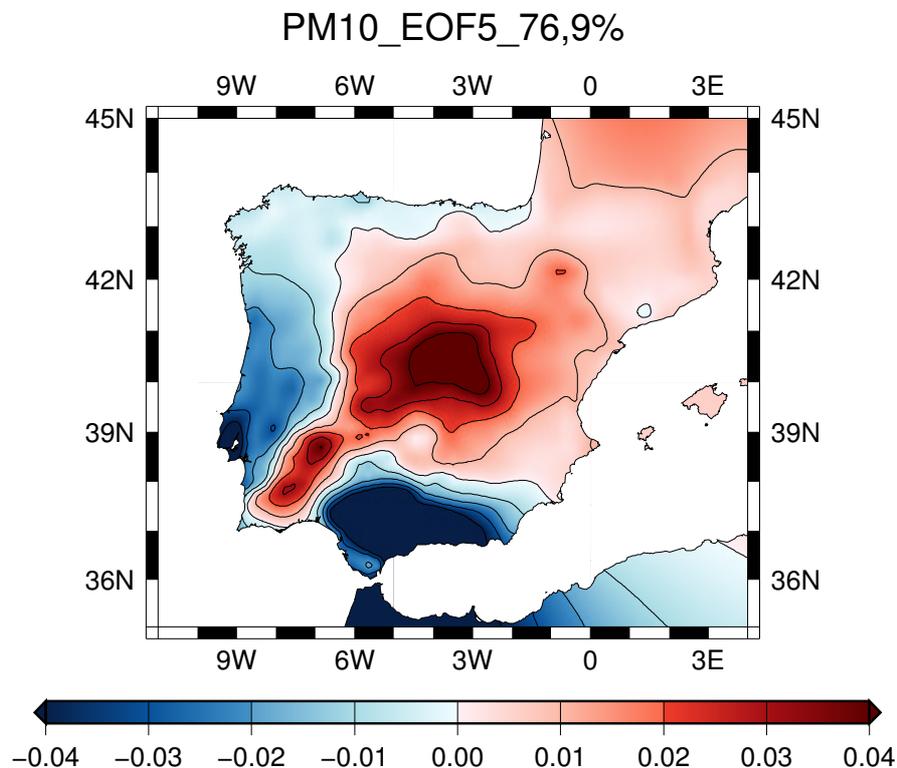


Figura 3.22. Representación espacial EOF5 PM10

La quinta componente, también muestra un aumento de la concentración de PM10 durante los 30 años de estudio, con un p-valor significativo (p-valor: $1,476e-09$). Teniendo en cuenta la tendencia temporal y espacial, se observa en la franja noreste-suroeste una tendencia positiva en las concentraciones de los episodios extremos siendo más intensos sobre el centro peninsular. Por otro lado, en el sur y oeste de la PI podemos observar un descenso de dicha concentración.

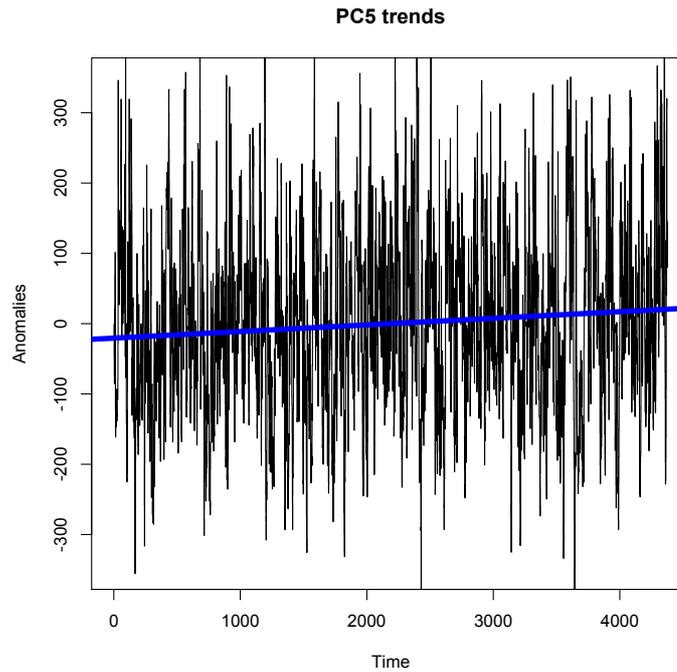


Figura 3.23. Tendencia temporal PC5 PM10

3.1.3.6. Sexta componente PM10

La EOF6 acumula una variabilidad explicada del 79,6%, por lo tanto solo explica un 2,7% (figura 3.24). La tendencia temporal de esta componente viene representada en la figura 3.25.

