



UNIVERSIDAD DE GRANADA

MÁSTER EN ESTADÍSTICA APLICADA

DEPARTAMENTO DE ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

UNIVERSIDAD DE GRANADA

TRABAJO FIN DE MÁSTER

ANÁLISIS DEL COMPORTAMIENTO DE LA CURVA POLÍNICA DEL TIPO *PLATANUS* MEDIANTE ANÁLISIS DE DATOS FUNCIONALES

Beatriz Lara Espinar

GRANADA, 2023

MÁSTER EN ESTADÍSTICA APLICADA
DEPARTAMENTO DE ESTADÍSTICA E INVESTIGACIÓN OPERATIVA
UNIVERSIDAD DE GRANADA

**ANÁLISIS DEL COMPORTAMIENTO DE LA CURVA POLÍNICA
DEL TIPO *PLATANUS* MEDIANTE ANÁLISIS DE DATOS
FUNCIONALES**

Trabajo Fin de Máster presentado por

Beatriz Lara Espinar

y dirigido por los profesores

Ana María Aguilera del Pino

Christian José Acal González

Declaración de autoría y originalidad del

TRABAJO FIN DE MÁSTER

Yo, Beatriz Lara Espinar, con documento nacional de identidad 03882270P, alumna del Máster de Estadística Aplicada de la Facultad de Ciencias de la Universidad de Granada, declaro que asumo la originalidad del Trabajo Fin de Máster presentado para su defensa y evaluación en el curso 2022-2023, entendida dicha originalidad en el sentido de que no se han empleado fuentes que no hayan sido citadas adecuadamente.

Granada, 05 de febrero de 2023

A handwritten signature in blue ink, consisting of several loops and a long horizontal stroke at the bottom.

Fdo: Beatriz Lara Espinar

Agradecimientos

En primer lugar quiero dar mis más sincero agradecimiento a la profesora Ana María Aguilera del Pino y al profesor Christian José Acal González por darme la oportunidad de desarrollar este trabajo con ellos. Gracias por vuestra buena disposición, vuestro optimismo ante el reto al que me enfrentaba y, en general, por vuestra labor como tutores. Para mi sois un ejemplo a seguir.

También quiero agradecer a la Red de Aerobiología de Castilla-La Mancha por cederme los datos necesarios para realizar este trabajo. A la profesora Rosa Pérez Badia, gracias por el apoyo y la confianza depositada.

No pueden faltar los agradecimientos a aquellas personas que me han mostrado su apoyo durante este tiempo, en concreto a mi madre, mi hermano, Juanjo, Cristina y Lola. En especial quiero agradecer a mi padre por su apoyo e insistencia en que retomase este trabajo, fruto de la promesa que te hice en su día y que ojalá pudieses ver cumplida.

ÍNDICE GENERAL

RESUMEN	1
1. INTRODUCCIÓN	4
2. DESCRIPCIÓN DE BASE DE DATOS.....	10
2.1. ÁREA DE ESTUDIO	10
2.2. DATOS AEROBIOLÓGICOS	11
2.3. DATOS METEOROLÓGICOS	16
3. METODOLOGÍA	17
3.1. REPRESENTACIÓN BÁSICA DE LAS CURVAS.....	17
3.2. RECONSTRUCCIÓN FUNCIONAL DE CURVAS ESTRICTAMENTE POSITIVAS	20
3.3. ANÁLISIS EN COMPONENTES PRINCIPALES FUNCIONAL	21
3.4. ANÁLISIS DE CORRELACIÓN.....	23
4. RESULTADOS.....	24
4.1. RECONSTRUCCIÓN DE LA FORMA FUNCIONAL	24
4.2. ANÁLISIS EN COMPONENTES PRINCIPALES FUNCIONAL	29
4.3. ANÁLISIS DE CORRELACIONES	35
5. CONCLUSIONES.....	41
6. REFERENCIAS	43
7. ANEXO	50

ÍNDICE DE FIGURAS

FIGURA 1. TEMPERATURA MEDIA REGISTRADA EN TOLEDO DURANTE EL AÑO 2016. DATOS DISCRETOS OBSERVADOS (IZQUIERDA) Y FORMA FUNCIONAL UTILIZANDO B-SPLINES (DERECHA) DE GRADO 3 CON UNA DIMENSIÓN DE 14 BASES.....	7
FIGURA 2. PLATANUS ORIENTALIS L. VAR. ACERIFOLIA DRYAND IN AITON.....	8
FIGURA 3. LOCALIZACIÓN DEL CAPTADOR DE POLEN Y DE LA ESTACIÓN METEOROLÓGICA	10
FIGURA 4. CAPTADOR VOLUMÉTRICO TIPO HIRST	11
FIGURA 5. CONCENTRACIÓN ATMOSFÉRICA DEL TIPO POLÍNICO PLATANUS (GRANOS DE POLEN/M ³) DURANTE EL PERIODO DE ESTUDIO. IZQUIERDA: CONCENTRACIÓN POLÍNICA DIARIA DURANTE EL AÑO COMPLETO. DERECHA: CONCENTRACIÓN POLÍNICA DIARIA DURANTE LOS MESES DE MARZO-MAYO CON ZOOM EN EL EJE Y. LÍNEA NEGRA INDICA EL COMPORTAMIENTO MEDIO DE LA ESTACIÓN POLÍNICA PARA LOS AÑOS 2005-2018.....	13
FIGURA 6. CALIDAD DE LA BASE DE DATOS DEL TIPO POLÍNICO PLATANUS.....	14
FIGURA 7. MÉTODOS DE INTERPOLACIÓN APLICADOS A LOS DATOS DE POLEN DE PLATANUS PARA EL AÑO 2012. LAS LÍNEAS ROJAS VERTICALES MUESTRAN EL PERIODO INTERPOLADO.....	15
FIGURA 8. EVOLUCIÓN DE LA CONCENTRACIÓN DIARIA DE POLEN DE PLATANUS EN LA CIUDAD DE TOLEDO DURANTE EL PERIODO 2005-2018.....	25
FIGURA 9. RECONSTRUCCIÓN FUNCIONAL DE LA EVOLUCIÓN DE LA CONCENTRACIÓN DIARIA DE POLEN DE PLATANUS EN LA CIUDAD DE TOLEDO DURANTE EL PERIODO 2005-2018.....	25
FIGURA 10. EVOLUCIÓN DE LA CONCENTRACIÓN POLÍNICA DE PLATANUS DURANTE LOS AÑOS 2005, 2009 Y 2018 EN LA CIUDAD DE TOLEDO. LÍNEA NEGRA REPRESENTA LOS DATOS DISCRETOS ORIGINALES Y LA LÍNEA AZUL REPRESENTA LOS DATOS DISCRETOS OBTENIDOS AL EVALUAR LA FORMA FUNCIONAL.	26
FIGURA 11. EVOLUCIÓN DE LOS DATOS DISCRETOS OBTENIDOS AL EVALUAR LOS DATOS FUNCIONALES DE LA VARIABLE POLEN. LÍNEA CONTINUA REPRESENTA LA MEDIA Y LAS LÍNEAS DISCONTINUAS ROJAS REPRESENTAN EL INTERVALO DE CONFIANZA.....	28
FIGURA 12. FUNCIÓN MEDIA (LÍNEA NEGRA) E INTERVALOS DE CONFIANZA (LÍNEAS ROJAS) DE LAS VARIABLES METEOROLÓGICAS ANALIZADAS. EN EL CASO DE LA PRECIPITACIÓN, SE REPRESENTA LA MEDIA DE LOS DATOS DISCRETOS OBTENIDOS AL EVALUAR LOS DATOS FUNCIONALES.....	29
FIGURA 13. REPRESENTACIÓN DE LAS CINCO PRIMERAS AUTOFUNCIONES ASOCIADAS A LAS COMPONENTES FUNCIONALES DEL LOGARITMO DE LA VARIABLE POLEN. LÍNEA CONTINUA REPRESENTA LA FUNCIÓN MEDIA	

MUESTRAL (ESCALA LOGARÍTMICA) Y LAS LÍNEAS DISCONTINUAS REPRESENTAN PERTURBACIONES RESPECTO A LA FUNCIÓN MEDIA MUESTRAL (ESCALA LOGARÍTMICA).	31
FIGURA 14. GRÁFICO DE DISPERSIÓN ENTRE LOS VALORES DE LA PRIMERA COMPONENTE (EJE X) Y LOS VALORES DE LA SEGUNDA COMPONENTE (EJE Y).	32
FIGURA 15. GRÁFICO DE DISPERSIÓN ENTRE LOS VALORES DE LA PRIMERA COMPONENTE (EJE X) Y LOS VALORES DE LA TERCERA COMPONENTE (EJE Y).	33
FIGURA 16. GRÁFICO DE DISPERSIÓN ENTRE LOS VALORES DE LA PRIMERA COMPONENTE (EJE X) Y LOS VALORES DE LA CUARTA COMPONENTE (EJE Y).	34
FIGURA 17. REPRESENTACIÓN DE LA PRIMERA COMPONENTE FUNCIONAL DEL POLEN (ESCALA LOGARÍTMICA) Y LAS COMPONENTES PRINCIPALES DE LAS VARIABLES METEOROLÓGICAS CON LAS QUE POSEE CORRELACIÓN SIGNIFICATIVA (P,VALOR<0.05). LÍNEA CONTINUA REPRESENTA LA FUNCIÓN MEDIA MUESTRAL Y LAS LÍNEAS DISCONTINUAS REPRESENTAN PERTURBACIONES RESPECTO A LA FUNCIÓN MEDIA MUESTRAL.	37
FIGURA 18. REPRESENTACIÓN DE LA SEGUNDA COMPONENTE FUNCIONAL DEL POLEN (ESCALA LOGARÍTMICA) Y LAS COMPONENTES PRINCIPALES DE LAS VARIABLES METEOROLÓGICAS CON LAS QUE POSEE SIGNIFICATIVA (P.VALOR<0,.05). LÍNEA CONTINUA REPRESENTA LA FUNCIÓN MEDIA MUESTRAL Y LAS LÍNEAS DISCONTINUAS REPRESENTAN PERTURBACIONES RESPECTO A LA FUNCIÓN MEDIA MUESTRAL.	38
FIGURA 19. REPRESENTACIÓN DE LA TERCERA COMPONENTE FUNCIONAL DEL POLEN (ESCALA LOGARÍTMICA) Y LAS COMPONENTES PRINCIPALES DE LAS VARIABLES METEOROLÓGICAS CON LAS QUE POSEE CORRELACIÓN SIGNIFICATIVA (P.VALOR<0,05). LÍNEA CONTINUA REPRESENTA LA FUNCIÓN MEDIA MUESTRAL Y LAS LÍNEAS DISCONTINUAS REPRESENTAN PERTURBACIONES RESPECTO A LA FUNCIÓN MEDIA MUESTRAL.	39
FIGURA 20. REPRESENTACIÓN DE LA CUARTA COMPONENTE FUNCIONAL DEL POLEN Y LAS COMPONENTES PRINCIPALES DE LAS VARIABLES METEOROLÓGICAS CON LAS QUE POSEE CORRELACIÓN SIGNIFICATIVA (P.VALOR<0.05). LÍNEA CONTINUA REPRESENTA LA FUNCIÓN MEDIA MUESTRAL Y LAS LÍNEAS DISCONTINUAS REPRESENTAN PERTURBACIONES RESPECTO A LA FUNCIÓN MEDIA MUESTRAL.	40
FIGURA 21. EVOLUCIÓN DE LAS VARIABLES ANALIZADAS DURANTE EL AÑO 2005 EN LA CIUDAD DE TOLEDO. LÍNEA NEGRA REPRESENTA LOS DATOS DISCRETOS ORIGINALES, LA LÍNEA AZUL REPRESENTA LOS DATOS	

DISCRETOS OBTENIDOS AL EVALUAR LA FORMA FUNCIONAL Y LA LÍNEA ROJA REPRESENTA LA FORMA FUNCIONAL.....	50
FIGURA 22. EVOLUCIÓN DE LAS VARIABLES ANALIZADAS DURANTE EL AÑO 2006 EN LA CIUDAD DE TOLEDO. LÍNEA NEGRA REPRESENTA LOS DATOS DISCRETOS ORIGINALES, LA LÍNEA AZUL REPRESENTA LOS DATOS DISCRETOS OBTENIDOS AL EVALUAR LA FORMA FUNCIONAL Y LA LÍNEA ROJA REPRESENTA LA FORMA FUNCIONAL.....	51
FIGURA 23. EVOLUCIÓN DE LAS VARIABLES ANALIZADAS DURANTE EL AÑO 2007 EN LA CIUDAD DE TOLEDO. LÍNEA NEGRA REPRESENTA LOS DATOS DISCRETOS ORIGINALES, LA LÍNEA AZUL REPRESENTA LOS DATOS DISCRETOS OBTENIDOS AL EVALUAR LA FORMA FUNCIONAL Y LA LÍNEA ROJA REPRESENTA LA FORMA FUNCIONAL.....	52
FIGURA 24. EVOLUCIÓN DE LAS VARIABLES ANALIZADAS DURANTE EL AÑO 2008 EN LA CIUDAD DE TOLEDO. LÍNEA NEGRA REPRESENTA LOS DATOS DISCRETOS ORIGINALES, LA LÍNEA AZUL REPRESENTA LOS DATOS DISCRETOS OBTENIDOS AL EVALUAR LA FORMA FUNCIONAL Y LA LÍNEA ROJA REPRESENTA LA FORMA FUNCIONAL.....	53
FIGURA 25. EVOLUCIÓN DE LAS VARIABLES ANALIZADAS DURANTE EL AÑO 2009 EN LA CIUDAD DE TOLEDO. LÍNEA NEGRA REPRESENTA LOS DATOS DISCRETOS ORIGINALES, LA LÍNEA AZUL REPRESENTA LOS DATOS DISCRETOS OBTENIDOS AL EVALUAR LA FORMA FUNCIONAL Y LA LÍNEA ROJA REPRESENTA LA FORMA FUNCIONAL.....	54
FIGURA 26. EVOLUCIÓN DE LAS VARIABLES ANALIZADAS DURANTE EL AÑO 2010 EN LA CIUDAD DE TOLEDO. LÍNEA NEGRA REPRESENTA LOS DATOS DISCRETOS ORIGINALES, LA LÍNEA AZUL REPRESENTA LOS DATOS DISCRETOS OBTENIDOS AL EVALUAR LA FORMA FUNCIONAL Y LA LÍNEA ROJA REPRESENTA LA FORMA FUNCIONAL.....	55
FIGURA 27. EVOLUCIÓN DE LAS VARIABLES ANALIZADAS DURANTE EL AÑO 2011 EN LA CIUDAD DE TOLEDO. LÍNEA NEGRA REPRESENTA LOS DATOS DISCRETOS ORIGINALES, LA LÍNEA AZUL REPRESENTA LOS DATOS DISCRETOS OBTENIDOS AL EVALUAR LA FORMA FUNCIONAL Y LA LÍNEA ROJA REPRESENTA LA FORMA FUNCIONAL.....	56
FIGURA 28. EVOLUCIÓN DE LAS VARIABLES ANALIZADAS DURANTE EL AÑO 2012 EN LA CIUDAD DE TOLEDO. LÍNEA NEGRA REPRESENTA LOS DATOS DISCRETOS ORIGINALES, LA LÍNEA AZUL REPRESENTA LOS DATOS DISCRETOS OBTENIDOS AL EVALUAR LA FORMA FUNCIONAL Y LA LÍNEA ROJA REPRESENTA LA FORMA FUNCIONAL.....	57

FIGURA 29. EVOLUCIÓN DE LAS VARIABLES ANALIZADAS DURANTE EL AÑO 2013 EN LA CIUDAD DE TOLEDO.	
LÍNEA NEGRA REPRESENTA LOS DATOS DISCRETOS ORIGINALES, LA LÍNEA AZUL REPRESENTA LOS DATOS DISCRETOS OBTENIDOS AL EVALUAR LA FORMA FUNCIONAL Y LA LÍNEA ROJA REPRESENTA LA FORMA FUNCIONAL.....	58
FIGURA 30. EVOLUCIÓN DE LAS VARIABLES ANALIZADAS DURANTE EL AÑO 2014 EN LA CIUDAD DE TOLEDO.	
LÍNEA NEGRA REPRESENTA LOS DATOS DISCRETOS ORIGINALES, LA LÍNEA AZUL REPRESENTA LOS DATOS DISCRETOS OBTENIDOS AL EVALUAR LA FORMA FUNCIONAL Y LA LÍNEA ROJA REPRESENTA LA FORMA FUNCIONAL.....	59
FIGURA 31. EVOLUCIÓN DE LAS VARIABLES ANALIZADAS DURANTE EL AÑO 2015 EN LA CIUDAD DE TOLEDO.	
LÍNEA NEGRA REPRESENTA LOS DATOS DISCRETOS ORIGINALES, LA LÍNEA AZUL REPRESENTA LOS DATOS DISCRETOS OBTENIDOS AL EVALUAR LA FORMA FUNCIONAL Y LA LÍNEA ROJA REPRESENTA LA FORMA FUNCIONAL.....	60
FIGURA 32. EVOLUCIÓN DE LAS VARIABLES ANALIZADAS DURANTE EL AÑO 2016 EN LA CIUDAD DE TOLEDO.	
LÍNEA NEGRA REPRESENTA LOS DATOS DISCRETOS ORIGINALES, LA LÍNEA AZUL REPRESENTA LOS DATOS DISCRETOS OBTENIDOS AL EVALUAR LA FORMA FUNCIONAL Y LA LÍNEA ROJA REPRESENTA LA FORMA FUNCIONAL.....	61
FIGURA 33. EVOLUCIÓN DE LAS VARIABLES ANALIZADAS DURANTE EL AÑO 2017 EN LA CIUDAD DE TOLEDO.	
LÍNEA NEGRA REPRESENTA LOS DATOS DISCRETOS ORIGINALES, LA LÍNEA AZUL REPRESENTA LOS DATOS DISCRETOS OBTENIDOS AL EVALUAR LA FORMA FUNCIONAL Y LA LÍNEA ROJA REPRESENTA LA FORMA FUNCIONAL.....	62
FIGURA 34. EVOLUCIÓN DE LAS VARIABLES ANALIZADAS DURANTE EL AÑO 2018 EN LA CIUDAD DE TOLEDO.	
LÍNEA NEGRA REPRESENTA LOS DATOS DISCRETOS ORIGINALES, LA LÍNEA AZUL REPRESENTA LOS DATOS DISCRETOS OBTENIDOS AL EVALUAR LA FORMA FUNCIONAL Y LA LÍNEA ROJA REPRESENTA LA FORMA FUNCIONAL.....	63

ÍNDICE DE TABLAS

TABLA 1. CARACTERÍSTICAS MÁS RELEVANTES DE LA ESTACIÓN POLÍNICA DEL TIPO PLATANUS	13
TABLA 2. VALORES PERDIDOS EN LA BASE HISTÓRICA DEL POLEN DE PLATANUS	15
TABLA 3. VALORES DE LAS COMPONENTES PRINCIPALES SELECCIONADAS PARA LA VARIABLE POLEN	29
TABLA 4. VALORES DE LAS COMPONENTES PRINCIPALES SELECCIONADAS PARA LAS VARIABLES METEOROLÓGICAS: PRECIPITACIÓN (PREC), TEMPERATURA MEDIA (TMED), HUMEDAD RELATIVA (HREL) Y HORAS DE SOL (HSOL).....	34
TABLA 5. RESUMEN DE LAS CORRELACIONES SIGNIFICATIVAS OBTENIDAS POR EL TEST NO PARAMÉTRICO DE SPEARMAN (P.VALOR<0,05) ENTRE LAS COMPONENTES PRINCIPALES DEL POLEN Y LAS COMPONENTES PRINCIPALES DE LAS VARIABLES METEOROLÓGICAS: PRECIPITACIÓN (PREC), TEMPERATURA MEDIA (TMED), TEMPERATURA MÍNIMA (TMIN), HUMEDAD RELATIVA (HREL) Y HORAS DE SOL (HSOL).	35

Resumen

En este Trabajo Fin de Máster se presenta un análisis en componentes principales funcional aplicado a datos aerobiológicos de concentración de polen de *Platanus*. Este tipo de polen procede de una especie ornamental ampliamente cultivada en los parques y jardines de las ciudades. Este tipo polínico ha sido seleccionado por presentar una intensa y explosiva floración que hace que se emitan elevadas concentraciones de polen a la atmósfera en un corto periodo de tiempo, por lo que su curva polínica tiene a presentar un marcado pico de concentración de polen que permite identificar adecuadamente los periodos relativos a la emisión y dispersión del polen.

El objetivo de este trabajo es analizar el comportamiento de la curva polínica aplicando análisis de datos funcionales y explicar la influencia que las variables meteorológicas pueden ejercer sobre esos comportamientos.

Las variables analizadas han sido la concentración de polen y diferentes variables meteorológicas, que posteriormente serán utilizadas para interpretar el comportamiento de la curva polínica. El periodo de análisis ha comprendido los años 2005-2018, acotando para cada año el periodo desde el inicio de la estación polínica de *Platanus* hasta los cuarenta días siguientes, concentrando así más del 90% la concentración polínica total anual.

La metodología seguida ha consistido en realizar una reconstrucción de la forma funcional tanto en la variable de polen como en las diferentes variables meteorológicas. Una vez obtenida la forma funcional de las variables, se ha aplicado un análisis en componentes principales a todas las variables para simplificar los comportamientos que presentan durante el periodo de estudio. Una vez seleccionado el número de componentes

principales adecuado, se ha analizado la influencia de las componentes principales de las variables meteorológicas sobre las componentes principales de la variable polen utilizando el test de correlación no paramétrico de Spearman.

Los resultados del análisis en componentes principales reflejan que, en la ciudad de Toledo, la curva polínica de *Platanus* presenta diferentes comportamientos relativos, por un lado, a la concentración de polen durante el periodo prepico o durante el periodo postpico; por otro lado, relativos a la amplitud de la curva de polen y, por último, a la fecha en la que se producen los picos de mayores concentraciones.

