



---

**Universidad de Valladolid**

Facultad de ciencias

# **TRABAJO FIN DE GRADO**

Grado en Física

**Análisis espacio-temporal de las concentraciones de distintos  
contaminantes**

**Autor: Álvaro Patricio Prieto Pérez**

**Tutor/es: Isidro Pérez Bartolomé  
María Ángeles García Pérez**





*En primer lugar, quiero agradecer a la Medical & Environmental Data Mash-up Infrastructure project por la disposición de sus datos para la realización de este estudio. Del mismo modo, me gustaría agradecer a mis tutores Isidro Pérez Bartolomé y María Ángeles García Pérez por toda la ayuda que me han prestado, pues han hecho que adquiriera nuevos puntos de vista y conocimientos.*

*Además, gracias a su ayuda he visto que este campo es al que me quiero dedicar en mi vida profesional. Por último, pero no por ello menos importante, quiero agradecer a mis amigos y seres queridos por toda la ayuda que me han prestado a lo largo de mi vida. Sin vosotros, el haber llegado hasta aquí y la realización de este trabajo no habría sido posible. Muchas gracias, no puedo encontrar palabras más adecuadas para expresar lo que siento.*





# Índice

<b>Resumen/Abstract</b>	<b>I</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Antecedentes . . . . .	1
1.2. Contaminantes . . . . .	1
1.2.1. Óxidos de nitrógeno . . . . .	1
1.2.2. Ozono . . . . .	2
1.2.3. Partículas $PM_{10}$ y $PM_{2,5}$ . . . . .	3
1.3. Normativa . . . . .	5
<b>2. Objetivos</b>	<b>7</b>
<b>3. Datos y metodología</b>	<b>8</b>
3.1. Análisis estadístico . . . . .	8
3.1.1. Cuantiles . . . . .	8
3.1.2. Posición . . . . .	8
3.1.3. Dispersión . . . . .	9
3.1.4. Simetría . . . . .	9
3.1.5. Aplanamiento . . . . .	10
3.1.6. Archivos con la media anual de concentración . . . . .	10
3.1.7. Archivos con la media diaria de concentración . . . . .	11
3.2. Análisis gráfico . . . . .	12
3.2.1. Diagramas de caja y bigotes . . . . .	13
3.2.2. Histogramas . . . . .	14
3.2.3. Suavizados de histogramas . . . . .	14
3.2.4. Coeficientes de eficiencia . . . . .	16

3.2.5. Zona de estudio . . . . .	17
3.3. Software complementario . . . . .	19
<b>4. Resultados</b>	<b>20</b>
4.1. Dióxido de nitrógeno . . . . .	20
4.2. Ozono . . . . .	26
4.3. Partículas $PM_{10}$ . . . . .	32
4.4. Partículas $PM_{2,5}$ . . . . .	37
<b>5. Conclusiones</b>	<b>44</b>
<b>6. Futuras líneas de estudio</b>	<b>45</b>
<b>7. Bibliografía</b>	<b>46</b>

## Resumen

En este trabajo hemos estudiado las distribuciones y concentraciones anuales y mensuales del  $NO_2$ ,  $O_3$ ,  $PM_{10}$  y  $PM_{2,5}$  en Inglaterra y Gales en el período 2007-2011. Para ello, hemos utilizado los datos dispuestos por la Medical & Environmental Data Mash-up Infrastructure project, que tiene medias anuales y diarias con una resolución espacial de un kilómetro cuadrado para cada contaminante estudiado.

Con los datos de media diaria, hemos aplicado algoritmos para calcular la media mensual para cada localización. Dado que los datos anuales ya tienen la media anual para cada localización, no ha habido necesidad de aplicarle algoritmos.

Para analizar las distribuciones y concentraciones mensuales, hemos representado diagramas de violín de la media mensual de todas las localizaciones por año de cada contaminante. También hemos representado mapas de contorno de las concentraciones de cada uno en el territorio de Inglaterra y Gales.

Para analizar las distribuciones y concentraciones anuales, hemos representado diagramas de caja y bigotes de la media anual de todas las localizaciones de cada contaminante y mapas de contorno de las concentraciones de cada uno en el territorio de Inglaterra y Gales. También hemos representado histogramas y sus suavizados de la concentración de cada contaminante y hemos calculado el porcentaje de territorio afectado por concentraciones por encima de los valores límites establecidos por la Organización Mundial de la Salud y por la Unión Europea.

Mediante estos análisis hemos visto que, en general para el período 2007-2011, la concentración de cada contaminante suele presentar máximos en determinadas estaciones del año, que para el  $NO_2$  suele haber una mayor concentración en casi la mayor parte de Inglaterra (salvo el norte) y que para el  $PM_{10}$  y  $PM_{2,5}$  el oeste del territorio de Inglaterra y Gales tiene mayor concentración que el este. Del mismo modo, la media anual de concentración de  $NO_2$  en algunas ciudades superó el valor límite establecido por la Organización Mundial de la Salud y la Unión Europea, mientras que la media anual de concentración del  $PM_{2,5}$  superó en casi la totalidad del territorio el valor límite establecido por la Organización Mundial de la Salud.

## Abstract

In this work we have studied the annual and monthly distributions and concentrations of  $NO_2$ ,  $O_3$ ,  $PM_{10}$  and  $PM_{2,5}$  in England and Wales in the period 2007-2011. For that, we have used the data provided by the Medical & Environmental Data Mash-up Infrastructure project, which has annual and daily means with a spatial resolution of one square kilometer for all the pollutants studied.

With the daily mean data, we have applicated algorithms to compute the monthly mean for each location. Since the annual data already have the annual mean for every location, there was no need to apply algorithms to it.

To analyse the monthly distributions and concentrations, we have plotted violin plots of the monthly mean concentrations of every location per year of each pollutant. We have also depicted

contour maps of the concentrations of each pollutant in the territory of England and Wales.

To analyse the annual distributions and concentrations, we have plotted boxplots of the annual mean concentrations of every location of each pollutant and the contour maps of the concentrations of each pollutant in the territory of England and Wales. We have also depicted histograms and their density smoothing of the concentration of each pollutant and calculated the percentage of the territory which is affected by concentrations over the value limits established by the World Health Organization and the Europe Union.

By performing these analysis we have concluded that, in general for the period 2007-2011, each pollutant's concentration tends to present maximums in particular seasons through the year, that  $NO_2$  tends to be more concentrated in almost all England's territory (except for the north) and that  $PM_{10}$  and  $PM_{2,5}$  concentrations tend to be more concentrated in the west of England and Wales's territory. Furthermore, the mean annual concentration of  $NO_2$  in some cities surpassed the World Health Organization and European Union's limit value, and the mean annual concentration of  $PM_{2,5}$  surpassed in almost the whole territory the World Health Organization's limit value.

# 1. Introducción

## 1.1. Antecedentes

La calidad del aire respecto a la concentración de contaminantes se ha vuelto una cuestión muy importante a lo largo de las últimas décadas. No solo tiene efectos en la salud de las personas, como varios estudios y sucesos como La Gran Niebla de 1952 prueban (Díaz et al., 2018; Met Office, 2020a; Rovira et al., 2020; Shaughnessy et al., 2015; Shi et al., 2018; WHO, 2006), sino que también tiene impactos en los ecosistemas, bienes materiales y en el cambio climático (EEA, 2019). Por ello, la monitorización de la concentración de contaminantes es importante para la protección de la salud humana, entre otras cosas.

En el Reino Unido, se tienen unos 170 sitios de monitorización horaria de la calidad del aire localizados tanto en zonas rurales y apartadas como en zonas altamente industrializadas y contaminadas. Ahora bien, con estos 170 sitios no es posible dar una estimación certera de la concentración de contaminantes en una localización cualquiera. Por ello, el modelo de *Air Quality Unified Model* (AQUM) es un modelo de dispersión atmosférico que puede dar una mejor estimación espacial de la concentración de distintos contaminantes, pero que tiene el inconveniente de que los datos obtenidos están sesgados negativamente.

Por este motivo, la **Medical & Environmental Data Mash-up Infrastructure project (MEDMI)** ha estimado las concentraciones de los cuatro contaminantes más perjudiciales para la salud humana, esto es, dióxido de nitrógeno ( $NO_2$ ), ozono ( $O_3$ ),  $PM_{10}$  y  $PM_{2,5}$ , en el territorio de Inglaterra y Gales para el período del 2007 al 2011 con una resolución espacial de un kilómetro cuadrado. Ello ha sido posible utilizando un modelo espacio-temporal bayesiano para corregir el sesgo que presenta el AQUM (Mukhopadhyay y Sahu, 2016). Los datos obtenidos están almacenados en archivos .txt y pueden ser descargados en <https://www.data-mashup.org.uk/research-projects/statistical-downscaling-of-gridded-air-quality-data/>.

A partir de los datos proporcionados por **MEDMI**, es posible realizar un estudio de la distribución de la concentración de estos contaminantes en el territorio de Inglaterra y Gales, tanto a escala anual como a escala mensual.

## 1.2. Contaminantes

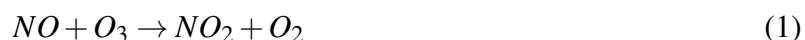
En esta sección se expondrán las principales características y reacciones en las que se encuentran involucrados los contaminantes que serán estudiados. Para ello, nos hemos basado sobre todo en el libro *Atmospheric Science. An introductory Survey* (Wallace y Hobbs, 2006) y en las pautas para la calidad del aire de la Organización Mundial de la Salud (WHO, 2006).

### 1.2.1. Óxidos de nitrógeno

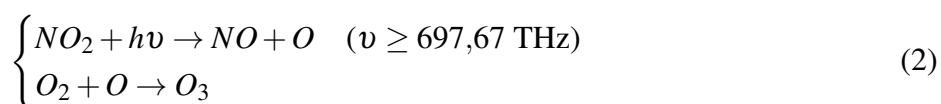
Los óxidos de nitrógeno (denominados  $NO_x$ ), esto es, el monóxido de nitrógeno ( $NO$ ) y el dióxido de nitrógeno ( $NO_2$ ), juegan un papel importante en la química de la atmósfera. Tienen fuentes naturales, como la intrusión de  $NO_x$  de la estratosfera, actividad volcánica y bacterial, etc. Ahora bien,

las mayores fuentes de óxidos de nitrógeno son antropogénicas, principalmente por procesos de combustión, ya sea de fuentes estacionarias (industria, generación de energía, etc.) o móviles (combustión interna de vehículos).

Sin embargo, el dióxido de nitrógeno no es emitido directamente por los procesos de combustión a la atmósfera, pues de los óxidos de nitrógeno emitidos por combustión, el 95 % se corresponde al monóxido de nitrógeno, y el 5 % restante se corresponde al dióxido de nitrógeno. Sin embargo, el monóxido de nitrógeno es capaz de reaccionar con el ozono presente en la troposfera de la siguiente manera:



Así, la mayor parte de dióxido de nitrógeno presente en la troposfera no es producto de la emisión directa de procesos de combustión (ya sea industrial, de vehículos, etc.), sino de la reacción con el ozono. Además, el dióxido de nitrógeno es capaz de descomponerse con luz solar (con  $\lambda \leq 430$  nm, muy próximo al ultravioleta) para volver a formar monóxido de nitrógeno, lo cual permite que pueda formarse ozono, véase:



De este modo, se establece un equilibrio entre la concentración del  $NO$  y la concentración del  $NO_2$  a través de las reacciones 1 y 2. Ahora bien, el dióxido de nitrógeno también puede reaccionar con otros compuestos. Durante el día, el principal sumidero del  $NO_2$  se produce a través de la reacción con el anión hidróxido<sup>1</sup>:



Por la noche, en el cual no se producen radicales hidróxidos, el dióxido de nitrógeno es oxidado por el ozono, formando  $NO_3$ . Este último reacciona con el  $NO_2$  presente para formar  $N_2O_5$ , y este último reacciona con agua para formar ácido nítrico. De ese modo, el tiempo de residencia en la atmósfera del dióxido de nitrógeno es de un día, aproximadamente.

Los impactos en la salud producidos solo por el  $NO_2$  son difíciles de determinar, pues suele venir acompañado de más contaminantes (la fuentes de  $NO_2$  son las mismas que gran parte del material particulado). Ahora bien, se sabe que puede afectar en el funcionamiento pulmonar (reduciendo su capacidad), tener impactos en el desarrollo fetal, aumentar la posibilidad de desarrollar cáncer, etc.

### 1.2.2. Ozono

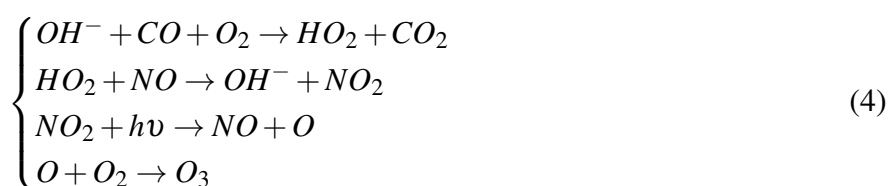
La alta reactividad del ozono cumple un papel importante en la capacidad oxidante de la troposfera, pero la mayor parte del  $O_3$  presente en la atmósfera se encuentra en la estratosfera: tan solo un

---

<sup>1</sup>El anión hidróxido se produce a través de la reacción de un átomo de oxígeno (producido por la descomposición de ozono troposférico debido a rayos UV) con una molécula de agua.

10% es ozono troposférico. Esta alta cantidad de ozono en la estratosfera es crucial, pues forma una barrera protectora que reduce la intensidad de la radiación UV proveniente del sol (con longitudes de onda entre 0,23 y 0,32  $\mu m$ ). Si bien la intrusión de ozono de la estratosfera a la troposfera mediante el descenso de corrientes de aire seco de ciclones de latitudes medias es una fuente de ozono troposférico (Lin et al., 2012), no es la única. Gases como el  $NO$  o el  $CO$  y compuestos orgánicos (todos ellos emitidos por actividad antropogénica) son capaces de producir ozono en la troposfera por reacciones fotoquímicas, por lo que es un contaminante secundario (no es emitido directamente a la atmósfera, sino que es formado por reacciones entre otros contaminantes).

Por ejemplo, en las reacciones 1 y 2 se encuentra involucrado el ozono. Dado que el monóxido de nitrógeno y el dióxido de nitrógeno se equilibran, no hay producción o creación neta de ozono en el proceso completo. Ahora bien, dada la alta reactividad de ozono, en atmósferas con una concentración suficiente de  $NO$ , pueden darse las siguientes reacciones:



Podemos ver que en la producción del  $NO_2$  no ha habido moléculas de ozono involucradas y, del mismo modo, que no ha habido variación en la concentración de  $OH^-$  y del  $NO$ . Por ello, esta cadena de reacciones genera moléculas de ozono, lo que implica que un aumento de la concentración de  $NO$  conlleva un aumento de la concentración de ozono troposférico. Del mismo modo, hay otra serie de reacciones relacionadas con moléculas orgánicas que, en presencia de  $NO$ , forma  $NO_2$  y da lugar a producción de ozono por las reacciones 2.

Hay numerosos factores que afectan a la concentración del ozono. Factores meteorológicos como la temperatura o la duración de la luz solar favorecen la concentración del ozono, mientras que la humedad o las precipitaciones lo desfavorecen. Así mismo, se ha visto que el ciclo solar de once años es también un factor importante en su formación. Por supuesto, otros contaminantes tienen un gran efecto en la concentración de  $O_3$  (como los óxidos de nitrógeno), y se ha visto que la reducción de visibilidad por contaminantes tiene también impacto (Chen et al., 2020a; 2020b; Goral et al., 2020).

Altos niveles de ozono en la troposfera pueden tener impacto en la salud, tanto a corto como a largo plazo. A corto plazo, puede producir alteraciones en el funcionamiento de los aparatos respiratorio y cardiovascular, e incluso elevar la tasa de mortalidad. A largo plazo, puede producir inflamación pulmonar, incentivar el desarrollo de asma y reducir la esperanza de vida. En España, se han encontrado evidencias de mortalidad por causas respiratorias y cardiovasculares por exposiciones a concentraciones altas de ozono (Díaz et al., 2018).

### 1.2.3. Partículas $PM_{10}$ y $PM_{2,5}$

Las partículas PM (del inglés *Particulate Matter*) son una mezcla compleja de partículas sólidas y líquidas en suspensión en el aire, con características físicas y químicas diversas. Debido a su alta heterogeneidad, las diferentes características de PM pueden ser determinantes para distintos efectos de la salud en las personas. Ahora bien, las partículas suelen ser clasificadas por sus propiedades aerodinámicas, pues estas últimas son las que determinan los procesos de disipación y transporte en

el aire y, por otro lado, de deposición y limpieza en los conductos respiratorios. Dado que el tamaño de las partículas es crítico en determinar dónde y con qué facilidad se depositan en los conductos respiratorios, las partículas PM se clasifican por su diámetro aerodinámico.

El  $PM_{10}$ , esto es, partículas de menos de  $10\ \mu m$ , por definición engloba también al  $PM_{2,5}$ . Por ello, las partículas entre  $2,5$  y  $10\ \mu m$  se les denomina fracción gruesa. Esta fracción gruesa se forma principalmente de manera mecánica, por la rotura de partículas sólidas, ya sea causado por levantamiento de polvo debido al viento (en tierra al aire libre o en procesos de agricultura), por procesos industriales o por procesos biológicos (granos de polen, por ejemplo). Estas partículas son lo suficientemente pequeñas como para penetrar en la región torácica.

El  $PM_{2,5}$  (partículas de menos de  $2,5\ \mu m$ ) es generado principalmente por gases, ya sea por condensación de metales y compuestos orgánicos vaporizados por procesos de combustión, o bien por condensación de gases producto de reacciones químicas en la atmósfera. En particular, la mayor parte del  $PM_{2,5}$  procede de sulfatos y sustancias orgánicas. En otras condiciones, los nitratos también pasan a ser los principales precursores, etc. Por ejemplo, a partir del dióxido de nitrógeno, este último puede reaccionar para formar  $HNO_3$  según la reacción 3, que forma parte de estas partículas (por ello, una parte del  $NO_2$  pasa a ozono, y otra se vuelve precursora de  $PM_{2,5}$ ). El  $PM_{2,5}$  es lo suficientemente pequeño como para depositarse en los conductos más pequeños del aparato respiratorio y en los alvéolos.

Los procesos meteorológicos pueden afectar a la concentración de partículas: fuertes vientos pueden favorecer la formación de material particulado pero, al mismo tiempo, favorece su transporte y dispersión. Del mismo modo, ambientes fríos con alta humedad relativa pueden favorecer la formación de partículas por condensación. Se ha visto que las temperaturas, velocidad del viento y las precipitaciones están inversamente relacionadas con la concentración de material particulado. En general, las partículas más pequeñas tienden a permanecer más tiempo en suspensión en la atmósfera y, por ende, pueden viajar mayores distancias que partículas más gruesas (de Jesús et al., 2020; Munir, 2017).

Los efectos del  $PM$  en la salud están asociados a un incremento de los síntomas en pacientes con asma, a incrementos en el riesgo de sufrir infartos de miocardio (alteran el aparato circulatorio), inflamación en órganos del aparato respiratorio e, incluso, la posibilidad de desarrollar COPD (enfermedad pulmonar obstructiva crónica). En el sureste asiático se ha visto un incremento de la mortalidad prematura por la concentración de  $PM_{2,5}$  (Shi et al., 2018), mientras que se ha visto que un cambio a corto plazo de concentraciones de  $PM$  en el ambiente conlleva a un cambio a corto plazo en problemas agudos relacionados con la respiración en jóvenes adultos y sanos (Shaughnessy et al., 2015). Además, en Tarragona se estima que la mortalidad asociada al  $PM_{2,5}$  podría ser reducida si se consigue bajar la concentración del  $PM_{2,5}$  por debajo del valor límite establecido por la OMS (Rovira et al., 2020).

Finalmente, a modo de recuento de los efectos de la salud que tienen los distintos contaminantes, puede decirse que hay evidencias de causar muertes prematuras, ya sea por enfermedades cardiovasculares y enfermedades del aparato respiratorio (incluido cáncer). Exposiciones a corto y largo plazo pueden reducir la capacidad pulmonar, incrementar las posibilidades de padecer infecciones en el aparato respiratorio y agravar el asma. Por otro lado, las exposiciones de mujeres embarazadas a la contaminación atmosférica pueden tener impacto en el desarrollo del feto, así como en el desarrollo de recién nacidos y niños. Además, los impactos de la contaminación atmosférica también dependen de la vulnerabilidad de cada persona: esta última puede aumentar con la edad, con condiciones de deterioro de salud preexistentes, etc.



Hemos visto, sobre todo, los efectos sobre la salud de las personas, pero los contaminantes también pueden afectar a los ecosistemas. Por ejemplo, los óxidos de nitrógeno ( $NO$  y  $NO_2$ ) alteran los ecosistemas terrestres y acuáticos al introducir un exceso de nutrientes de nitrógeno, lo cual puede llevar a la eutrofización del ecosistema. Del mismo modo, contribuyen a la acidificación del suelo. Otro ejemplo es el efecto que produce el ozono presente a nivel del suelo, que reduce el ritmo de crecimiento de la vegetación, lo cual perjudica tanto a los ecosistemas como a las plantaciones agrícolas.

Por otro lado, ciertos contaminantes, como el ozono troposférico y el carbono negro (constituyente de  $PM$ ), contribuyen al calentamiento global; mientras que otros contaminantes (constituyentes del  $PM$  como carbono orgánico, etc.) contribuyen a su enfriamiento. Los cambios en el tiempo atmosférico por el cambio climático pueden cambiar la formación, transporte, deposición y dispersión de contaminantes en la atmósfera.