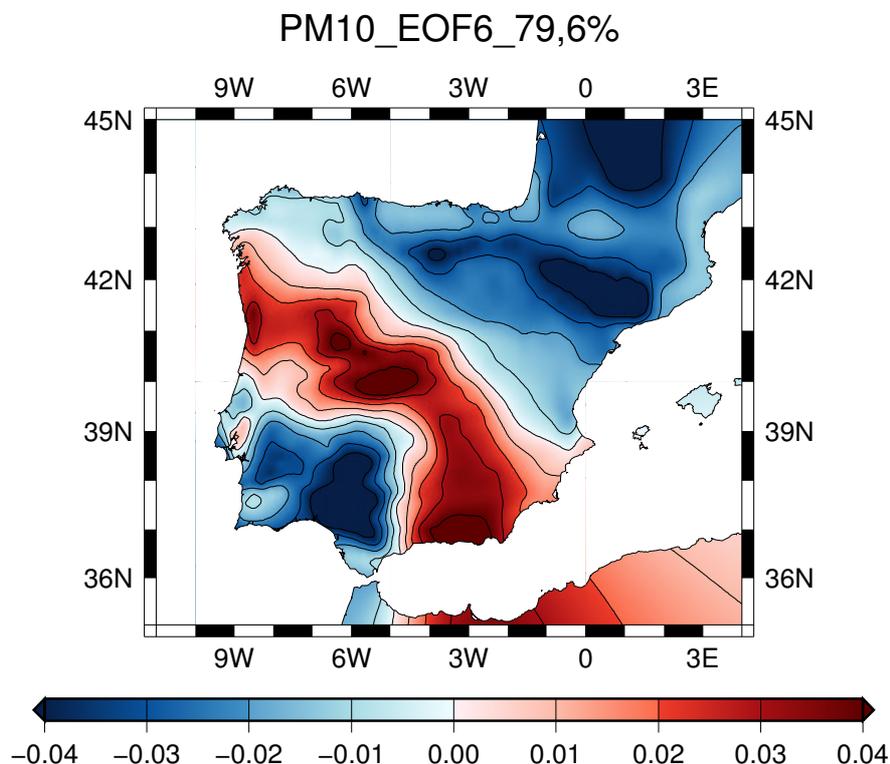


Figura 3.24. Representación espacial EOF6 PM10

Por último, la sexta componente también muestra que se ha producido un aumento de la concentración de PM10 durante los 30 años de estudio. El test de significación dio un p-valor de $1,869e-3$, siendo significativo. Teniendo en cuenta la tendencia temporal y espacial, se observa un gradiente norte-sur, con una tendencia positiva en las concentraciones de los episodios extremos sobre el norte de la PI. Sin embargo, podemos observar un descenso de dicha concentración en el suroeste peninsular. Esta última componente es el patrón menos frecuente de los comentados (sólo explica el 2,7% de los casos).

Por último cabe destacar que los 4378 episodios quedan reducidos a unos pocos patrones que representan la distribución espacial de la concentración del PM10 en la PI.

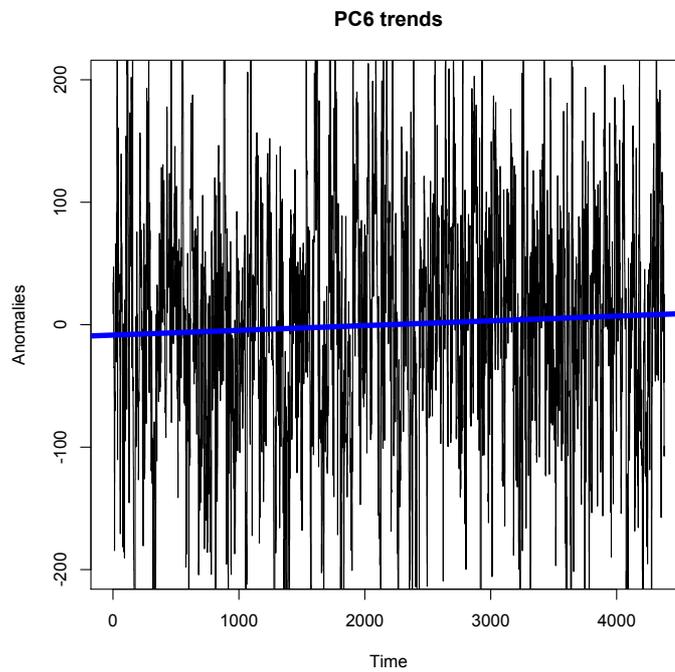


Figura 3.25. Tendencia temporal PC6 PM10

3.1.4. ANÁLISIS CLUSTER DE LOS EPISODIOS EXTREMOS DE PM10

En el caso del PM10, se ha partido del archivo original de la matriz de 4378 filas por 1595 columnas y hemos calculado el dendrograma o diagrama de árbol (figura 3.26) con el objetivo de seleccionar el número de clústeres para el ozono.