Respecto a la influencia de las variables meteorológicas, los resultados obtenidos permiten identificar a la temperatura media y a la precipitación como las variables meteorológicas que producen un efecto en la concentración de polen durante el periodo prepico o durante el periodo postpico. Así, los años en los que el polen registrado se concentra en el periodo prepico se debe a la ausencia de precipitaciones durante este periodo y a un aumento de la temperatura media durante el periodo de mayor emisión polínica. Por el contrario, la concentración de polen se produce principalmente en el periodo pico y postpico cuando se produce eventos de precipitación en el inicio de la estación polínica y un aumento brusco de la temperatura media durante el periodo postpico.

Respecto a la amplitud de la curva polínica, los resultados indican que está principalmente relacionada con temperatura media y las horas de sol. La presencia de mayores valores de temperatura media y horas de sol, con respecto a la función media, durante el periodo inicial y final de la estación polínica favorece la emisión de concentraciones de polen generando curvas polínicas más amplias.

Por último, la fecha en la que se produce el pico de máxima concentración de polen está relacionado negativamente con la precipitación que se registra durante el inicio de la estación polínica. Lo que se traduce en que los eventos de precipitación al inicio de la estación polínica producen un retraso en el registro de los picos de concentración de polen.

1. Introducción

El Análisis de Datos Funcionales considera los datos objeto de análisis como un conjunto de curvas, en lugar de una secuencia de observaciones individuales (Ramsay & Silverman, 2002). Por tanto, un conjunto de datos funcionales proporciona información sobre funciones continuas obtenidas a partir de observaciones de un proceso estocástico continuo en el tiempo. La principal ventaja del análisis de datos funcionales frente a las técnicas multivariantes clásicas reside en la posibilidad de analizar el comportamiento completo de una función en lugar de trabajar con un conjunto de observaciones discretas en el tiempo, de forma que no se pierde información esencial como puede ser la continuidad o la suavidad de las curvas.

Según Ramsay & Silverman (2005) los objetivos del análisis de datos funcionales son:

- Representar los datos de forma que se facilite su estudio.
- Mostrar los datos de forma que se pongan de manifiesto sus características.
- Estudiar las principales fuentes de patrones y variabilidad entre los datos.
- Explicar la variación de un resultado o variable dependiente a partir de una entrada o variable independiente.
- Comparar dos o más conjuntos de datos respecto a ciertos tipos de variación.

Para la consecución de estos objetivos es necesario aplicar técnicas apropiadas a cada uno de ellos, que pueden dividirse en técnicas exploratorias, técnicas confirmatorias o técnicas predictivas. En el modo exploratorio, las técnicas aplicadas se centran en los datos disponibles y deben permitir relevar aspectos nuevos e interesantes de los datos. Los principales métodos exploratorios más empleados son los componentes principales funcionales y la correlación canónica. Los análisis confirmatorios poseen unas hipótesis de partida y las técnicas aplicadas buscan dar respuesta a dichas hipótesis. Un ejemplo de

estas técnicas son los modelos lineales funcionales. Por otro lado, los estudios predictivos utilizan técnicas que permitan, con los datos disponibles, proporcionar información sobre datos no observados como son los datos a futuro.

Asumiendo que los datos funcionales proceden de una cantidad finita de observaciones discretas, es necesario reconstruir la forma funcional de manera que permita su evaluación en cualquier instante de tiempo (Ramsay & Silverman, 2005). La reconstrucción de la forma funcional puede llevarse a cabo mediante técnicas no paramétricas (Ferraty and Vieu, 2006) o a través de la expansión básica o representación en bases de funciones (Ramsay and Silverman, 2005), siendo esta segunda la más conocida y la que será aplicada en este trabajo.

Según la naturaleza de las curvas y las observaciones se dispone de un gran número de bases que pueden emplearse para realizar la reconstrucción de la forma funcional por el método propuesto por Ramsay & Silverman (2005). Las más utilizadas son:

- Funciones Bases de Fourier. En este caso las bases utilizadas son senos y cosenos de frecuencia creciente.

$$\{1, \text{sen}(wt), \text{cos}(wt), \text{sen}(2wt), \text{cos}(2wt), \dots\}$$

Esta base es periódica y w determina el periodo $2\pi/w$. Este tipo de bases son muy útiles cuando las trayectorias que se quieren aproximar son muy estables, no presentan un comportamiento local fuerte y la curvatura de la función es estable y constante en todo el intervalo analizado.

- Funciones Bases de B-Splines. El sistema de bases B-Spline más popular es el desarrollado por De Boor (2001). Un B-Spline está formado por trozos de polinomios conectados entre sí de manera suave y donde las derivadas hasta orden

$q - 1$ son continuas en los puntos de unión. Según este algoritmo, la función spline se construye a partir de una combinación lineal de B-Spline base $B_{j,q}$. Denotando los nodos de definición por $\tau_0 < \dots < \tau_r$, y extendiendo esta partición en el dominio como $\tau_{-q} < \dots < \tau_{-2} < \tau_{-1} < \tau_0 < \dots < \tau_r < \tau_{r+1} < \tau_{r+2} < \dots < \tau_{r+q}$, los B-Spline de orden $q + 1$ (grado q) se definen iterativamente como:

$$B_{j,q+1}(\tau) = \frac{\tau - \tau_{j-2}}{\tau_{j+q-2} - \tau_{j-2}} B_{j,q}(\tau) + \frac{\tau_{j+p-1} - \tau}{\tau_{j+q-1} - \tau_{j-1}} B_{j+1,q}(\tau),$$

$$q = 1, 2, \dots; j = -1, 0, 1, \dots, r - q + 4,$$

con

$$B_{j,1}(\tau) = \begin{cases} 1 & \text{si } \tau_{j-2} \leq \tau \leq \tau_{j-1}, j = -1, 0, 1, \dots, r + 4, \\ 0 & \text{en caso contrario} \end{cases}$$

siendo $r + 4$ el número de nodos en el periodo de observación T .

Este tipo de bases se emplea principalmente cuando se trabaja con datos no periódicos y presentan buenos resultados cuando el comportamiento local es suave.

En este trabajo se trabajará con una base B-spline cúbica ($q = 3$).

Otro tipo de bases son las funciones indicadoras, las funciones polinómicas, las funciones exponenciales, las funciones potenciales o las funciones bases Wavelets. Un ejemplo de la obtención de la forma funcional a partir de datos discretos se muestra en la Figura 1.

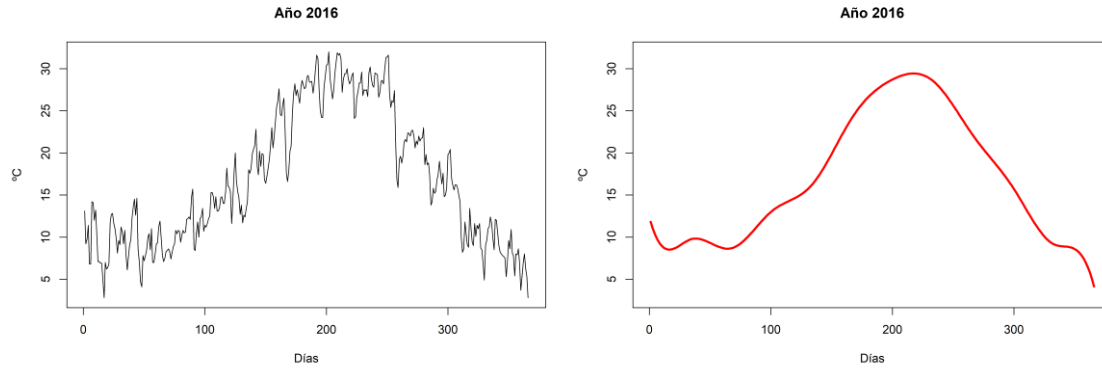


Figura 1. Temperatura media registrada en Toledo durante el año 2016. Datos discretos observados (izquierda) y forma funcional utilizando B-Splines (derecha) de grado 3 con una dimensión de 14 bases.

El primer trabajo que aplica el análisis de datos funcionales se publicó en 1995 y analiza la fuerza aplicada durante un pellizco. La justificación de este tipo de análisis reside en que un gráfico de impulsos de fuerza se parece a una función de densidad logarítmica (Ramsay et al., 1995). Desde esta publicación, el 75% de los artículos publicados que utilizan técnicas de análisis de datos funcionales se han publicado desde 2005 (Ullah and Finch, 2013). El análisis de datos funcionales tiene aplicación en todos aquellos campos científicos que analicen la evolución de una variable a lo largo del tiempo. Algunos ejemplos de campos científicos donde se han empleado estas técnicas son: biomecánica (Acal and Aguilera, 2022; Aguilera-Morillo and Aguilera, 2019; Crane et al., 2010), biomedicina (Hongtu Zhu et al., 2010; Wu and Müller, 2010), electrónica (Aguilera-Morillo et al., 2019), química (Hutchinson et al., 2004), medioambiente (Acal et al., 2022), meteorología (Curceac et al., 2019; Sözen and Öner, 2022) y aerobiología (Escabias et al., 2013; Valderrama et al., 2010), entre otros.

Por todo lo comentado anteriormente, en este trabajo, se ha planteado la aplicación del análisis de datos funcionales en datos de concentración de polen. El seguimiento, a largo plazo, de las concentraciones atmosféricas de polen permite identificar las principales fuentes de emisión (Rojo et al., 2015), obtener información sobre la fenología de las

plantas y su productividad (Fernández-Llamazares et al., 2014; García-Mozo, 2011), conocer la influencia de las variables meteorológicas sobre las concentraciones polínicas y generar modelos de predicción.

En este trabajo se ha seleccionado el tipo polínico *Platanus*, este tipo de polen procede de la especie mayoritaria del género *Platanus* en España, *Platanus orientalis* L. var. *acerifolia* Dryand in Aiton (plátano de sombra) (Figura 2). Se trata de una especie ampliamente utilizada como especie ornamental y puede encontrarse en paseos y avenidas o en parques y jardines.

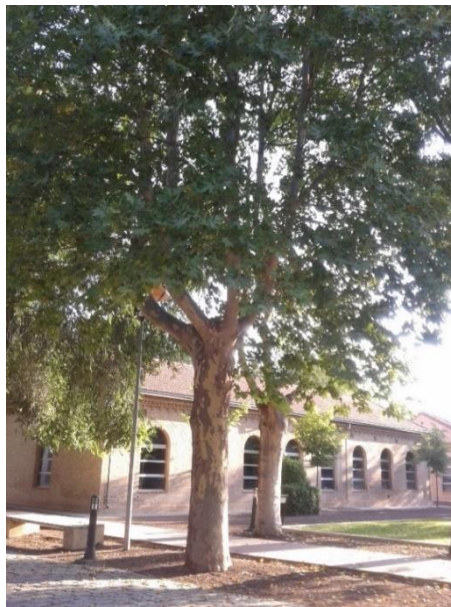


Figura 2. *Platanus orientalis* L. var. *acerifolia* Dryand in Aiton

El tipo polínico *Platanus* es un importante polen alergénico en áreas mediterráneas como España y el sur de Francia durante los meses de marzo hasta mayo (D'Amato et al., 2007). Dos importantes ejemplos del uso masivo de esta especie son las ciudades de Madrid y Barcelona, donde se pueden encontrar hasta 60000 y 150000 árboles de esta especie, respectivamente (Cariñanos and Casares-Porcel, 2011). El polen de *Platanus* se ha definido como un importante causante de polinosis en Madrid (Subiza et al., 1995; Varela et al., 1997) y Barcelona (Enrique et al., 2002). Respecto a la prevalencia de este

tipo de polen, en la ciudad de Madrid existe una prevalencia del 52% (Subiza et al., 1995) mientras que en Barcelona es del 37% (Puiggròs et al., 2015). En la Península Ibérica, la sensibilización al género *Platanus* es mayor en el centro que en otras zonas de la Península, con una prevalencia del 35%, siendo las personas comprendidas entre los 21 y los 30 años el grupo más afectado (Pereira et al., 2006). En Toledo, a pesar de la corta temporada de polen de *Platanus*, que comienza a finales de marzo y termina en abril, durante varios días se supera el umbral de “alto riesgo” (Pérez-Badia et al., 2010) que, según los criterios de la Red Española de Aerobiología, se alcanza cuando se supera una concentración de 200 granos de polen/m³.

Este tipo polínico presenta una intensa y explosiva floración, lo que se traduce en una elevada emisión de granos de polen a la atmósfera en un corto periodo de tiempo. La característica de la curva polínica es que suele presentar un pico de máxima concentración que permite dividir la estación polínica en dos periodos: el periodo prepico en el que las variables meteorológicas ejercen su influencia sobre la emisión y transporte de los granos de polen y un periodo postpico, tras el día de máxima concentración, que se caracteriza porque el máximo de emisión polínica ha terminado y es menos dependiente de las variables meteorológicas, que influyen principalmente en el transporte de polen.

Por todo lo comentado anteriormente, el objetivo principal de este trabajo es analizar la dinámica de la curva de polen de *Platanus* mediante el análisis en componentes principales funcional, así como identificar las variables meteorológicas que afectan a dicho comportamiento a través de un estudio de correlación.

2. Descripción de base de datos

2.1. Área de estudio

Este trabajo se ha llevado a cabo en la ciudad de Toledo, situada en el centro de la península Ibérica (39° 51' N, 4° 02' W) a una altitud de 529 metros sobre el nivel del mar. Toledo presenta un clima mediterráneo de tendencia continental. La temperatura media anual es de 15,8 °C, que oscila entre los 6,4 °C del mes más frío (enero) y los 26,8 °C del más cálido (julio). Respecto a la precipitación, la media anual es de 342 mm y se concentra durante los meses de abril a mayo y de noviembre a diciembre (AEMET, 2010). El bioclima se clasifica como mediterráneo pluviestacional-oceánico, con un termotipo mesomediterráneo y un ombroclima seco, caracterizado por un marcado déficit hídrico (Rivas-Martínez et al., 2011).

Las bases de datos utilizadas en este trabajo proceden de dos fuentes diferentes. Por un lado, disponemos de datos aerobiológicos pertenecientes a la Red de Aerobiología de Castilla-La Mancha (AEROCAM) y, por otro lado, los datos meteorológicos que pertenecen a la Agencia Estatal de Meteorología (AEMET) (Figura 3).

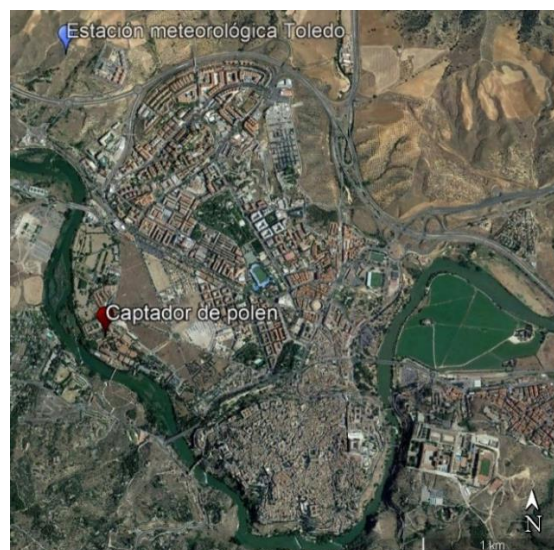


Figura 3. Localización del captador de polen y de la estación meteorológica

2.2. Datos aerobiológicos

El muestreo de polen aerotransportado se ha realizado durante el periodo 2005-2018 utilizando un captador volumétrico tipo Hirst (Hirst, 1952) situado a 17 metros sobre el nivel del suelo (Figura 4), en la terraza de uno de los edificios de la Universidad de Castilla-La Mancha en el Campus de Toledo (Figura 3). Este captador tiene un volumen de succión de 10 L/min, volumen de succión similar al de inhalación de aire del pulmón humano. El material particulado, aspirado por una bomba de succión, queda retenido al impactar contra una cinta Melinex (recubierta por una sustancia adherente) que tiene un avance diario de 2 mm/hora, lo que permite realizar el recuento diario de las partículas aerobiológicas.



Figura 4. Captador volumétrico tipo Hirst

La recogida, preparación y el recuento de las muestras se ha llevado a cabo semanalmente siguiendo la metodología propuesta por la Red Española de Aerobiología (REA) y los requerimientos propuestos por la Asociación Internacional de Aerobiología (Galán et al., 2014, 2007). El análisis de las muestras se realiza al microscopio óptico con el objetivo de 40x. La Red Española de Aerobiología especifica que, para cada

preparación, deben realizarse los recuentos en cuatro barridos horizontales continuos a lo largo de toda la preparación, lo que representa el análisis de un 10% del total de la superficie. El recuento de los granos de polen al microscopio requiere de personal especializado, capaz de identificar los diferentes tipos polínicos que puedan aparecer en la muestra.

La serie de datos aerobiológicos contiene los valores diarios de la concentración de polen de *Platanus* expresados en granos de polen por metro cúbico de aire (granos de polen/m³) para el periodo 2005-2018. El número de días analizados en la serie de datos aerobiológicos se ha reducido al periodo de polinización principal, para lo cual se ha calculado la fecha de inicio de la estación polínica de *Platanus*. El inicio de la estación polínica se ha considerado como el primer día con al menos 1 grano de polen/m³ de *Platanus* seguido de cinco días con una concentración igual o superior (García-Mozo et al., 1999).

El periodo principal de polinización o estación polínica, que es el periodo del año en el que se alcanzan concentraciones relevantes de un tipo polínico en el aire, para el tipo polínico *Platanus* en la ciudad de Toledo (Figura 5, Tabla 1) se inicia en torno al 21 de marzo ($\pm 5,8$ días) y finaliza el 14 de mayo ($\pm 20,5$ días). Presenta una duración media de 55,4 ($\pm 20,7$) días y el pico de polen o día en el que se registra la máxima concentración polínica tiene lugar el 01 de abril ($\pm 6,5$ días).

Para el análisis de series de tiempo funcional es necesario disponer de datos con la misma duración para todos los años, por tanto, a partir del inicio de la estación polínica se han considerado los 40 días siguientes para todos los años estudiados. Este periodo de tiempo permite concentrar los niveles de polen relevantes registrados durante el transcurso de la estación polínica (Lara et al., 2019), en concreto, se concentra más del

90% la concentración total anual registrada o Índice Polínico Anual en 12 de los 14 años estudiados (Tabla 1).

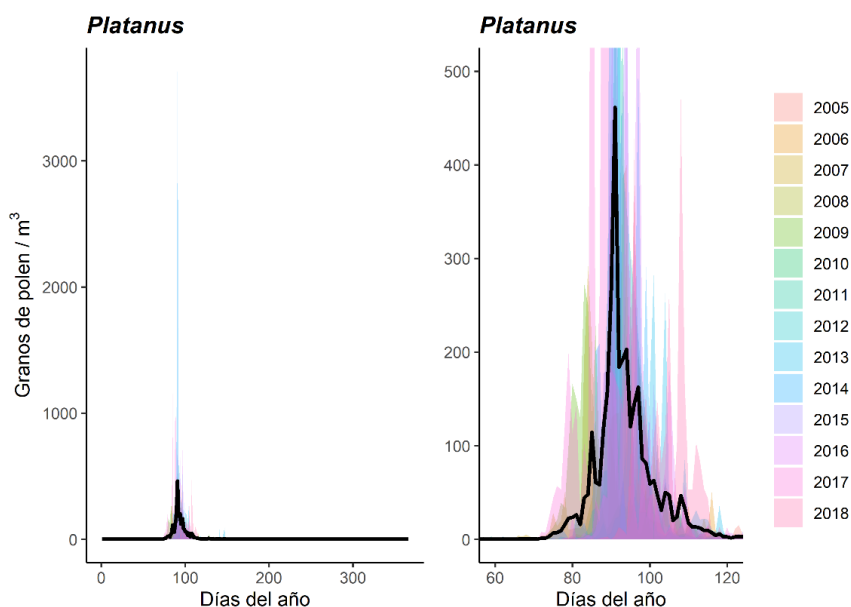


Figura 5. Concentración atmosférica del tipo polínico *Platanus* (granos de polen/m³) durante el periodo de estudio. Izquierda: concentración polínica diaria durante el año completo. Derecha: concentración polínica diaria durante los meses de marzo-mayo con zoom en el eje y. Línea negra indica el comportamiento medio de la estación polínica para los años 2005-2018.

Tabla 1. Características más relevantes de la estación polínica del tipo *Platanus*

Año	Inicio estación polínica	Duración estación polínica	Fecha día pico	Valor día pico	Índice Polínico Anual	Polen durante el periodo de 40 días considerado (%)
2005	2005-03-21	55	2005-04-01	402	2190	94.29
2006	2006-03-24	35	2006-04-04	400	1916	99.69
2007	2007-03-12	55	2007-03-25	292	2203	93.01
2008	2008-03-18	26	2008-03-30	331	1934	96.23
2009	2009-03-15	84	2009-03-24	273	2171	91.57
2010	2010-03-27	28	2010-04-05	248	1440	97.01
2011	2011-03-20	52	2011-04-01	645	2508	96.37
2012	2012-03-23	62	2012-04-01	679	2716	88.40
2013	2013-03-28	65	2013-04-11	282	1935	80.72
2014	2014-03-16	45	2014-04-01	3710	7679	99.35
2015	2015-03-23	31	2015-04-01	767	4401	99.07
2016	2016-03-18	69	2016-03-30	721	4383	97.49
2017	2017-03-13	88	2017-03-26	1152	6225	98.07
2018	2018-04-01	80	2018-04-18	470	3031	95.08

Respecto a la calidad de los datos utilizados en este trabajo, existen valores atípicos en la concentración registrada durante el día pico o día de máxima concentración para los años 2014 y 2017 con valores de 3710 granos de polen/m³ y 1152 granos de polen/m³, respectivamente. Por otro lado, un problema frecuente a la hora de trabajar con datos aerobiológicos es no disponer de datos debido a problemas con el captador polínico. La calidad de los datos se ha evaluado aplicando la función “quality_control” implementada en el paquete de R “AeRobiology” (Rojo et al., 2019b) que considera la falta de datos para el cálculo de la estación polínica y el porcentaje de datos en blanco dentro del periodo principal de polinización. Este análisis de la calidad de los datos indica que el año 2012 presenta un riesgo (Figura 6) que se debe principalmente a que los datos perdidos se encuentran cercanos al día de máxima concentración.