Finalmente, la contaminación atmosférica puede dañar los materiales, edificios y obras artísticas. Por ejemplo, el  $PM_{10}$  es un factor de riesgo para el ensuciamiento de la piedra caliza y del vidrio. Otro ejemplo es el ozono, que desgasta y desvanece los colores (EEA, 2019).

### 1.3. Normativa

En este apartado se establecerá un resumen de los criterios seguidos por la Organización Mundial de la Salud (OMS) y por la Unión Europea (UE) para la protección de la salud humana. Tanto el análisis de la distribución de la concentración de contaminantes como la clasificación del territorio en función de la concentración de cada uno tendrán en cuenta los criterios de ambos organismos para las medias anuales. Los criterios seguidos por la OMS y por la UE que se tienen en cuenta en este trabajo vienen recogidos en [Air quality guidelines. Global update 2005. Particulate matter, ozone, nitrogen dioxide and sulfur dioxide](#) (WHO, 2006) y [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](#) (Parlamento Europeo y Consejo de la Unión Europea, 2008), respectivamente.

Respecto al  $NO_2$ , tanto la OMS como la UE coinciden en establecer como valor límite la media anual de  $40 \mu g/m^3$ . Este valor límite está establecido para la protección de la salud pública de los efectos del  $NO_2$ .

Para el  $PM_{10}$  y el  $PM_{2,5}$ , la OMS establece el valor límite de media anual de 20 y  $10 \mu g/m^3$ , respectivamente; mientras que la UE establece de valor límite la media anual de 40 y  $25 \mu g/m^3$ , respectivamente. Los valores límites que establece la OMS son los niveles más bajos en los que la mortalidad por cáncer al pulmón y cardiovascular se incrementan por exposición a largo plazo al  $PM$ .

Finalmente, para el ozono no se establece un valor límite para la media anual, sino un valor límite de la máxima media para medias octohorarias móviles que no debe superarse por más de un determinado número de días al año. La OMS establece como valor límite la media octohoraria de  $100 \mu g/m^3$ , mientras que la UE establece como valor límite la media octohoraria de  $120 \mu g/m^3$ . El valor límite que establece la OMS da una protección adecuada para la salud pública, aunque exposiciones a estas concentraciones pueden aumentar de un 1 al 2 % la mortalidad diaria (WHO, 2006).

En la tabla 1, se pueden ver los valores límites de la OMS y de la UE para el  $NO_2$ , el  $O_3$ , el  $PM_{10}$  y el  $PM_{2,5}$ .

Tabla 1: Valores límite de concentración de contaminantes establecidos por la OMS y la UE.

Contaminante	Normativa OMS	Normativa UE
$NO_2$	40 $\mu g/m^3$ media anual	40 $\mu g/m^3$ media anual
$O_3$	100 $\mu g/m^3$ media octohoraria	120 $\mu g/m^3$ media octohoraria
$PM_{10}$	20 $\mu g/m^3$ media anual	40 $\mu g/m^3$ media anual
$PM_{2,5}$	10 $\mu g/m^3$ media anual	25 $\mu g/m^3$ media anual

## 2. Objetivos

Dado que nosotros partimos de un punto en el que no es necesario realizar las mediciones de la concentración de contaminantes, sino que tomamos los datos proporcionados por MEDMI, en este trabajo nos centraremos en el análisis de estos datos para poder obtener información de la distribución de la concentración de contaminantes en Inglaterra y Gales. Con los datos de la media anual será posible estudiar su distribución año a año, mientras que con los datos de la media diaria hallaremos la media mensual y, una vez hecho esto, estudiar la distribución mensual durante un año dado durante el período 2007-2011. Así, podemos establecer como objetivos:

- Estudiar la distribución espacial de la media mensual de concentración de contaminantes a lo largo de un año, de cara a encontrar patrones en su distribución y explicarlos.
- Estudiar la distribución espacial de la media anual de concentración de contaminantes para establecer si, en el período 2007-2011, se cumplió la normativa de la OMS y de la UE.
- Estudiar la distribución espacial de la media anual de concentración de contaminantes con el objetivo de encontrar tendencias en su evolución a lo largo del período 2007-2011.

## 3. Datos y metodología

### 3.1. Análisis estadístico

Antes de utilizar y analizar los datos proporcionados por **MEDMI**, es necesario hacer un breve repaso de las herramientas estadísticas que pueden utilizarse para analizar distribuciones de datos. Para realizar este apartado, nos hemos basado sobre todo en el libro *Statistical Methods in the Atmospheric Sciences* (Wilks, 2019).

A la hora de observar y estudiar la atmósfera o aspectos de ella, se obtienen numerosos datos de todas las variables que se analizan. Además, se suele estudiar en un período más o menos largo en el tiempo, lo cual hace que haya un conjunto enorme de datos a analizar para poder extraer conclusiones. El análisis estadístico permite tratar, clasificar y entender la información que proporcionan estos datos. Ahora bien, los datos extraídos de la atmósfera no siguen el comportamiento que exigen las asunciones de los métodos estadísticos clásicos (lo que implica que no podamos obtener información fiable aplicando esos métodos), lo cual hace que los métodos que utilicemos deban ser robustos y resistentes. Un método *robusto* es aquel que no es sensible a las asunciones acerca de la naturaleza de un conjunto de datos: se desempeña bien en la mayoría de los casos (aunque no de manera óptima); mientras que un método *resistente* es aquel que no se ve influenciado por los extremos de un conjunto de datos.

#### 3.1.1. Cuantiles

Supongamos que tenemos un conjunto numérico de datos con  $n$  elementos. Un **cuantil**  $q_p$  es un número que tiene las mismas unidades que los datos y que el conjunto que forman los números menores que él son la proporción de los datos que denota el subíndice  $p$  (con  $0 < p < 1$ ). Otra forma equivalente de expresar el cuantil  $q_p$  es mediante el **percentil**, definido como  $p \cdot 100$ . Por supuesto, para poder determinar un cuantil o percentil cualquiera, es necesario ordenar de menor a mayor el conjunto de datos  $\{x_1, x_2, x_3, x_4, \dots, x_n\}$ , siendo  $x_{i-1} \leq x_i \leq x_{i+1}$ .

Algunos cuantiles suelen ser bastante utilizados en los análisis de datos, como el cuantil  $q_{0,5}$  o mediana (explicado más adelante), o los cuantiles  $q_{0,25}$  y  $q_{0,75}$  (denominados *cuartiles inferior* y *superior*, respectivamente). Estos tres cuantiles separan el conjunto de datos en cuartos. Del mismo modo, los *terciles inferior* y *superior* ( $q_{0,333}$  y  $q_{0,667}$ , respectivamente) separan al conjunto de datos en tercios. Hay más cuantiles con nombres propios, como los *deciles* ( $q_{0,1}$ ,  $q_{0,2}$ , ...,  $q_{0,9}$ ), etc.

Con los cuantiles es posible hallar estadísticos que aportan información acerca de la naturaleza de los datos sin necesidad de representación gráfica. A continuación, se expondrán estadísticos que dan información acerca de la **posición**, **dispersión**, **simetría** y **aplanamiento**.

#### 3.1.2. Posición

Los estadísticos que atienden a la posición revelan la tendencia central de los datos estudiados. El estadístico más conocido de esta categoría es la **media**. Suponiendo que tenemos un conjunto de  $n$  datos  $\{x_1, x_2, x_3, x_4, \dots, x_n\}$ , la media se expresa como:

$$\hat{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (5)$$

El inconveniente que presenta la media es que no es resistente, pues es muy sensible a los valores extremos en un conjunto de datos. Por ejemplo, los conjuntos  $\{10; 12; 9; 7; 8\}$  y  $\{2; 1; 3; 2; 38\}$  tienen la misma media (9,2). Por ello, se suele utilizar como estadístico robusto y resistente al cuantil  $q_{0,5}$ , también conocido como **mediana**. Es el valor central del conjunto de datos: divide a este en dos partes iguales. En caso de que el número de elementos del conjunto de datos sea impar, la mediana es simplemente el elemento en el centro. En caso de que el número de elementos sea par, entonces hay dos valores centrales, por lo que se suele tomar la media de ambos elementos. Así, la mediana puede definirse como:

$$\begin{cases} q_{0,5} = x_{(n+1)/2} & \text{si } n \text{ es impar} \\ q_{0,5} = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} & \text{si } n \text{ es par} \end{cases} \quad (6)$$

Si volvemos a los dos conjuntos de datos utilizados de ejemplos anteriormente, vemos que la mediana de cada uno de ellos es distinta: la mediana de  $\{10; 12; 9; 7; 8\}$  es 9, mientras que la mediana de  $\{2; 1; 3; 2; 38\}$  es 2.

### 3.1.3. Dispersión

Los estadísticos que atienden a la dispersión muestran la amplitud de los datos alrededor del centro del conjunto de datos estudiados. El estadístico robusto y resistente más usado y simple es el **rango intercuartílico** (RIC), que no es más que la diferencia entre el cuartil superior y el cuartil inferior:

$$RIC = q_{0,75} - q_{0,25} \quad (7)$$

Podemos ver que el RIC da una idea general de la amplitud en el rango central del conjunto de datos estudiados, y dado que omite los extremos superiores e inferiores, es resistente a datos extremos.

### 3.1.4. Simetría

Los estadísticos que atienden a la simetría revelan la forma en la que los datos están agrupados respecto de su centro. El estadístico robusto y resistente que da cuenta de esto es el **índice de Yule-Kendall**, que se define como:

$$\gamma_{YK} = \frac{(q_{0,75} - q_{0,50}) - (q_{0,50} - q_{0,25})}{RIC} = \frac{q_{0,75} - 2q_{0,50} + q_{0,25}}{RIC} \quad (8)$$

Así, vemos que el índice de Yule-Kendall expresa la proporción entre, por un lado, la diferencia de la dispersión entre el cuartil superior y la mediana y la mediana y el cuartil inferior y, por otro lado, el RIC. Un conjunto de datos que tenga una simetría perfecta (en el rango central del conjunto

de datos), esto es, que  $q_{0,75} - q_{0,50} = q_{0,50} - q_{0,25}$ , dará  $\gamma_{YK} = 0$ . Por otro lado, un caso extremo en el que los datos estén agrupados tan a la derecha que  $q_{0,75} = q_{0,50}$ , entonces  $\gamma_{YK} = -1$ . Del mismo modo, en un caso extremo de datos agrupados a la izquierda en el que  $q_{0,25} = q_{0,50}$ , tendremos que  $\gamma_{YK} = 1$ . Por ello, tenemos que  $-1 \leq \gamma_{YK} \leq 1$ .

### 3.1.5. Aplanamiento

El aplanamiento da cuenta del peso que tienen las colas superiores e inferiores de un conjunto de datos respecto a su centro. Un estadístico de aplanamiento que es robusto y resistente es la curtosis robusta (Sachs, 1984), que se define como:

$$K_r = \frac{RIC}{2(q_{0,9} - q_{0,1})} \quad (9)$$

Así, vemos que la curtosis robusta expresa la proporción que existe entre el 50% de la información en el centro del conjunto de datos (esto es, entre el cuartil inferior y el cuartil superior) y entre la diferencia de los deciles extremos  $d_{z_{0,9}}$  y  $d_{z_{0,1}}$ . El rango de  $K_r$  está entre 0 y 0,5. Cuanto más cercano se halle a cero, más peso tienen las colas, y viceversa. Una distribución normal tiene  $K_r = 0,263$ .

Ahora que hemos visto los estadísticos robustos y resistentes que podemos utilizar para el análisis de diversos datos, veamos cómo tratar los datos de los que disponemos. Para cada contaminante que se desea estudiar, se tienen dos tipos de archivos con la estimación de la media de la concentración del contaminante. Uno de ellos atiende a la media anual de la concentración, mientras que el otro tipo de archivo atiende a la media diaria de concentración del contaminante a estudiar. Utilizando ambos tipos de archivos se podrá realizar un análisis exhaustivo de la concentración de contaminantes a lo largo del período 2007-2011.

### 3.1.6. Archivos con la media anual de concentración

Utilizando este tipo de archivos, es posible estudiar la distribución anual de contaminación a lo largo del período del 2007 al 2011 en el que ha sido realizada la estimación. La estructura de datos es la siguiente:

Tabla 2: Ejemplo de la estructura de datos de los archivos con la media anual de concentración en un año dado.

Índice	Concentración
1	$P_1$
2	$P_2$
3	$P_3$
.	.
.	.
.	.
n	$P_n$

Se usan los índices para representar cada kilómetro cuadrado de la superficie de Inglaterra y

Gales. Para cada índice hay asociadas unas coordenadas determinadas (dadas en longitud y latitud) y una concentración  $P_i$  dadas en  $\mu g/m^3$ . Con esta estructura de datos, el análisis de distribución de contaminantes alrededor del territorio puede realizarse directamente. Más adelante se dirá la manera de proceder.

### 3.1.7. Archivos con la media diaria de concentración

Con este tipo de archivos, es posible calcular la media mensual para cada localización y, por ende, poder realizar un estudio de la distribución de la contaminación en Inglaterra y Gales a lo largo de un año dado. La estructura de los datos es la siguiente:

Tabla 3: Ejemplo de la estructura de datos de los archivos con la media diaria de concentración en un año dado.

Índice	Mes	Día	Concentración
1	1	1	$P_1$
1	1	2	$P_2$
.	.	.	.
.	.	.	.
.	.	.	.
1	1	31	$P_{31}$
1	2	1	$P_{32}$
.	.	.	.
.	.	.	.
.	.	.	.
1	12	31	$P_{365}$
2	1	1	$P_{366}$
.	.	.	.
.	.	.	.
.	.	.	.
n	12	31	$P_{365-n}$

Como puede verse en la tabla 3, para cada índice se encuentra, de manera ordenada, la concentración de contaminantes en  $\mu g/m^3$  para cada día del año. Sin embargo, para realizar el análisis que hemos descrito, necesitamos una estimación mensual de concentración para cada contaminante. Así, si se implementa un algoritmo que calcula la media mensual de cada localización, se obtiene la tabla 4. Ahora, se tiene la media mensual de concentración (representado por  $\mu_i$ ) en  $\mu g/m^3$  para cada índice. Sin embargo, el orden de la tabla no permite ver la distribución a lo largo del territorio para un mes dado. Por ello, hay que aplicar un segundo algoritmo que ordene los datos por el mes, como puede verse en la tabla 5.

Tabla 4: Ejemplo de la estructura de datos de los archivos con la media mensual de concentración en un año dado.

Índice	Mes	Concentración
1	1	$\mu_1$
1	2	$\mu_2$
.	.	.
.	.	.
.	.	.
1	12	$\mu_{12}$
2	1	$\mu_{13}$
.	.	.
.	.	.
.	.	.
n	12	$\mu_{12-n}$

Tabla 5: Las concentraciones están ordenadas mes a mes.

Índice	Mes	Concentración
1	1	$\mu_1$
2	1	$\mu_{13}$
.	.	.
.	.	.
.	.	.
n	1	$\mu_i$
1	2	$\mu_2$
.	.	.
.	.	.
.	.	.
n	12	$\mu_{12-n}$

Con esta nueva estructura en los datos, ya es posible realizar un análisis mes a mes de la distribución de la concentración de contaminantes en el territorio de Inglaterra y Gales en un año dado.

Tanto para los archivos con la media anual de concentración de contaminantes como a los archivos con la media mensual hemos hallado los estadísticos de posición, dispersión, simetría y aplanamiento que cumplen los requisitos de ser robustos y resistentes, esto es, hemos hallado la mediana, el RIC, el índice de Yule-Kendall y la curtosis robusta (definidas en las ecuaciones 6, 7, 8 y 9, respectivamente).

### 3.2. Análisis gráfico

Aunque los estadísticos que hemos descrito en los apartados anteriores den información numérica de las características del conjunto de datos que queremos analizar de manera rápida y sencilla de calcular, no dan demasiados detalles ni permiten sacar conclusiones a simple vista, ni es sencillo comparar distintos datos entre sí solo utilizando estadísticos. Por ello, diversas técnicas de representación de los datos existen para poder, de un vistazo, ver la distribución que tienen los conjuntos de datos y sus características.



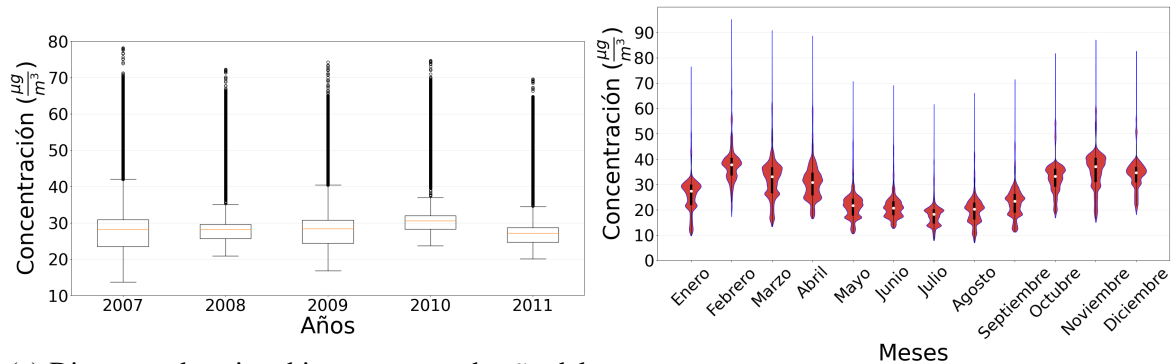
### 3.2.1. Diagramas de caja y bigotes

Una de estas herramientas gráficas es el diagrama de caja y bigotes, compuesta por cinco cuantiles que dan información de la distribución del conjunto de datos: el mínimo, el cuartil inferior, la mediana, el cuartil superior y el máximo. En una caja se representan el cuartil inferior, la mediana y el cuartil superior, y lo que se encuentra fuera de ella se representa mediante bigotes. Ahora bien, para tener más detalles en las colas, se suele establecer un límite de representación de los bigotes: todo lo que se halle fuera de ese límite se consideran datos anómalos y se representan individualmente con puntos.

$$\begin{cases} \text{Límite superior} = q_{0,75} + 1,5RIC \\ \text{Límite inferior} = q_{0,25} - 1,5RIC \end{cases} \quad (10)$$

Además, se puede representar los diagramas de violín, una combinación sinérgica de los diagramas de caja y bigotes y densidad de probabilidad. Las principales modificaciones respecto a los diagramas de caja y bigotes son que la mediana se representa como un punto y que los datos anómalos no se representan individualmente. Además, la anchura de la caja se vuelve proporcional a la densidad de probabilidad, de manera que su forma muestra la distribución de los datos. De ese modo, en una sola gráfica podemos ver información sobre la posición, dispersión, simetría, aplanamiento y distribución de los datos, con lo cual resulta más sencillo comparar entre varios grupos (Hintze y Nelson, 1998).

Utilizando los datos de media anual de concentración, se han representado diagramas de caja y bigotes con los datos de cada año desde el 2007 hasta el 2011 (de cara a estudiar la evolución de la distribución espacial en el tiempo). Con los datos de media mensual de concentración, se han representado en diagramas de violín los datos de cada mes de cada año, con el objetivo de estudiar la evolución de la distribución espacial de la concentración a lo largo de un año. Las figuras 1a y 1b son los diagramas de caja y bigotes y violín de la distribución anual y mensual del  $NO_2$  a modo de ejemplo, respectivamente.



(a) Diagrama de caja y bigotes para cada año del  $NO_2$ .

(b) Diagrama de violín de los meses del 2007 para el  $NO_2$ .

Figura 1: Ejemplos de diagramas de caja y bigotes y de violín.

### 3.2.2. Histogramas

Los histogramas representan la cantidad que tiene cierto rango de valores de un conjunto de datos. Para ello, se divide el conjunto de datos en regiones de igual ancho (normalmente) y se cuenta el número de elementos que contiene cada una. Así, el histograma consiste en celdas cuya altura da cuenta del número de elementos que contiene, por lo que aporta información acerca de la distribución de los datos como la posición, dispersión, simetría y aplanamiento de un solo vistazo. Del mismo modo, es posible ver si la distribución de los datos es unimodal o multimodal rápidamente.

En particular, una cuestión que ayuda mucho de los histogramas es que el área de cada celdilla es proporcional a la probabilidad de obtener un elemento en ese rango de valores (en caso de que el histograma esté normalizado, entonces el área de cada celdilla es la probabilidad de obtener un elemento en ese rango de valores).

Ahora bien, lo complicado a la hora de representar un histograma es escoger la anchura de la celdilla, pues dependiendo de ella la información que se extraiga será útil o no: en caso de ser celdillas demasiado anchas, entonces los detalles importantes de la distribución de los datos no podrán apreciarse; por otro lado, si las celdillas son muy delgadas, entonces el histograma resultante será muy irregular y, por ende, difícil de interpretar. Por ello, una aproximación para establecer el ancho de la celdilla es mediante:

$$anchura = c \frac{RIC}{n^{1/3}} \quad (11)$$

Donde  $n$  es el número total de elementos que tiene nuestro conjunto de datos y  $c$  es una constante que puede tomar valor entre 2,0-2,6. Los valores cercanos a 2,6 suelen ser óptimos para distribuciones gaussianas, mientras que los más cercanos a 2,0 son más apropiados para distribuciones asimétricas o multimodales. Hay que tener en cuenta que la anchura que da esta ecuación es una guía, una aproximación al valor óptimo del ancho de la celdilla. El valor óptimo depende de la naturaleza y particularidades de nuestro conjunto de datos.