Cluster Dendrogram

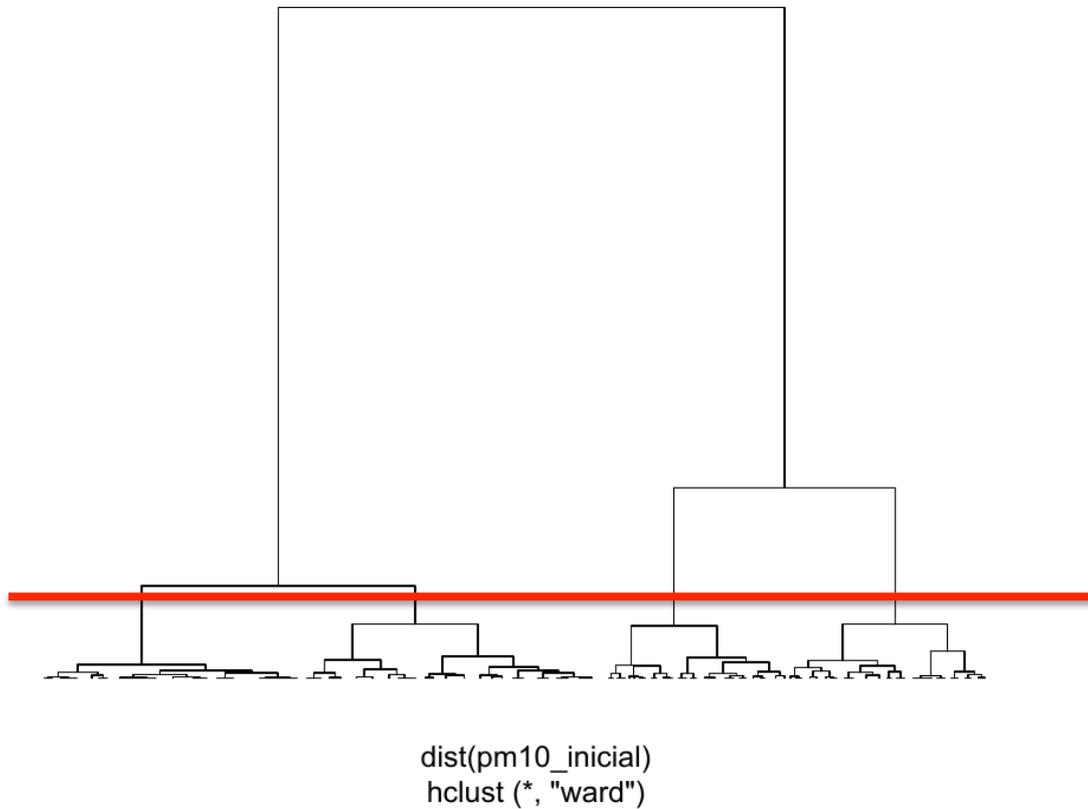


Figura 3.26. Clúster de PM10

A la vista de la figura 3.27, se han seleccionado cuatro clústers. Una vez sabíamos el número de clústers, los representamos gráficamente (figura 3.28).

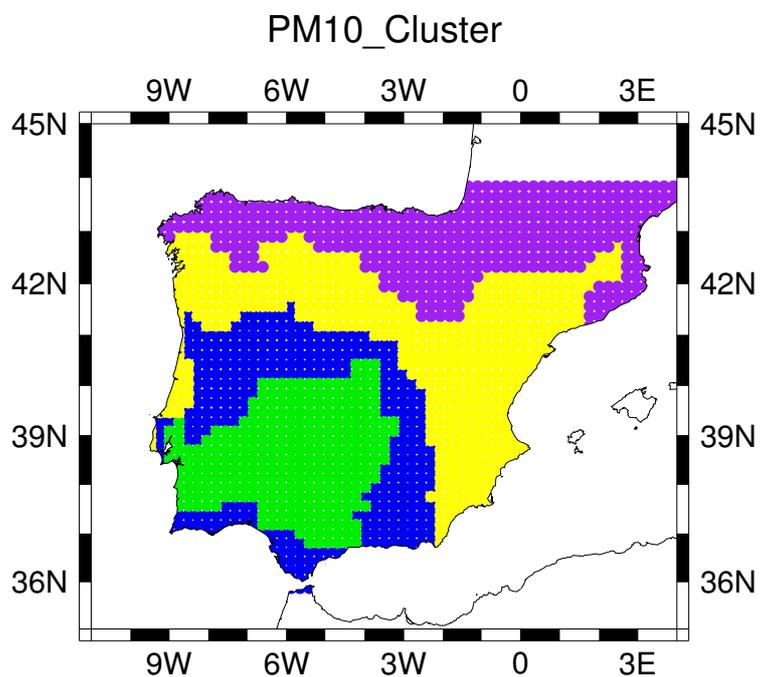


Figura 3.27. Representación espacial Clúster PM10

Al igual que ocurría con el ozono, el resultado más destacable es que la representación espacial de la EOF1 coincide con la representación espacial obtenida en el análisis clúster, tal y como se puede observar en la figura 3.28.