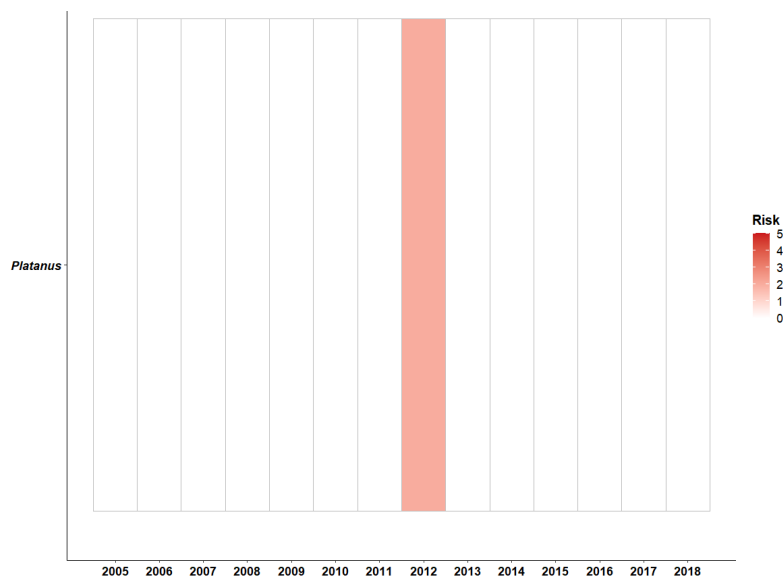


Figura 6. Calidad de la base de datos del tipo polínico *Platanus*

Si analizamos en detalle los datos perdidos dentro de la base de datos histórica utilizada en este trabajo, encontramos que los años 2005, 2011, 2012 y 2014 presentan datos en blanco (Tabla 2).

Tabla 2. Valores perdidos en la base histórica del polen de *Platanus*

Año	Nº días	Periodo principal de polinización	Periodo 40 días
2005	3	Postpico	Sí
2011	6	Postpico	No
2012	9	Pico	Sí
	5	Postpico	Sí
2014	5	Postpico	Sí

Para subsanar los problemas asociados a la presencia de valores perdidos en las bases de datos se ha realizado una interpolación de dichos datos siguiendo el procedimiento del trabajo de Picornell et al. (2021), donde se comparan diferentes métodos para interpolar datos aerobiológicos (Figura 7).

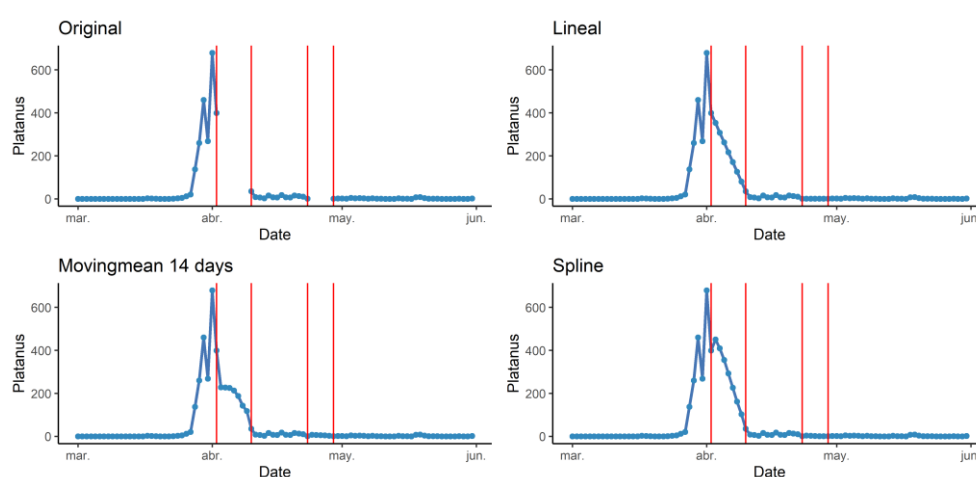


Figura 7. Métodos de interpolación aplicados a los datos de polen de *Platanus* para el año 2012. Las líneas rojas verticales muestran el periodo interpolado.

Los datos aerobiológicos se han interpolado utilizando el método de la media móvil considerando las conclusiones obtenidas por Picornell et al. (2021) que destacan la media móvil por ser el método que genera menos errores relativos, ser independiente de la disponibilidad y extensión de la base de datos y ser menos sensible a las variaciones de las concentraciones de polen durante días consecutivos. Este método consiste en calcular la media móvil de las concentraciones diarias de polen aplicando un tamaño de ventana

del tamaño del hueco que no presenta datos multiplicado por el argumento del factor y reemplaza los datos que faltan con la media móvil de esos días (Rojo et al., 2019b).

2.3. Datos meteorológicos

Los datos meteorológicos han sido obtenidos de la estación meteorológica de Toledo (39° 53' N, 04° 02, W) situada a una altitud de 515 metros sobre el nivel del mar (Figura 3).

Para el análisis de datos funcional, se han considerado los datos diarios registrados durante los 40 días siguientes al inicio de la estación polínica de *Platanus*. Se dispone de los datos diarios para el periodo 2005-2018 relativos a las variables meteorológicas de temperatura media (°C), precipitación diaria acumulada (mm), humedad relativa (%) y horas de sol (h).

3. Metodología

Como se ha descrito en la sección anterior, las bases de datos están compuestas por las observaciones diarias de los primeros 40 días desde el inicio de la estación polínica de *Platanus* para el periodo de estudio 2005-2018 de las variables: concentración de polen (granos de polen/m³), temperatura media (°C), precipitación diaria acumulada (mm), humedad relativa (%) y horas de sol (h).

La metodología seguida para el análisis del comportamiento de la curva de polen ha consistido en obtener la reconstrucción de la forma funcional de todas las variables objeto de estudio y aplicar un análisis en componentes principales funcional a dichas variables. Una vez definidas las componentes principales que explican una variabilidad adecuada de la forma funcional de las variables, se han determinado las componentes principales de las variables meteorológicas que están correlacionadas con cada componente principal de la variable polen. Por último, se ha analizado la influencia que tienen las componentes principales de las variables meteorológicas sobre cada componente principal del polen.

3.1. Representación básica de las curvas

En cada una de las variables de estudio, partimos de una muestra aleatoria simple de trayectorias denotadas por $x_1(t), \dots, x_n(t)$ de un proceso estocástico $\{X(t): t \in T\}$ que satisface las siguientes condiciones:

- El proceso estocástico definido es de segundo orden.
- El proceso estocástico definido es continuo en media cuadrática.
- El proceso estocástico definido tiene trayectorias que pertenecen al espacio de Hilbert $L^2[T]$ de las funciones de cuadrado integrable con la métrica usual en $L^2[T]$ definidas por:

$$L^2[T] = \left\{ f: T \rightarrow \mathbb{R} : \int_T |f(t)|^2 dt < \infty \right\},$$

con producto escalar usual:

$$\langle f, g \rangle = \int_T f(t)g(t)dt, \forall f, g \in L^2[T],$$

de forma que la norma asociada a este producto escalar tiene estructura de espacio de Hilbert.

La reconstrucción de la forma funcional de las curvas a partir de los datos discretos observados se ha realizado siguiendo el enfoque conocido como expansión básica o representación en bases de funciones (Ramsay and Silverman, 2005). Para representar esta aproximación es necesario un sistema de base de funciones, es decir, una base de funciones $\{\phi_1(t), \dots, \phi_p(t)\}$ que sean independientes entre sí y con la propiedad de aproximar arbitrariamente cualquier función tomando una suma ponderada o combinación lineal de un número p de estas funciones. Por tanto, las curvas pueden expresarse como:

$$x_i(t) = \sum_{j=1}^p a_{ij}\phi_j(t); i = 1, \dots, n,$$

donde $\mathbf{a}_i = (a_{i1}, \dots, a_{ip})^t$ es el vector que representa los coeficientes básicos de la reconstrucción de la curva muestral.

En este caso, los coeficientes básicos han sido estimados minimizando el error de mínimos cuadrados penalizados.

$$PMSE_d(\mathbf{a}_i|x_i) = (x_i - \Phi_i\mathbf{a}_i)^T(x_i - \Phi_i\mathbf{a}_i) + \lambda\mathbf{a}_i^T P_d \mathbf{a}_i,$$

donde $x_i = (x_{i1}, x_{i2}, \dots, x_{ik_i})^T$ es el vector de medidas discretas de la curva registrada $x_i(t)$, $\Phi_i = (\Phi_j(x_{ij}))_{k_j \times q}$ es la matriz de valores de las funciones en los puntos de muestreo, λ el parámetro de penalización de suavizado y $P_d = (\Delta^d)^T \Delta^d$ siendo Δ^d la representación matricial del operador diferencia de orden d . Por tanto, los coeficientes básicos de cada curva son estimados por

$$\mathbf{a}_i = (\Phi_i^T \Phi_i + \lambda P_d)^{-1} \Phi_i^T x_i.$$

Existen diferentes bases de funciones que pueden aplicarse, pero debido a la naturaleza de los datos que se van a analizar en este trabajo nos centraremos en la base B-Spline. La base B-Spline es la opción más común para aproximar curvas no periódicas con comportamiento local suave. Un B-Spline de grado k está compuesto por polinomios conectados entre sí de manera suave y las derivadas hasta el orden $k - 1$ son continuas en los puntos de unión, es decir, un spline es una función construida a trozos, donde cada trozo es un polinomio de un determinado grado.

Otro aspecto importante es determinar el número de funciones básicas que es adecuado para la representación de las curvas. Este es un factor clave para evitar el sobre-ajuste que produciría la utilización de un número elevado de funciones base o la pérdida de información que proporcionaría el caso contrario. En la literatura no se define un criterio estándar para la selección del número de funciones base por lo que se ha optado por probar con diferentes dimensiones hasta definir la óptima. Sin embargo, también se pueden usar métodos numéricos tales como la validación cruzada o validación cruzada generalizada a través de una penalización por rugosidad (ver, por ejemplo, la función “optim.basis” implementada en el paquete “fda.usc” de R y las referencias ahí contenidas).

En este estudio se han empleado B-Spline de grado 3, que son adecuadas para el caso de trayectorias regulares, con una dimensión de 14 bases para las variables

meteorológicas de temperatura media, humedad relativa y horas de sol. Este procedimiento se ha realizado utilizando el paquete de R “fda” utilizando las funciones “create.bspline.basis” y “smooth.basis”. Una explicación más detallada del paquete “fda” puede consultarse en Ramsay et al. (2009).

3.2. Reconstrucción funcional de curvas estrictamente positivas

En el caso concreto de las variables concentración de polen y precipitación, dada su naturaleza positiva, es necesario realizar una reconstrucción de las curvas estrictamente positiva para lo cual se ha aplicado una función de suavizado positivo (Ramsay et al., 2009).

Una función de suavizado positivo $x_i(t)$ se puede definir como la exponencial de una función $W_i(t)$ sin restricciones (Ramsay & Silverman, 2005), por tanto:

$$x_i(t) = e^{W_i(t)},$$

donde W_i es el logaritmo de x_i . Como resultado, esta función de ajuste de datos no tiene restricciones en cuanto a su signo y se garantiza que la reconstrucción de la curva tenga valores positivos (Ramsay et al., 2009). Aplicando esta función de ajuste, el logaritmo de las curvas quedaría expresado de la siguiente manera:

$$W_i(t) = \sum_{k=1}^K c_{ik} \phi_k(t); i = 1, \dots, n.$$

Para la reconstrucción funcional de las variables concentración de polen y precipitación se han empleado B-Spline de grado 3 con una dimensión de 12 bases. Este procedimiento se ha realizado utilizando las funciones “create.bspline.basis” y “smooth.pos” del paquete “fda”.

Mediante este procedimiento, la forma funcional obtenida es el logaritmo de la variable, por tanto, se pierde la expresión simple para la curva de suavizado (Ramsay et al., 2009). Para ello se ha utilizado la función “eval.posfd” del paquete “fda”.

3.3. Análisis en Componentes Principales Funcional

Para reducir la dimensión del problema y explicar los patrones de variabilidad en los datos de polen, se ha realizado un análisis en componentes principales funcional para todas las variables de estudio empleando como criterio para la selección del número de componentes que la variabilidad explicada por las componentes principales debe ser aproximadamente del 95 %.

Sea $x_1(t), \dots, x_n(t)$ una muestra de funciones muestrales del proceso $\{X(t): t \in T\}$. Consideramos sin pérdida de generalidad que el proceso es centrado, es decir, $\bar{x}(t) = 0$. Las componentes principales funcionales son variables incorreladas de media cero calculadas como combinaciones lineales generalizadas del proceso con máxima varianza. En particular, la j -ésima componente principal adopta la siguiente expresión:

$$\xi_{ij} = \int_T x_i(t) f_j(t) dt; i = 1, \dots, n,$$

donde las funciones peso $f_j(t)$ que definen a las componentes son las $n - 1$ soluciones de la ecuación de segundo orden

$$\widehat{C}(f) = \int_T \widehat{C}(t, s) f(s) ds = \lambda f(t), \quad \forall t \in T,$$

siendo $\widehat{C}(f)$ el operador de covarianza muestral, $\widehat{C}(t, s)$ la función covarianza y λ_j una secuencia de autovalores no nulos asociados a las autofunciones tal que $\lambda_j = VAR [x_i]$.

La función covarianza puede expresarse de la siguiente manera:

$$\widehat{C}(t, s) = \sum_{j=1}^{n-1} \lambda_j f_j(s) f_j(t).$$

Las trayectorias del proceso estocástico $\{X(t): t \in T\}$ se pueden representar en términos de las componentes principales mediante la expansión de Karhunen-Loeve de la siguiente forma:

$$x_i(t) = \sum_{j=1}^{n-1} \xi_{ij} f_j(t); \quad i = 1, \dots, n.$$

La representación anterior es óptima porque es la mejor aproximación de las curvas muestrales en el sentido de mínimos cuadrados. Además, el desarrollo de Karhunen-Loeve truncado en términos de las primeras q componentes principales nos da la siguiente representación:

$$x_i^q(t) = \sum_{j=1}^q \xi_{ij} f_j(t),$$

dando lugar a que la varianza explicada por las primeras q componentes sea $\sum_{j=1}^q \lambda_j$.

Considerando que las curvas son expresadas en términos de las funciones base de la siguiente manera

$$x_i(t) = \sum_{j=1}^p a_{ij} \phi_j(t); \quad i = 1, \dots, n,$$

y denotando a la matriz de coeficientes básicos como $A = (a_{ij})_{n \times p}$, la función de peso de la componente principal f_j admite la expansión básica

$$f_j(t) = \sum_{k=1}^p b_{jk} \phi_k(t),$$

y el análisis en componentes principales funcional es equivalente al análisis de componentes principales multivariado de la matriz $A\Psi^{\frac{1}{2}}$ (Ocaña et al., 2007), con $\Psi^{\frac{1}{2}}$ siendo la raíz cuadrada de la matriz de productos internos entre las funciones bases $\Psi = (\Psi_{ij})_{p \times p} = \int_T \phi_i(t)\phi_j(t) dt$.

Entonces, el vector b_j de coeficientes básicos es la j -ésima función principal de peso que viene dada por $b_j = \Psi^{-\frac{1}{2}}t_j$, donde el vector t_j se calcula como la solución al problema de autovalores $n^{-1}\Psi^{\frac{1}{2}}A^T A\Psi^{\frac{1}{2}}t_j = \lambda_j t_j$, donde $n^{-1}\Psi^{\frac{1}{2}}A^T A\Psi^{\frac{1}{2}}$ es la matriz de covarianza muestral de $A\Psi^{\frac{1}{2}}$.

Este procedimiento se ha realizado utilizando la función “pca.fd” del paquete “fda”.

3.4. Análisis de correlación

La influencia de las variables funcionales meteorológicas sobre la variable funcional de polen de *Platanus* se ha analizado mediante un análisis de correlación de las componentes principales funcionales del polen y las componentes principales funcionales de las variables meteorológicas.

El test de correlación seleccionado es el test no paramétrico de Spearman debido a que las componentes principales funcionales del polen de *Platanus* no siguen una distribución normal. La distribución de las componentes principales funcionales de *Platanus* ha sido previamente analizada utilizando el test de normalidad de Shapiro-Wilks.

4. Resultados

En este apartado se muestran los resultados obtenidos tras aplicar la metodología descrita en el apartado anterior. En todos los casos, la variable concentración de polen de *Platanus*, al ser la variable respuesta cuyo comportamiento se pretende analizar, será descrita en primer lugar y con mayor detalle que las variables meteorológicas.

En primer lugar, se reflejan los resultados obtenidos tras la reconstrucción de la forma funcional para las diferentes variables. A continuación, se muestran los resultados obtenidos tras el análisis en componentes principales funcional para todas las variables. Por último, se muestran las correlaciones obtenidas en el análisis de la influencia de las componentes principales de las variables meteorológicas sobre las componentes principales de la variable concentración de polen.

4.1. Reconstrucción de la forma funcional

Partiendo de los datos de concentración diaria de polen de *Platanus* registrados en la ciudad de Toledo durante el periodo 2005-2018 (Figura 8), la reconstrucción de la forma funcional con una dimensión de 12 bases permite representar adecuadamente la forma de la curva polínica (Figura 9).

Tomando alguno de los años de estudio, se puede observar en detalle la representación de la curva polínica que se ha obtenido (Figura 10).

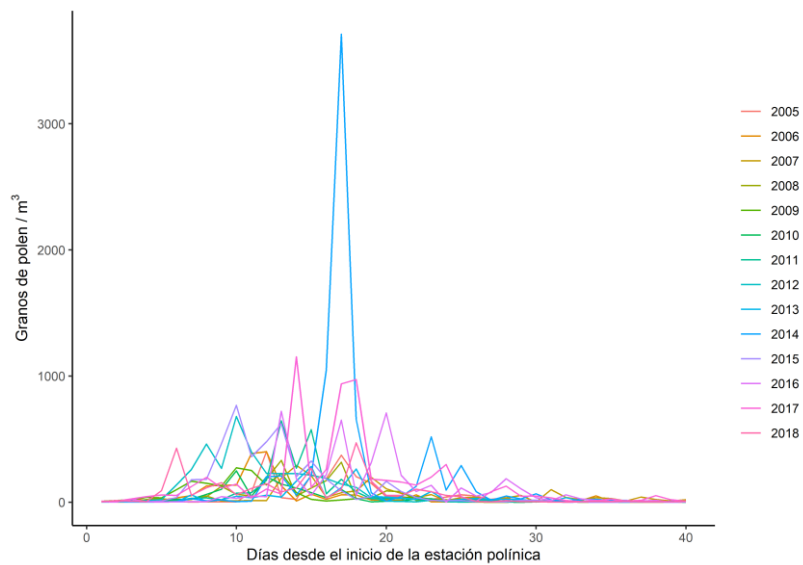


Figura 8. Evolución de la concentración diaria de polen de *Platanus* en la ciudad de Toledo durante el periodo 2005-2018.

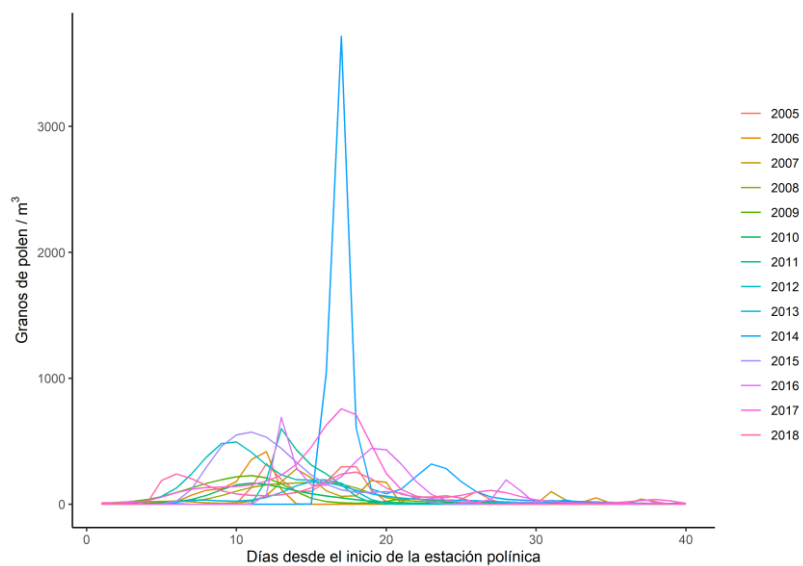


Figura 9. Reconstrucción funcional de la evolución de la concentración diaria de polen de *Platanus* en la ciudad de Toledo durante el periodo 2005-2018.

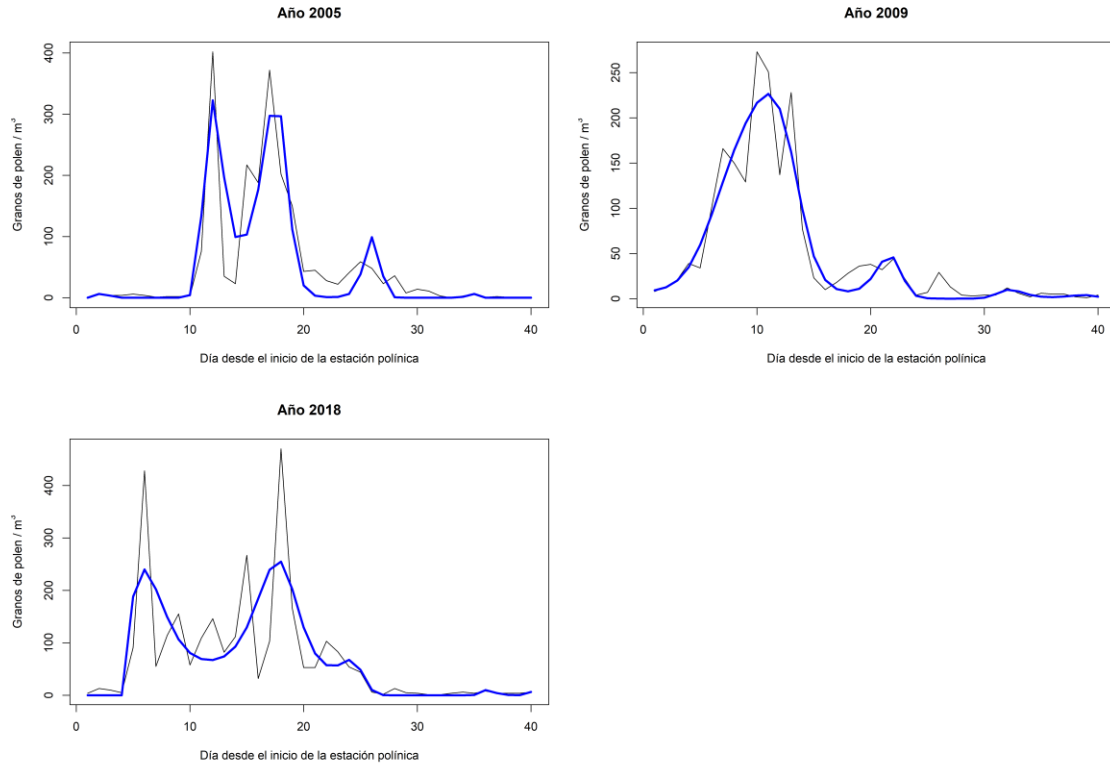


Figura 10. Evolución de la concentración polínica de Platanus durante los años 2005, 2009 y 2018 en la ciudad de Toledo. Línea negra representa los datos discretos originales y la línea azul representa los datos discretos obtenidos al evaluar la forma funcional.