### 3.2.3. Suavizados de histogramas

Los histogramas pueden dar información muy útil acerca de la distribución de datos que tenemos, aunque tienen algunos inconvenientes. Por ejemplo, dónde colocamos el centro de cada celdilla es un proceso arbitrario (no hay un punto 'correcto' donde ponerlos) y, del mismo modo, hay que redondear los valores a los extremos de cada celdilla para establecer en qué rango entran. Por ello, una alternativa para el histograma es su suavizado, en el cual no está el inconveniente de tener que redondear los valores de cada elemento y tiene una estructura suave. En los histogramas, cada celda es un 'bloque unidad' que se apila junto con otros para poder formar la gráfica. En cambio, en los suavizados se emplean formas características llamadas 'núcleos' que son, en general, más suaves que los rectángulos. En la tabla 6 podemos ver unos ejemplos (Fernández-Duque et al., 2019).

Tabla 6: Núcleos de suavizado utilizados normalmente.

Nombre	Núcleo $h(t)$	Rango de $t$
Triangular	$1 -  t $	$-1 < t < 1$
Rectangular	$\frac{1}{2}$	$-1 < t < 1$
Cuadrático (Epanechnikov)	$\frac{3}{4}(1 - t^2)$	$-1 < t < 1$
Cuártico (Biweight)	$\frac{15}{16}(1 - t^2)^2$	$-1 < t < 1$
Tricúbico	$\frac{70}{81}(1 -  t ^3)^3$	$-1 < t < 1$
Gaussiano	$(2\pi)^{-1/2} \exp(-t^2/2)$	$-\infty < t < \infty$

Así, los análogos a las celdas serían las funciones de suavizado, definidas en la ecuación 12 como:

$$\hat{f}(x_0) = \frac{1}{nw} \sum_{i=1}^n h\left(\frac{x_0 - x_i}{w}\right) \quad (12)$$

Donde  $n$  es el número de elementos de la distribución de datos y  $w$  es un parámetro arbitrario.

De este modo, el resultado sería la suma de la altura de todas las funciones de suavizado (hay tantas funciones de suavizado como elementos tiene la distribución de datos).

A la hora de realizar el suavizado, no es tan importante el núcleo que utilicemos para ello, sino el parámetro  $w$ , pues en función de este último, el suavizado puede estar 'aplanado' y que no sea posible que apreciemos detalles importantes de la distribución de datos; o al contrario, que tenga muchos picos y detalles y, por ende, su interpretación resulte difícil. Una buena aproximación inicial para el parámetro  $w$  viene definida como (Donnelly et al., 2011):

$$w = \frac{0,9s}{n^{1/5}} \quad (13)$$

Donde  $s$  es la desviación estándar de los datos. Como ejemplo de histogramas y de sus suavizados, la figura 2 es el histograma y su suavizado para la concentración anual de ozono en el año 2007. Claramente, la distribución de datos es multimodal.

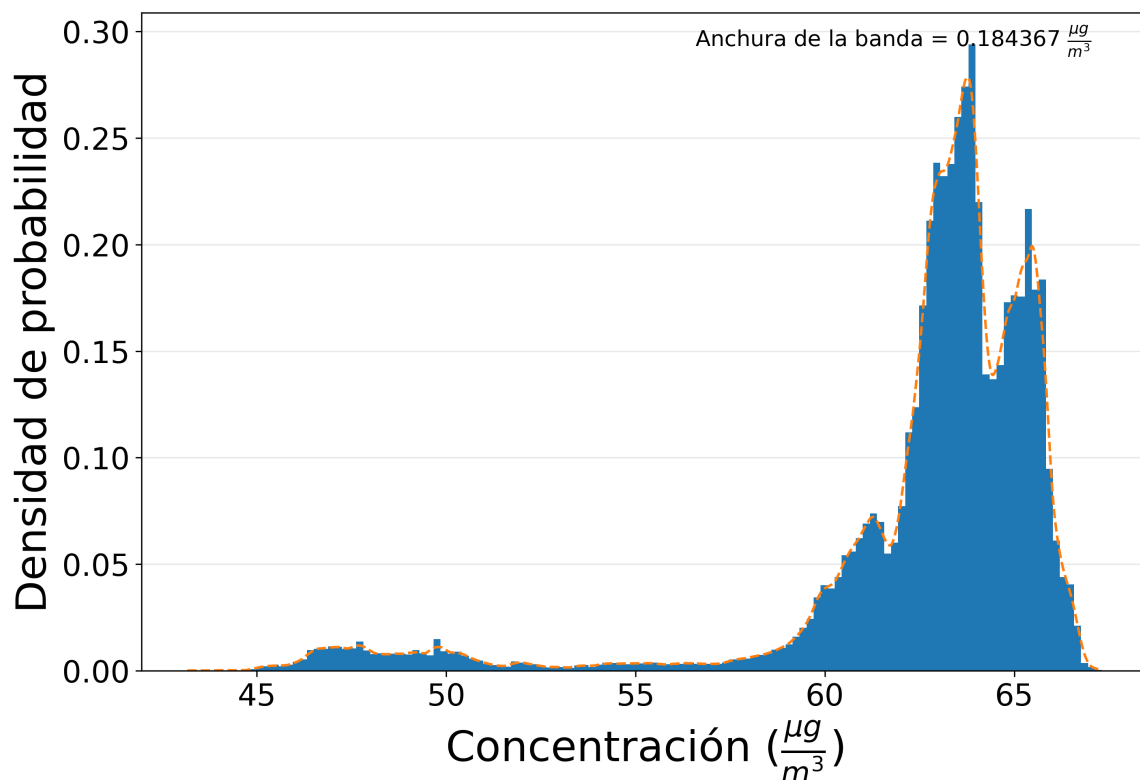


Figura 2: Histograma y su respectivo suavizado de la concentración anual de  $O_3$  en el año 2007.

### 3.2.4. Coeficientes de eficiencia

Para poder establecer el acuerdo que hay entre los suavizados calculados de los histogramas y los histogramas en sí, podemos utilizar índices de correlación que nos indiquen el grado de acierto. Podemos utilizar cinco coeficientes para estudiar la correlación entre los histogramas y sus suavizados: coeficiente de determinación, índice de Nash-Sutcliffe, índice de concordancia y los índices de Nash-Sutcliffe y concordancia modificados (Krause et al., 2005).

El coeficiente de determinación  $r^2$  se define como el cuadrado del coeficiente de correlación de Pearson, esto es:

$$r^2 = \left( \frac{\sum_{i=1}^n (O_i - \hat{O})(P_i - \hat{P})}{\sqrt{\sum_{i=1}^n (O_i - \hat{O})^2} \sqrt{\sum_{i=1}^n (P_i - \hat{P})^2}} \right)^2 \quad (14)$$

Donde  $O$  son los valores observados,  $P$  son los valores calculados y  $\hat{O}$  y  $\hat{P}$  son la media de los valores observados y calculados, respectivamente.

El rango de  $r^2$  está entre cero y uno. Un valor de cero indica que no hay ninguna correlación entre los valores observados y calculados, mientras que un valor de uno indica que la dispersión de los valores calculados es igual a la de los valores observados. Ahora bien, ello no indica necesariamente que haya correlación entre los valores observados y predichos pues, al solo tener en cuenta la dispersión, si nuestro suavizado tiene unos valores muy distintos a su histograma correspondiente pero tiene su forma, un valor cercano a uno puede darse.

El coeficiente de Nash-Sutcliffe se define como uno menos el sumatorio de la diferencia absoluta al cuadrado de los valores observados y calculados dividido por la varianza de los valores observados, esto es:

$$E = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \hat{O})^2} \quad (15)$$

El rango de  $E$  es entre 1 (una correlación perfecta) y  $-\infty$ . La mayor desventaja de este coeficiente es que, dado que las diferencias entre los valores observados y calculados están elevadas al cuadrado, los valores grandes tienen una contribución muy grande y, en cambio, los valores pequeños tienen una contribución nimia.

El índice de concordancia representa la proporción entre el error medio al cuadrado (numerador) y el error potencial (denominador), esto es, se define como:

$$d = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (|P_i - \hat{O}| + |P_i - \hat{P}|)^2} \quad (16)$$

Dado que el error medio al cuadrado se encuentra también en el numerador, el índice de concordancia presenta el mismo inconveniente que  $E$ .

Para poder reducir el gran o poco peso que tienen los valores grandes y pequeños, respectivamente, en los coeficientes  $E$  y  $d$ , se establecen estas modificaciones:

$$E_j = 1 - \frac{\sum_{i=1}^n |O_i - P_i|^j}{\sum_{i=1}^n |O_i - \hat{O}|^j} \text{ con } j \in N \quad (17)$$

$$d = 1 - \frac{\sum_{i=1}^n |O_i - P_i|^j}{\sum_{i=1}^n (|P_i - \hat{O}| + |P_i - \hat{P}|)^j} \text{ con } j \in N \quad (18)$$

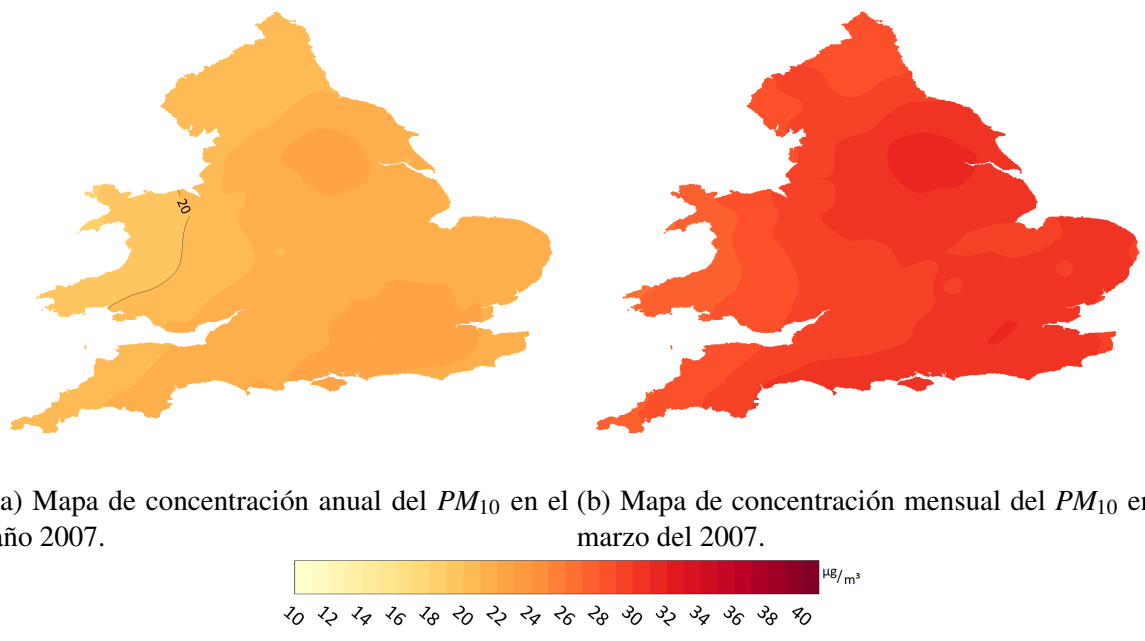
Con  $j = 1$ , el peso que tienen los valores grandes disminuye, mientras que el peso de los valores pequeños crece. Así, cuanto más grande sea  $j$ , más notorio se hace el aumento o la disminución del peso de los valores grandes o pequeños, respectivamente (que, en algunos casos, puede ser interesante hacerlo, como en el caso de eventos extremos).

### 3.2.5. Zona de estudio

Mediante la interpolación de datos y su suavizado, es posible representar los mapas de concentración de contaminantes, lo cual nos permite visualizar directamente las zonas en las que se da la mayor concentración, así como su distribución espacial a lo largo de Inglaterra y Gales. En la figura 3 podemos ver el contorno que hemos utilizado en este trabajo para representar los mapas de contaminación con las ciudades con mayor concentración de contaminantes. En las figuras 4a y 4b tenemos dos ejemplos de mapas de concentración del  $PM_{10}$ , el primero anual y el segundo mensual.



Figura 3: Contorno de Inglaterra y Gales con sus ciudades más contaminadas en el período de estudio.



(a) Mapa de concentración anual del  $PM_{10}$  en el año 2007. (b) Mapa de concentración mensual del  $PM_{10}$  en marzo del 2007.

Figura 4: Mapas de concentración anual y mensual del  $PM_{10}$ .

Por otro lado, dado que la resolución de la concentración de cada contaminante es de un kilómetro

cuadrado, es posible catalogar el territorio en función de la concentración presente. Ello permite obtener la superficie afectada por distintos intervalos de concentración y puede ayudar a estudiar el número de personas afectadas por ella.

Para representar la concentración de contaminantes en los mapas, se ha utilizado una escala de colores que va del amarillo (valor más bajo) hasta el rojo oscuro (valor más alto). Tanto para los mapas de media de concentración anual como para los mapas de media de concentración mensual, se ha establecido como valor máximo en la escala de colores de cada contaminante el valor límite establecido por la UE para cada uno (salvo para el ozono, cuyo valor límite es la media de concentración de ocho horas).

### **3.3. Software complementario**

De cara a realizar este análisis de datos, hay que tener en cuenta que la resolución de concentración de contaminantes es de un kilómetro cuadrado, esto es, cada archivo consta de 151 428 filas de datos. Teniendo en cuenta que hay que analizar cuatro contaminantes en un periodo de cinco años (tanto para la media anual como para la media mensual), ello implica que, en total, hay que clasificar, tratar y analizar más de mil millones de filas. Por supuesto, para implementar los algoritmos descritos anteriormente es necesario recorrer los datos de cada archivo uno a uno; y para el cálculo de estadísticos, representación en histogramas y diagramas de caja y bigotes, etc., es necesario ordenar los datos de cada uno de los archivos de menor a mayor.

Naturalmente, realizar estas tareas manualmente resultan muy tediosas y complicadas, por no decir imposibles. Utilizando programas como Excel y otros, es posible realizar todas estas operaciones, pero aún así resulta muy complicado y lento aplicar cada cuestión que hemos mencionado a cada archivo (que además, debido al gran tamaño que tienen, estos programas no son capaces de abrir). Es por ello que se ha escrito un código en *Python 3.7* que se encarga de abrir, tratar, calcular los estadísticos y realizar todas las gráficas comentadas anteriormente para cada uno de los archivos de los contaminantes que se desean estudiar. Ello permite que todas estas cuestiones se realicen de manera mucho más rápida que si se hiciera 'manualmente' utilizando un programa externo.

Por otro lado, se ha utilizado el programa *Surfer* para poder representar los mapas de concentración de contaminantes de Inglaterra y Gales. Este programa se encargaba de la interpolación y el suavizado de los datos de los que disponíamos.

## 4. Resultados

En este apartado serán presentados las gráficas, estadísticos e información obtenidas mediante el código escrito en Python 3.7. Del mismo modo, se intentará explicar la distribución espacial de la concentración de los distintos contaminantes y, de ser posible, ver cuál es la tendencia para cada uno de ellos.

### 4.1. Dióxido de nitrógeno

Como vimos en apartados anteriores, la mayor parte del  $NO_2$  no es emitido directamente a la atmósfera, sino que es producido mediante la reacción del  $NO$  con el  $O_3$ , como puede verse en la reacción 1. Ahora bien, el ozono puede disociarse con luz solar y reaccionar con el  $NO_2$  para volver a formar  $NO$ , como puede verse en la reacción 2.

Por ello, la explicación de lo que ocurre será propuesta al final del apartado del ozono, cuando se haya analizado la información de ambos contaminantes. Dado que Inglaterra y Gales son territorios muy húmedos, una parte del  $NO_2$  puede reaccionar con los aniones hidróxido según la reacción 3, por lo que habrá que tenerlo en cuenta a la hora de analizar la interrelación entre el ozono y el dióxido de nitrógeno.

Empezando con los diagramas de violín con la media mensual de concentración del dióxido de nitrógeno (figura 5), vemos en todos los años que la máxima concentración se da en meses fríos de otoño e invierno (sobre todo febrero y noviembre, aunque hay años en los que también hay mayor concentración en enero, diciembre e incluso en octubre), mientras que la concentración mínima se da en meses cálidos de verano (sobre todo en junio, julio y agosto). Dependiendo del año, estas diferencias son más o menos pronunciadas, pero en todos los años del período analizado, se cumple esta regla. Por otro lado, vemos que los datos anómalos de todos los años son concentraciones altas, con una densidad de probabilidad muy pequeña. Esto quiere decir que las grandes concentraciones se producen en pequeñas superficies que, como veremos en los mapas de concentración, corresponden a ciertas ciudades. Finalmente, si analizamos las cajas, vemos que la dispersión del conjunto central de los datos es pequeña en comparación con el rango total de los datos.

Si se analizan los mapas de concentración de dióxido de nitrógeno (figuras 6, 7, 8, 9 y 10) es posible determinar la distribución de la concentración de  $NO_2$  en Inglaterra y Gales.

Las zonas más contaminadas a lo largo del año son las ciudades de Londres, Oxford, Bristol y Cardiff (localizadas al sur de Inglaterra y Gales), la ciudad de Birmingham (en el centro de Inglaterra), la ciudad de Mánchester (localizada en la costa oeste de Inglaterra) y las ciudades de Kingston Upon Hull y Sunderland (al noreste de Inglaterra). Alrededor de estas ciudades, la concentración de  $NO_2$  va disminuyendo progresivamente. Sin embargo, en los meses de mayor concentración (febrero, noviembre y diciembre) puede que se superen los  $40 \mu g/m^3$  (el valor límite de media anual establecida por la OMS y la UE, tabla 1) en zonas muy extensas del territorio estudiado, ya sea el sur de Inglaterra o, directamente, la mayor parte de este mismo país.

Por otro lado, las zonas menos contaminadas con diferencia son el oeste de Gales y el norte de Inglaterra (con excepción de las ciudades de Sunderland y Kingston Upon Hull). En los meses con mayor concentración de  $NO_2$  también aumentó aquí la concentración, pero nunca de la manera en la



que se concentraba en otras zonas del territorio.

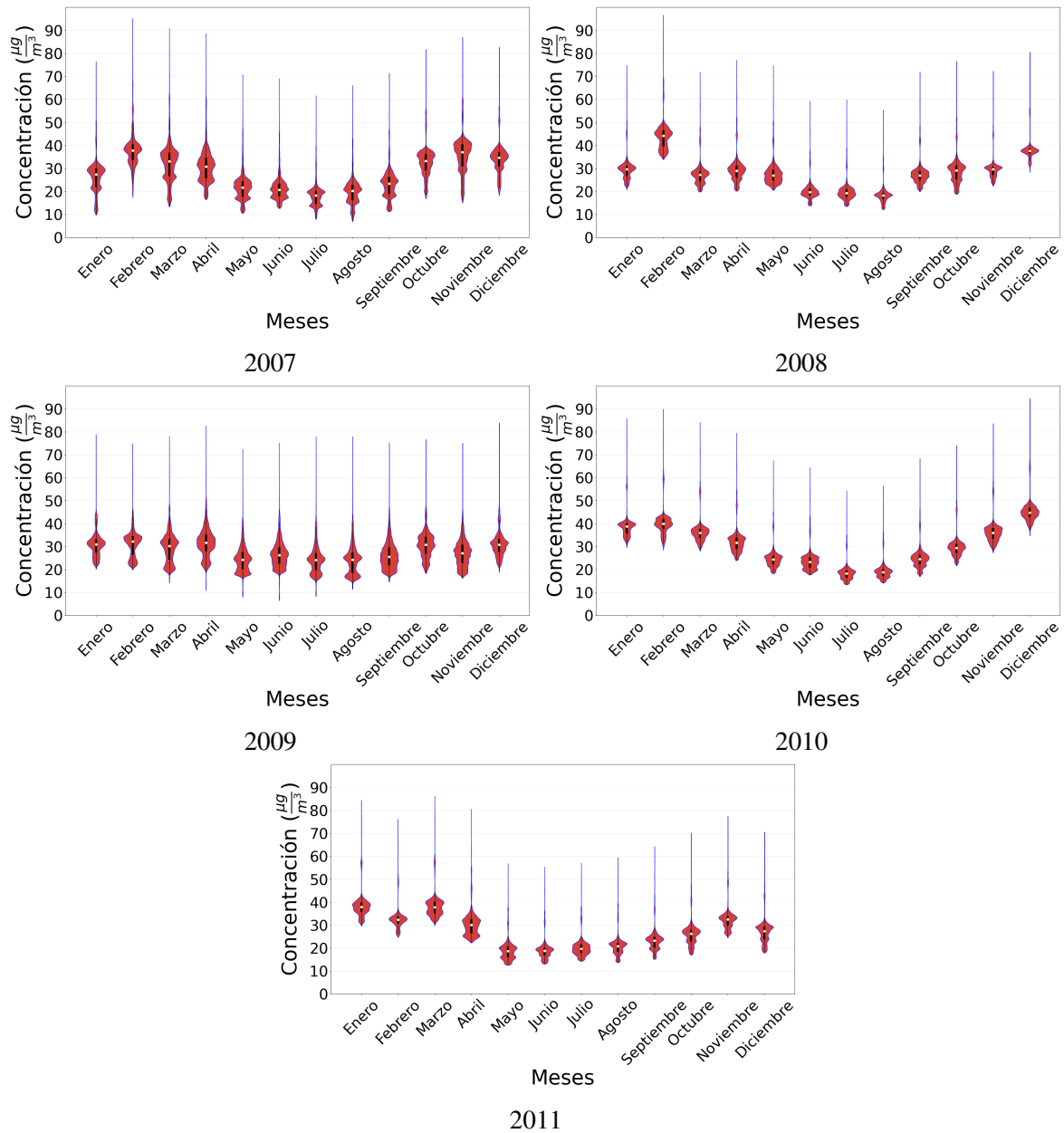


Figura 5: Diagramas de violín de la media mensual de concentración de  $NO_2$  en el período 2007-2011.