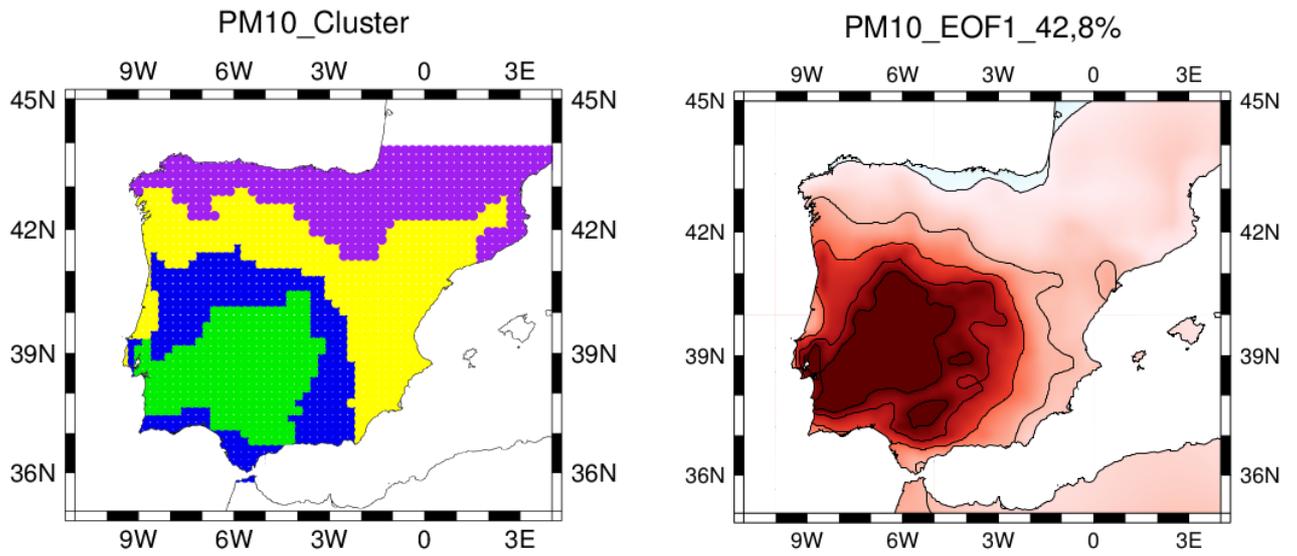


Figura 3.28. Comparativa Clúster y EOF1 PM10

4. CONCLUSIONES

Recordando los objetivos propuestos en el presente trabajo, se marcó como principal caracterizar los patrones espacio-temporales de episodios extremos de contaminación de la PI. Tras la realización de este trabajo, los patrones espaciales de los episodios extremos sobre la PI han sido identificados junto con un análisis de la tendencia temporal para los contaminantes ozono y PM10. Así mismo, se ha realizado un análisis clúster tras el cual se han agrupado esos episodios extremos con el fin de resumirlos en un número menor.

Las conclusiones más destacables obtenidas tras el análisis llevado a cabo son las siguientes:

- Patrones espaciales y temporales: La primera componente tanto para ozono como para PM10 explica alrededor del 50% de la variabilidad de la señal de la concentración. En cuanto al ozono, el patrón espacial que mayor varianza explica (49%) muestra que ha habido un incremento de los niveles de ozono sobre el sur de la PI a lo largo de los 30 años de estudio. Por otro lado, el patrón espacial que más varianza explica del PM10 (42,1%) no ha mostrado una tendencia positiva durante los 30 años de estudio. Sin embargo, el segundo patrón espacial (el cual sólo explica un 16% de la varianza) sí muestra tendencia positiva, observándose un incremento de los niveles de PM10 en el área Mediterránea para el periodo de estudio escogido.
- Análisis Clúster: Tras realizar el análisis, se ha podido evaluar la consistencia de los resultados obtenidos con el PCA, en donde la representación espacial del clúster, tanto para ozono como PM10, es muy similar al patrón espacial de la primera componente. En el caso del ozono se han obtenido 3 regiones y para el PM10 se obtuvieron 4.

Por tanto, la conclusión más importante obtenida en este trabajo es que se ha reducido el problema de dimensionalidad al que nos enfrentábamos, ya que se han encontrado aquellos patrones que agrupan el inmenso número de episodios modelizados para la Península Ibérica en el periodo de 30 años de estudio.

La contaminación atmosférica continúa representando un riesgo para la salud de la población; por tanto, estudios como el nuestro, pueden ir encaminados a ayudar a una mejor definición de los planes de control y mitigación de la contaminación atmosférica sobre la Península Ibérica.

5. BIBLIOGRAFÍA

Barros, N., Toll, I., Soriano, C., Jiménez, P., Borrego, P., Baldasano, J.M. (2003). Urban Photochemical Pollution in the Iberian Peninsula: the Lisbon and Barcelona Airsheds. *Journal of the Air & Waste Management Association* 53, 347-359.

Basart, S., Pay, M.t., Jorba, O., Pérez, C., Jiménez-Guerrero, P., Schulz, M., and Baldasano, J.M (2011). Aerosols in the CALIOPE air quality modelling system: evaluation and analysis of PM levels, optical depths and chemical composition over Europe.

Basic Air Pollution Meteorology, APTI Course SI: 409 (Rev. April, 2005). Air Pollution Training Institute. Environmental Protection Agency.

Bessagnet, B., Hodzic, A., Vautard, R., Beekmann, M., Cheinet, S., Honore, C., Liousse, C., and Rouil, L. (2004). Aerosol modeling with chimere – preliminary evaluation at the continental scale. *Atmospheric Environment*, 2803-2817.

Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245-276.

DIRECTIVA 2008/50/CE DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 21 de mayo de 2008 relativa a la calidad del aire ambiente y a una atmósfera más limpia en Europa. DOUE L, num. 152,11-6-2008, 1-44.

Dueñas, C., Fernández, M.C., Cañete, S., Carretero, J., Liger, E. (2002). Assessment of ozone variations and meteorological effects in an urban area in the Mediterranean coast. *The Science of the Total Environment* 299, 97-113.

Environmental European Agency, EEA (2012). Air quality in Europe. EEA report, No 4/2012. doi:10.2800/55823.

Escudero, M., Stein, A., Draxler, R.R., Xerol, X., Alasutey, A., Castillo, S., and Avila, A. (2006). Determination of the contribution of northern Africa dust source areas to PM10 concentrations over the central Iberian Peninsula using the Hybrid Single-Particle Lagrangian Integrated Trajectory model (HYSPLIT) model. *Journal of Geophysical Research: Atmospheres*. 111, D06210.