Por ejemplo, para los datos pertenecientes al año 2005, la evaluación de la forma funcional obtenida permite observar la evolución de la curva polínica identificando correctamente los dos picos de elevada concentración de polen (Figura 10). Respecto al año 2009, se puede observar que la curva polínica original presenta un periodo de elevada concentración situado pocos días después de iniciar la estación polínica. Este periodo se divide en tres picos de elevada concentración, siendo el pico central el día de máxima concentración de la estación polínica. La evaluación de la forma funcional permite identificar el periodo de elevada concentración, con un único pico que coincide con el día de máxima concentración. En el caso del año 2018, se puede observar la presencia de dos marcados picos de concentración polínica que la evaluación de la forma funcional representa adecuadamente. Las figuras referentes a la reconstrucción funcional de todas las variables analizadas por años pueden verse en el Anexo del presente trabajo.

La curva media de la evolución de la concentración del polen y el intervalo de confianza (Figura 11) muestra que el pico de máxima concentración se produce de media aproximadamente el día 18 desde el inicio de la estación polínica, con un intervalo de confianza muy amplio debido a la variabilidad que se produce en la concentración registrada durante ese día. Además, se puede observar un pico que se produce el día 12 y que presenta menor concentración y un menor intervalo de confianza que el día pico de máxima concentración. Si consideramos el día de máxima concentración como una división de la estación polínica en periodo prepico, periodo que considera desde el inicio de la estación polínica hasta incluir el día en el que se registra el día de máxima concentración, y periodo postpico considerando el periodo comprendido desde el día siguiente al día de máxima concentración hasta el día 40; se observa que el intervalo de confianza es más amplio en el periodo prepico que en el periodo postpico. En general, el periodo comprendido entre el inicio de la estación polínica hasta el día de máxima concentración se caracteriza por representar el principal periodo de liberación y transporte de granos de polen. La amplitud del intervalo de confianza refleja que las curvas polínicas presentan sus mayores diferencias en este periodo como consecuencia de la complejidad asociada al proceso de emisión de los granos de polen. La liberación de los granos de polen desde las anteras y su posterior transporte es característica de cada grupo taxonómico de plantas y, en general, los procesos implicados son dependientes de los factores meteorológicos. Por el contrario, durante el periodo posterior se asumen que la mayoría de la concentración polínica producida se ha liberado y transportado por lo que las concentraciones remanentes de polen son menos dependientes de las variables meteorológicas.

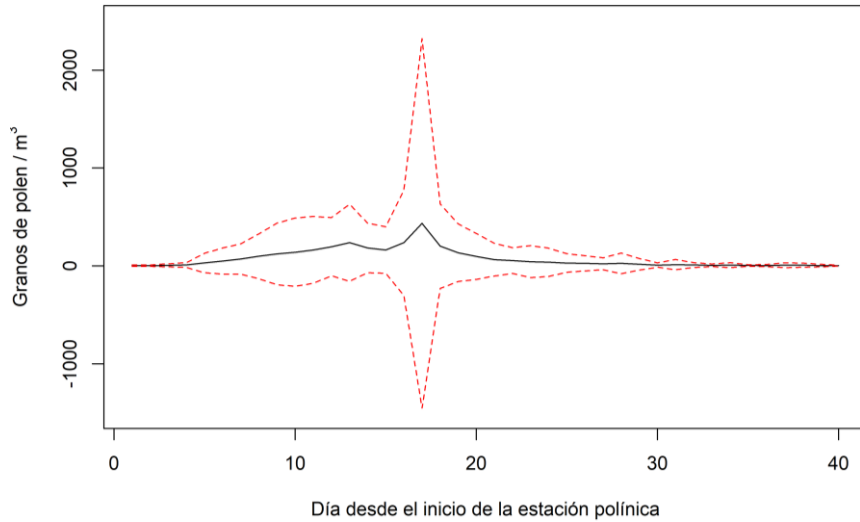


Figura 11. Evolución de los datos discretos obtenidos al evaluar los datos funcionales de la variable polen. Línea continua representa la media y las líneas discontinuas rojas representan el intervalo de confianza.

Respecto a las variables meteorológicas (Figura 12), la función media permite observar que las variables de temperatura media y horas de sol presentan una ligera pendiente positiva durante los años 2005-2018 para el periodo de 40 días de estudio. La humedad relativa se mantiene más o menos constante durante el periodo de estudio. Por otro lado, la evaluación de la forma funcional de la precipitación muestra eventos de precipitación puntuales a lo largo del periodo de estudio. En todos los casos, el intervalo de confianza respecto a la función media presenta gran variabilidad.

Las figuras referentes a la reconstrucción funcional de todas las variables analizadas por años pueden verse en el Anexo del presente trabajo.

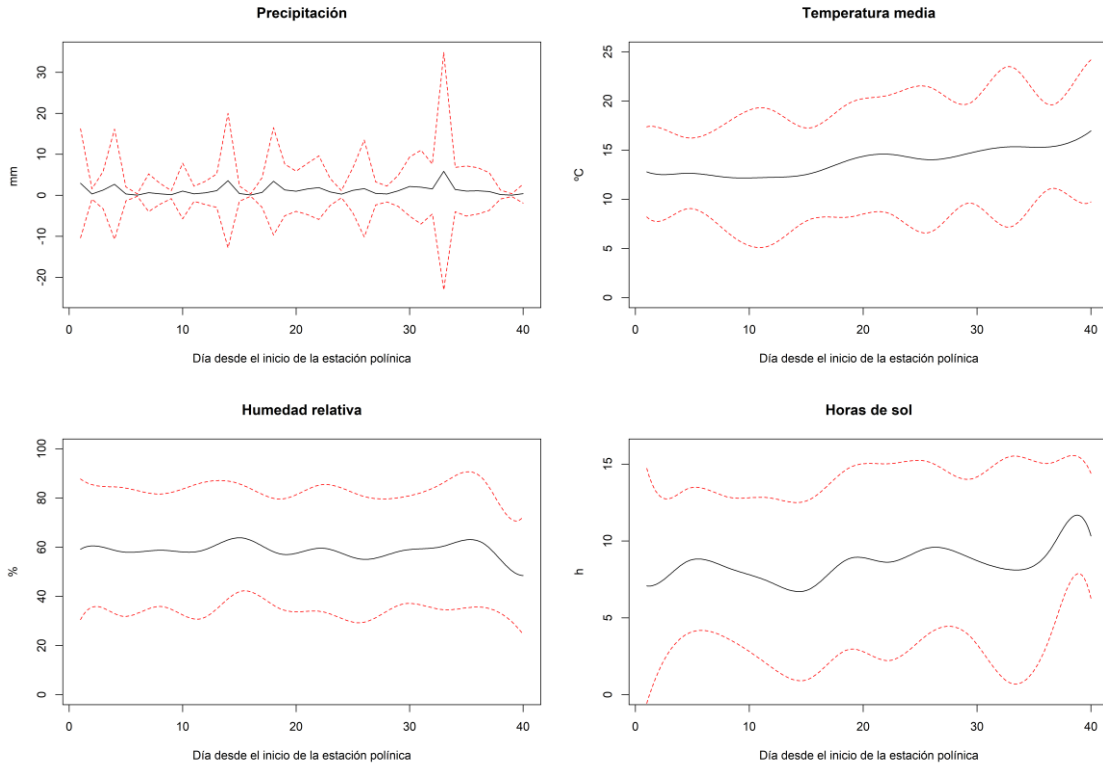


Figura 12. Función media (línea negra) e intervalos de confianza (líneas rojas) de las variables meteorológicas analizadas. En el caso de la precipitación, se representa la media de los datos discretos obtenidos al evaluar los datos funcionales.

4.2. Análisis en Componentes Principales Funcional

El análisis de componentes principales funcional realizado sobre los datos funcionales de polen, considerando que la variabilidad explicada por las componentes principales debe ser aproximadamente del 95 %, refleja que el un número adecuado de componentes principales que deben extraerse es 5, explicando una variabilidad acumulada de casi el 94% (Tabla 3).

Tabla 3. Valores de las Componentes Principales seleccionadas para la variable polen

	CP1	CP2	CP3	CP4	CP5
Variabilidad (%)	32,83	24,43	19,6	11,5	5,37
Variabilidad acumulada (%)	32,83	57,26	76,86	88,36	93,73

Para interpretar las funciones peso de cada componente principal (autofunciones) en términos de variación respecto a la media, se representa la curva media y las funciones obtenidas al sumar y restar una pequeña cantidad de cada autofunción (Ramsay et al., 2009; Ramsay and Silverman, 2005).

En la primera componente (Figura 13) se observa un cambio de tendencia en los primeros 12 días desde el inicio de la estación polínica con respecto a la segunda mitad de la estación polínica, destacando una marcada perturbación respecto a la función media en el periodo de tiempo situado entre los días 22 y 30 desde el inicio de la estación polínica. Por tanto, esta primera componente hace referencia al periodo en el que se centran las concentraciones de polen con respecto a la función media del periodo analizado, siendo un valor positivo para aquellas curvas en las que la concentración de polen se registra principalmente durante el periodo del pico y el postpico y un valor negativo en la componente principal en aquellas curvas en las que la concentración de polen se registra mayoritariamente en el prepico y el pico.

En la segunda componente, se observan perturbaciones en el periodo del inicio y del final de la estación polínica; en concreto, en los periodos comprendidos entre el día de inicio y el día 10 y, por otro lado, entre el día 30 y el final de la estación polínica (Figura 13). Esta componente principal resume la amplitud del registro de concentraciones elevadas, es decir, un valor positivo en esta componente refleja las curvas que presentan concentraciones por encima de la media tanto en el periodo prepico como en el postpico.

En la tercera componente se observan perturbaciones pequeñas a lo largo de la función media, por tanto, esta componente captura los años en los que la evolución fue constante con respecto a la media (Figura 13).

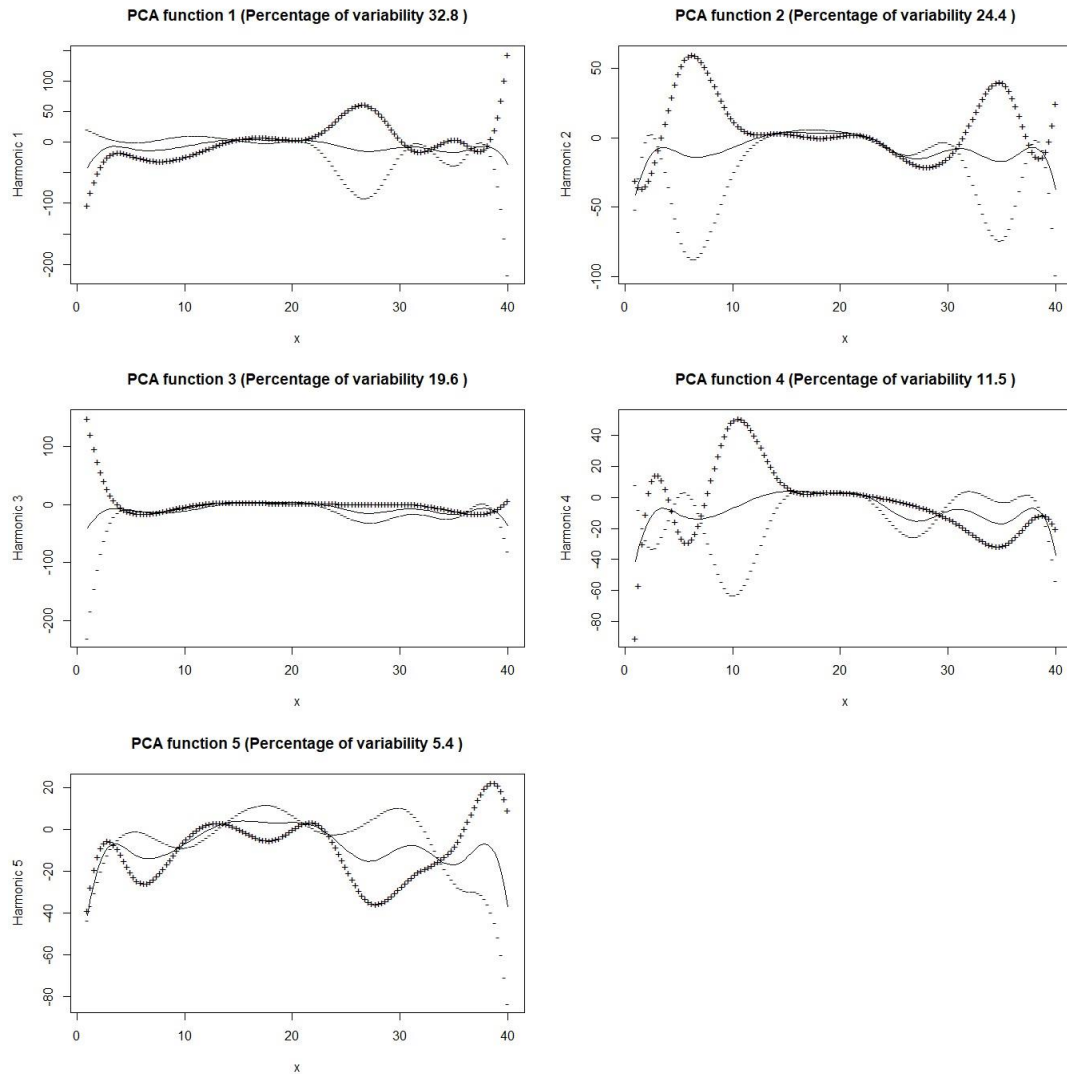


Figura 13. Representación de las cinco primeras autofunciones asociadas a las componentes funcionales del logaritmo de la variable polen. Línea continua representa la función media muestral (escala logarítmica) y las líneas discontinuas representan perturbaciones respecto a la función media muestral (escala logarítmica).

En la componente cuatro se observa que las perturbaciones se localizan principalmente en el tramo inicial, siendo más elevadas alrededor del día 12, y otra perturbación en el periodo comprendido entre el día 22 y el día 30 que cambia de signo en el periodo 30-40 (Figura 13). La cuarta componente recoge el comportamiento de los picos de polen, considerando que la función media presenta dos picos de elevada concentración polínica, valores negativos en esta componente reflejan un retraso en el registro de estos picos de concentración y, por el contrario, valores positivos reflejan un adelanto.

La forma en la que cargan las diferentes curvas a cada componente se puede visualizar mediante un gráfico de dispersión que enfrente las diferentes componentes. Enfrentando la primera componente, que hace referencia al periodo en el que se registran las mayores concentraciones, y la segunda componente, que representa la amplitud de la curva, se observan dos agrupamientos (Figura 14). Por un lado, se observa el agrupamiento de los años 2011, 2014 y 2016 que se caracterizan por el registro de las concentraciones de polen principalmente en el periodo pico y en el periodo posterior al día de máxima concentración, es decir, las emisiones de polen importantes se encuentran en la franja comprendida entre el día 10 y el día 30. Y, por otro lado, se observa el agrupamiento de los años 2009, 2010, 2012, 2017 y 2018 que poseen valores positivos en ambas componentes lo que se traduce en que su curva polínica presenta concentraciones por encima de la función media en todos los tramos definidos por las componentes principales, se trata de curvas muy amplias y, en general, la duración de la estación polínica es mayor que el resto de los años.

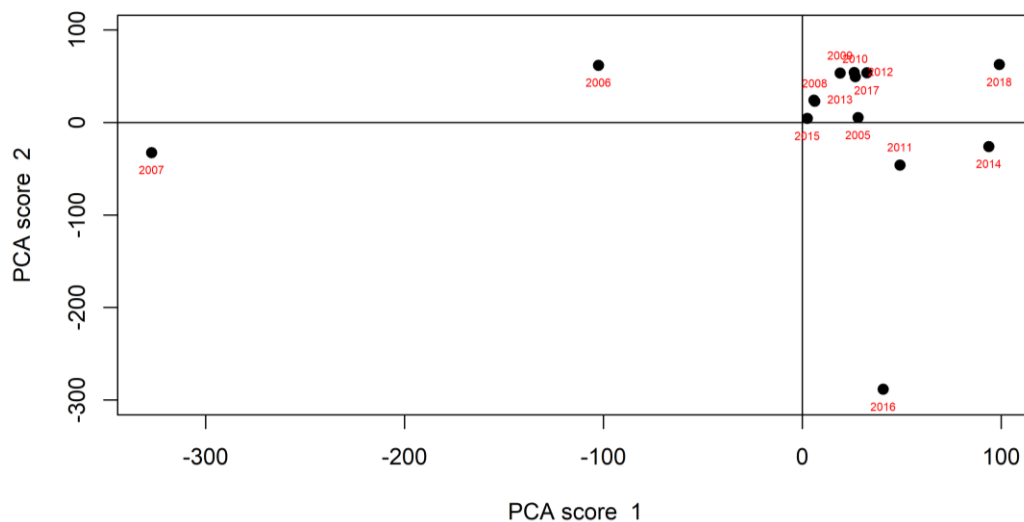


Figura 14. Gráfico de dispersión entre los valores de la primera componente (eje x) y los valores de la segunda componente (eje y).

El gráfico de dispersión entre la primera componente principal, que hace referencia al periodo en el que se concentran las emisiones de polen, y la tercera componente, que refleja una evolución constante con respecto a la función media (Figura 15), muestra que en general la mayoría de los años presenta un comportamiento que se ajusta a la función media. Destacan por presentar un valor negativo en la tercera componente los años 2007 y 2018, estos años se caracterizan porque los picos de elevada concentración no coinciden con los picos de la función media.

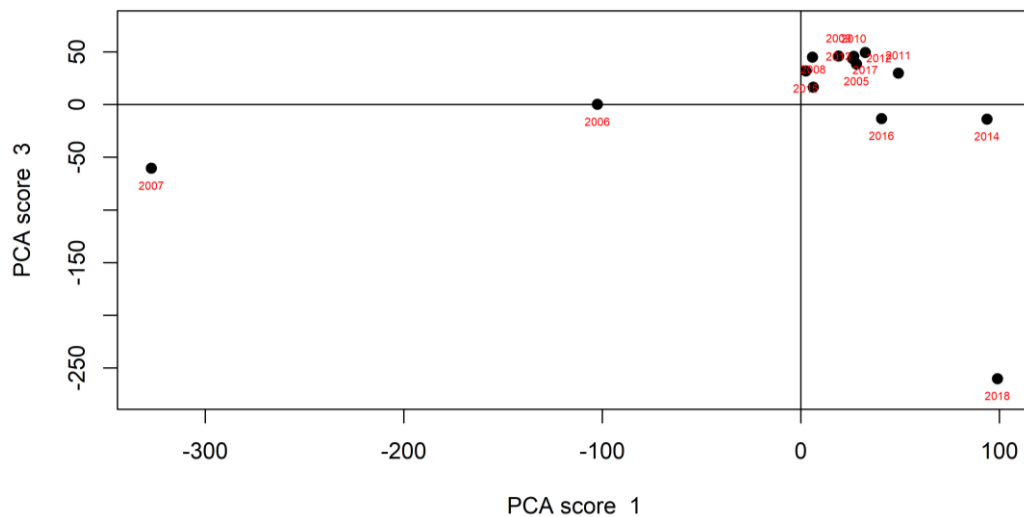


Figura 15. Gráfico de dispersión entre los valores de la primera componente (eje x) y los valores de la tercera componente (eje y).

Por último, el gráfico que representa los valores obtenidos en la primera componente frente a la cuarta componente, que refleja la coincidencia de los picos con respecto a la función media, muestra dos agrupamientos (Figura 16). Por un lado, los años 2007 y 2014 que se caracterizan por registrar los picos de mayor concentración con un poco de retraso respecto a la función media. En el caso contrario se encuentran los años 2016 y 2018, años en los que al menos un pico de elevada concentración se adelanta. Sin embargo, la mayoría de los años no tienen un peso significativo en la cuarta componente por lo que

no existe una gran variación en la fecha en la que se registran los días con mayores concentraciones.