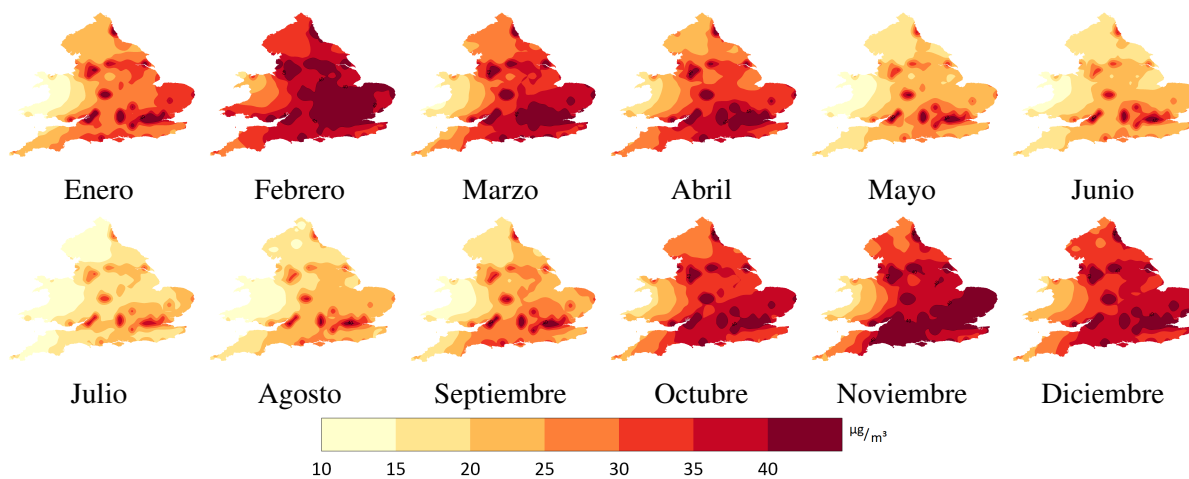


Figura 6: Media mensual de concentración del  $\text{NO}_2$  en el año 2007.

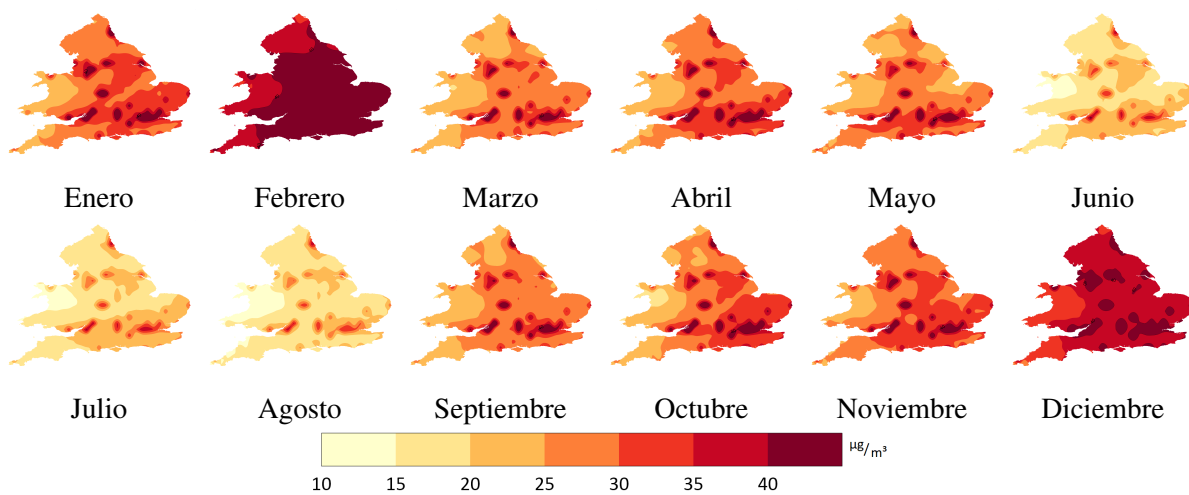


Figura 7: Media mensual de concentración del  $\text{NO}_2$  en el año 2008.

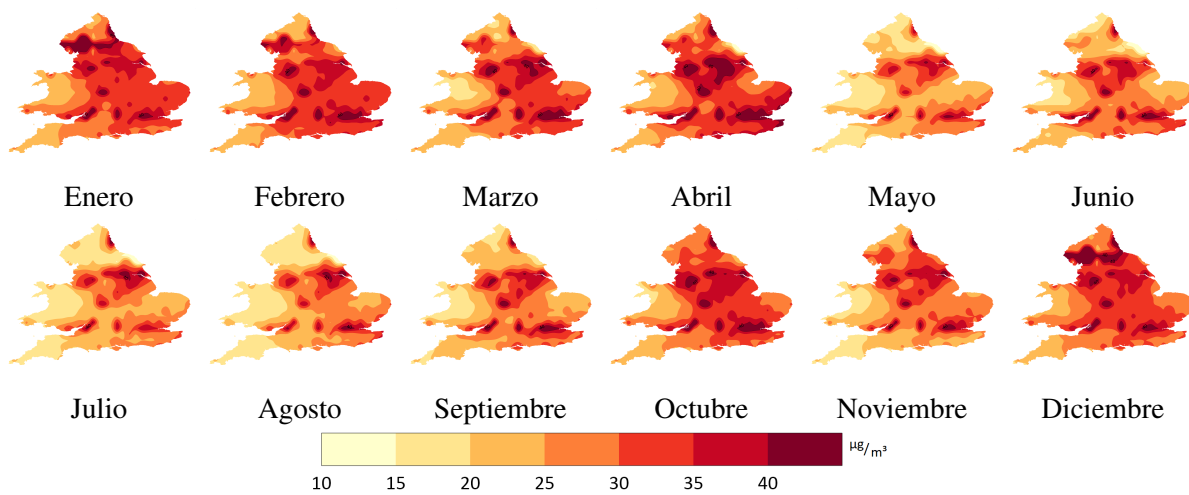


Figura 8: Media mensual de concentración del  $\text{NO}_2$  en el año 2009.

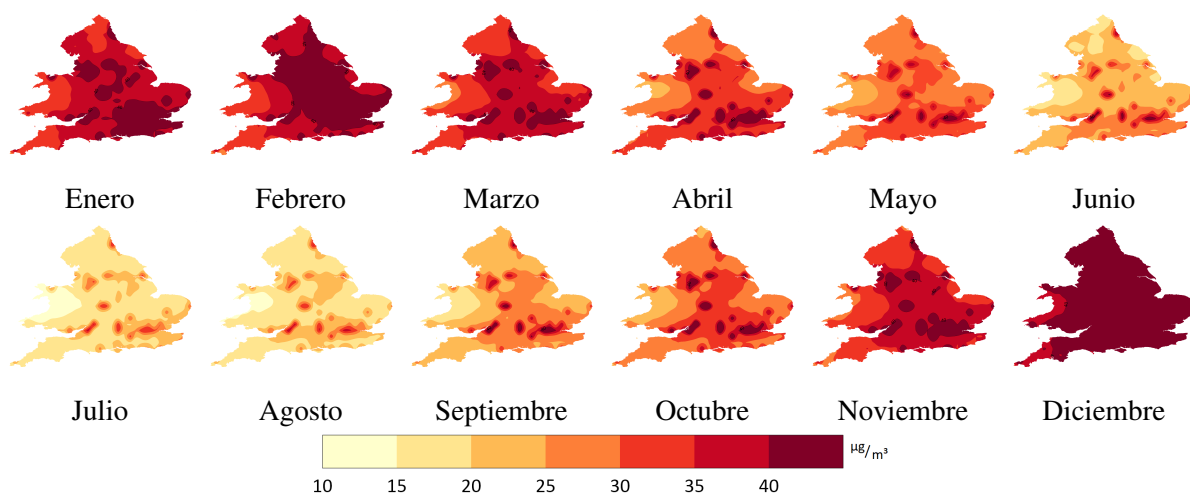


Figura 9: Media mensual de concentración del  $NO_2$  en el año 2010.

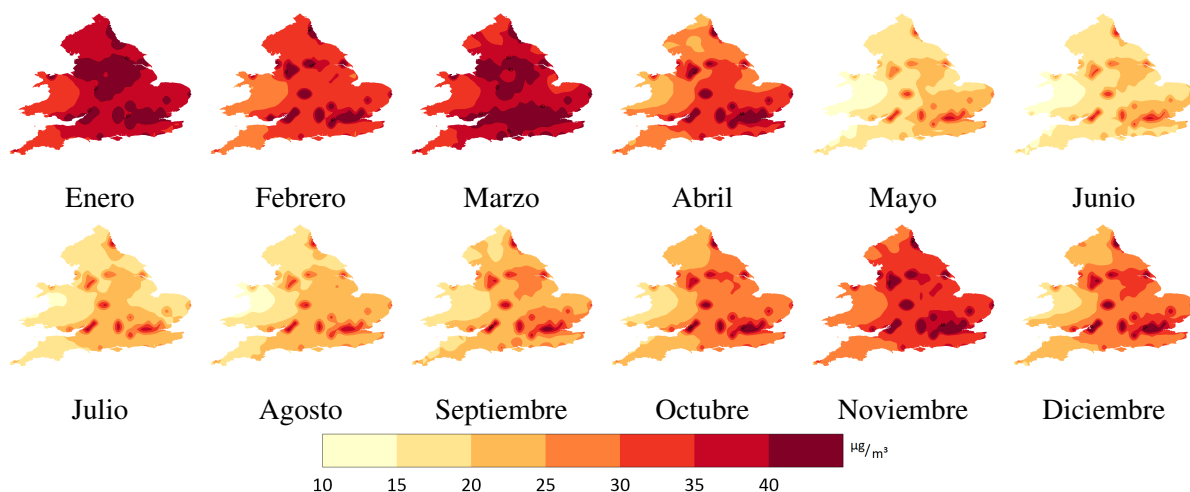


Figura 10: Media mensual de concentración del  $NO_2$  en el año 2011.

En cuanto a la concentración anual de dióxido de nitrógeno, vemos en la figura 12 que se supera el valor límite de  $40 \mu\text{g}/\text{m}^3$  en las ciudades y alrededores de Londres, Oxford, Bristol, Birmingham, Mánchester, Kingston Upon Hull y Sunderland. El territorio a través del cual se extiende esta concentración aumenta o disminuye en función del año, pero se encuentra en torno al 5%, como puede comprobarse en la tabla 7 (los intervalos de concentraciones corresponden a las concentraciones por debajo del valor límite establecidos por la OMS y la UE y los valores que superan ese valor límite). Datos del período 2010-2017 de dos estaciones de Londres muestran que, en una de ellas, la media anual superaba  $40 \mu\text{g}/\text{m}^3$ , mientras que en la otra no se superaba (con muy poco margen) este valor (Stirling et al., 2020), por lo que vemos que hay cierta concordancia con los resultados que hemos obtenido.

En la figura 11 puede observarse en los diagramas de caja y bigotes que, en el período estudiado, no hay casi variación de la concentración anual. Además, los datos anómalos son concentraciones muy altas (correspondientes a las ciudades más contaminadas) y la dispersión de los datos centrales es mucho menor que el rango total del conjunto de datos. Dada la insuficiente cantidad de datos anuales de los que disponemos, no es posible estimar una tendencia de la evolución de la concentración de  $NO_2$ .

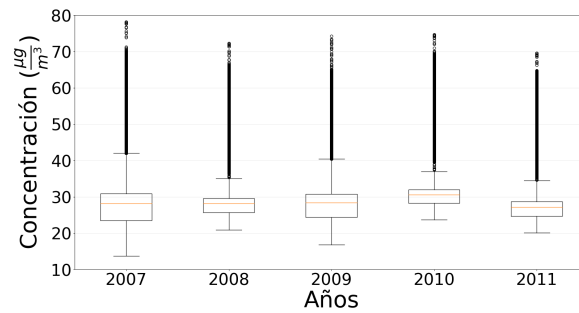


Figura 11: Diagramas de caja y bigotes de la media anual de concentración de  $NO_2$  en el período 2007-2011.

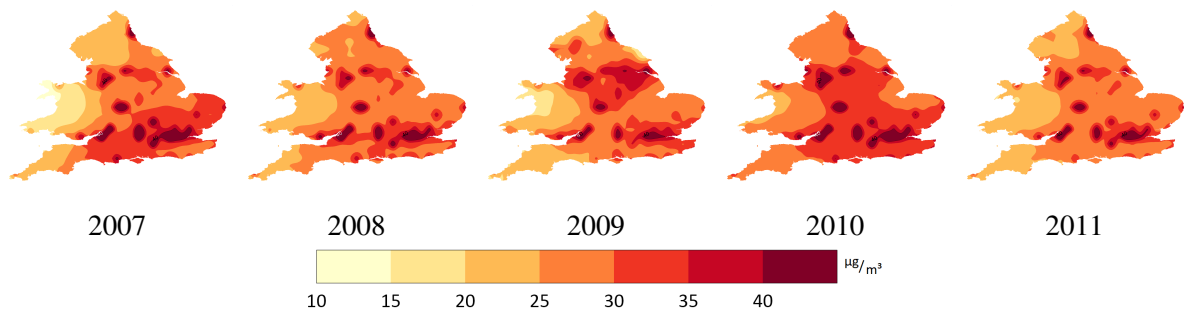


Figura 12: Media anual de concentración del  $NO_2$  en el período 2007-2011.

Tabla 7: Porcentaje de la superficie de Inglaterra y Gales contaminada de acuerdo al rango de concentración de  $NO_2$ .

Año	$(0, 40) / \frac{\mu g}{m^3}$	$(40, \infty) / \frac{\mu g}{m^3}$
2007	94,48	5,52
2008	94,03	5,97
2009	96,16	3,84
2010	93,87	6,13
2011	94,33	5,67

Tabla 8: Coeficientes de eficacia del suavizado del histograma para el  $NO_2$ .

Año	$r$	$d$	$E$	$d_1$	$E_1$
2007	1,0000	0,9892	0,9461	0,8961	0,7681
2008	0,9999	0,0000	-5,0220	0,0000	-1,4486
2009	1,0000	1,0000	0,9998	0,9950	0,9899
2010	1,0000	0,9989	0,9954	0,9682	0,9344
2011	1,0000	0,0000	-0,5338	0,0000	-0,2383

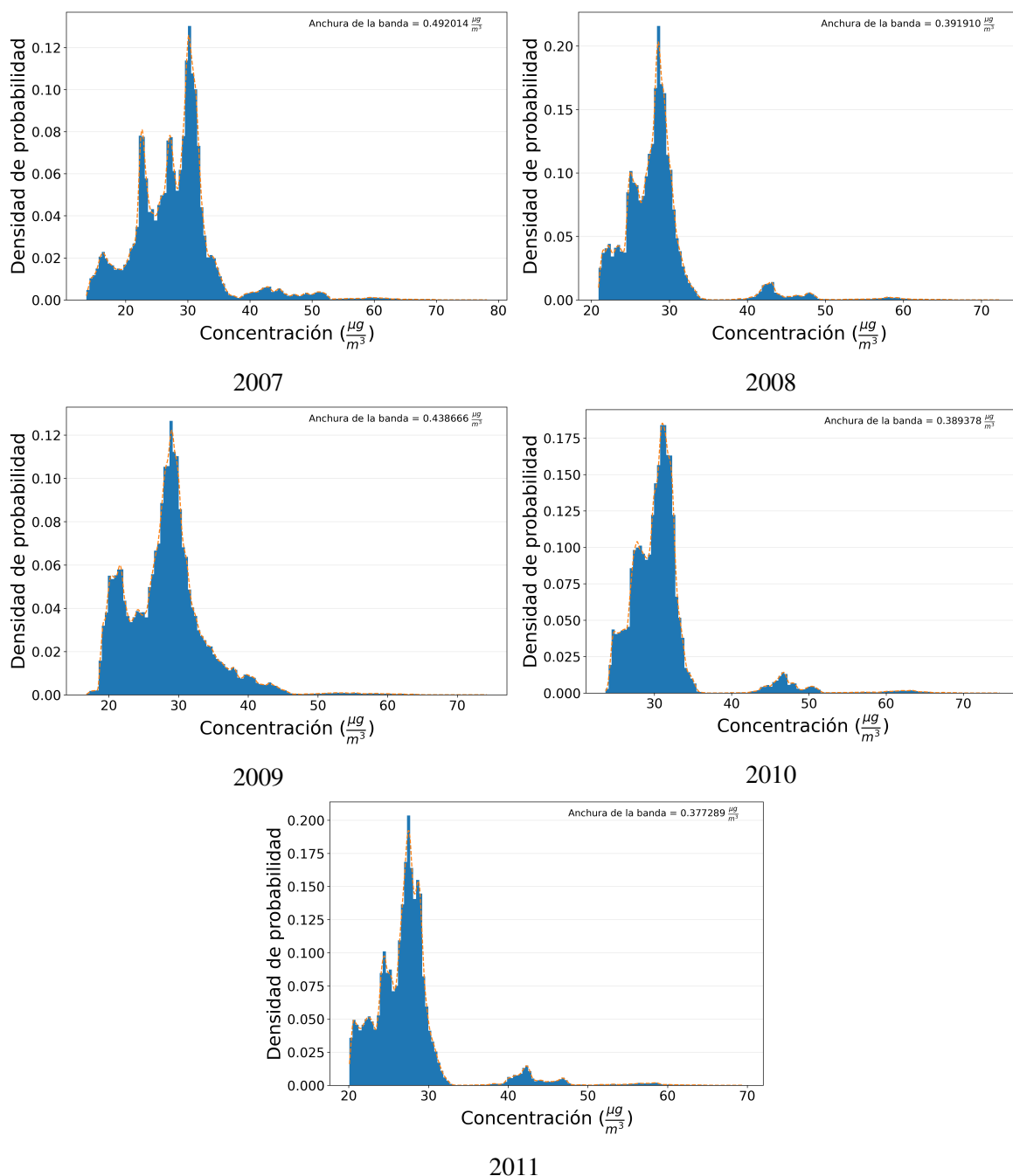


Figura 13: Histogramas de la media anual de concentración de  $NO_2$  en el período 2007-2011.

Los histogramas y sus suavizados representados en la figura 13 nos dan más información acerca de la distribución espacial anual del  $NO_2$  en Inglaterra y Gales. Vemos que en casi todos los años, una gran parte de los valores se hallan concentrados entre 20 y 30  $\mu g/m^3$  (correspondiente a los niveles de concentración del campo y de las ciudades pequeñas), mientras que una pequeña parte se halla a partir de 40  $\mu g/m^3$ , llegando algunos años hasta el límite de 60  $\mu g/m^3$ . Esta última parte corresponde a las grandes ciudades contaminadas que vimos antes.

Finalmente, vemos que los histogramas y sus suavizados concuerdan muy bien, pero si miramos la tabla 8, vemos que para los años 2008 y 2011 el índice de acuerdo, el índice de acuerdo modificado, el coeficiente de Nash-Sutcliffe y el coeficiente de Nash-Sutcliffe modificado dan valores de muy poco acuerdo (en algunos casos, nulo). Si vemos estos años en la figura 13, vemos que el comienzo del histograma es muy abrupto (la primera barra tiene un valor relativamente alto respecto a los otros

años). Dado que los coeficientes  $d$ ,  $E$ ,  $d_1$  y  $E_1$  tienen en el numerador el cuadrado del error medio y el error medio, respectivamente, la diferencia entre este comienzo abrupto del histograma y el comienzo paulatino del suavizado es alta, lo que hace que diverjan estos coeficientes. Respecto a los otros años, todos los coeficientes muestran un buen acuerdo, siendo  $d_1$  y  $E_1$  los más bajos.

## 4.2. Ozono

Dado que el  $NO_2$  y el  $O_3$  están muy relacionados, solo analizando conjuntamente los datos del dióxido de nitrógeno y del ozono se podrá dar una explicación de la distribución de la concentración de ambos contaminantes en Inglaterra y Gales. Veamos, primero, los diagramas de violín con la media mensual de concentración de ozono. Como puede apreciarse en la figura 14, los meses de mayor concentración de ozono se producen en los meses de abril, mayo y junio, mientras que los meses de menor concentración suelen ser los meses de otoño e invierno (sobre todo, enero y febrero). Los datos anómalos de las concentraciones de ozono, para casi todos los años, son concentraciones muy pequeñas y con una densidad de probabilidad muy baja. Esto quiere decir que las concentraciones bajas se dan en superficies muy pequeñas del territorio estudiado que, como veremos más adelante en los mapas de concentración, corresponde a las ciudades más contaminadas por  $NO_2$ . Finalmente, la dispersión central de los datos es muy pequeña en comparación con el rango total de los datos y, del mismo modo, el rango de los datos presenta variaciones dependiendo de año, aunque no muy pronunciadas.

Con ello, se puede comprobar que tiene relación con el dióxido de nitrógeno, pues en la figura 5 se comprobaba que la mayor concentración se produce en meses de otoño e invierno, mientras que en el caso del ozono, esa mayor concentración se produce en meses más cálidos y con mayor horas de luz solar. Del mismo modo, la menor concentración de dióxido de nitrógeno se produce en los meses de verano, mientras que la menor concentración de ozono se produce en los meses de invierno: están inversamente relacionados.