ESPAÑA. Ley 34/2007, de 15 de noviembre, de calidad del aire y protección de la atmósfera. BOE, num 275, de 16 de noviembre de 2007, 46962-46987.

ESPAÑA. Real Decreto 102/2011, de 28 de enero, relativo a la mejora de la calidad del aire. BOE, num. 25, de 29 de enero de 2011, pag 9574-9626.

Giorgi, F. (2006). Climate change hot-spots. *Geophysical Research Letters*, 33, L08707, doi: 10.1029/2006GL025734.

Gómez Navarro, J.J. Montavez, J.P., Jiménez-Guerrero, P., Jerez Rodríguez,

S., Garcia Valero, J.A., Gonzalez Rouco, J. (2010). Warming patterns in regional climate change projections over the Iberian Peninsula. *Meteorologische Zeitschrift*, 19, 275-285.

Grell, G.A., Dudhia, J., and Stauffer, D.R. (1994). A description of the fifth generation penn state/ncar mesoscale model (MM5). Technical Report NCAR/TN-398+STR, National Center for Atmospheric Research.

Jolliffe, I.T. (2002). *Principal Component Analysis*. Springer Series in Statistics. 489.

Jiménez-Guerrero, P., Gómez Navarro, J.J., Jerez Rodriguez, S., Lorente Plazas, R., Garcia Valero, J.A., Montavez, J.P. (2011). Isolating the effects of climate change in the variation of secondary inorganic aerosols (SIA) in Europe for the 21st century (1991-2100). *Atmospheric Environment*, 45, 1059-1063.

Jiménez-Guerrero, P., Montávez, J.P., Gómez-Navarro, J.J., Jerez, S., Lorente-Plazas, R. (2012). Impacts of climate change on ground level gas-phase pollutants and aerosols in the Iberian Peninsula for the late XXI century. *Atmospheric Environment* 55 483-495.

Jorba, O. (2005). Simulación de los campos de viento de la Península Ibérica y el área geográfica de Catalunya con alta resolución espacial para distintas situaciones meteorológicas típicas. PhD thesis, Universidad Politécnica de Cataluña. 238.

Lesniok, M., Malarzewski, L., and Niedzwiedz, T. (2010). Classification of circulation types for Southern Poland with an application to air pollution concentration in Upper Silesia. *Physics and Chemistry of the Earth* 35: 516-522.

Lorenz, E., 1956. Technical report. Statistical forecast project report 1. Department of Meteorology. MIT 49, 1. Massachusetts Institute of Technology, chap Empirical orthogonal functions and statistical weather prediction, 52.

Martín-Vide, J. y Olcina, J (2001): *Climas y tiempos de España*, Madrid, Alianza Editorial, 258.

Palacios, M., Kirchner, F., Martilli, A., Clappier, A., Martín F., and Rodríguez, M.E. (2002). Summer Ozone Episodes in the Greater Madrid Area: Analysis of the Ozone Response to Abatement Strategies by Modeling. *Atmospheric Environment* , 36, 5323-5333.

Palacios, M., Martín, F., and Aceña, B. (2005). Estimate of potentially high ozone concentration areas in the centre of the Iberian Peninsula, *International Journal of Environment and Pollution*. 24, Nos. 1/2/3/4, 260–271.

Peña, D. 2002. *Análisis de datos multivariantes*. McGraw-Hill. 515.

Pielke, R.A., and Segal, M. (1986). *Mesoscale Meteorology and Forecasting*. American Meteorology Society. 793.

Querol, X., Alastuey, J.A., Puigercus, E., Mantilla, J.V., Miró, A., López-Soler, F., Plana and Artíñano, B (1998). Seasonal evolution of suspended particles around a large coal-fired power station: Particle levels and sources. *Atmospheric Environmen*, 32, 1963-1978.

Ribas, A., Peñuelas, J. (2004). Temporal patterns of surface ozone levels in different habitats of the North Western Mediterranean basin. *Atmospheric Environment* 38, 985-992.

Rodríguez, S., Querol, X., Alastuey, A., Plana, F., 2002. Sources and processes affecting levels and composition of atmospheric aerosol in the Western Mediterranean. *Journal of Geophysical Research* 107 (D24), 4777.

Seaman, N.L (2003). Future directions of meteorology related to air-quality research. *Environment International* 29, 245–252.

Serrano, A., García, J.A., Mateos, V.L., Cancillo, M.L., Garrido, J. (1998). Monthly Modes of Variation of Precipitation over the Iberian Peninsula. *Journal of Climate* 12 2894-2919.

Stensrud, D. (2007). *Parametrization schemes: Keys to understanding Numerical Weather Prediction Models*. Cambridge University press, Cambridge CB2 8RU, UK.

Terradellas, E., Camino, C., Alonso-Pérez, Basart, S., Baldasano, J.M. & Cuevas, E (2012). Distribución espacial y temporal de polvo mineral atmosférico en el norte de África y Oriente medio estimada a partir de la visibilidad horizontal. XXXII Jornadas Científicas Asociación Meteorológica Española “Meteorología y Calidad del Aire”, 1-8.

Wilks, D.S. (1995). *Statistical Methods in the Atmospheric Sciences*. International Geophysics Series. 91.

WHO (2005). World Health Organization. *Air Quality Guidelines Global Update*.