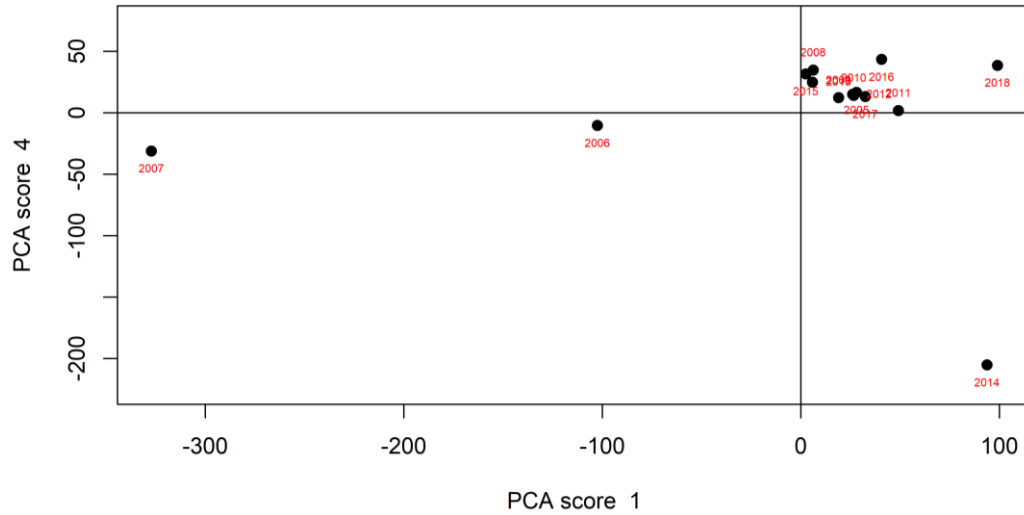


Figura 16. Gráfico de dispersión entre los valores de la primera componente (eje x) y los valores de la cuarta componente (eje y).

Respecto a las variables meteorológicas, el análisis de componentes principales funcionales realizado indica que para explicar una variabilidad de aproximadamente el 95% mediante componentes principales es necesario extraer 7 componentes principales en todas las variables meteorológicas salvo en la precipitación, en la que se alcanza aproximadamente el 95% con 5 componentes principales (Tabla 4).

Tabla 4. Valores de las Componentes Principales seleccionadas para las variables meteorológicas: precipitación (Prec), temperatura media (TMed), humedad relativa (HRel) y horas de sol (HSol)

		PC1	PC2	PC3	PC4	PC5	PC6	PC7
Prec	Variabilidad (%)	36,03	26,83	12,75	11,53	7,83	2,26	2,06
	Variabilidad acumulada (%)	36,03	62,86	75,61	87,14	94,97	97,23	99,29
TMed	Variabilidad (%)	30,66	21,91	18,71	8,52	7,95	3,83	3,44
	Variabilidad acumulada (%)	30,66	52,57	71,28	79,80	87,75	91,58	95,02
HRel	Variabilidad (%)	31,22	21,26	12,37	10,53	9,5	5,9	4,6
	Variabilidad acumulada (%)	31,22	52,48	64,85	75,38	84,88	90,78	95,38
HSol	Variabilidad (%)	26,86	21,92	17,76	12,01	7,67	6,23	3,55
	Variabilidad acumulada (%)	26,86	48,78	66,54	78,55	86,22	92,45	96,00

4.3. Análisis de correlaciones

Los resultados obtenidos en el análisis de correlación entre las componentes principales del polen y las componentes principales de las variables meteorológicas indican que en cada componente principal del polen influyen diferentes componentes principales de las variables meteorológicas (Tabla 5). Sin embargo, no se ha encontrado ninguna correlación significativa entre la quinta componente del polen y las diferentes componentes principales de las variables meteorológicas.

Tabla 5. Resumen de las correlaciones significativas obtenidas por el test no paramétrico de Spearman (p .valor<0,05) entre las componentes principales del polen y las componentes principales de las variables meteorológicas: precipitación (Prec), temperatura media (TMed), temperatura mínima (TMin), humedad relativa (HRel) y horas de sol (HSol).

	polen_PC1	polen_PC2	polen_PC3	polen_PC4
Prec_PC2	----	----	----	-0,741
Prec_PC4	0,543	----	----	----
TMed_PC1	0,543	----	----	----
TMed_PC7	----	0,684	----	----
HRel_PC4	----	----	-0,591	----
HSol_PC1	----	0,600	----	----
HSol_PC3	----	0,547	----	----

En general, los resultados de las correlaciones obtenidas (Tabla 5) indican que la primera componente principal del polen, que refleja el periodo en el que se concentran los valores de polen entorno al día pico, se explica principalmente mediante la temperatura media y la precipitación. La segunda componente principal del polen, que recoge la amplitud de la curva polínica, se explica por la temperatura media y las horas de sol. La tercera componente principal del polen, que captura el comportamiento de la curva de polen siguiendo la función media, está explicada por la humedad relativa siendo esta relación negativa. Por último, la cuarta componente del polen, que recoge el comportamiento de la curva de polen, se explica mediante la variable precipitación.

En la primera componente de la variable polen los resultados obtenidos indican que está correlacionada positivamente con la cuarta componente de la precipitación y la primera componente de la temperatura media (Tabla 5, Figura 17). Considerando que la primera componente captura aquellos años en los que se produce un cambio de tendencia tras el día 20, se puede observar que la correlación positiva y significativa con la primera componente principal de la temperatura media se debe a un cambio de tendencia brusco en el periodo de los días 9 al 15. Siguiendo la perturbación positiva en la temperatura media, la curva polínica comienza a elevarse por encima de la media como consecuencia de un descenso en la temperatura media durante un periodo relativamente prolongado continuado por un aumento brusco de la temperatura media. El caso contrario ocurre con la perturbación negativa en la temperatura media, en la que un aumento en la temperatura media durante el periodo del día 9 al día 15 seguido por un descenso brusco en la temperatura media propicia una emisión rápida de las concentraciones de polen que luego descienden como consecuencia del descenso de la temperatura media. A su vez, la cuarta componente de la precipitación representa los años en los que existe una gran perturbación en el inicio de la estación polínica, con respecto a la función media. Así, en los años en los que la perturbación es positiva, es decir, se produce una mayor cantidad de precipitación durante el inicio de la estación polínica se produce una menor emisión polínica durante ese periodo, que aumenta cuando desciende la cantidad de precipitación y viceversa, los años en los que se produce mayor emisión polínica durante el inicio genera una menor concentración de polen durante el inicio de la estación polínica.

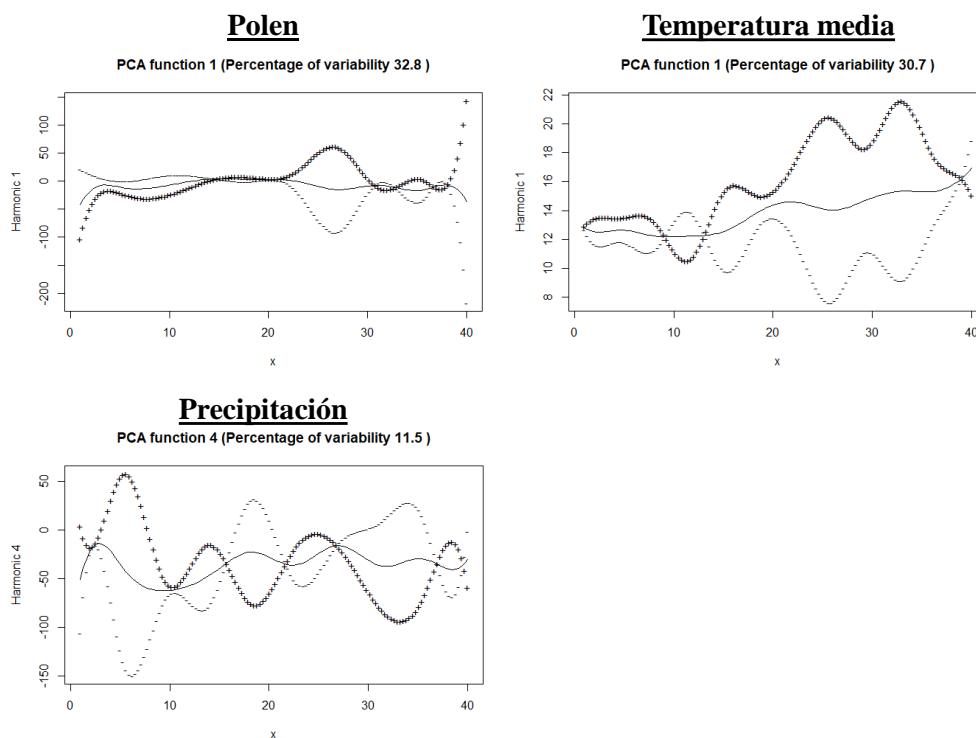


Figura 17. Representación de la primera componente funcional del polen (escala logarítmica) y las componentes principales de las variables meteorológicas con las que posee correlación significativa ($p, \text{valor} < 0.05$). Línea continua representa la función media muestral y las líneas discontinuas representan perturbaciones respecto a la función media muestral.

La segunda componente del polen está correlacionada positivamente con la séptima componente de la temperatura media, con la primera y tercera componente de la variable horas de sol (Tabla 5, Figura 18). La segunda componente principal del polen recoge la presencia de perturbaciones en el inicio y en el final de la estación polínica, es decir, en la amplitud de la curva de polen. La presencia de mayores concentraciones de polen en el inicio y final de la estación polínica con respecto a la función media están favorecidas por un aumento en los valores de temperatura media y horas de sol con respecto a la media de cada una de estas variables.

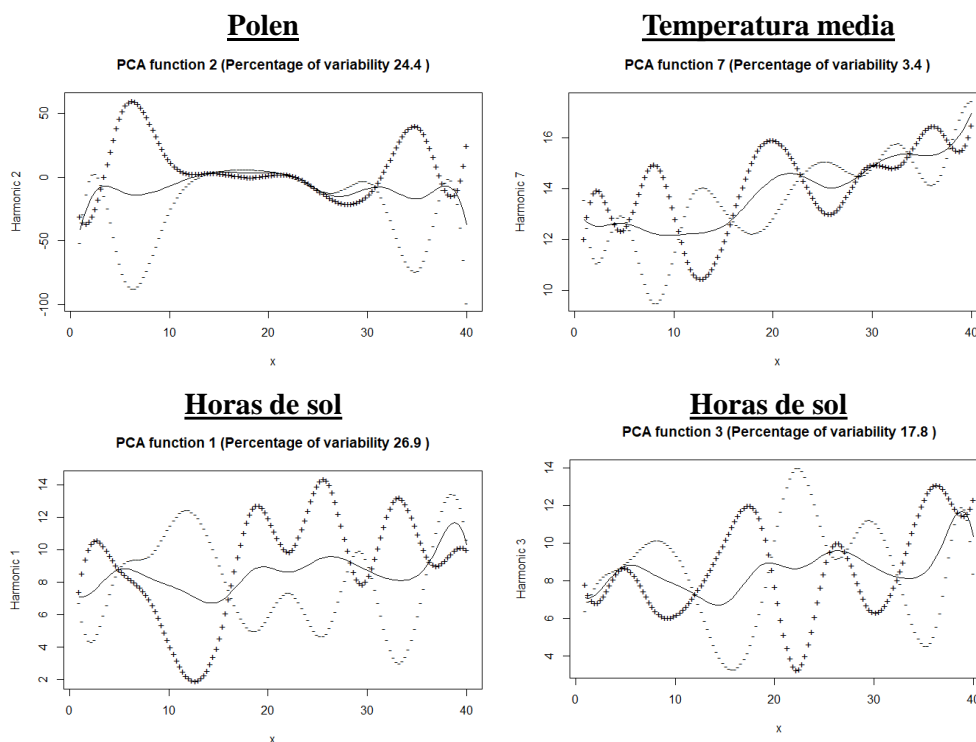


Figura 18. Representación de la segunda componente funcional del polen (escala logarítmica) y las componentes principales de las variables meteorológicas con las que posee significativa ($p.\text{valor} < 0.05$).

Línea continua representa la función media muestral y las líneas discontinuas representan perturbaciones respecto a la función media muestral.

La tercera componente del polen está correlacionada negativamente con la cuarta componente de la humedad relativa (Tabla 5, Figura 19). Considerando que la tercera componente refleja el comportamiento de la curva polínica en aquellos años en los que no se producen grandes perturbaciones, la evolución de la concentración polínica siguiendo la función media está influenciada negativa y significativamente con la humedad relativa debido a que una mayor cantidad de humedad relativa dificulta la emisión y el transporte del polen. La cuarta componente en la humedad relativa recoge los años en los que la humedad relativa presenta grandes fluctuaciones a partir del día 13, por tanto, el primer periodo en el que la humedad relativa desciende coincide con el pico de máxima concentración en la función media debido a que una menor humedad relativa favorece la emisión de polen. A partir de este momento, de forma general, el registro de

las concentraciones de polen depende de las variables que influyen en el transporte; por tanto, las fluctuaciones en la humedad relativa generan periodos con mayor humedad relativa que dificulta el transporte seguidos por periodos que presentan un descenso en la humedad relativa que facilitan ese transporte.

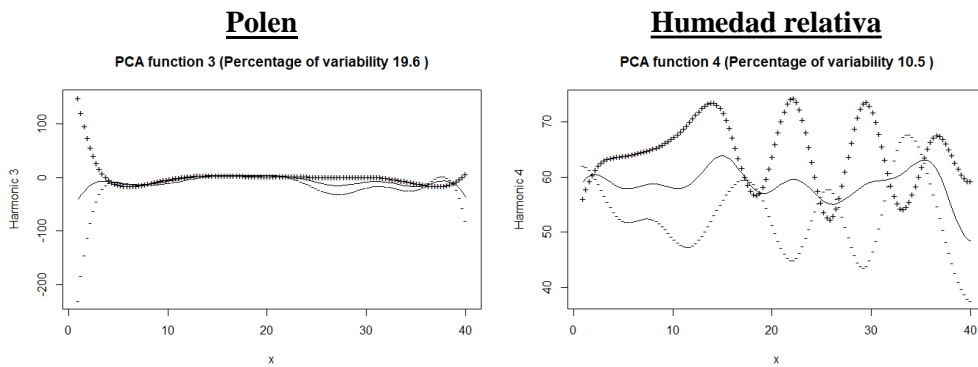


Figura 19. Representación de la tercera componente funcional del polen (escala logarítmica) y las componentes principales de las variables meteorológicas con las que posee correlación significativa ($p.\text{valor} < 0,05$). Línea continua representa la función media muestral y las líneas discontinuas representan perturbaciones respecto a la función media muestral.

La cuarta componente del polen, que recoge el comportamiento de los picos de polen, está correlacionada negativamente con la segunda componente de la variable precipitación (Tabla 5, Figura 20). Se puede observar durante el periodo de emisión de polen, que abarca el periodo desde el primer día hasta el día 18, una perturbación negativa en la componente del polen coincide con una perturbación positiva en la componente de precipitación y viceversa. Por tanto, la presencia de precipitación durante el inicio de la estación polínica retrasa la emisión de polen y, como consecuencia, los picos de máxima concentración se registran más tarde.

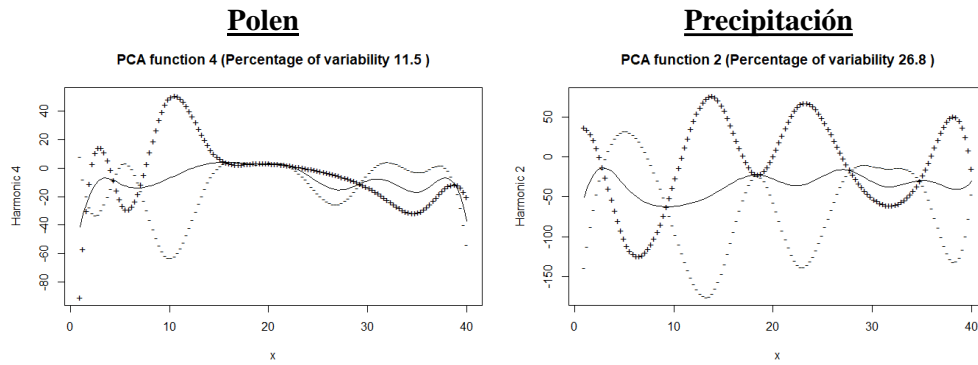


Figura 20. Representación de la cuarta componente funcional del polen y las componentes principales de las variables meteorológicas con las que posee correlación significativa ($p.\text{valor} < 0.05$). Línea continua representa la función media muestral y las líneas discontinuas representan perturbaciones respecto a la función media muestral.

5. Conclusiones

Las partículas biológicas que se registran en un momento determinado dependen de las condiciones meteorológicas que prevalecen a lo largo de días anteriores y las distintas variables meteorológicas pueden ejercer una influencia diferente en el proceso aerobiológico de cada partícula (Cresti and Linskens, 2000; Maya-Manzano et al., 2022; Rojo et al., 2015). La liberación de los granos de polen desde la antera se ve favorecida por las altas temperaturas y la baja humedad relativa. Además, la influencia de las variables meteorológicas continúa durante el proceso de dispersión.

La estación polínica del polen de *Platanus* se caracteriza por presentar una intensa y explosiva floración que produce la emisión de elevadas concentraciones de polen en un corto periodo de tiempo (González-Minero and Candau, 1997). Debido a que este polen procede de una especie ornamental, ampliamente cultivada en las zonas verdes de las ciudades, el transporte del polen de *Platanus* está principalmente afectado por las condiciones ambientales locales (Adams-Groom et al., 2017; Rojo et al., 2019a).

El análisis de datos funcionales realizado en este trabajo ha permitido identificar diferentes comportamientos en la curva de polen. El análisis de las variables meteorológicas se ha centrado en el periodo de emisión y dispersión del polen con el fin de determinar los comportamientos que ejercen una influencia directa sobre la curva polínica. Los resultados reflejan que, en la ciudad de Toledo, la curva polínica de *Platanus* presenta diferentes comportamientos relativos, por un lado, a la concentración de polen durante el periodo prepico o durante el periodo postpico; por otro lado, relativos a la amplitud de la curva de polen y, por último, a la fecha en la que se producen las principales concentraciones.

Además, este trabajo aporta información sobre la influencia del comportamiento de las variables meteorológicas sobre el comportamiento de la curva polínica, destacando la influencia de la temperatura media y la precipitación. Las variables meteorológicas que mayor influencia ejercen durante los procesos de emisión y dispersión del polen de *Platanus* son las variables de temperatura, la precipitación y la humedad relativa (Lara et al., 2023).

En concreto, los resultados obtenidos permiten identificar a la temperatura media y a la precipitación como las variables meteorológicas que afectan en la concentración de polen durante el periodo prepico o durante el periodo postpico. Respecto a la amplitud de la curva polínica, los resultados indican que está principalmente relacionada la temperatura media y las horas de sol. Por último, la fecha en la que se produce el pico de máxima concentración de polen está relacionado negativamente con la precipitación que se registra durante el inicio de la estación polínica.

Hasta la fecha no existen trabajos aerobiológicos que apliquen el análisis en componentes principales funcional para la identificación del comportamiento de la curva de polen y la determinación del comportamiento de las variables meteorológicas que ejercen una influencia en dicho comportamiento. Por tanto, este trabajo proporciona información relevante en el entendimiento de las dinámicas aerobiológicas, así como una nueva metodología que puede implementarse para analizar las series de datos polínicos.

6. Referencias

- Acal, C., Aguilera, A.M., 2022. Basis expansion approaches for functional analysis of variance with repeated measures. *Adv. Data Anal. Classif.* <https://doi.org/10.1007/s11634-022-00500-y>
- Acal, C., Aguilera, A.M., Sarra, A., Evangelista, A., Di Battista, T., Palermi, S., 2022. Functional ANOVA approaches for detecting changes in air pollution during the COVID-19 pandemic. *Stoch. Environ. Res. Risk Assess.* 36, 1083–1101. <https://doi.org/10.1007/S00477-021-02071-4/TABLES/4>
- Adams-Groom, B., Skjøth, C.A., Baker, M., Welch, T.E., 2017. Modelled and observed surface soil pollen deposition distance curves for isolated trees of *Carpinus betulus*, *Cedrus atlantica*, *Juglans nigra* and *Platanus acerifolia*. *Aerobiologia* (Bologna). 33, 407–416. <https://doi.org/10.1007/s10453-017-9479-1>
- AEMET, 2010. Guía resumida del clima en España 1981-2010. Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente, Madrid.
- Aguilera-Morillo, M.C., Aguilera, A.M., 2019. Multi-class classification of biomechanical data: A functional LDA approach based on multi-class penalized functional PLS. *Stat. Modelling* 20, 1471082X1787115. <https://doi.org/10.1177/1471082X17871157>
- Aguilera-Morillo, M.C., Aguilera, A.M., Jiménez-Molinos, F., Roldán, J.B., 2019. Stochastic modeling of Random Access Memories reset transitions. *Math. Comput. Simul.* 159, 197–209. <https://doi.org/10.1016/j.matcom.2018.11.016>
- Cariñanos, P., Casares-Porcel, M., 2011. Urban green zones and related pollen allergy: A

- review. Some guidelines for designing spaces with low allergy impact. *Landsc. Urban Plan.* 101, 205–214. <https://doi.org/10.1016/j.landurbplan.2011.03.006>
- Crane, E.Ae. of registration on cyclical kinematic data, Cassidy, R.B., Rothman, E.D., Gerstner, G.E., 2010. Effect of registration on cyclical kinematic data. *J. Biomech.* 43, 2444–2447. <https://doi.org/10.1016/j.jbiomech.2010.04.024>
- Cresti, M., Linskens, H.F., 2000. Pollen-allergy as an ecological phenomenon: A review. *Plant Biosyst. - An Int. J. Deal. with all Asp. Plant Biol.* 134, 341–352. <https://doi.org/10.1080/11263500012331350495>
- Curceac, S., Ternynck, C., Ouarda, T.B.M.J., Chebana, F., Niang, S.D., 2019. Short-term air temperature forecasting using Nonparametric Functional Data Analysis and SARMA models. *Environ. Model. Softw.* 111, 394–408. <https://doi.org/10.1016/j.envsoft.2018.09.017>
- D'Amato, G., Cecchi, L., Bonini, S., Nunes, C., Annesi-Maesano, I., Behrendt, H., Liccardi, G., Popov, T., van Cauwenberge, P., 2007. Allergenic pollen and pollen allergy in Europe. *Allergy* 62, 976–990. <https://doi.org/10.1111/j.1398-9995.2007.01393.x>
- De Boor, C., 2001. *A Practical Guide to Splines*, Revised Ed. ed. Springer, New York.
- Deville, 1974. Méthodes statistiques et numériques de l'analyse harmonique. *Ann. Insee.* 15, 3. <https://doi.org/10.2307/20075177>
- Enrique, E., Cisteró-Bahíma, A., Bartolomé, B., Alonso, R., San Miguel-Moncín, M., Bartra, J., Martínez, A., 2002. *Platanus acerifolia* pollinosis and food allergy. *Allergy* 57, 351–356. <https://doi.org/10.1034/j.1398-9995.2002.1s3509.x>