Si se analiza la distribución territorial de la concentración de ozono, se podrán sacar conclusiones de qué es lo que ocurre entre el  $O_3$  y el  $NO_2$ .

Como puede observarse en las figuras 15, 16, 17, 18 y 19, la concentración de ozono suele ser más alta en el campo de Inglaterra y Gales que en las ciudades, donde siempre, sea el mes que sea, la concentración de ozono es menor a la mediana del mes. Y además, muchas ciudades en las que se vio que se concentraba el  $NO_2$  son las mismas en las que la concentración de  $O_3$  es menor que el resto del territorio: Londres, Oxford, Birmingham, Mánchester, Kingston Upon Hull y Sunderland, entre otras.

Precisamente, la distribución de  $O_3$  se comporta de manera contraria a la distribución de  $NO_2$  en Inglaterra y Gales: mientras el dióxido de nitrógeno se concentra especialmente en ciudades grandes o con mucha industria (con una concentración más baja en el campo), el ozono se concentra especialmente en el campo, con las ciudades grandes como puntos en los que su concentración es mínima.

Dada la reacciones químicas que hemos descrito en 1 y 2, estas dos reacciones se producen hasta que se halla un equilibrio entre el dióxido de nitrógeno y el ozono. Así, es comprensible que, en los meses de invierno, la concentración de  $O_3$  sea mínima y la de  $NO_2$  máxima. Ello es debido a que hay una menor cantidad de horas de luz al día, lo que provoca que una menor cantidad de  $NO_2$  se descomponga y, en consecuencia, se forme ozono. Sin embargo, a partir de abril, la concentración de

ozono pasa a ser máxima, mientras que la del dióxido de nitrógeno pasa a ser mínima, precisamente porque hay una mayor cantidad de horas de luz al día, por lo que se descompone una mayor cantidad de dióxido de nitrógeno y, en consecuencia, se forma más ozono. Ahora bien, también hay que tener en cuenta que una de las reacciones que puede tener el  $NO_2$  es a ácido nítrico (reacción 3), por lo que no todo el  $NO_2$  contribuye a la formación de ozono, sino también a la del ácido nítrico.

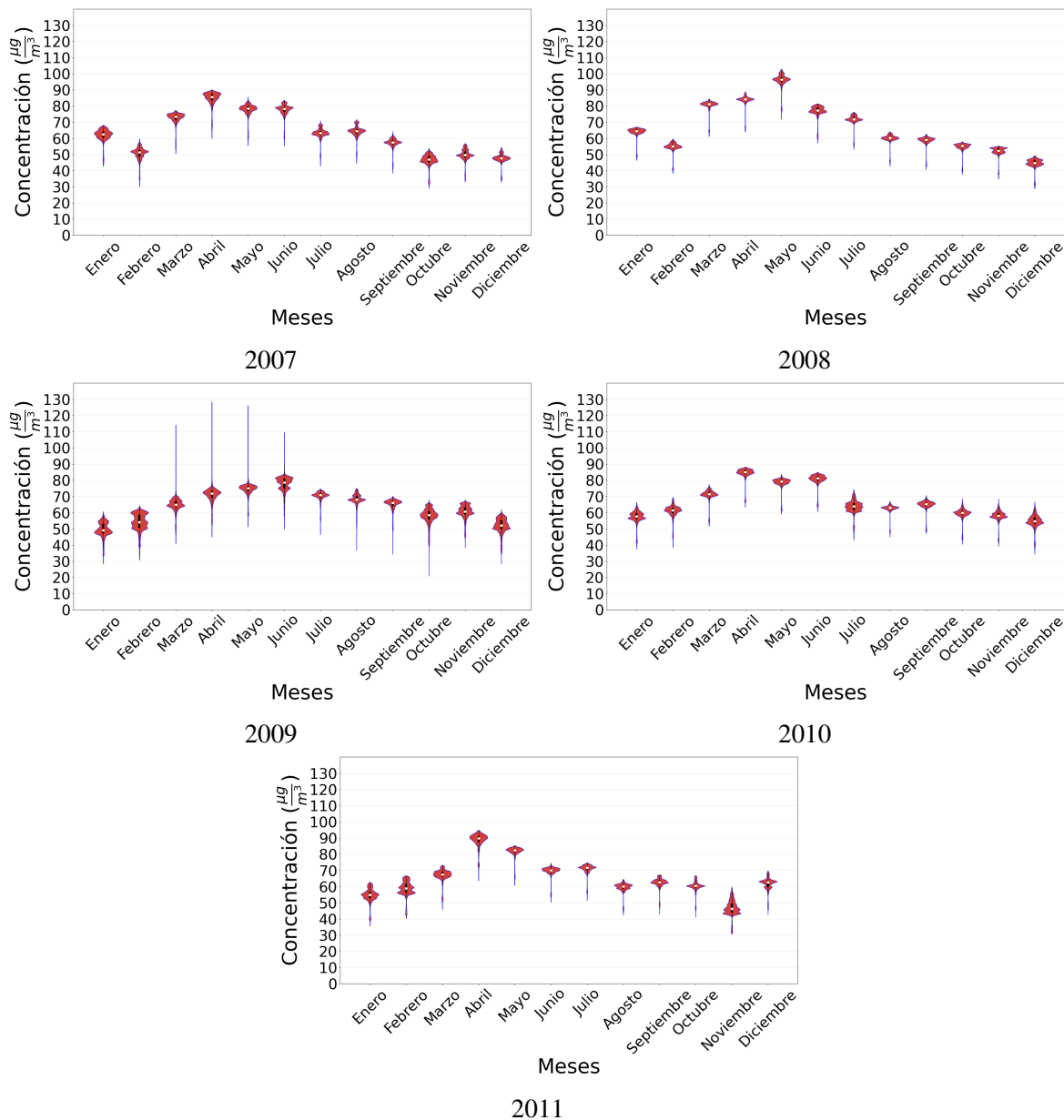


Figura 14: Diagramas de violín de la media mensual de concentración de  $O_3$  en el período 2007-2011.

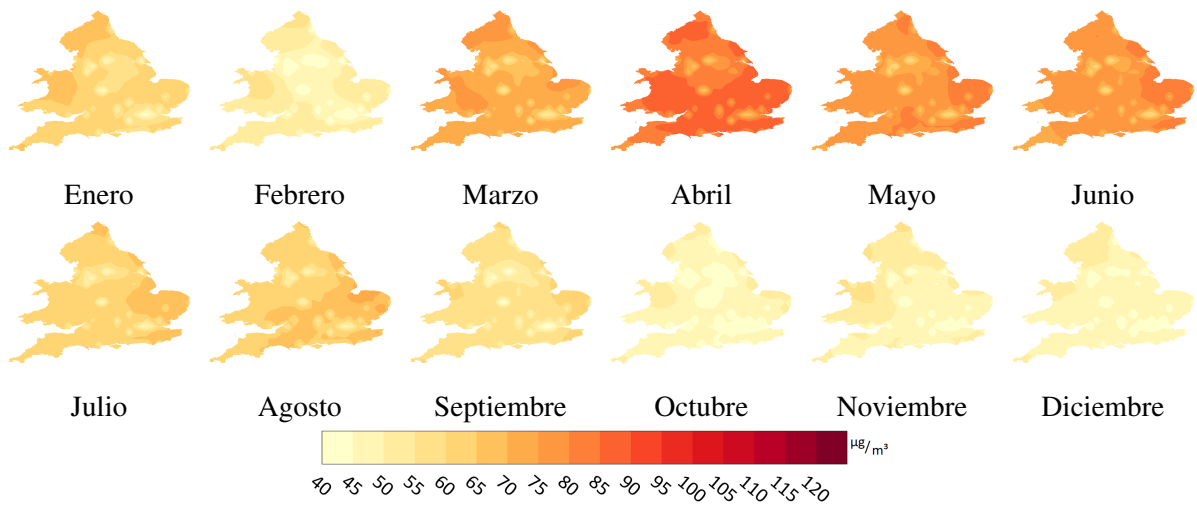


Figura 15: Media mensual de concentración del  $O_3$  en el año 2007.

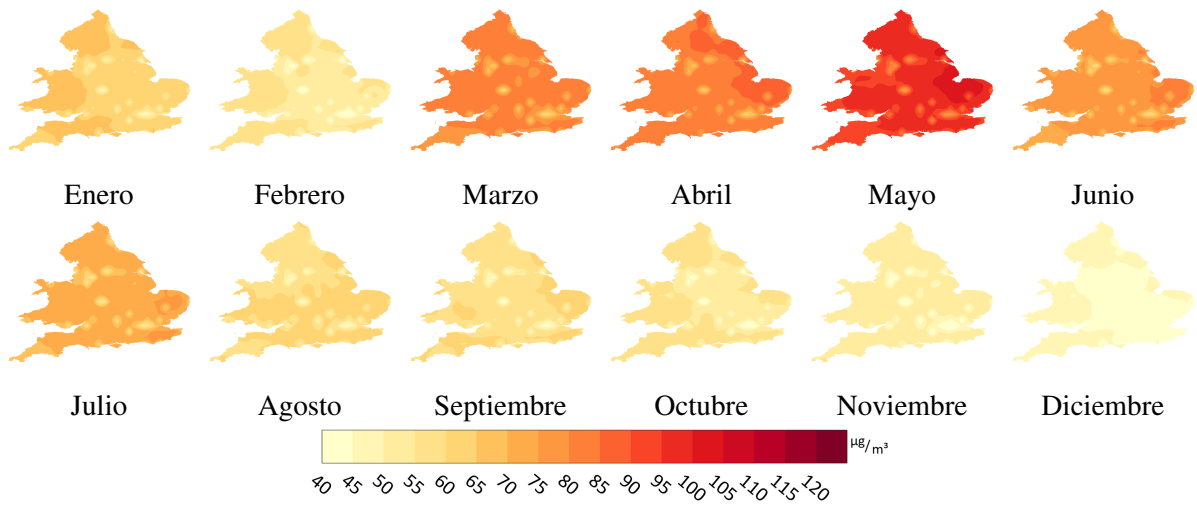


Figura 16: Media mensual de concentración del  $O_3$  en el año 2008.

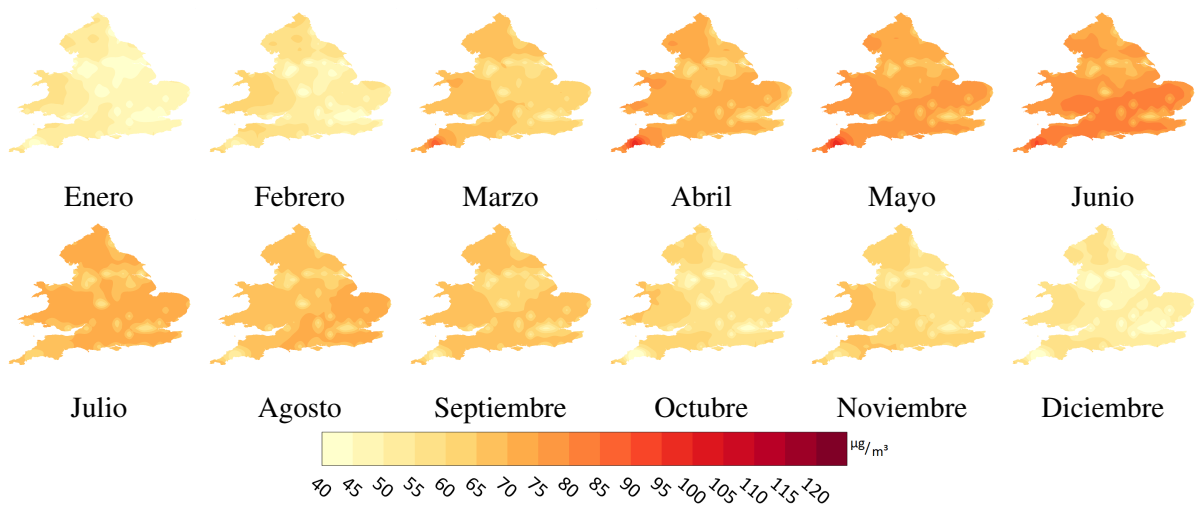


Figura 17: Media mensual de concentración del  $O_3$  en el año 2009.



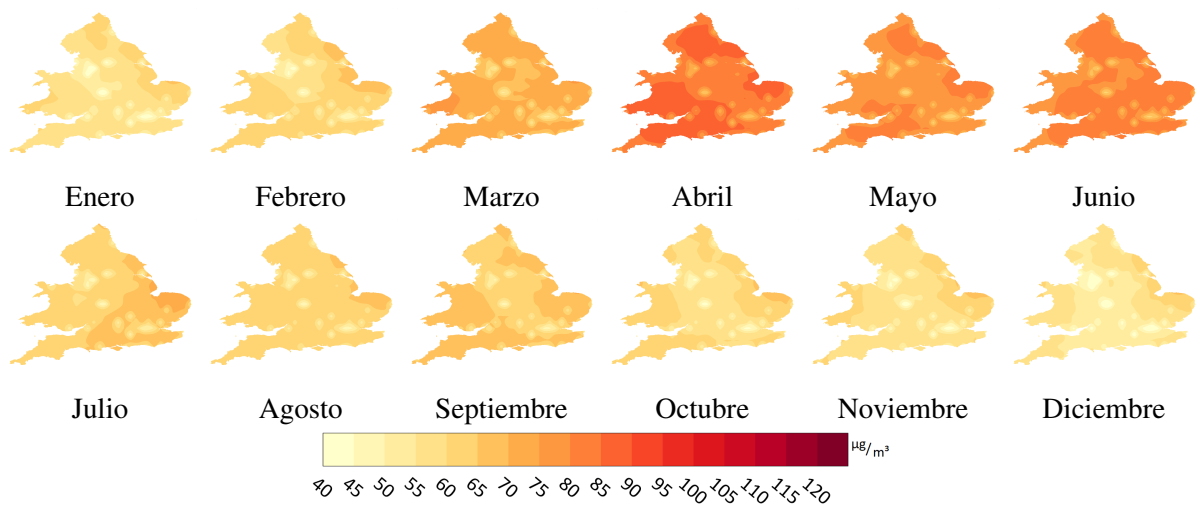


Figura 18: Media mensual de concentración del  $O_3$  en el año 2010.

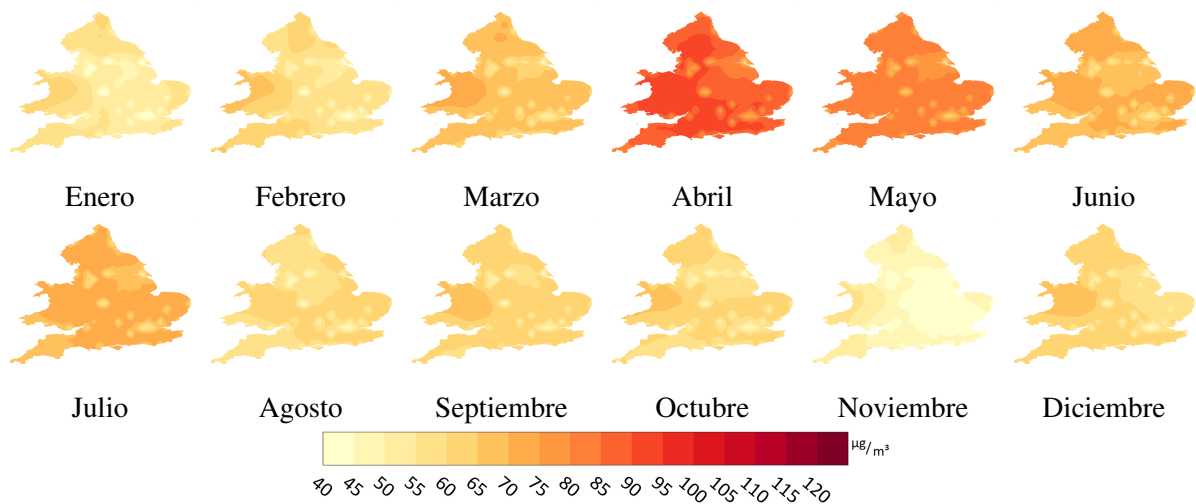


Figura 19: Media mensual de concentración del  $O_3$  en el año 2011.

Respecto a la concentración anual de ozono, en el período estudiado no es posible observar una tendencia de la concentración a lo largo de Inglaterra y Gales, como puede observarse en las figuras 20 y 21. Si analizamos la figura 20, vemos que la mediana no tiene tendencia a cambiar, así como el rango total de los datos. Remitiendonos a otros estudios, uno mostró que, en el período 2005-2015, la columna troposférica de  $NO_2$  en ciudades contaminadas de Inglaterra tenía una tendencia a disminuir con los años. Del mismo modo, la subcolumna de ozono, en el territorio de Inglaterra y Gales, tenía una muy ligera tendencia a aumentar con los años (Pope et al., 2018), aunque lamentablemente no podemos compararlo con nuestro estudio.

Por otro lado, los datos anómalos son concentraciones muy bajas que, como puede comprobarse en la figura 21, corresponden a las concentraciones de las ciudades. Finalmente, la dispersión del rango central de los datos es pequeña en comparación con el rango total de datos.

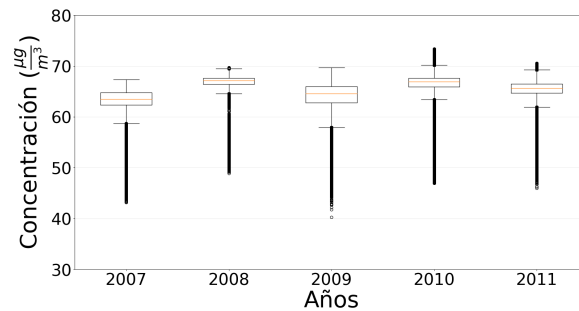


Figura 20: Diagramas de caja y bigotes de la media anual de concentración de  $O_3$  en el período 2007-2011.

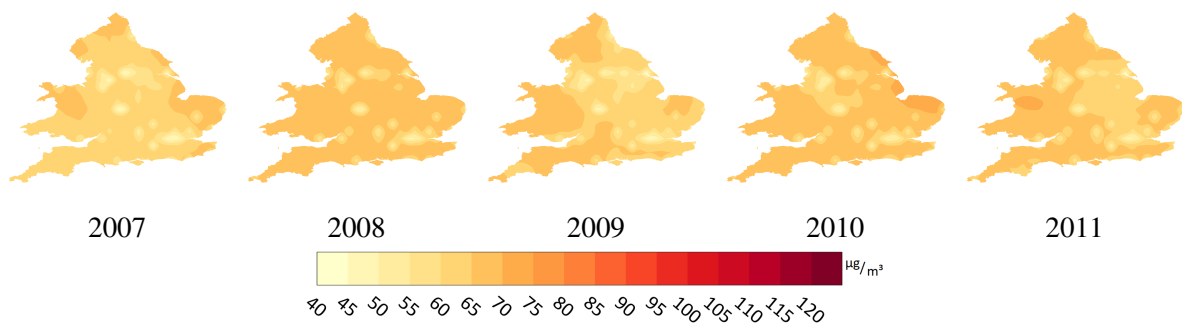


Figura 21: Media anual de concentración del  $O_3$  en el período 2007-2011.