Ziomas, I.C., Gryning, S.E., Borsteing, R.D. (1998). The Mediterranean campaign of photochemical tracers-transport and chemical evolution (MEDCAPHOT-TRACE). *Atmospheric Environment* 32, 2043-2326.

ANEXOS

ANEXO A

Código análisis de OZONO

```
x=read.table("03.txt",na.strings='-9999')

#Creamos la matriz descartes donde metemos los valores NA
descartes=NULL;
temp2=NULL
for (i in 1:3417){
  temp=length(which(is.na(x[i,])))
  temp2=c(temp2,temp)
  if (temp>500){
    descartes=c(descartes,i);
  }
}

#Obtenemos un vector donde aparece a qué índice se corresponde a
cada puntito con valores (los no "NA"). #Porque luego los
quitaremos y cambiarán de numeración en la nueva matriz, ya no
seria la "malla IP"

vector=1:3417
vector=vector[-descartes]

#Creamos una nueva matriz sin los NA:

xconv<- x[-descartes,]
ozono<-data.frame(xconv)

#Eliminamos la columna row.names
row.names(ozono) <- NULL

# ANALISIS DE COMPONENTES PRINCIPALES
#Aplicamos componentes principales

ozono.pc<-prcomp(ozono)
summary (ozono.pc)

#Lo que aparece tras este comando es entre otras cosas el
porcentaje de variabilidad acumulada.
#Con esa información nos quedaríamos con 4 PC porque acumulan
>80% de la variabilidad.

#Pintamos el diagrama de codos
plot(ozono.pc$sdev^2, type="b", main="Gráfico de
sedimentación",xlab="Numero de componentes", ylab="Valores
propios")

#Queremos las puntuaciones (EOF's) porque queremos separar por
```

casos (que en nuestro estudio son las #localizaciones, puestas en filas) y no por variables (que en nuestro estudio son los eventos, puestos en columnas).

```
#Puntuaciones: EOFs, nos da la tendencia espacial
puntuaciones<-ozono.pc$x
```

```
#Creamos los archivos por cada EOF que nos quedamos para su
posterior representacion grafica
#"vector" es un vector de los índices "no descartes"
```

```
localizaciones=1:3417
puntuaciones1=rep(NA,3417)
puntuaciones1[vector]=ozono.pc$x[,1]
puntuacionesPC1<-data.frame(puntuaciones1)
write.table(puntuacionesPC1,file="EOF1_03.txt")
```

```
puntuaciones2=rep(NA,3417)
puntuaciones2[vector]=ozono.pc$x[,2]
puntuacionesPC2<-data.frame(puntuaciones2)
write.table(puntuacionesPC2,file="EOF2_03.txt")
```

```
puntuaciones3=rep(NA,3417)
puntuaciones3[vector]=ozono.pc$x[,3]
puntuacionesPC3<-data.frame(puntuaciones3)
write.table(puntuacionesPC3,file="EOF3_03.txt")
```

```
puntuaciones4=rep(NA,3417)
puntuaciones4[vector]=ozono.pc$x[,4]
puntuacionesPC4<-data.frame(puntuaciones4)
write.table(puntuacionesPC4,file="EOF4_03.txt")
```

```
#Sacamos las PCs, nos dan la tendencia temporal
PC_ozono<-ozono.pc$rotation
```

```
#Vamos seleccionando las PCs y las representamos graficamente
PC1_03=ozono.pc$rotation[,1]
PC1_03_df<-data.frame(PC1_03)
c<-(1:782)
p<-(PC1_03_df$PC1_03)
plot(c,p,main="PC1 trends", xlab="Excedences", ylab="pc",
type="s", col="black")
abline(lm(p~c), col="blue", lwd=5) #pinta la linea de tendencia
modelo=lm(p~c) # con esto sacamos el pvalor para ver si es
significativo
summary(modelo)
```

```
PC2_03=ozono.pc$rotation[,2]
PC2_03_df<-data.frame(PC2_03)
p2<-(PC2_03_df$PC2_03)
plot(c,p2,main="PC2 trends", xlab="Excedences", ylab="pc",
type="s", col="red")
abline(lm(p2~c), col="blue",lwd=5) #pinta la linea de tendencia
```

```

modelo2=lm(p2~c) # con esto sacamos el pvalor para ver si es
significativo
summary(modelo2)

```

```

PC3_03=ozono.pc$rotation[,3]
PC3_03_df<-data.frame(PC3_03)
p3<-(PC3_03_df$PC3_03)
plot(c,p3,main="PC3 trends", xlab="Excedences", ylab="pc",
type="s", col="red")
abline(lm(p3~c), col="blue", lwd=5) #pinta la linea de tendencia
modelo3=lm(p3~c) # con esto sacamos el pvalor para ver si es
significativo
summary(modelo3)

```

```

PC4_03=ozono.pc$rotation[,4]
PC4_03_df<-data.frame(PC4_03)
p4<-(PC4_03_df$PC4_03)
plot(c,p4,main="PC4 trends", xlab="Excedences", ylab="pc",
type="s", col="red")
abline(lm(p4~c), col="blue",lwd=5) #pinta la linea de tendencia
modelo4=lm(p4~c) # con esto sacamos el pvalor para ver si es
significativo
summary(modelo4)