- Escabias, M., Valderrama, M.J., Aguilera, A.M., Santofimia, M.E., Aguilera-Morillo, M.C., 2013. Stepwise selection of functional covariates in forecasting peak levels of olive pollen. *Stoch. Environ. Res. Risk Assess.* 27, 367–376. <https://doi.org/10.1007/s00477-012-0655-0>
- Ferraty, F., Vieu, P., 2006. *Nonparametric functional data analysis*. Springer-Verlag, Berlin.
- Galán, C., Cariñanos, P., Alcázar, P., Domínguez, E., 2007. *Manual de calidad y gestión de la Red Española de Aerobiología*. Servicio de publicaciones de la Universidad de Córdoba, Córdoba.
- Galán, C., Smith, M., Thibaudon, M., Frenguelli, G., Oteros, J., Gehrig, R., Berger, U., Clot, B., Brandao, R., 2014. Pollen monitoring: minimum requirements and reproducibility of analysis. *Aerobiologia (Bologna)*. 30, 385–395. <https://doi.org/10.1007/s10453-014-9335-5>
- García-Mozo, H., Galán, C., Cariñanos, P., Alcázar, P., Méndez, J., Vendrell, M., Alba, F., Sáenz, C., Fernández, D., Cabezudo, B., Domínguez, E., 1999. Variations in the *Quercus* sp. pollen season at selected sites in Spain. *Polen* 10, 59–69.
- González-Minero, F.J., Candau, P., 1997. Study on *Platanus hispanica* Miller pollen content in the air of Seville, southern Spain. *Aerobiologia (Bologna)*. 13, 109–115.
- Hirst, J.M., 1952. An automatic volumetric spore trap. *Ann. Appl. Biol.* 39, 257–265.
- Hongtu Zhu, Styner, M., Niansheng Tang, Zhexing Liu, Weili Lin, Gilmore, J.H., 2010. FRATS: Functional Regression Analysis of DTI Tract Statistics. *IEEE Trans. Med. Imaging* 29, 1039–1049. <https://doi.org/10.1109/TMI.2010.2040625>

- Hutchinson, R.A., McLellan, P.J., Ramsay, J.O., Sulieman, H., Bacon, D.W., 2004. Investigating the impact of operating parameters on molecular weight distributions using functional regression. *Macromol. Symp.* 206, 495–508. <https://doi.org/10.1002/MASY.200450238>
- Lara, B., Rojo, J., Costa, A.R., Burgos-Montero, A.M., Antunes, C.M., Pérez-Badia, R., 2023. Atmospheric pollen allergen load and environmental patterns in central and southwestern Iberian Peninsula. *Sci. Total Environ.* 858, 159630. <https://doi.org/10.1016/j.scitotenv.2022.159630>
- Lara, B., Rojo, J., Fernández-González, F., Pérez-Badia, R., 2019. Prediction of airborne pollen concentrations for the plane tree as a tool for evaluating allergy risk in urban green areas. *Landsc. Urban Plan.* 189, 285–295. <https://doi.org/10.1016/j.landurbplan.2019.05.002>
- Maya-Manzano, J.M., Oteros, J., Rojo, J., Traidl-Hoffmann, C., Schmidt-Weber, C., Buters, J., 2022. Drivers of the release of the allergens Bet v 1 and Phl p 5 from birch and grass pollen. *Environ. Res.* 214. <https://doi.org/10.1016/J.ENVRES.2022.113987>
- Ocaña, Francisco A, Aguilera, Ana M, Escabias, Manuel, 2007. Computational considerations in functional principal component analysis. *Comput. Stat.* 22, 449–465. <https://doi.org/10.1007/s00180-007-0051-2>
- Pearson, K., 1901. LIII. On lines and planes of closest fit to systems of points in space. *London, Edinburgh, Dublin Philos. Mag. J. Sci.* 2, 559–572. <https://doi.org/10.1080/14786440109462720>
- Pereira, C., Valero, A., Loureiro, C., Dávila, I., Martínez-Cócera, C., Murio, C., Rico, P.,

- Palomino, R., 2006. Sensitisation in Allergic Rhinitis. *Eur. Ann. Allergy Clin. Immunol.* 38, 186–194.
- Pérez-Badia, R., Rapp, A., Morales, C., Sardinero, S., Galán, C., García-Mozo, H., 2010. Pollen spectrum and risk of pollen allergy in central Spain. *Ann Agric Env. Med* 17, 139–151.
- Picornell, A., Oteros, J., Ruiz-Mata, R., Recio, M., Trigo, M.M., Martínez-Bracero, M., Lara, B., Serrano-García, A., Galán, C., García-Mozo, H., Alcázar, P., Pérez-Badia, R., Cabezudo, B., Romero-Morte, J., Rojo, J., 2021. Methods for interpolating missing data in aerobiological databases. *Environ. Res.* 200, 111391. <https://doi.org/10.1016/j.envres.2021.111391>
- Puigròs, A., Muñoz-Cano, R., Roger, R.A., Raga, E., Belmonte, J., Valero, A., 2015. Prevalence of sensitization to pollen from trees planted in Barcelona city. *J. Investig. Allergol. Clin. Immunol.* 25, 133–162.
- Ramsay, J., Hooker, G., Graves, S., 2009. *Functional Data Analysis with R and MATLAB, Applied Spatial Data Analysis with R.* Springer New York, New York, NY. <https://doi.org/10.1007/978-0-387-98185-7>
- Ramsay, J., Silverman, B.W., 2005. *Functional Data Analysis.* 2nd edn. Springer, New York.
- Ramsay, J., Silverman, B.W., 2002. *Applied Functional Data Analysis: Methods and Case Studies.* Springer-Verlag, New York.
- Ramsay, J.O., Wang, X., Flanagan, R., 1995. A Functional Data Analysis of the Pinch Force of Human Fingers. *Appl. Stat.* 44, 17. <https://doi.org/10.2307/2986192>

- Rivas-Martínez, S., Aguiar, C., Aguilera, A., Alonso, R., Alvarez, M., Amich, F., Arnaiz, C., Baccheta, G., Barbero, M., Barbour, M.G., Bartolome, C., 2011. Map of series, geoseries and geopermaseries of vegetation in Spain. *Itinera Geobot.* 18, 1–764.
- Rojó, J., Oteros, J., Pérez-Badia, R., Cervigón, P., Ferencova, Z., Gutiérrez-Bustillo, A.M., Bergmann, K.C., Oliver, G., Thibaudon, M., Albertini, R., Rodríguez-De la Cruz, D., Sánchez-Reyes, E., Sánchez-Sánchez, J., Pessi, A.-M., Reiniharju, J., Saarto, A., Calderón, M.C., Guerrero, C., Berra, D., Bonini, M., Chiodini, E., Fernández-González, D., García, J., Trigo, M.M., Myszkowska, D., Fernández-Rodríguez, S., Tormo-Molina, R., Damialis, A., Kolek, F., Traidl-Hoffmann, C., Severova, E., Caeiro, E., Ribeiro, H., Magyar, D., Makra, L., Udvardy, O., Alcázar, P., Galán, C., Borycka, K., Kasprzyk, I., Newbiggin, E., Adams-Groom, B., Apangu, G.P., Frisk, C.A., Skjøth, C.A., Radišić, P., Šikoparija, B., Celenk, S., Schmidt-Weber, C.B., Buters, J., 2019a. Near-ground effect of height on pollen exposure. *Environ. Res.* 174, 160–169. <https://doi.org/10.1016/j.envres.2019.04.027>
- Rojó, J., Picornell, A., Oteros, J., 2019b. AeRobiology: the computational tool for biological data in the air. *Methods Ecol. Evol.* 10, 2041–210X.13203. <https://doi.org/10.1111/2041-210X.13203>
- Rojó, J., Rapp, A., Lara, B., Fernández-González, F., Pérez-Badia, R., 2015. Effect of land uses and wind direction on the contribution of local sources to airborne pollen. *Sci. Total Environ.* 538, 672–682. <https://doi.org/10.1016/j.scitotenv.2015.08.074>
- Sözen, Ç., Öner, Y., 2022. The investigation of temperature data in Turkey's Black Sea Region using functional data analysis. *J. Appl. Stat.* 49, 2403–2415. <https://doi.org/10.1080/02664763.2021.1896683>
- Subiza, J., Jerez, M., Jimenez, J., Narganes, M., Cabrera, M., Varela, S., Subiza, E., 1995.

- Allergenic pollen and pollinosis in Madrid. *J. Allergy Clin. Immunol.* 96, 15–23.
[https://doi.org/10.1016/S0091-6749\(95\)70028-5](https://doi.org/10.1016/S0091-6749(95)70028-5)
- Ullah, S., Finch, C.F., 2013. Applications of functional data analysis: A systematic review. *BMC Med. Res. Methodol.* 13, 43. <https://doi.org/10.1186/1471-2288-13-43/TABLES/1>
- Valderrama, M.J., Ocaña, F.A., Aguilera, A.M., Ocaña-Peinado, F.M., 2010. Forecasting Pollen Concentration by a Two-Step Functional Model. *Biometrics* 66, 578–585.
<https://doi.org/10.1111/j.1541-0420.2009.01293.x>
- Varela, S., Subiza, J., Subiza, J., Rodríguez, R., García, B., Jerez, M., Jiménez, J., Panzani, R., 1997. *Platanus* pollen as an important cause of pollinosis. *J. Allergy Clin. Immunol.* 100, 748–754. [https://doi.org/10.1016/S0091-6749\(97\)70268-9](https://doi.org/10.1016/S0091-6749(97)70268-9)
- Wu, P.-S., Müller, H.-G., 2010. Functional embedding for the classification of gene expression profiles. *Bioinformatics* 26, 509–517.
<https://doi.org/10.1093/bioinformatics/btp711>

7. Anexo

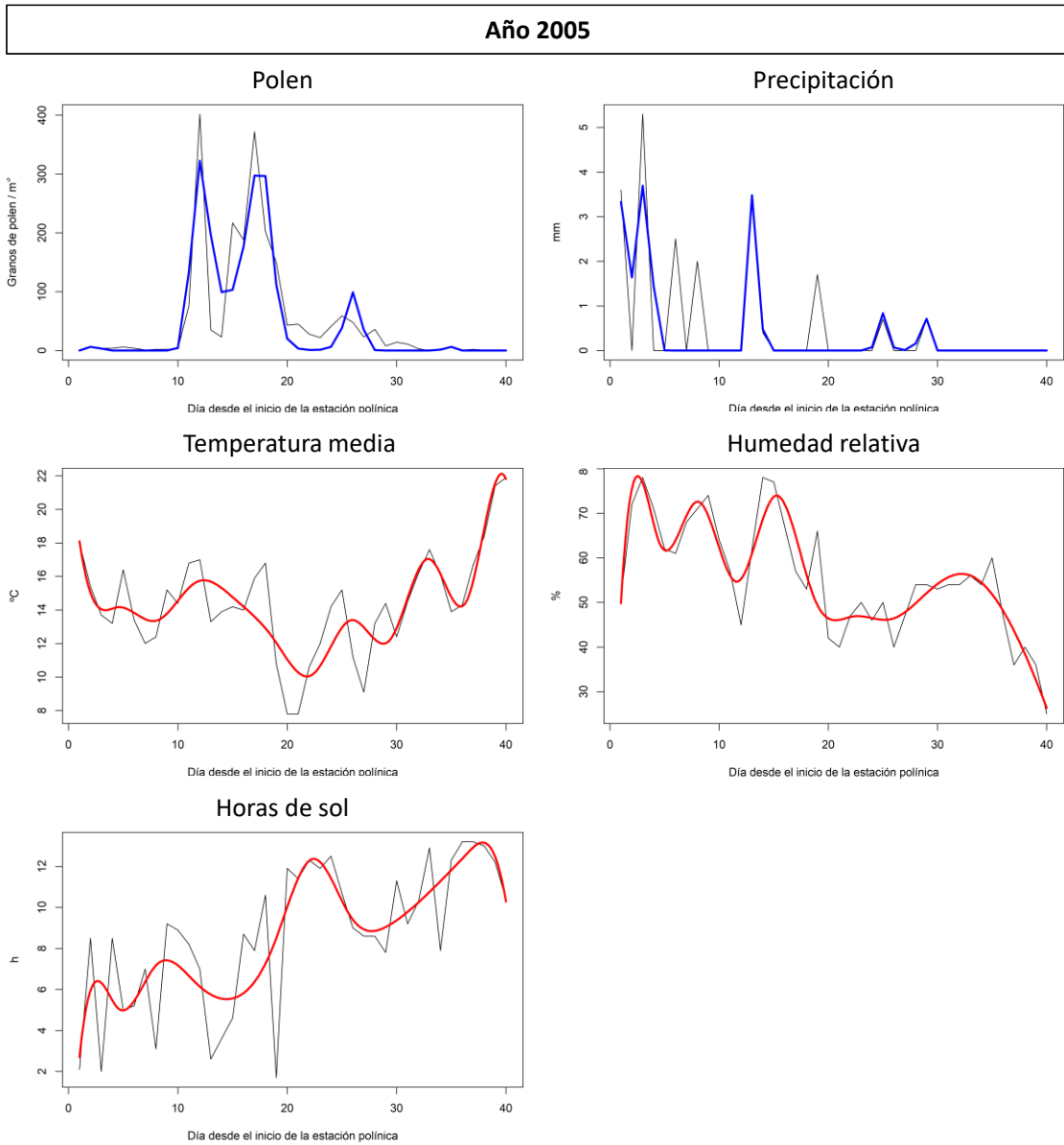


Figura 21. Evolución de las variables analizadas durante el año 2005 en la ciudad de Toledo. Línea negra representa los datos discretos originales, la línea azul representa los datos discretos obtenidos al evaluar la forma funcional y la línea roja representa la forma funcional.

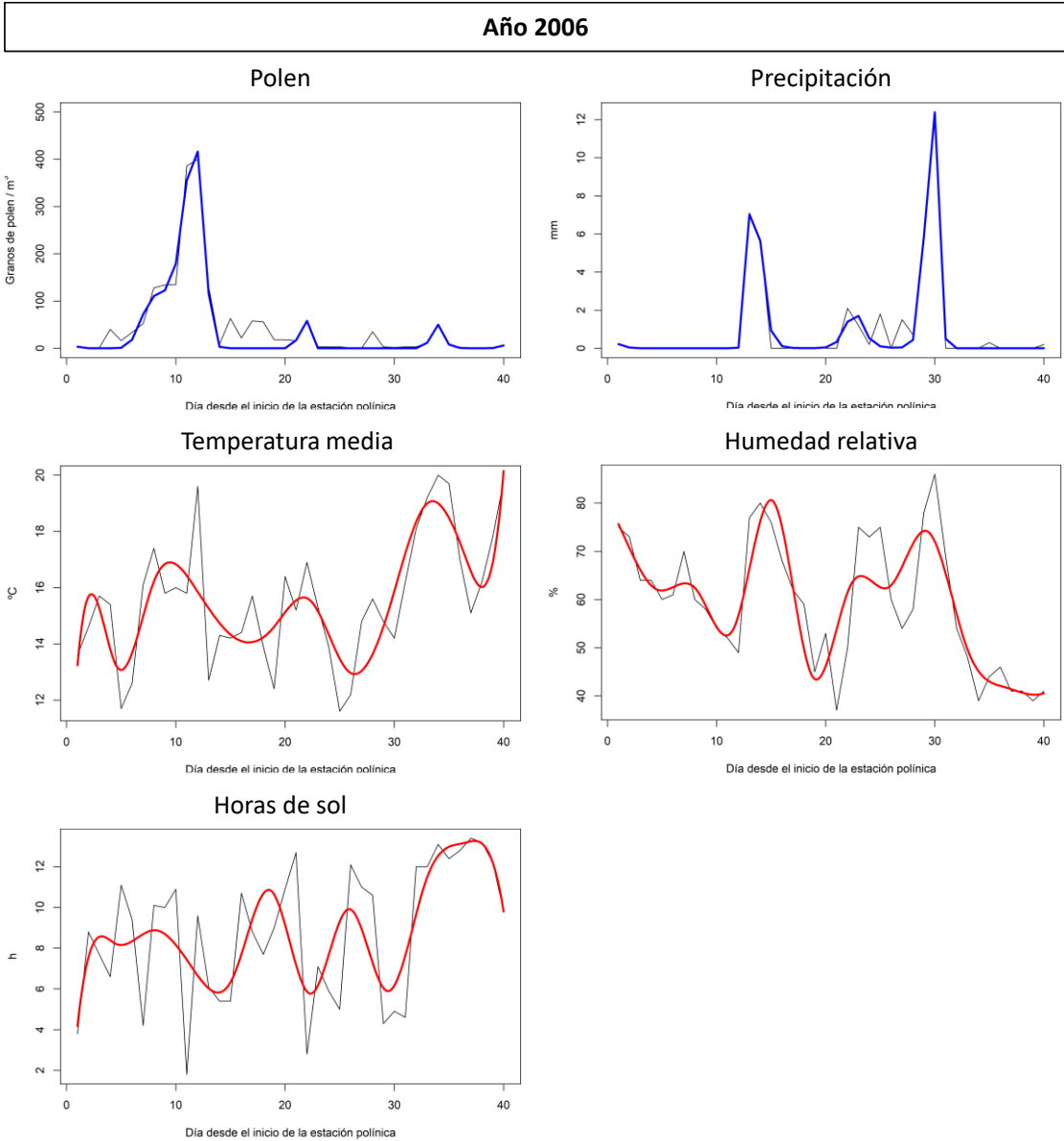


Figura 22. Evolución de las variables analizadas durante el año 2006 en la ciudad de Toledo. Línea negra representa los datos discretos originales, la línea azul representa los datos discretos obtenidos al evaluar la forma funcional y la línea roja representa la forma funcional.

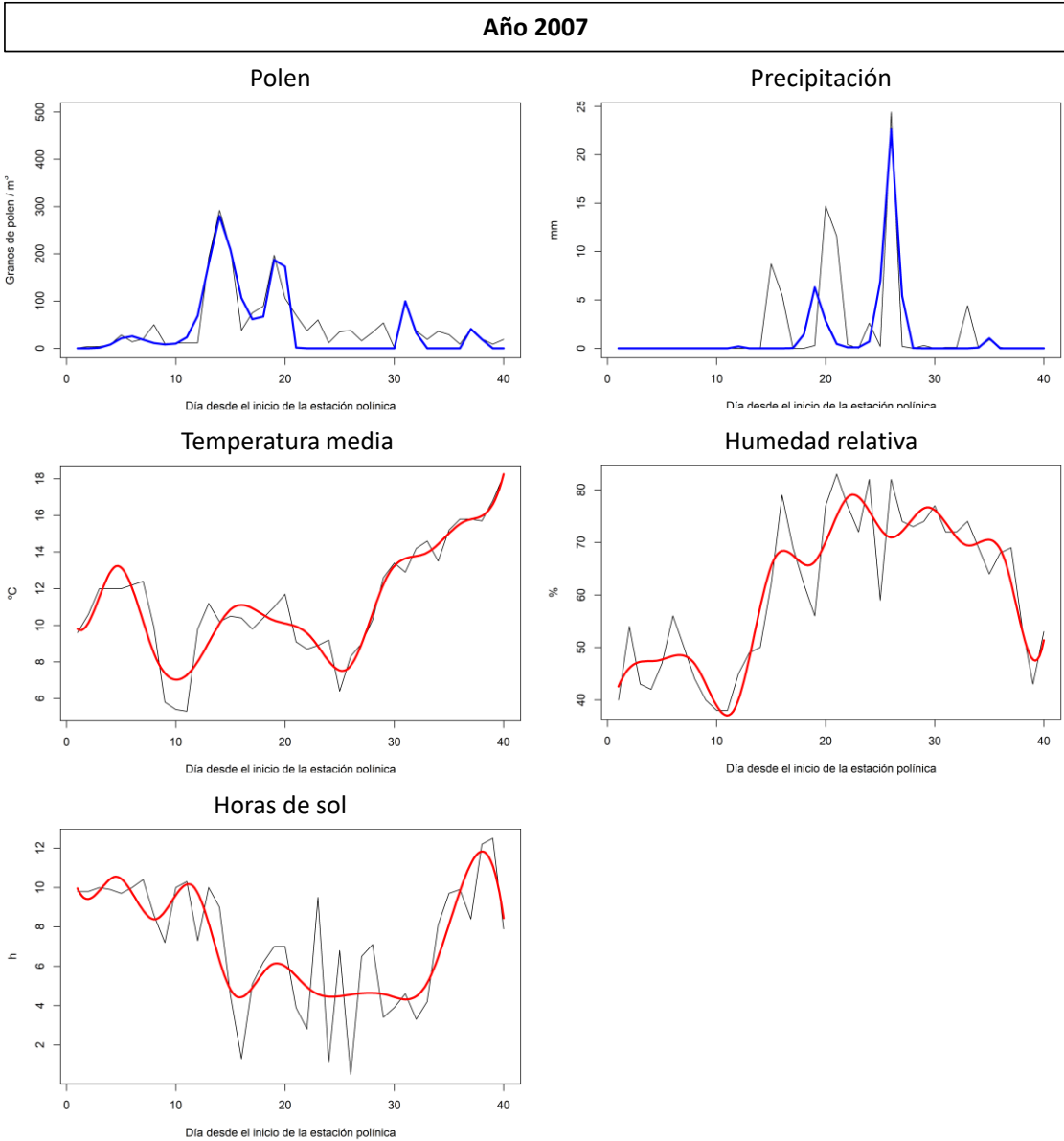


Figura 23. Evolución de las variables analizadas durante el año 2007 en la ciudad de Toledo. Línea negra representa los datos discretos originales, la línea azul representa los datos discretos obtenidos al evaluar la forma funcional y la línea roja representa la forma funcional.