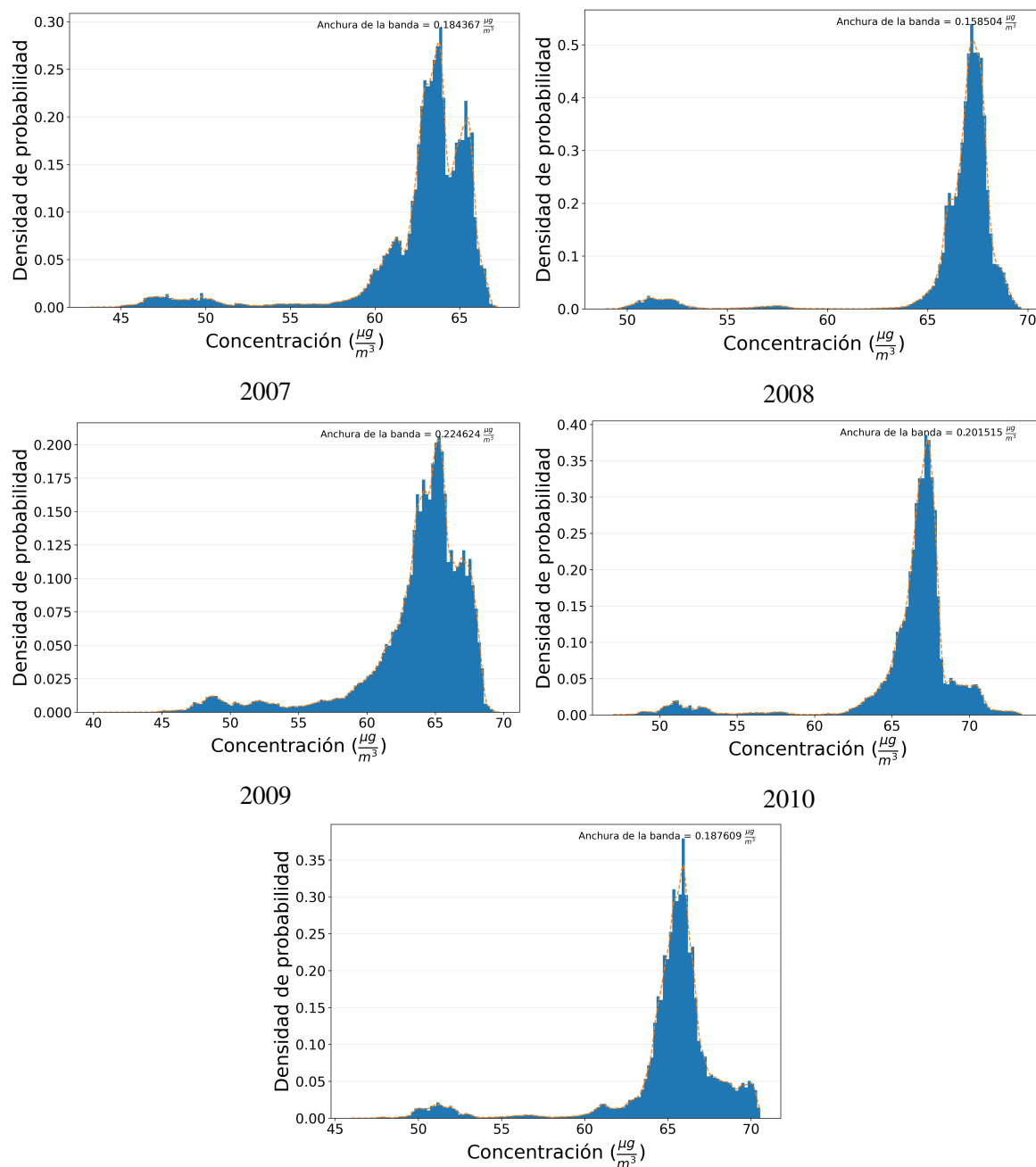
Como es de esperar, los histogramas y sus suavizados de la concentración anual del ozono (figura 22) presentan un comportamiento inverso al del dióxido de nitrógeno. Vemos que gran parte de las concentraciones en el territorio se agrupan entre los valores de 60 y 70  $\mu\text{g}/\text{m}^3$ , que corresponden a los valores de concentración en el campo y ciudades pequeñas de Inglaterra y Gales. Ahora bien, entre 45 y 55  $\mu\text{g}/\text{m}^3$  hay una agrupación de valores que corresponden a la concentración en las grandes ciudades, justo donde la concentración de  $\text{NO}_2$  es mayor.

Así, podemos deducir que, en las grandes ciudades, donde hay una gran emisión de  $\text{NO}$ , el ozono presente reacciona mediante la reacción 1. A medida que se va transportando el dióxido de nitrógeno al resto del territorio, va disminuyendo su concentración, ya sea por la reacción 2 o por su transformación en ácido nítrico. El ozono del campo reacciona con el monóxido de nitrógeno transportado desde las ciudades y es reestablecido por la reacción 2. Muy probablemente, la concentración de  $\text{NO}$  en el campo será menor que en las grandes ciudades, por lo que una menor cantidad de ozono reacciona con él para formar  $\text{NO}_2$ , lo cual hace que en el campo haya una mayor concentración de ozono.

Finalmente, podemos ver que los histogramas y sus suavizados coinciden muy bien en la figura 22. Además, en la tabla 9 podemos apreciar que todos los coeficientes de eficiencia dan un acuerdo casi absoluto entre los histogramas y suavizados. En general, los más bajos son el índice de acuerdo concordancia y el coeficiente de Nash-Sutcliffe modificado, pero por muy poco, salvo en el año 2008, en el que los índices más bajos fueron el índice de concordancia y el coeficiente de Nash-Sutcliffe.

Tabla 9: Coeficientes de eficacia del suavizado del histograma para el  $O_3$ .

Año	$r$	$d$	$E$	$d_1$	$E_1$
2007	1,0000	1,0000	1,0000	0,9991	0,9982
2008	1,0000	0,9978	0,9901	0,9997	0,9994
2009	1,0000	1,0000	1,0000	0,9997	0,9994
2010	1,0000	1,0000	0,9999	0,9955	0,9910
2011	1,0000	1,0000	1,0000	0,9998	0,9996



2011

Figura 22: Histogramas de la media anual de concentración de  $O_3$  en el período 2007-2011.

### 4.3. Partículas $PM_{10}$

Como puede verse en la figura 23, cada año es muy disparatejo respecto a la amplitud de la concentración. Por ejemplo, en el año 2009, las concentraciones de cada mes están muy agrupadas en ciertos valores, mientras que las concentraciones de los meses del año 2011 son muy dispersas. Ahora bien, a pesar de las diferencias en concentración año a año, la distribución a lo largo de un año dado parece seguir un patrón. Por ejemplo, en los meses de verano, la concentración de  $PM_{10}$  es mínima, mientras que en los meses de invierno y primavera es máxima. Además, no parece haber cierta tendencia a que los valores anómalos sean concentraciones bajas o altas. Finalmente, dependiendo del año, la dispersión de los valores centrales es más o menos pequeña en comparación con el rango total de los datos (compárense, por ejemplo, los años 2009 y 2011).

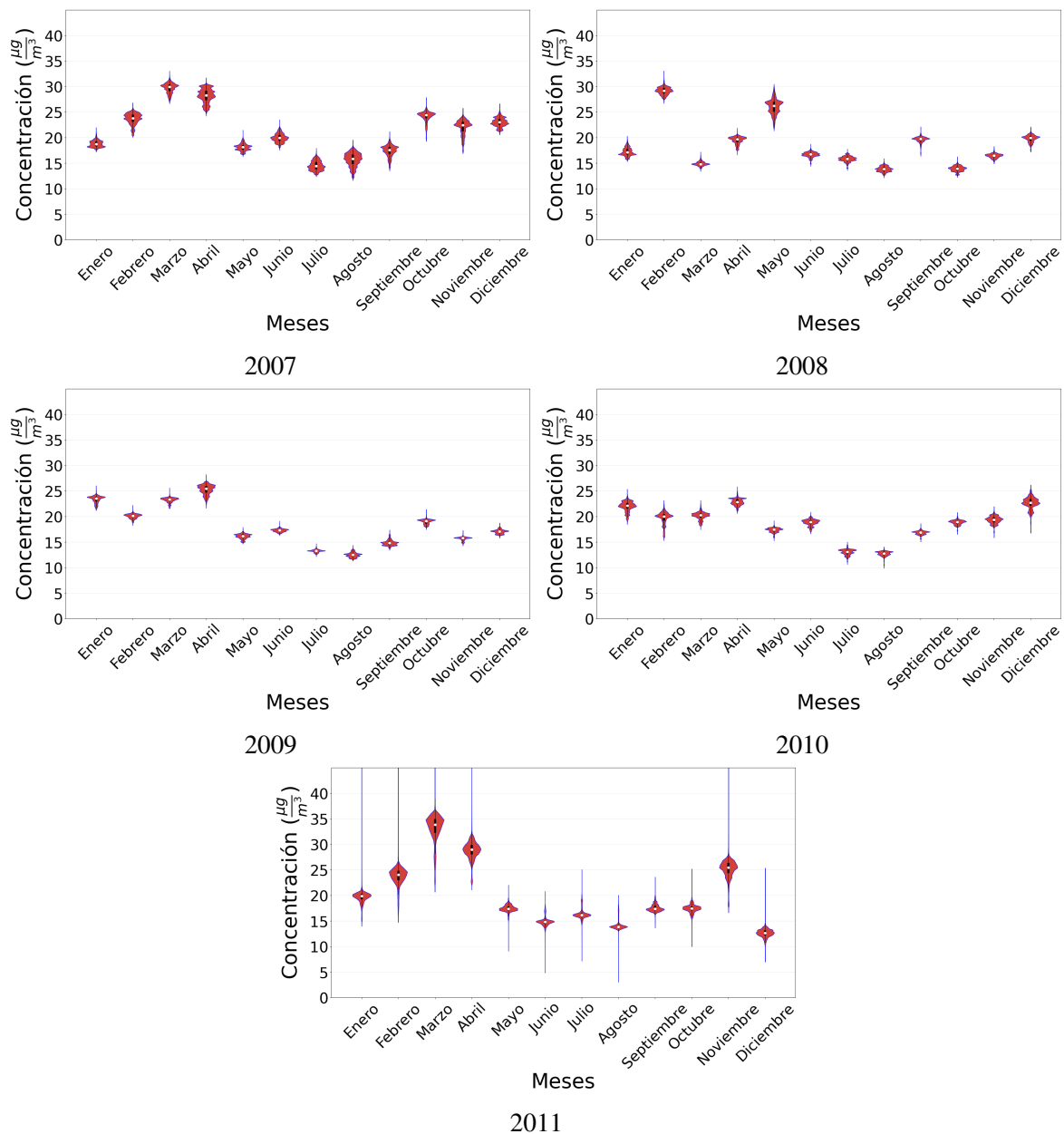


Figura 23: Diagramas de violín de la media mensual de concentración de  $PM_{10}$  en el período 2007-2011.

Analizando las figuras 24, 25, 26, 27 y 28, puede observarse que la distribución de  $PM_{10}$  suele ser

bastante homogénea. Sin embargo, hay una tendencia en que la concentración de este contaminante sea mayor al este de Inglaterra que en Gales. Esto es debido, probablemente, a la lluvia que cae en cada territorio: el este de Inglaterra es mucho más seco que Gales (Met Office, 2020b), lo cual favorece que se acumule más  $PM_{10}$ .

Si se analiza la figura 28, podemos observar que las concentraciones en mayo y en abril del 2011 son más altas en gran parte del territorio si las comparamos con las concentraciones de esos meses en otros años. Es probable que ello sea debido a que los meses de marzo, abril y mayo de 2011 fueron extremadamente secos con respecto a la media, sobre todo en Inglaterra y Gales. Las precipitaciones en el centro y este de Inglaterra fueron, por ejemplo, menores al 20% de la media (Met Office, 2016). Ello hace que el  $PM_{10}$  no sea depositado tan rápidamente en el suelo, por lo se favorece así su concentración. En febrero de 2008 podemos observar también una concentración mayor respecto a otros años (figura 25). Es posible que sea debido a razones parecidas a las concentraciones del 2011, pero no hemos encontrado datos que corroboren esto, por lo que no puede afirmarse de manera fehaciente.

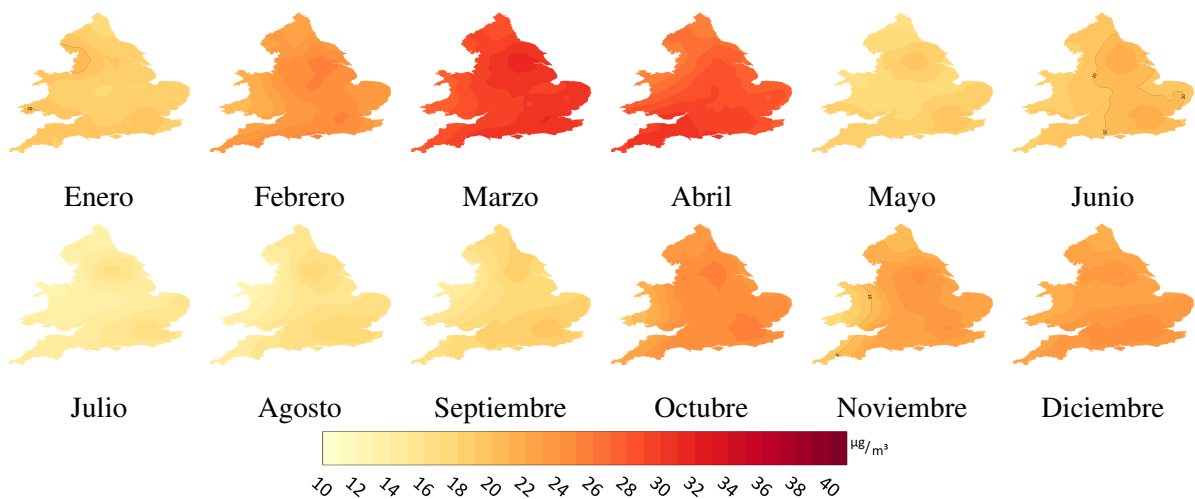


Figura 24: Media mensual de concentración del  $PM_{10}$  en el año 2007.

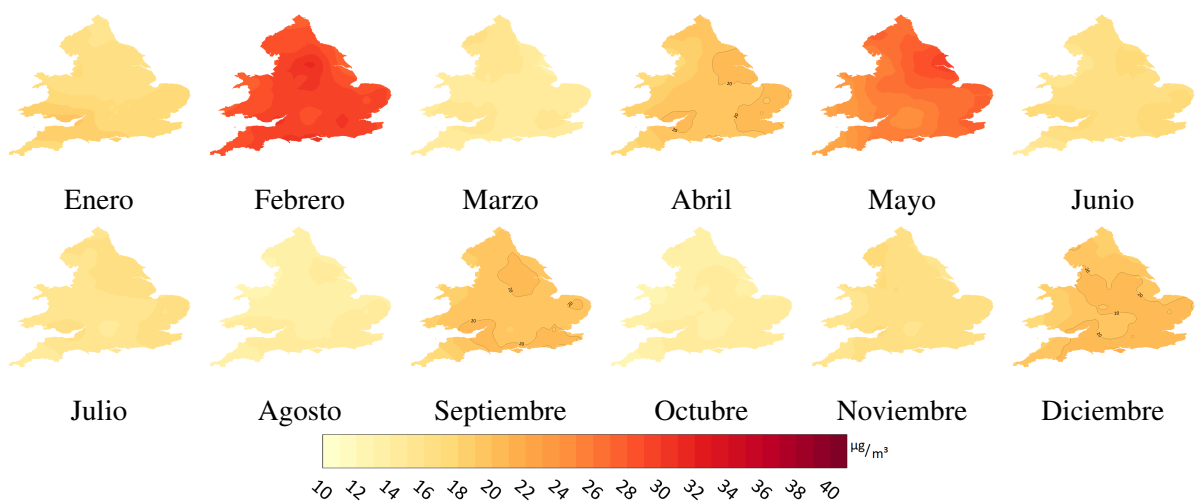


Figura 25: Media mensual de concentración del  $PM_{10}$  en el año 2008.

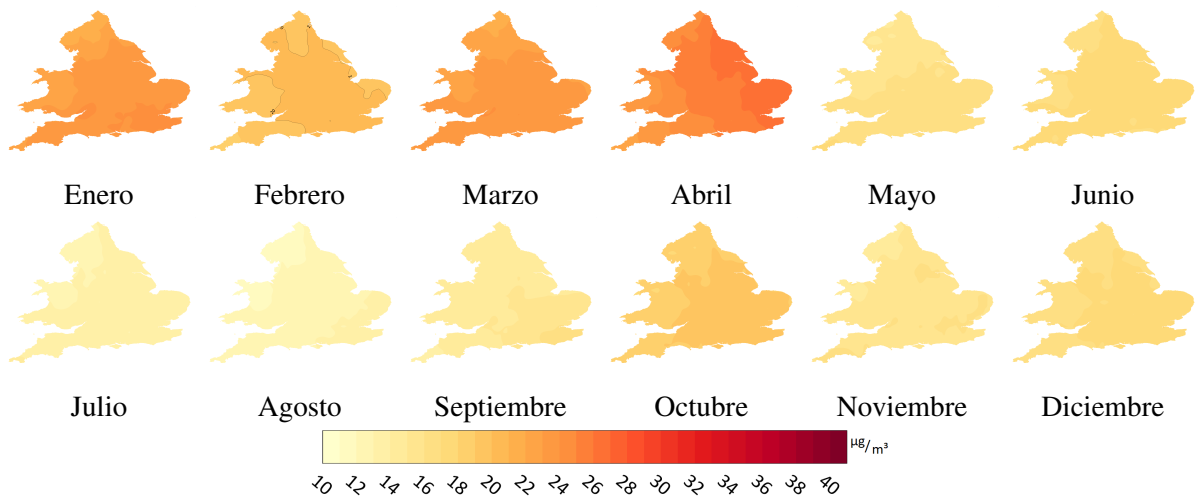


Figura 26: Media mensual de concentración del  $PM_{10}$  en el año 2009.

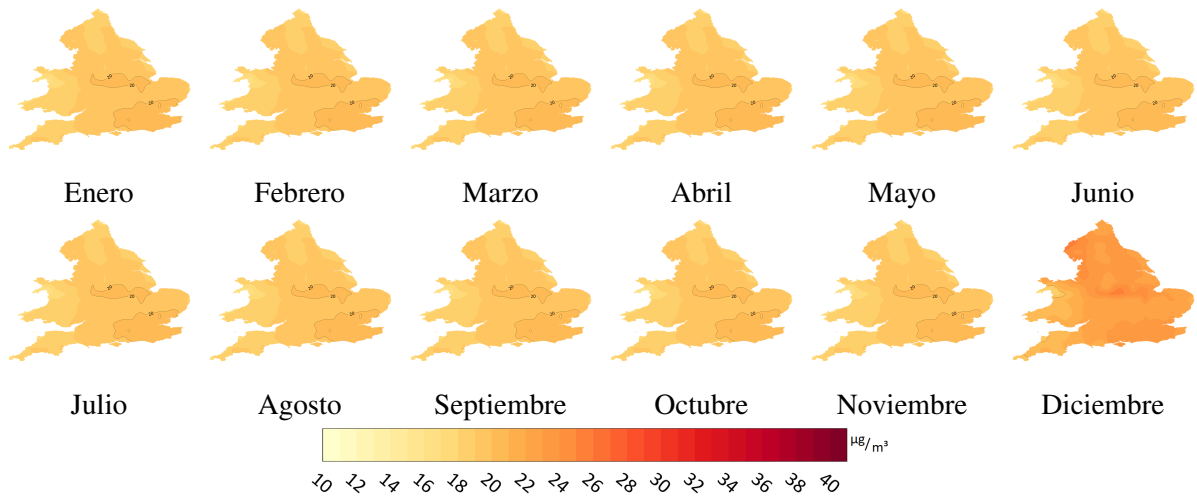


Figura 27: Media mensual de concentración del  $PM_{10}$  en el año 2010.

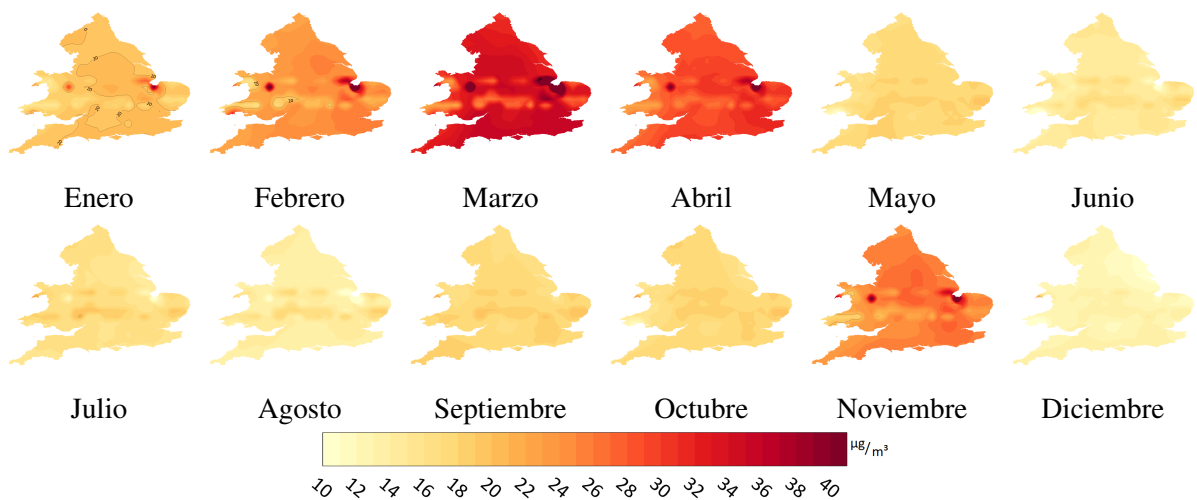


Figura 28: Media mensual de concentración del  $PM_{10}$  en el año 2011.

Si se analiza la distribución de la concentración anual de  $PM_{10}$  a lo largo del territorio de Inglaterra y Gales en el período 2007-2011 (figura 30), se observa que sigue una distribución parecida a la

analizada anteriormente: en el oeste de Gales hay una menor concentración, mientras que en el centro y este de Inglaterra se concentra más. Sin embargo, no es posible establecer una tendencia en la evolución de su concentración. Si atendemos a la figura 29, vemos que, en general, las concentraciones anuales son muy parecidas, salvo en el año 2011, en el que hay una mayor dispersión espacial que en otros años. Del mismo modo, los valores anómalos son tanto concentraciones bajas como altas y, además, la dispersión de los valores centrales no es tan pequeña en comparación con el rango total de los datos (como veíamos en el  $NO_2$  y  $O_3$ ), salvo para el 2011. Respecto al porcentaje de territorio que cumple la normativa (tabla 10), se puede decir que en relación al  $PM_{10}$ , el territorio de Inglaterra y Gales está poco contaminado: en todo el periodo del 2007-2011, ningún porcentaje del territorio ha superado el límite establecido por la UE. Ahora bien, respecto al límite establecido por la OMS, en la mayoría de los años no ha sido superado por una fracción significativa del territorio, salvo en los años 2007 (un 92% del territorio) y 2011 (un 58% del territorio). Los intervalos que hemos establecido corresponden a aquellos valores que están por debajo del valor límite establecido por la OMS, aquellos valores entre los valores límite establecidos por la OMS y la UE, respectivamente, y aquellos valores que superan el límite establecido por la UE.