```

```

#Si separamos 2 a 2
#plot(ozono.pc$x[,c(1,2)],col='red')
#text(ozono.pc$x[,c(1,2)],labels=row.names(ozono.pc$x),cex
#=0.7, lwd=2)

```

```

# ANALISIS CLUSTER

```

```

#Primero probamos el metodo de las Kmeans, pero no conseguimos
que nos de un numero de cluters

```

```

ozono.cl2=kmeans(ozono,2,20)
scdg2=sum(ozono.cl2$withinss)
scdg2

```

```

ozono.cl3=kmeans(ozono,3,20)
scdg3=sum(ozono.cl3$withinss)
scdg3

```

```

f23=(scdg2-scdg3)/(scdg3/(dim(ozono)[1]-2-1))
f23

```

```

ozono.cl4=kmeans(ozono,4,20)
scdg4=sum(ozono.cl4$withinss)
scdg4

```

```

f34=(scdg3-scdg4)/(scdg4/(dim(ozono)[1]-3-1))
f34

```

```
ozono.cl5=kmeans(ozono,5,20)
scdg5=sum(ozono.cl5$withinss)
scdg5

f45=(scdg4-scdg5)/(scdg5/(dim(ozono)[1]-4-1))
f45

ozono.cl6=kmeans(ozono,6,20)
scdg6=sum(ozono.cl6$withinss)
scdg6

f56=(scdg5-scdg6)/(scdg6/(dim(ozono)[1]-5-1))
f56
#Tendría que salirnos menor de 10

#Probamos el metodo jerarquico
ozono.hcl=hclust(dist(ozono),method="ward") #ojo con las
comillas, se queja
plot(ozono.hcl) #dendograma, nos quedariamos con 3 clusters

prueba=cutree(ozono.hcl,k=3) #crea un vector

#Generamos el archivo con las localizaciones y a que cluster
corresponde para pintar mapas.
localizaciones=1:3417
puntuaciones.clusters=rep(NA,3417)
puntuaciones.clusters[vector]=prueba
cluster<-data.frame(puntuaciones.clusters)
write.table(cluster, file="cluster.nc")
```

Código análisis de PM10

```

x=read.table("PM10.txt",na.strings='-9999')

#Creamos la matriz descartes donde metemos los valores NA
descartes=NULL;
temp2=NULL
for (i in 1:3417){
  temp=length(which(is.na(x[i,])))
  temp2=c(temp2,temp)
  if (temp>500){
    descartes=c(descartes,i);
  }
}

#Obtenemos un vector donde aparece a qué índice se corresponde
cada puntito con valores (los no "NA"). #Porque luego los
quitaremos y cambiarán de numeración en la nueva matriz, ya no
seria la "malla IP"

vector=1:3417
vector=vector[-descartes]

#Creamos una nueva matriz sin los NA:

xconv<- x[-descartes,]

#hago la transpuesta, por lo que las PCs seran las EOFs y
viceversa
pm10<-data.frame(t(xconv))

#Eliminamos la columna row.names
row.names(pm10) <- NULL

# ANALISIS DE COMPONENTES PRINCIPALES
#Aplicamos componentes principales

pm10.pc<-prcomp(pm10)
summary (pm10.pc)

#Lo que aparece tras este comando es entre otras cosas el
porcentaje de variabilidad acumulada.
#Con esa información nos quedaríamos con 6 PC porque acumulan
aprox 80% de la variabilidad.

#Pintamos el diagrama de codos
plot(pm10.pc$sdev^2, type="b", main="Gráfico de
sedimentación",xlab="Numero de componentes", ylab="Valores
propios")

#Como hemos hecho la transpuesta, para obtener las EOFs (o