Año 2008

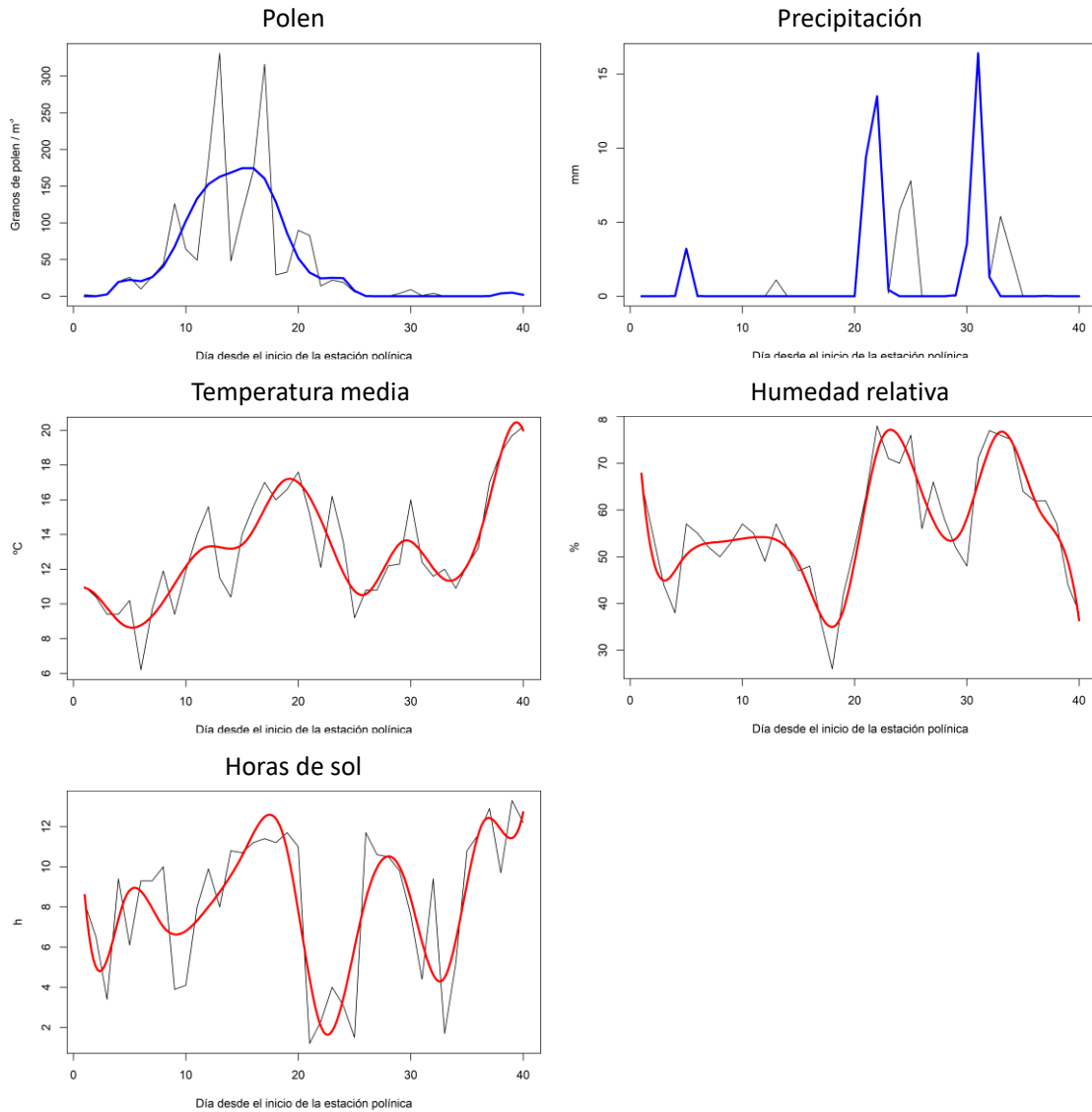


Figura 24. Evolución de las variables analizadas durante el año 2008 en la ciudad de Toledo. Línea negra representa los datos discretos originales, la línea azul representa los datos discretos obtenidos al evaluar la forma funcional y la línea roja representa la forma funcional.

Año 2009

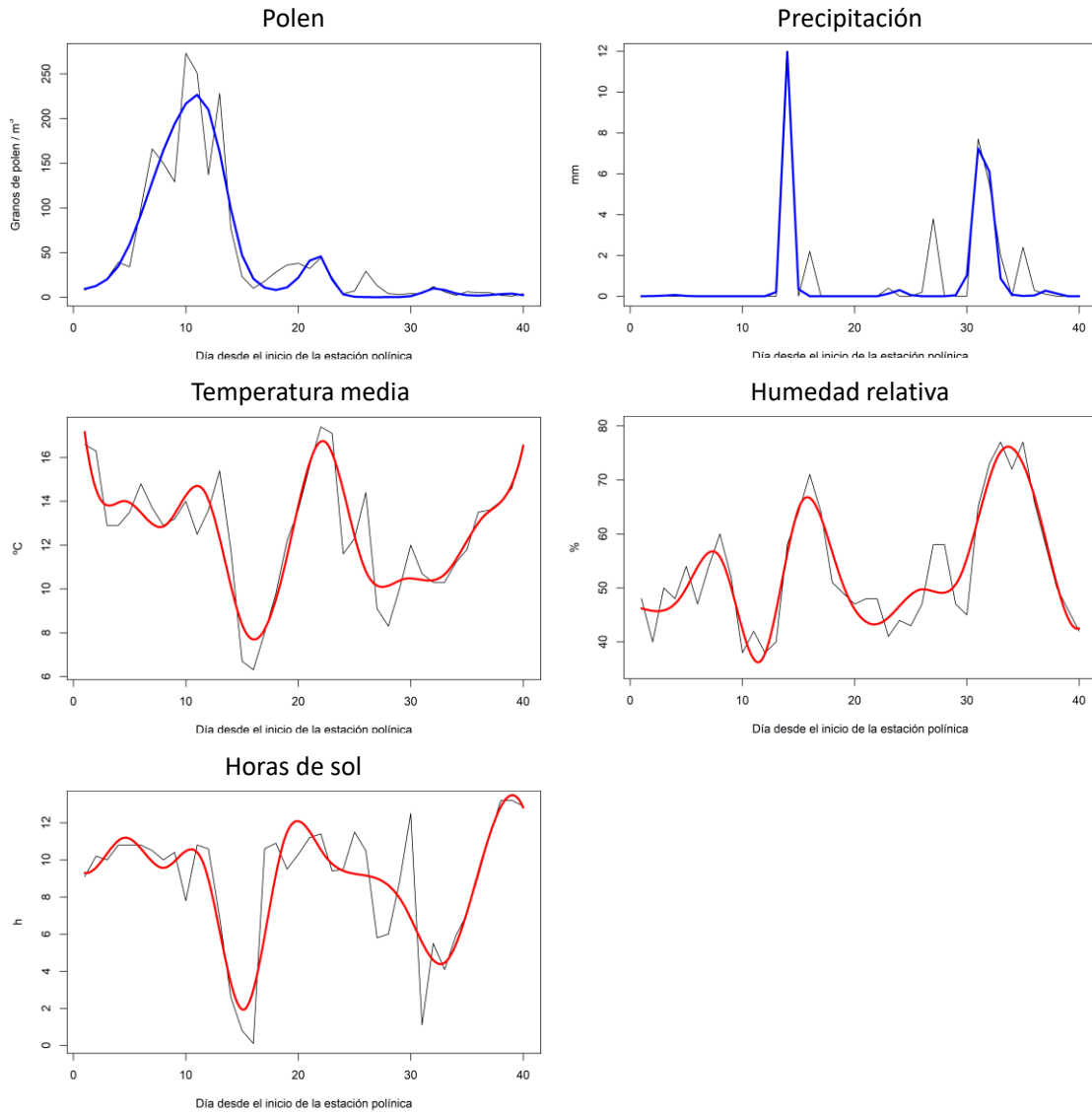


Figura 25. Evolución de las variables analizadas durante el año 2009 en la ciudad de Toledo. Línea negra representa los datos discretos originales, la línea azul representa los datos discretos obtenidos al evaluar la forma funcional y la línea roja representa la forma funcional.

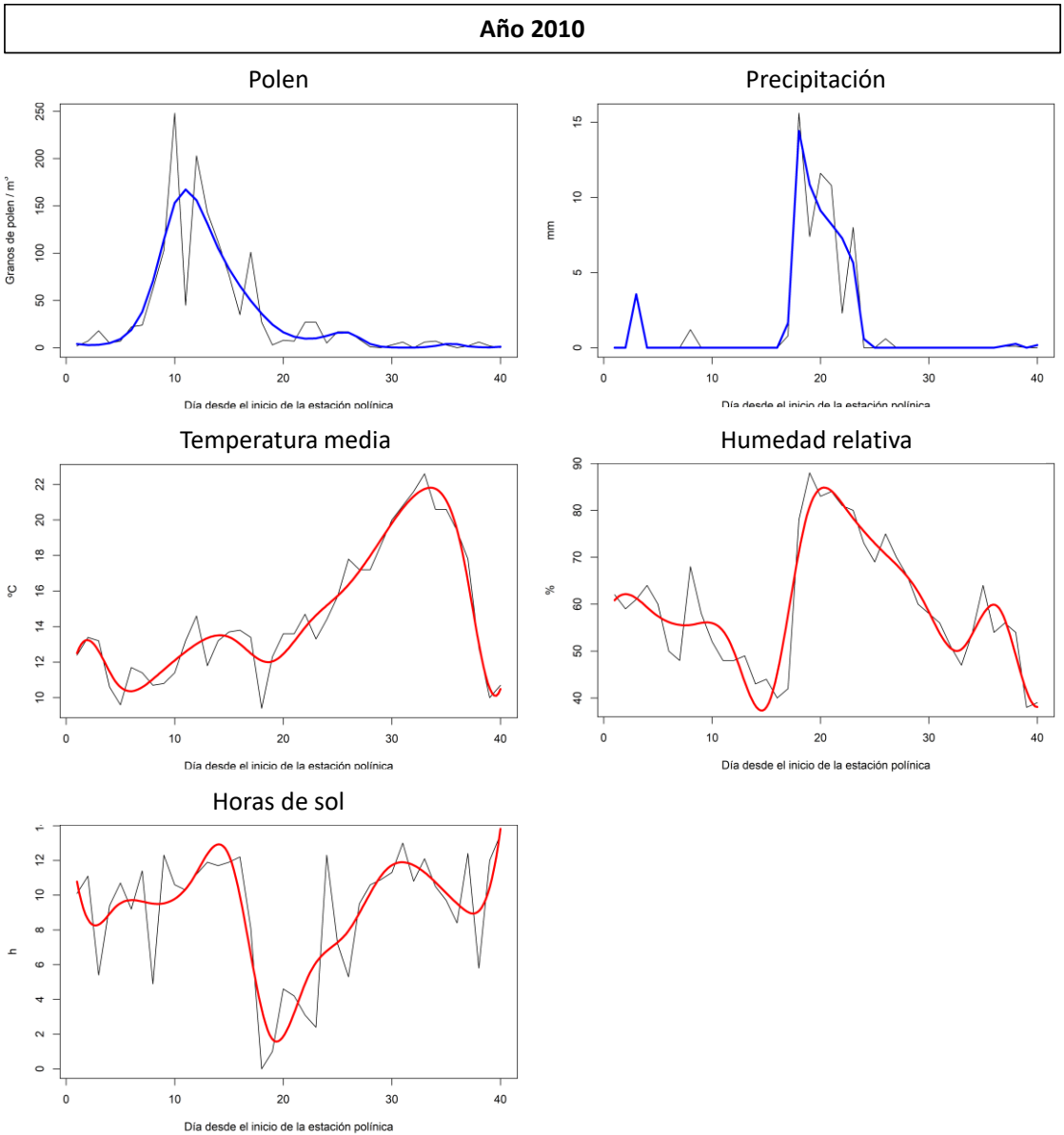


Figura 26. Evolución de las variables analizadas durante el año 2010 en la ciudad de Toledo. Línea negra representa los datos discretos originales, la línea azul representa los datos discretos obtenidos al evaluar la forma funcional y la línea roja representa la forma funcional.

Año 2011

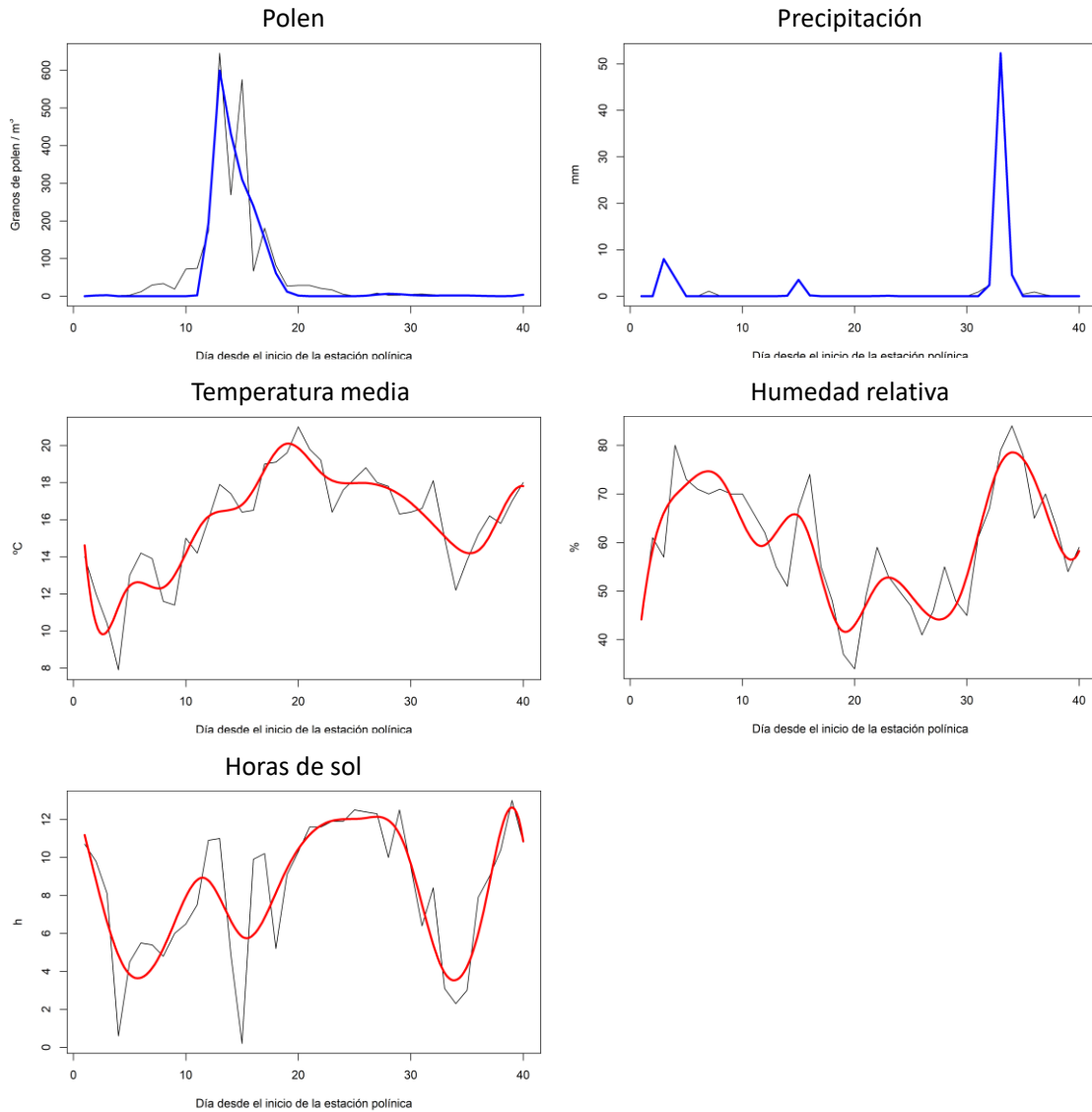


Figura 27. Evolución de las variables analizadas durante el año 2011 en la ciudad de Toledo. Línea negra representa los datos discretos originales, la línea azul representa los datos discretos obtenidos al evaluar la forma funcional y la línea roja representa la forma funcional.

Año 2012

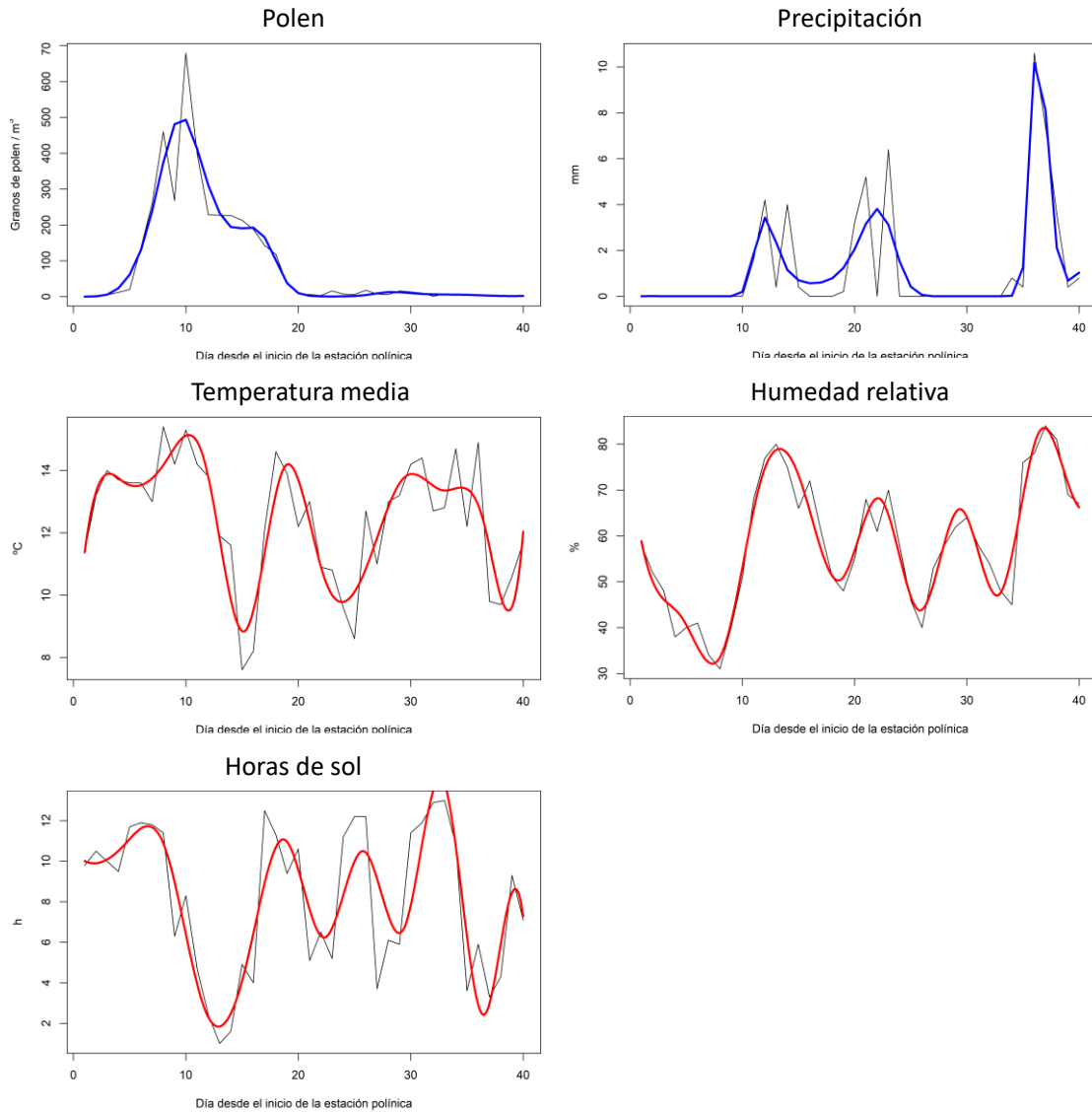


Figura 28. Evolución de las variables analizadas durante el año 2012 en la ciudad de Toledo. Línea negra representa los datos discretos originales, la línea azul representa los datos discretos obtenidos al evaluar la forma funcional y la línea roja representa la forma funcional.

Año 2013

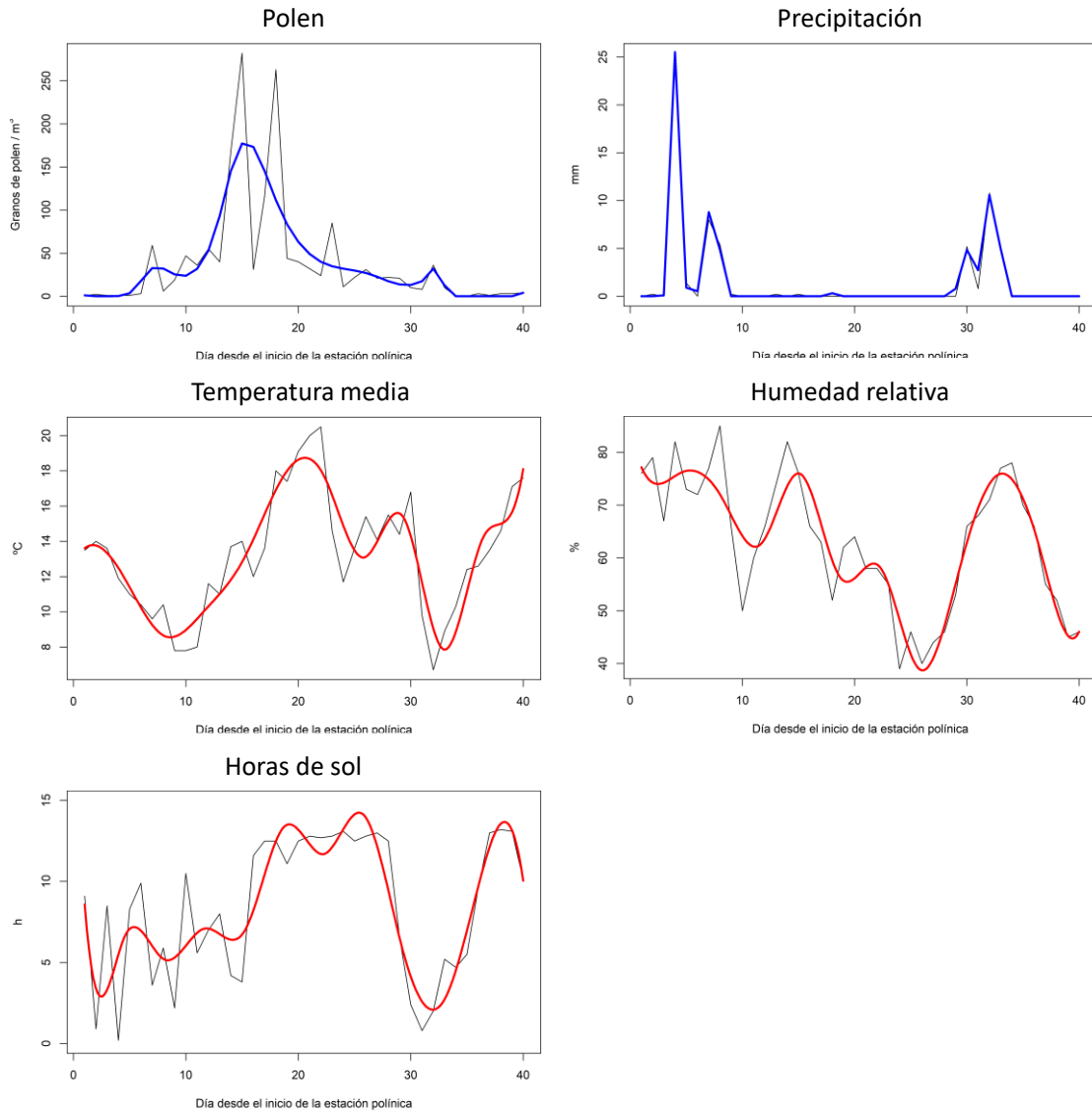


Figura 29. Evolución de las variables analizadas durante el año 2013 en la ciudad de Toledo. Línea negra representa los datos discretos originales, la línea azul representa los datos discretos obtenidos al evaluar la forma funcional y la línea roja representa la forma funcional.