Ahora bien, si analizamos los histogramas y sus respectivos suavizados obtenidos en la figura 31, vemos que la mayor parte de concentración de  $PM_{10}$  está agrupada en torno a los  $20 \mu g/m^3$ . Así, en los años en los que la mayor parte del territorio cumple los valores recomendados por la OMS (2008, 2009 y 2010), lo hace solo por un margen muy pequeño. Por último, vemos que el suavizado del histograma y el histograma en sí coinciden muy bien, y en la tabla 11 vemos que hay muy buen acuerdo con todos los coeficientes calculados, siendo el índice de concordancia modificado y  $E$  modificado los más bajos, pero por muy poco.

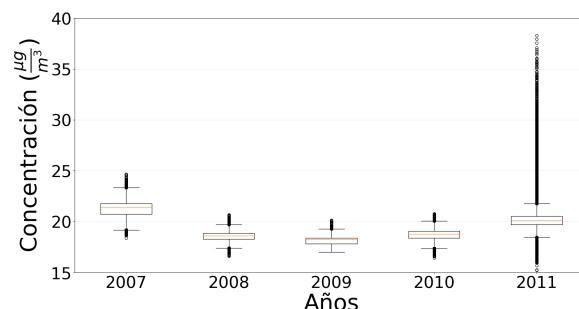


Figura 29: Diagramas de caja y bigotes de la media anual de concentración de  $PM_{10}$  en el período 2007-2011.

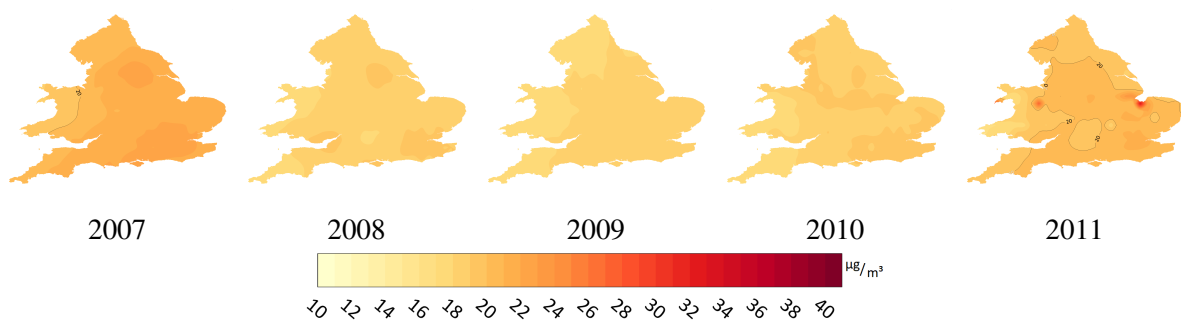


Figura 30: Media anual de concentración del  $PM_{10}$  en el período 2007-2011.

Tabla 10: Porcentaje de la superficie de Inglaterra y Gales contaminada de acuerdo al rango de concentración de  $PM_{10}$ .

Año	$(0, 20) / \frac{\mu g}{m^3}$	$(20, 40) / \frac{\mu g}{m^3}$	$(40, \infty) / \frac{\mu g}{m^3}$
2007	8,16	91,84	0,00
2008	99,96	0,04	0,00
2009	99,99	0,01	0,00
2010	99,89	0,11	0,00
2011	42,33	57,67	0,00

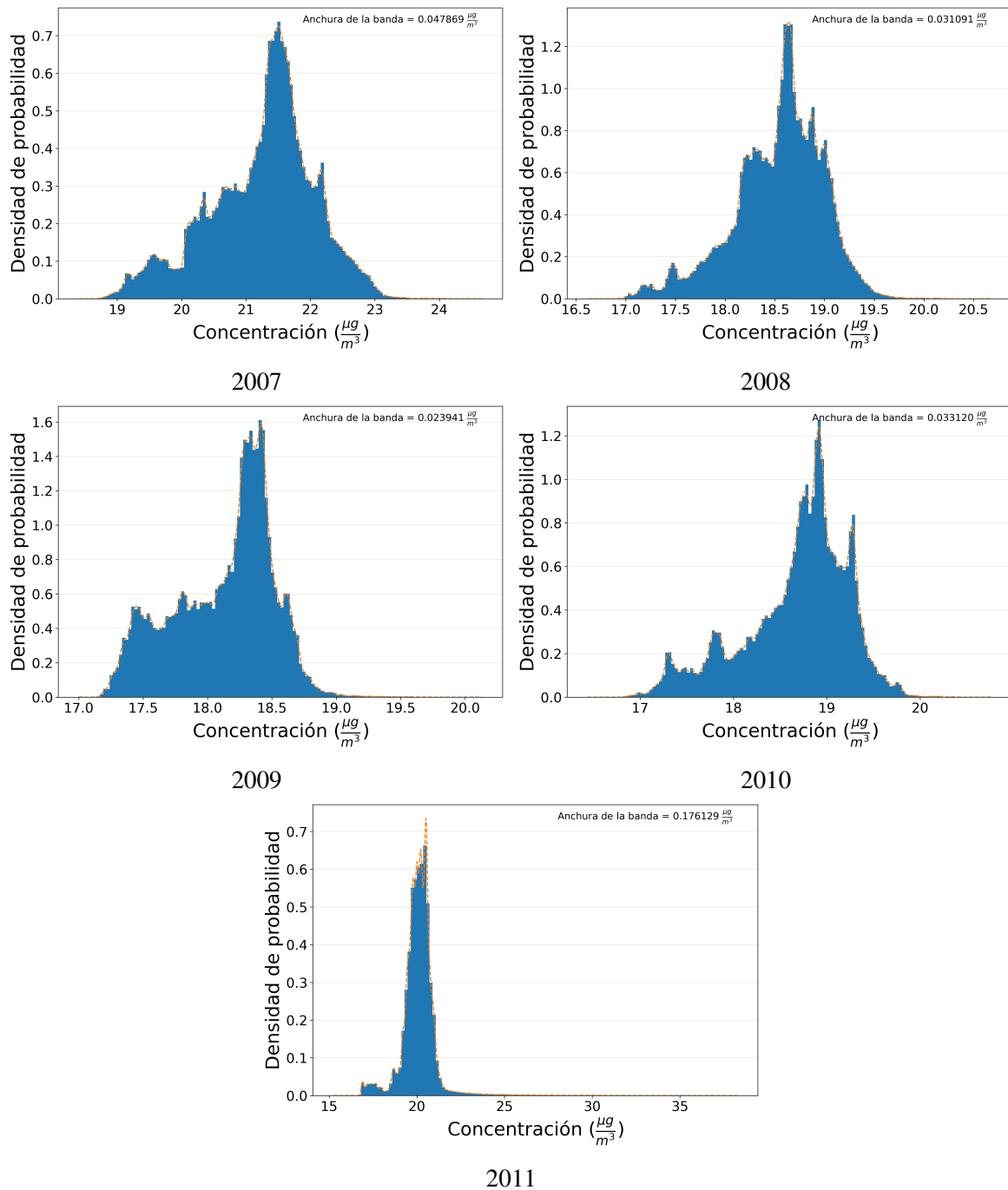


Figura 31: Histogramas de la media anual de concentración de  $PM_{10}$  en el período 2007-2011.



Tabla 11: Coeficientes de eficacia del suavizado del histograma para el  $PM_{10}$ .

Año	$r$	$d$	$E$	$d_1$	$E_1$
2007	1,0000	1,0000	1,0000	0,9996	0,9992
2008	1,0000	1,0000	1,0000	0,9999	0,9997
2009	1,0000	1,0000	1,0000	0,9988	0,9975
2010	1,0000	1,0000	1,0000	0,9997	0,9995
2011	1,0000	1,0000	1,0000	0,9989	0,9977

#### 4.4. Partículas $PM_{2,5}$

Como puede observarse en la figura 32, el  $PM_{2,5}$  varía mucho según el año: en algunos está muy concentrado en ciertos valores (como en el 2009), mientras que en otros el rango de concentración de  $PM_{2,5}$  es muy amplio (como en el 2008). Sin embargo, en todos puede verse una tendencia de haber menor concentración en los meses de verano, mientras que la concentración es más alta a principios de año. Es posible que esta tendencia esté relacionada con la mayor concentración de  $NO_2$  en los meses de invierno y primavera (pues los nitratos son precursores del  $PM_{2,5}$ ); sin embargo, dada la amplia gama de precursores de este contaminante, hacen falta más datos para poder afirmar esto. Además, parece haber tendencia en que los datos anómalos (los años en los que hay) sean concentraciones altas con una densidad de probabilidad muy pequeña, lo que indica que se hallan en superficies muy pequeñas del territorio estudiado. Finalmente, la dispersión de los datos centrales varía mucho dependiendo del año y del mes. Parece ser que hay cierta tendencia en que sea mayor en invierno y otoño que en verano, pero no se cumple esto todos los años. Además, esta dispersión es más o menos pequeña respecto al rango total de datos dependiendo del año y del mes.

Si se observan las figuras 33, 34, 35, 36 y 37, se puede comprobar que la distribución de  $PM_{2,5}$  a lo largo de Inglaterra y Gales es distinta año a año, no parece seguir ningún patrón de distribución. Lo único que puede decirse es que parece que en el territorio de Gales tiende a concentrarse menos  $PM_{2,5}$  que en Inglaterra, pero esto no se cumple todos los años: hacen falta más datos para poder corroborar esto.

Ahora bien, en la figura 37 podemos observar que los meses de marzo y abril tienen unas concentraciones mucho mayores si las comparamos con el resto de años (sobre todo en marzo). Es, además, en los mismos meses en los que la concentración de  $PM_{10}$  es mayor de lo usual en el período estudiado. El motivo de este aumento de la concentración probablemente sea debido a, una vez más, los meses excepcionalmente secos en Inglaterra y Gales que causaron que no pudiera depositarse tan rápidamente como en otros meses el  $PM_{2,5}$ . Además, en el mes de abril hubo una incursión de polvo proveniente del Sáhara en Irlanda y en el Reino Unido, lo que puede haber contribuido también a una mayor concentración de  $PM_{2,5}$  en ese mes (NASA, 2011).

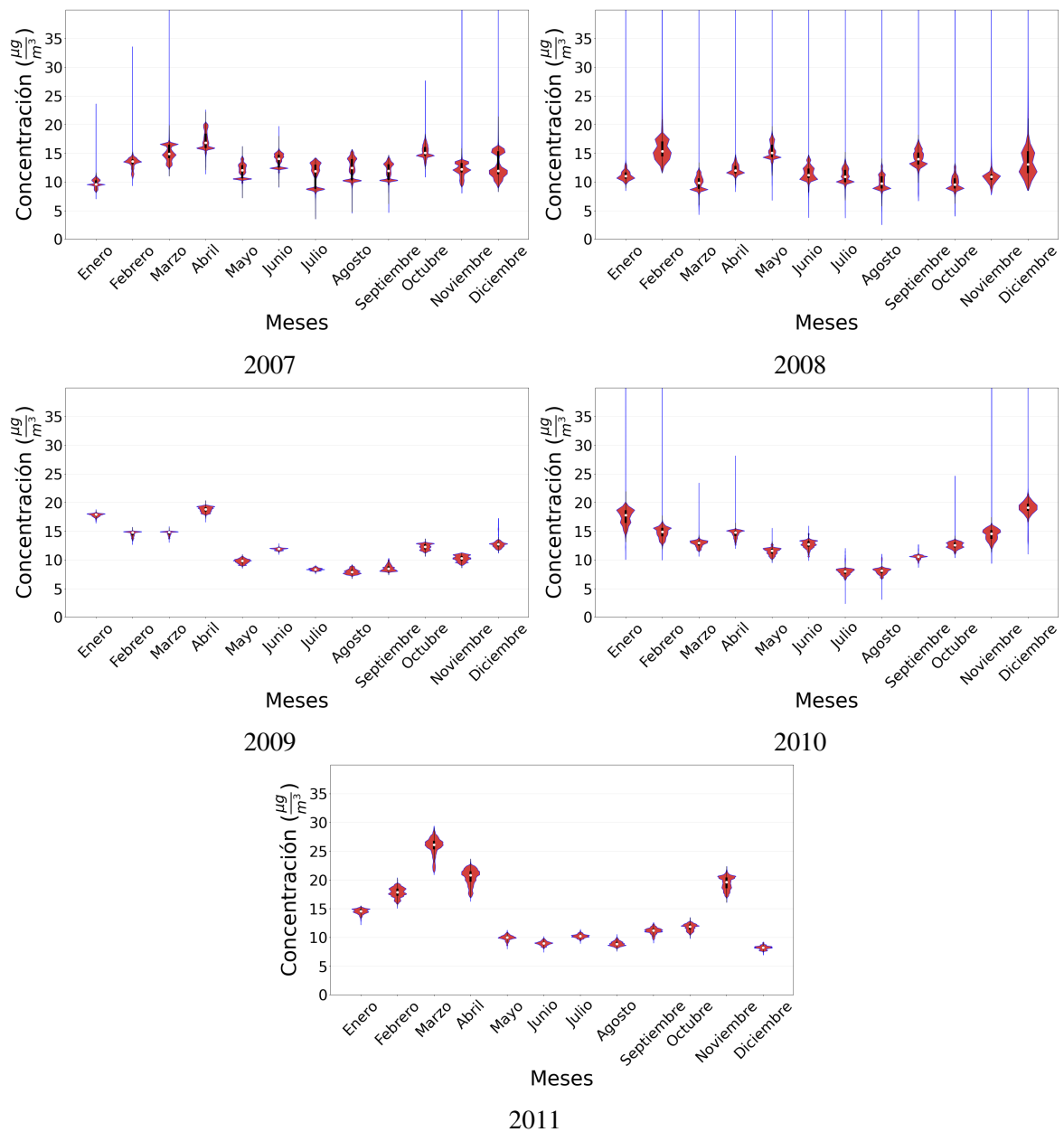


Figura 32: Diagramas de violín de la media mensual de concentración de  $PM_{2.5}$  en el período 2007-2011.

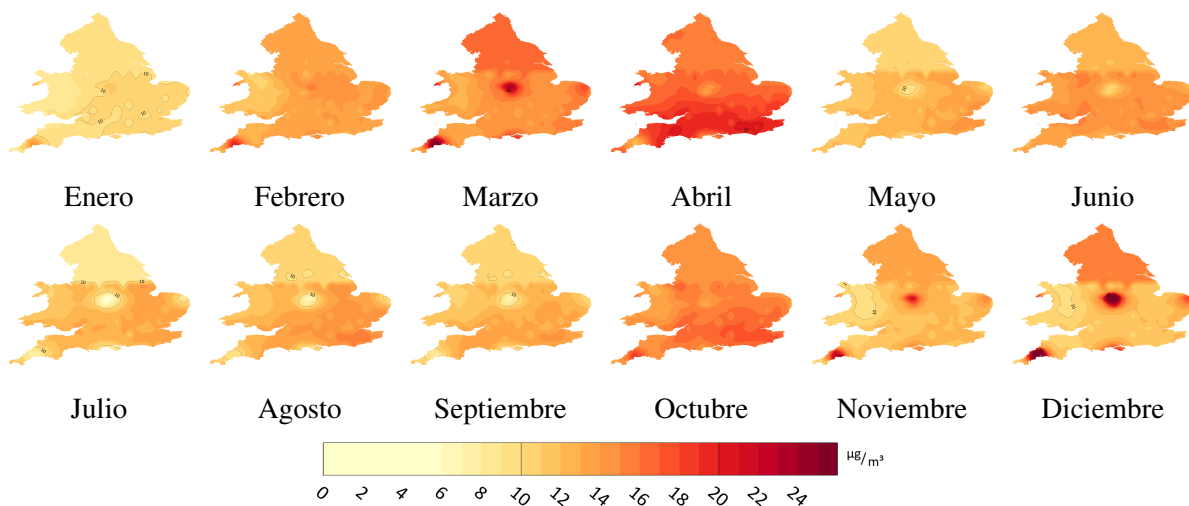


Figura 33: Media mensual de concentración del  $PM_{2,5}$  en el año 2007.

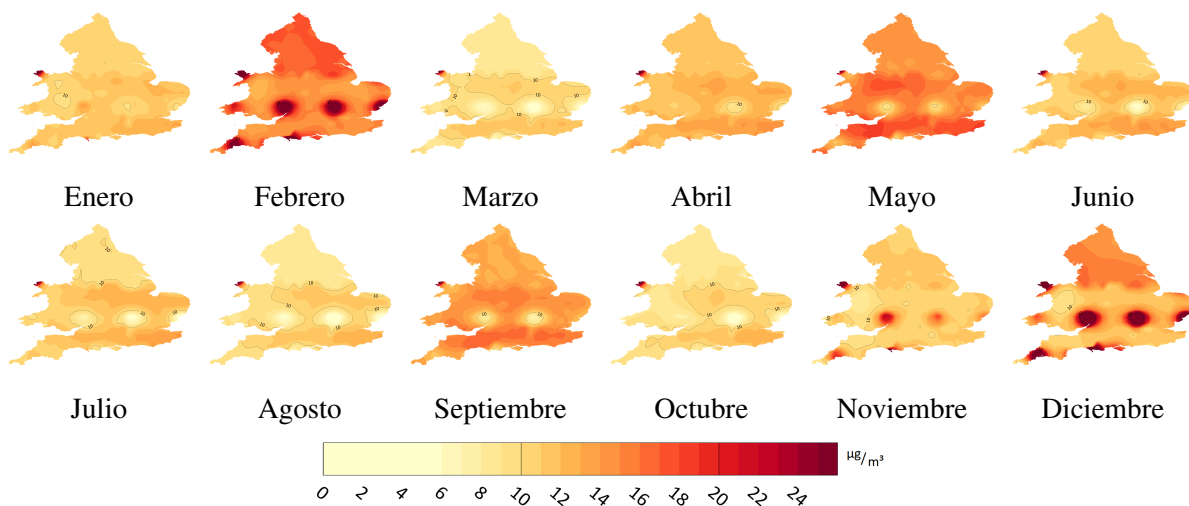


Figura 34: Media mensual de concentración del  $PM_{2,5}$  en el año 2008.

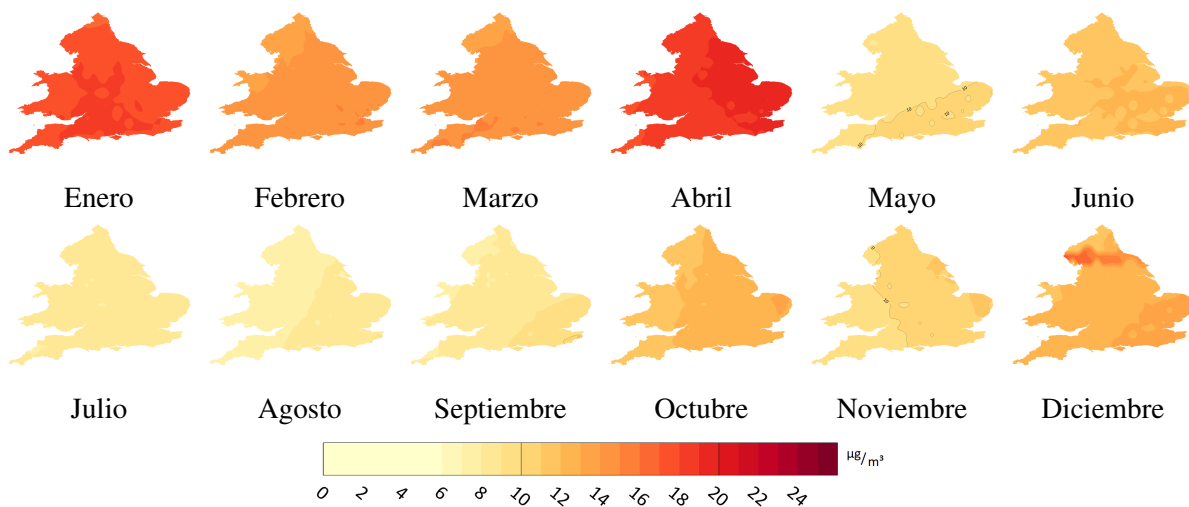


Figura 35: Media mensual de concentración del  $PM_{2,5}$  en el año 2009.

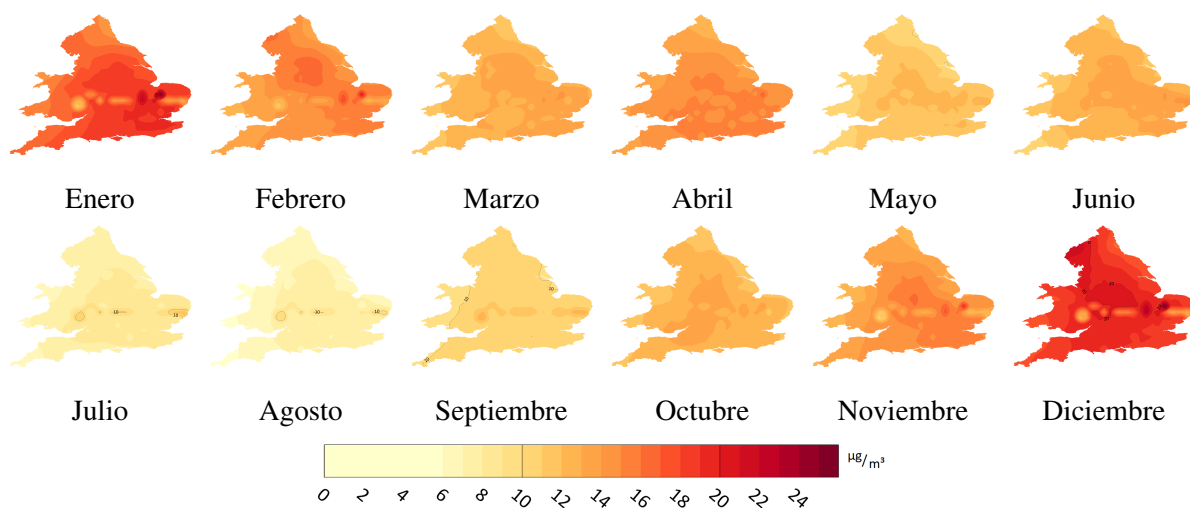


Figura 36: Media mensual de concentración del  $PM_{2,5}$  en el año 2010.