```

puntuaciones) tenemos que usar el comando \$rotation, (que serian las PCs si no fuera la transpuesta).

```
PC_PM10<-pm10.pc$rotation
```

```
#Generamos los archivos de las 6 EOF y obtenemos la tendencia espacial.
```

```
#"vector" es un vector de los índices "no descartes"
```

```
localizaciones=1:3417
```

```
PC_PM10_1=rep(NA,3417)
```

```
PC_PM10_1[vector]=pm10.pc$rotation[,1]
```

```
PC_PM10_1<-data.frame(PC_PM10_1)
```

```
write.table(PC_PM10_1,file="EOF1_PM10.txt")
```

```
PC_PM10_2=rep(NA,3417)
```

```
PC_PM10_2[vector]=pm10.pc$rotation[,2]
```

```
PC_PM10_2<-data.frame(PC_PM10_2)
```

```
write.table(PC_PM10_2,file="EOF2_PM10.txt")
```

```
PC_PM10_3=rep(NA,3417)
```

```
PC_PM10_3[vector]=pm10.pc$rotation[,3]
```

```
PC_PM10_3<-data.frame(PC_PM10_3)
```

```
write.table(PC_PM10_3,file="EOF3_PM10.txt")
```

```
PC_PM10_4=rep(NA,3417)
```

```
PC_PM10_4[vector]=pm10.pc$rotation[,4]
```

```
PC_PM10_4<-data.frame(PC_PM10_4)
```

```
write.table(PC_PM10_4,file="EOF4_PM10.txt")
```

```
PC_PM10_5=rep(NA,3417)
```

```
PC_PM10_5[vector]=pm10.pc$rotation[,5]
```

```
PC_PM10_5<-data.frame(PC_PM10_5)
```

```
write.table(PC_PM10_5,file="EOF5_PM10.txt")
```

```
PC_PM10_6=rep(NA,3417)
```

```
PC_PM10_6[vector]=pm10.pc$rotation[,6]
```

```
PC_PM10_6<-data.frame(PC_PM10_6)
```

```
write.table(PC_PM10_6,file="EOF6_PM10.txt")
```

```
#Para obtener las PCs, hay que sacar las EOFs usando el comando $x. Asi obtendremos la tendencia temporal
```

```
PC_PM10<-PM10.pc$x
```

```
PC1_PM10=pm10.pc$x[,1]
```

```
PC1_PM10_df<-data.frame(PC1_PM10)
```

```
c<-(1:4378)
```

```
m<-(PC1_PM10_df$PC1_PM10)
```

```
xrange=c(0,4378)
```

```
yrange=c(-500,500)
```

```
plot(c,m,main="PM10 PC1", xlab="Excedences", ylab="pc",
```

```
type="s", xlim=xrange, ylim=yrange, col="black")
```

```
abline(lm(m~c), col="blue", lwd=5)
```

```
modelo=lm(m~c) # con esto sacamos el pvalor para ver si es
```

```

significativo
summary(modelo)

PC2_PM10=pm10.pc$x[,2]
PC2_PM10_df<-data.frame(PC2_PM10)
m2<-(PC2_PM10_df$PC2_PM10)
#plot(c,m2,main="PC2 trends", xlab="Excedences", ylab="pc",
type="s", col="black")
#abline(lm(m2~c), col="blue", lwd=5)

PC3_PM10=pm10.pc$x[,3]
PC3_PM10_df<-data.frame(PC3_PM10)
m3<-(PC3_PM10_df$PC3_PM10)
#plot(c,m3,main="PC3 trends", xlab="Excedences", ylab="pc",
type="s", col="black")
#abline(lm(m3~c), col="blue", lwd=5)
modelo=lm(m3~c) # con esto sacamos el pvalor para ver si es
significativo
summary(modelo)

PC4_PM10=pm10.pc$x[,4]
PC4_PM10_df<-data.frame(PC4_PM10)
m4<-(PC4_PM10_df$PC4_PM10)
#plot(c,m4,main="PC4 trends", xlab="Time", ylab="Anomalies",
type="s",xlim=xrange, ylim=yrange, col="black")
#abline(lm(m4~c), col="blue", lwd=5)
modelo=lm(m4~c) # con esto sacamos el pvalor para ver si es
significativo
summary(modelo)

PC5_PM10=pm10.pc$x[,5]
PC5_PM10_df<-data.frame(PC5_PM10)
m5<-(PC5_PM10_df$PC5_PM10)
#plot(c,m5,main="PC5 trends", xlab="Excedences", ylab="pc",
type="s", col="black")
#abline(lm(m5~c), col="blue", lwd=5)
modelo=lm(m5~c) # con esto sacamos el pvalor para ver si es
significativo
summary(modelo)

PC6_PM10=pm10.pc$x[,6]
PC6_PM10_df<-data.frame(PC6_PM10)
m6<-(PC6_PM10_df$PC6_PM10)
#plot(c,m6,main="PC6 trends", xlab="Excedences", ylab="pc",
type="s", col="black")
#abline(lm(m6~c), col="blue", lwd=5)
modelo=lm(m6~c) # con esto sacamos el pvalor para ver si es
significativo
summary(modelo)

```

```

#CLUSTER
#Primero probamos el metodo de las Kmeans, pero no conseguimos
que nos de un numero de cluters

pm10.cl2=kmeans(pm10,2,20)
scdg2=sum(pm10.cl2$withinss)
scdg2

pm10.cl3=kmeans(pm10,3,20)
scdg3=sum(pm10.cl3$withinss)
scdg3

f23=(scdg2-scdg3)/(scdg3/(dim(pm10)[1]-2-1))
f23

ozono.cl4=kmeans(pm10,4,20)
scdg4=sum(pm10.cl4$withinss)
scdg4

f34=(scdg3-scdg4)/(scdg4/(dim(pm10)[1]-3-1))
f34

pm10.cl5=kmeans(pm10,5,20)
scdg5=sum(pm10.cl5$withinss)
scdg5

f45=(scdg4-scdg5)/(scdg5/(dim(pm10)[1]-4-1))
f45

pm10.cl6=kmeans(pm10,6,20)
scdg6=sum(pm10.cl6$withinss)
scdg6

f56=(scdg5-scdg6)/(scdg6/(dim(pm10)[1]-5-1))
f56
#Tendría que salirnos menor de 10

#Probamos metodo jerarquico
pm10_inicial<-data.frame(t(pm10))
pm10.hcl=hclust(dist(pm10_inicial),method='ward') #ojo con las
comillas, se queja
plot(pm10.hcl) #dendograma, nos quedariamos con 3 clusters

#segun el dendograma, me quedaria 4 clusters.
cluster.vect=cutree(pm10.hcl,k=4) #crea un vector

#Generamos el archivo con las localizaciones y a que cluster
corresponde para pintar mapas.
localizaciones=1:3417
PM10.clusters=rep(NA,3417)
PM10.clusters[vector]=cluster.vect
cluster.PM10<-data.frame(PM10.clusters)
write.table(cluster.PM10, file="cluster4_PM10.nc")

```