Año 2014

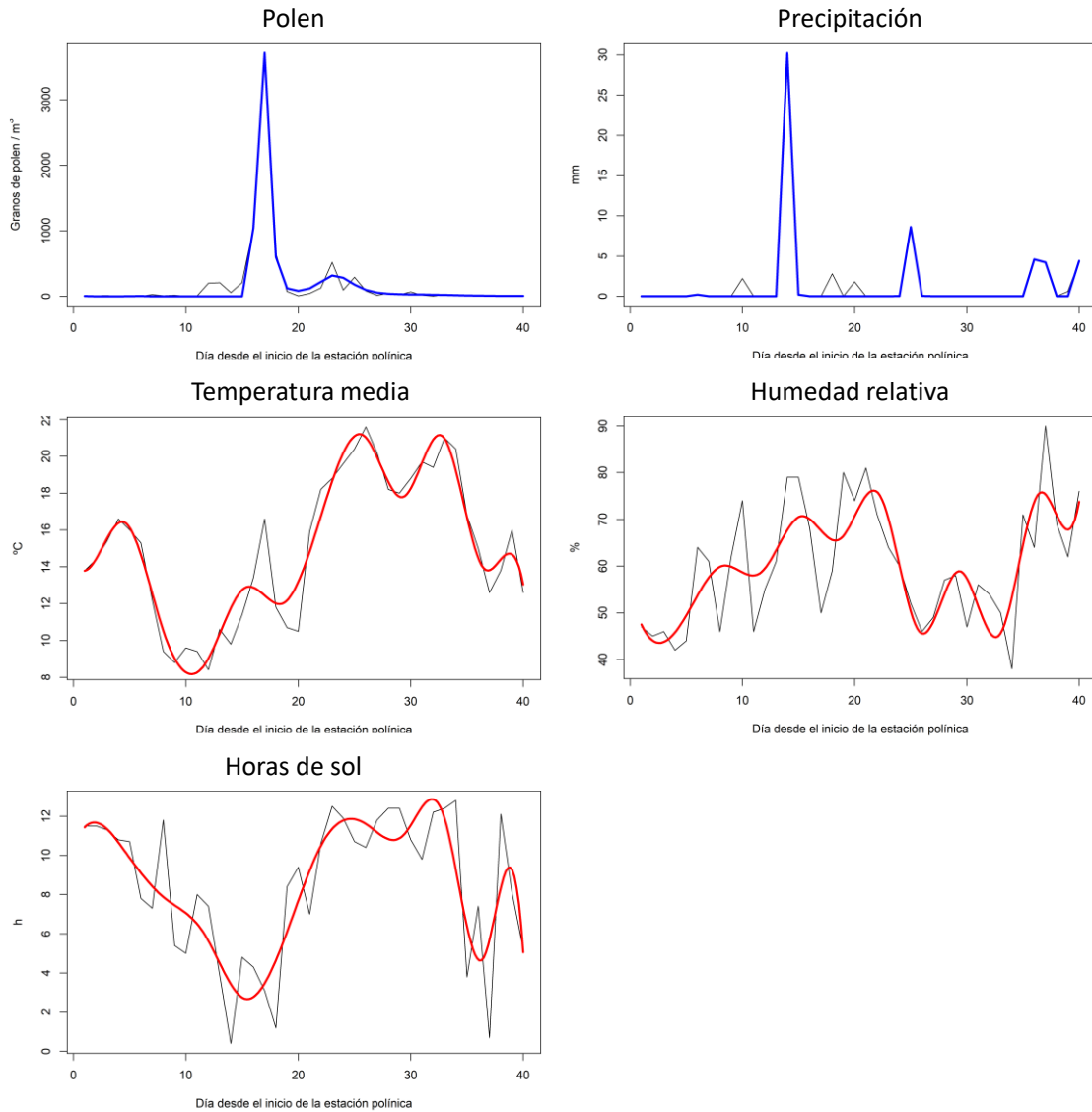


Figura 30. Evolución de las variables analizadas durante el año 2014 en la ciudad de Toledo. Línea negra representa los datos discretos originales, la línea azul representa los datos discretos obtenidos al evaluar la forma funcional y la línea roja representa la forma funcional.

Año 2015

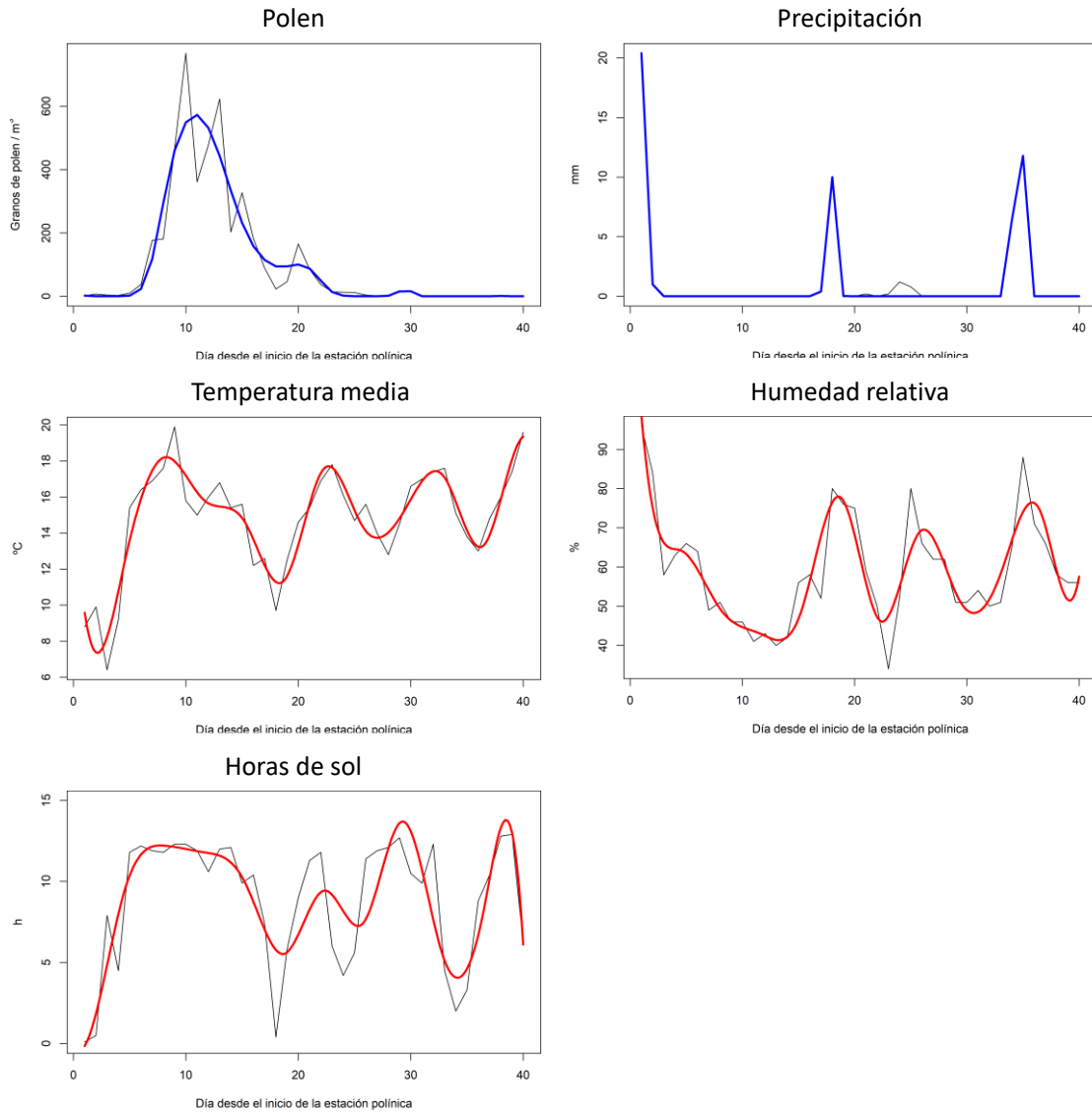


Figura 31. Evolución de las variables analizadas durante el año 2015 en la ciudad de Toledo. Línea negra representa los datos discretos originales, la línea azul representa los datos discretos obtenidos al evaluar la forma funcional y la línea roja representa la forma funcional.

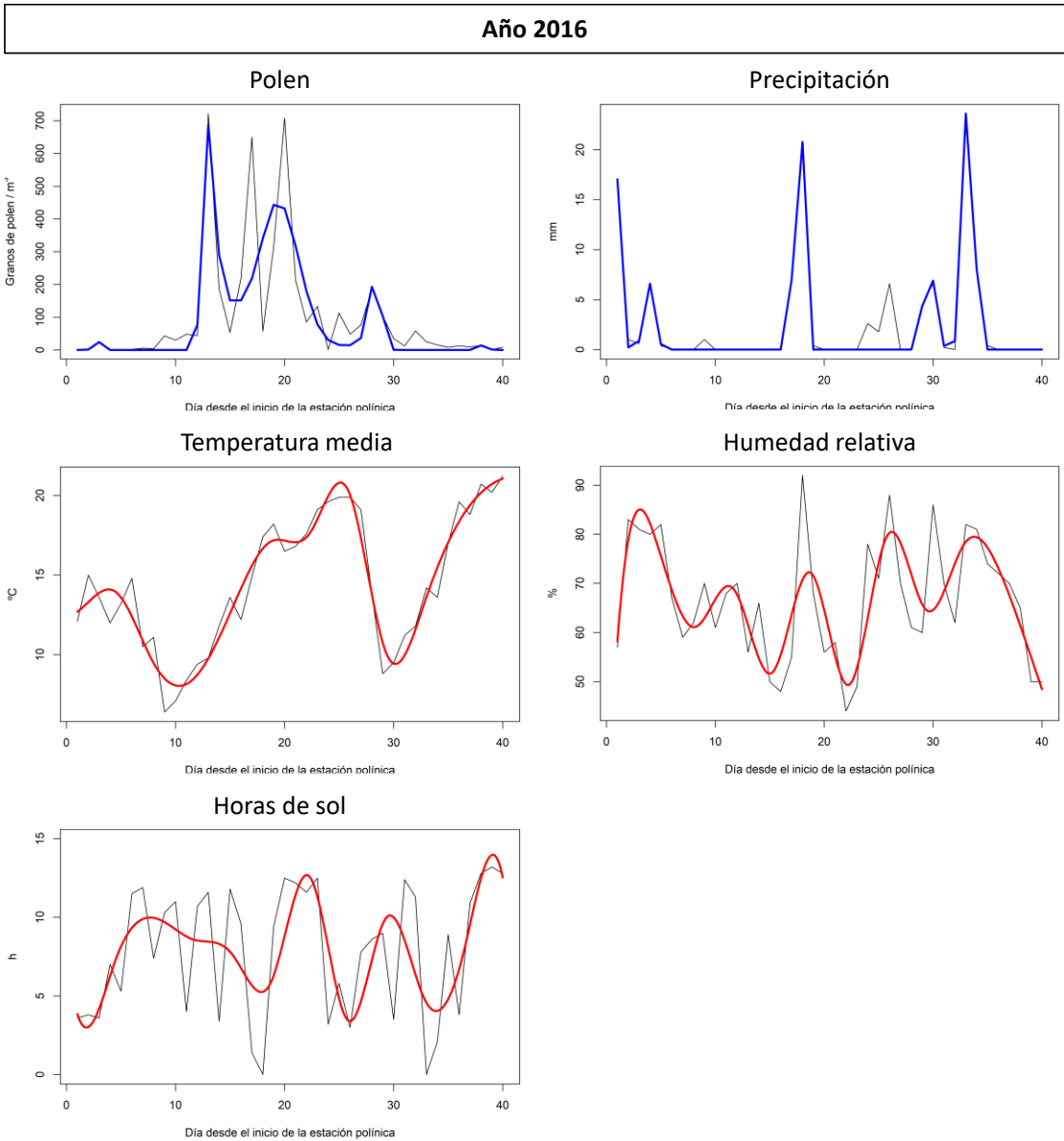


Figura 32. Evolución de las variables analizadas durante el año 2016 en la ciudad de Toledo. Línea negra representa los datos discretos originales, la línea azul representa los datos discretos obtenidos al evaluar la forma funcional y la línea roja representa la forma funcional.

Año 2017

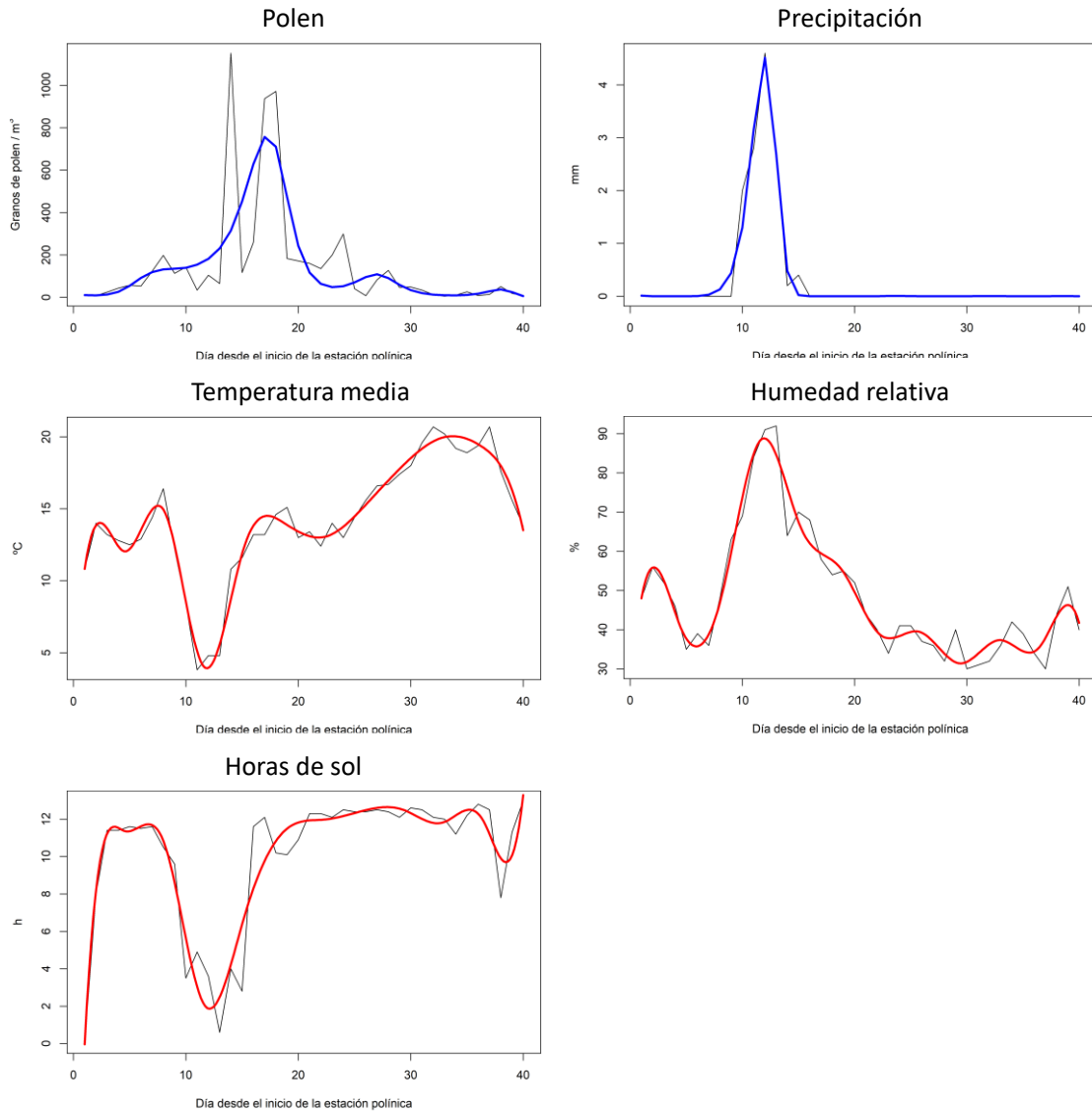


Figura 33. Evolución de las variables analizadas durante el año 2017 en la ciudad de Toledo. Línea negra representa los datos discretos originales, la línea azul representa los datos discretos obtenidos al evaluar la forma funcional y la línea roja representa la forma funcional.

Año 2018

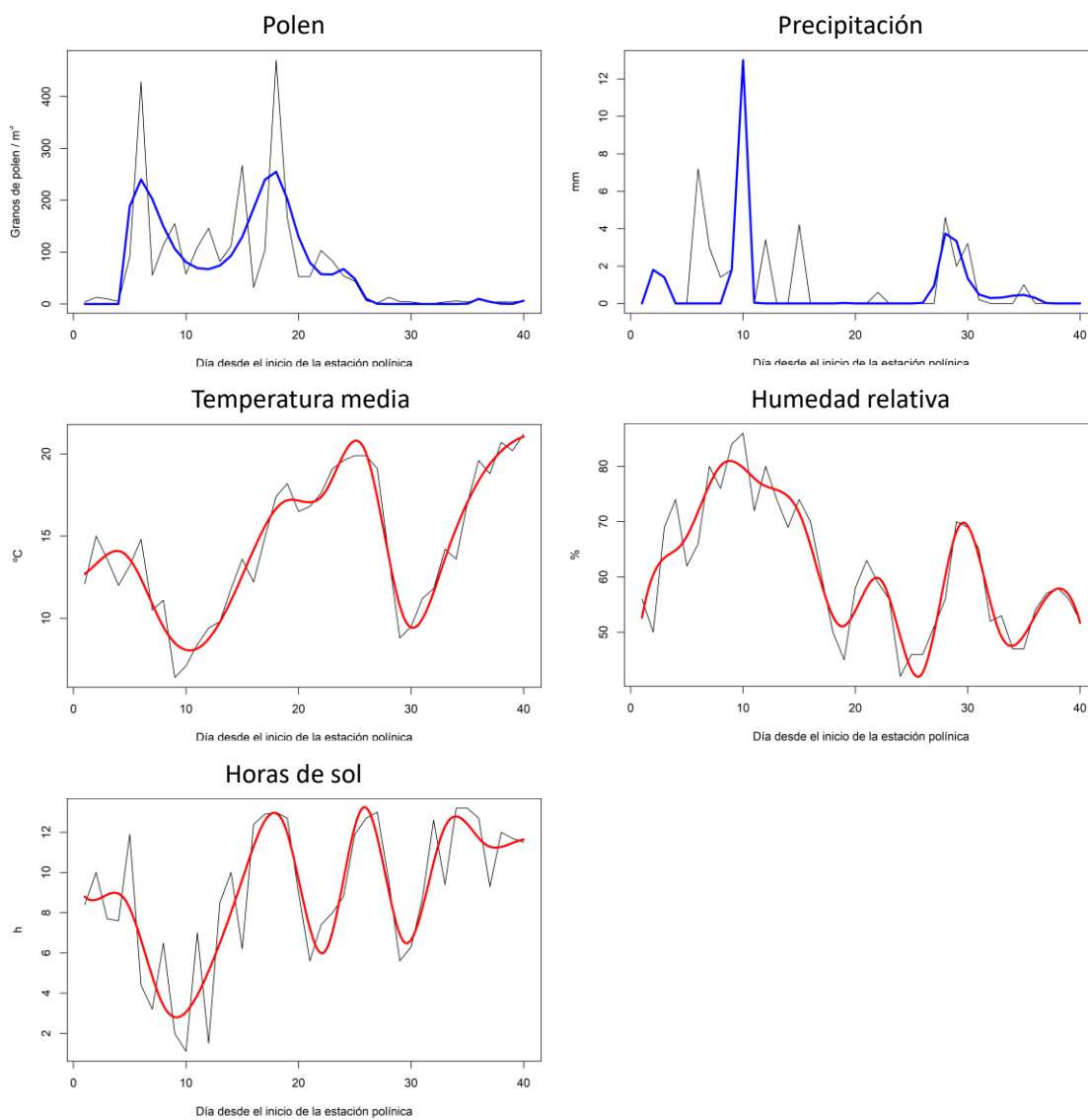


Figura 34. Evolución de las variables analizadas durante el año 2018 en la ciudad de Toledo. Línea negra representa los datos discretos originales, la línea azul representa los datos discretos obtenidos al evaluar la forma funcional y la línea roja representa la forma funcional.