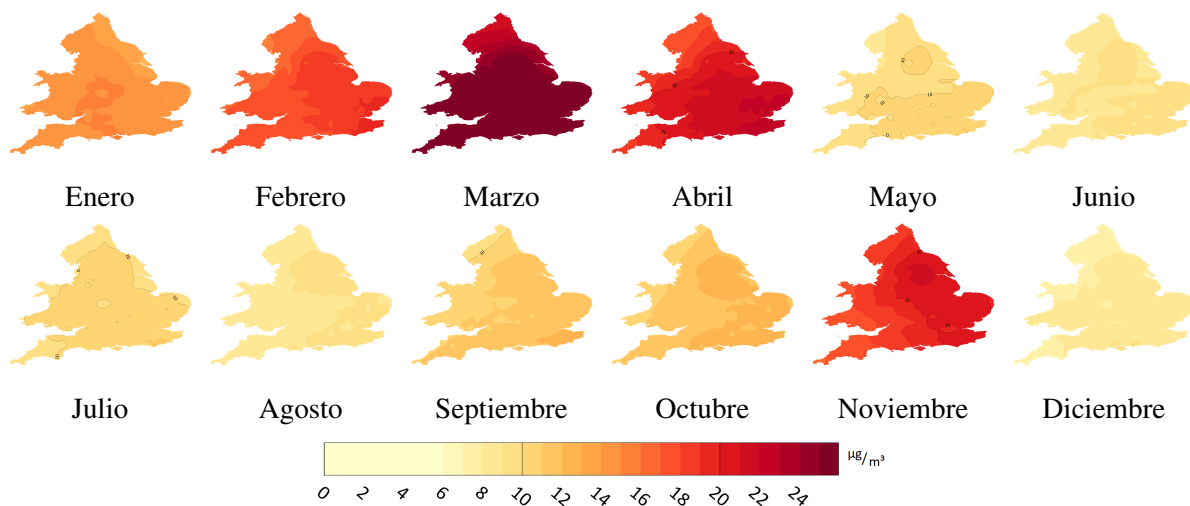


Figura 37: Media mensual de concentración del  $PM_{2,5}$  en el año 2011.

Finalmente, respecto a la distribución anual de  $PM_{2,5}$  en Inglaterra y Gales, parece que hay tendencia en haber menos concentración en Gales (sobre todo en el este) y el norte de Inglaterra, aunque no se puede afirmar con total rotundidad esto. Dada la insuficiente cantidad de datos y su heterogeneidad espacial y temporal (figuras 38 y 39), no es posible establecer una tendencia en la evolución de la concentración del  $PM_{2,5}$ . Si analizamos en detalle la figura 38, vemos que los valores centrales de los datos no tienen muchas diferencias en función del año. Para ciertos años, se dan valores anómalos de concentraciones altas y bajas (aunque hay una mayor cantidad de datos anómalos de concentraciones altas). Finalmente, la dispersión de los datos centrales es, en general, no muy pequeña en comparación con el rango total de los datos (salvo para los años 2007 y 2008).

Durante los cinco años que se han analizado, casi todo el territorio de Inglaterra y Gales ha cumplido la normativa de la UE respecto a la concentración de  $PM_{2,5}$  (máximo de  $25 \mu g/m^3$  de media anual). Sin embargo, todo el territorio no ha cumplido el margen establecido por la OMS (con un máximo de  $10 \mu g/m^3$  de media anual). Todo esto puede comprobarse en la figura 39 y en la tabla 12. Los intervalos establecidos corresponden a valores que se hallan por debajo del límite establecido por la OMS, por valores entre los límites de la OMS y la UE, respectivamente, y valores por encima del límite establecido por la UE.

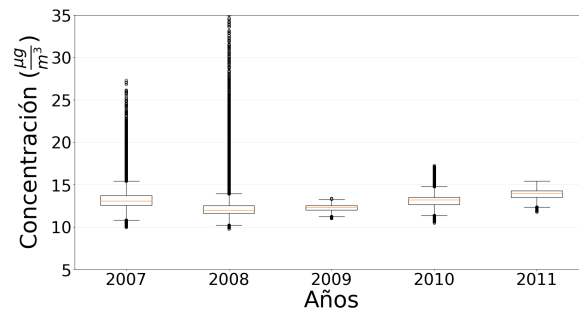


Figura 38: Diagramas de caja y bigotes de la media anual de concentración de  $PM_{2,5}$  en el período 2007-2011.

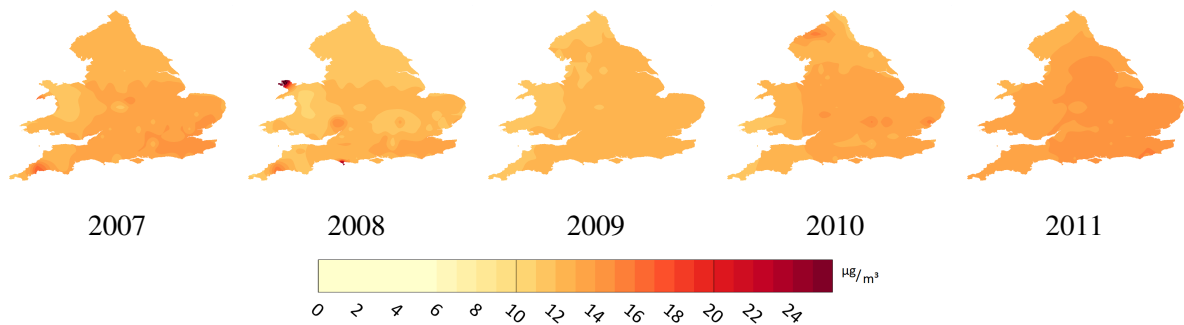


Figura 39: Media anual de concentración del  $PM_{2,5}$  en el período 2007-2011.

Tabla 12: Porcentaje de la superficie de Inglaterra y Gales contaminada de acuerdo al rango de concentración de  $PM_{2,5}$ .

Año	$(0, 10) / \frac{\mu g}{m^3}$	$(10, 20) / \frac{\mu g}{m^3}$	$(20, \infty) / \frac{\mu g}{m^3}$
2007	0,00	99,90	0,10
2008	0,00	99,57	0,43
2009	0,0	100,00	0,00
2010	0,0	100,00	0,00
2011	0,0	100,00	0,00

Si analizamos los histogramas y sus respectivos suavizados obtenidos en la figura 40, vemos que la mayor parte de concentración de  $PM_{2,5}$  está agrupada entre los  $10 \mu g/m^3$  y  $17 \mu g/m^3$ . Así, aunque en todos los años casi todo el territorio de Inglaterra y Gales cumpliera la legislación de la UE, en varios años una fracción considerable se hallaba en torno a los  $15 \mu g/m^3$ , acercándose al valor límite de la UE. Por último, vemos que el suavizado del histograma y el histograma en sí coinciden muy bien, y en la tabla 13 vemos que hay muy buen acuerdo con todos los coeficientes calculados, siendo el índice de concordancia modificado y  $E$  modificado los más bajos, pero por muy poco.

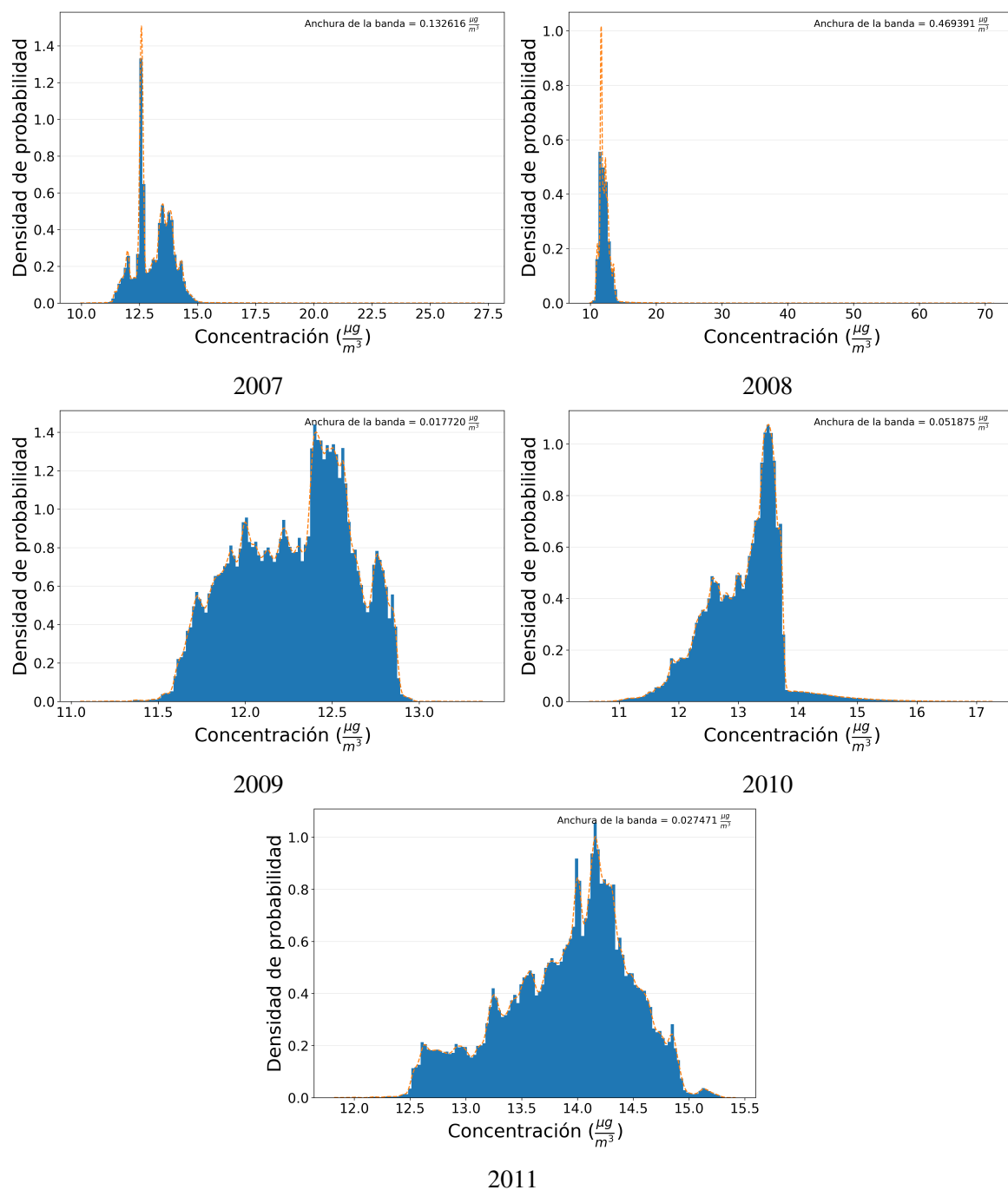


Figura 40: Histogramas de la media anual de concentración de  $PM_{2,5}$  en el período 2007-2011.

Tabla 13: Coeficientes de eficacia del suavizado del histograma para el  $PM_{2,5}$ .

Año	$r$	$d$	$E$	$d_1$	$E_1$
2007	1,0000	1,0000	1,0000	0,9991	0,9983
2008	1,0000	0,9978	0,9901	0,9530	0,9015
2009	1,0000	1,0000	1,0000	0,9986	0,9972
2010	1,0000	1,0000	1,0000	0,9997	0,9994
2011	1,0000	1,0000	1,0000	0,9984	0,9968

Por último, una cuestión que afecta a todos los contaminantes es que se ha visto que el transporte de estos provenientes de Europa afectan a las concentraciones en Reino Unido. En general, se

ha comprobado que los vientos provenientes del este, sureste y sur (provenientes de Europa) traen aparejados mayores concentraciones de  $NO_2$ ,  $O_3$  y  $PM_{2,5}$ , mientras que los vientos del oeste son más limpios (provenientes del Atlántico). Del mismo modo, los vientos que soplan del este son más frecuentes en la primavera que en otras estaciones, mientras que los vientos que soplan del oeste son más frecuentes en invierno y verano. Esto podría explicar, en parte, las concentraciones altas de  $O_3$ , de  $NO_2$  y de  $PM_{2,5}$  en los meses de primavera y su baja concentración en verano, pero no explicarían los valores altos observados en los meses de invierno (Baker, 2010; Stirling et al., 2020; Graham et al., 2020).

## 5. Conclusiones

Tras el análisis de los resultados obtenidos, hemos podido explicar la distribución de las concentraciones de  $NO_2$  y de  $O_3$  mes a mes durante el período 2007-2011. Sin embargo, dado que los datos que analizamos del  $PM_{10}$  y  $PM_{2,5}$  no siguen, en apariencia, un patrón definido, no ha sido posible dar una explicación satisfactoria de ello (necesitamos datos de más años para poder establecer esto).

Del mismo modo, se ha podido analizar la distribución y concentración anual del  $NO_2$ ,  $O_3$ ,  $PM_{10}$  y  $PM_{2,5}$  durante el período 2007-2011 y hallar el porcentaje del territorio que superan los valores límite establecidos por la OMS y la UE. Para el caso del  $NO_2$ , aunque el territorio en el que superaba el valor límite era inferior al 7% en todos los años, se encontraba localizado en las grandes ciudades, esto es, donde reside una gran parte de la población. En el caso del  $PM_{10}$ , durante gran parte del período 2007-2011 casi todo el territorio ha estado por niveles inferiores al valor límite de  $20 \mu g/m^3$  establecido por la OMS, salvo para los años 2007 y 2011, en los que una gran parte del territorio sobrepasó este valor. Finalmente, la media anual de concentración del  $PM_{2,5}$  ha sobrepasado el valor límite de  $10 \mu g/m^3$  en la totalidad del territorio de Inglaterra y Gales durante el período analizado.

Aunque se ha podido analizar la distribución y concentración anual de estos cuatro contaminantes en el período del 2007-2011, cinco años de distribución es muy poco tiempo para poder establecer una tendencia sobre la concentración de contaminantes. Por ende, no hemos podido observar una tendencia en la concentración y distribución de estos contaminantes en Inglaterra y Gales.

Vemos así que la mayor limitación que se ha tenido en la elaboración de este trabajo ha sido el corto período de tiempo de datos para analizar. No solo es el motivo por el que no se haya podido observar una tendencia en la evolución de la distribución y concentración de los contaminantes, sino que también ha establecido dificultades para poder observar un patrón en la evolución mensual año a año de la distribución y concentración del  $PM_{10}$  y del  $PM_{2,5}$ . Además, aunque se haya podido explicar satisfactoriamente los patrones mes a mes de la distribución y concentración del  $NO_2$  y del  $O_3$ , el corto período de tiempo da cierta incertidumbre sobre la solidez de esta explicación.



## 6. Futuras líneas de estudio

Una vez vistas las conclusiones y limitaciones de este trabajo y, teniendo en cuenta la gran variedad de estudios que pueden realizarse respecto a la concentración de contaminantes, podemos hacer un breve desglose de las futuras líneas de estudio en las que puede ir encaminado.

La más directa y que tiene que ver con las limitaciones de este trabajo es de disponer de datos como los que hemos analizado durante un período de tiempo mucho más amplio. Esto permitiría, por un lado, dar una mayor solidez a las explicaciones que hemos dado para los patrones mes a mes de las distribuciones y concentraciones del  $NO_2$  y del  $O_3$  (en caso de que sean, efectivamente, de esa manera) y, del mismo modo, permitiría encontrar los patrones mes a mes de la distribución y la concentración del  $PM_{10}$  y del  $PM_{2,5}$ . Por otro lado, dispondríamos de una gran cantidad de años para observar las distribuciones y concentraciones de los contaminantes estudiados, lo que permitiría determinar una tendencia en sus distribuciones, concentraciones y porcentajes del territorio afectados por ciertos niveles de concentración. Del mismo modo, en este trabajo no se han tenido en cuenta los contaminantes transportados de Europa a Inglaterra y Gales y los contaminantes emitidos y/o producidos en la zona de estudio ni, del mismo modo, el transporte en el territorio de Inglaterra y Gales. Por ello, esta es otra línea de estudio que podemos establecer: el efecto del transporte de estos contaminantes en la concentración en otras regiones, ya sea dentro de Inglaterra y Gales o en otras regiones de Reino Unido y de Europa. Además, dado que en este trabajo no se han distinguido las concentraciones debidas a actividades naturales de las antropogénicas, otra futura línea de estudio es la investigación en las emisiones antropogénicas de estos contaminantes, intrusiones de ozono estratosférico, etc. de cara a determinar la influencia de cada proceso en la concentración de cada contaminante.

Otras líneas futuras en las que este trabajo puede resultar de ayuda es el estudio estadístico de casos de enfermedades vinculadas por exposición a largo plazo de estos contaminantes durante el período 2007-2011. Del mismo modo, dado que el  $NO_2$  puede convertirse en ácido nítrico a través de la reacción 3, puede realizarse un estudio de las concentraciones del  $HNO_3$ , sus precipitaciones, etc. durante este período de tiempo.

Por otro lado, la contaminación atmosférica también afecta a los ecosistemas y agricultura, por lo que pueden realizarse estudios a lo largo de este período de posibles casos de eutrofización en los ecosistemas, ritmo de crecimiento de la vegetación, impacto en el rendimiento de actividades agrícolas, etc.

## 7. Bibliografía

### Referencias

- Baker, J. (2010). A cluster analysis of long range air transport pathways and associated pollutant concentrations within the UK. *Atmospheric Environment*, 44: 563-571.
- Chen, X., Zhong, B., Huang, F., Wang, X., Sarkar, S., Jia, S., Deng, X., Chen, D. y Shao, M. (2020a). The role of natural factors in constraining long-term tropospheric ozone trends over Southern China. *Atmospheric Environment*, 220.
- Chen, Z., Li, R., Chen, D., Zhuang, Y., Gao, B. y Yang, L. (2020b). Understanding the casual influence of major meteorological factors on ground ozone concentrations across China. *Journal of Cleaner Production*, 242: 107-116.
- de Jesús, A. L., Thompson, H., Knibbs, L. D., Kowalski, M., Cyrus, J., Niemi, J. V., Kousa, A., Timonen, H., Luoma, K., Petäjä, T., Beddows, D., Harrison, R. M., Hopke, P. y Morawska, L. (2020). Long-term trends in  $PM_{2.5}$  mass and particle number concentrations in urban air: The impacts of mitigation measures and extreme events due to changing climates. *Environmental Pollution*, 263: 114500.
- Díaz, J., Ortiz, C., Falcón, I., Salvador, C. y Linares, C. (2018). Short-term effect of tropospheric ozone on daily mortality in Spain. *Atmospheric environment*, 187: 107-116.
- Donnely, A., Misstear, B. y Broderick, B. (2011). Applications of nonparametric regression methods to study the relationship between  $NO_2$  concentrations and local wind direction speed at background sites. *Science of the total environment*, 409: 1134-1144.
- EEA (2019). Air quality in Europe - 2019 report. *European Environment Agency*.
- Fernández-Duque, B., Pérez, I. A., García, M. A., Pardo, N. y Sánchez, M. L. (2019). Annual and seasonal cycles of  $CO_2$  and  $CH_4$  in a Mediterranean Spanish environment using different kernel functions. *Stochastic Environmental Research and Risk Assessment*, 33: 915-930.
- Goral, A. K., Raval, S., Patra, A. K. (2020). Path analysis approach to quantify the casual factors of ground-level ozone concentration near coal-mining regions. *International Journal of Environmental Science and Technology*, 17: 645-660.
- Graham, A. M., Pringle, K. J., Arnold, S. R., Pope, R. J., Vieno, M., Butt, E. W., Conibear, L., Stirling, E. L. y McQuaid, J. B. (2020). Impact of weather types on UK ambient particulate matter concentrations. *Atmospheric Environment: X*, 5: 100061.
- Hintze, J. L. y Nelson, R. D. (1998). Violin Plots: A Box Plot-Density Trace Synergism. *The American Statistician*, 52, No. 2.
- Krause, P., Boyle, D. P. y Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assesment. *Advances in geosciences*, 5: 89-97.
- Lin, M., Fiore, A. M., Cooper, O. R., Horowitz, L. W., Langford, A. O., Il, H. L., Johnson, B. J., Naik, V., Oltmans, S. J. y Senff, C. J. (2012). Springtime high surface ozone events over the western United States: Quantifying the role of stratospheric intrusions. *Journal of geophysical research*, 117: D00V22.

- Met Office (2016). Exceptionally warm and dry Spring 2011.
- Met Office (2020b). How much does it rain in the UK? En: <https://www.metoffice.gov.uk/weather/learn-about/weather/types-of-weather/rain/how-much-does-it-rain-in-the-uk>
- Met Office (2020a). The Great Smog of 1952. En: <https://www.metoffice.gov.uk/weather/learn-about/weather/case-studies/great-smog>
- Mukhopadhyay, S. y Sahu, S. K. (2016). A Bayesian spatio-temporal model to estimate long term exposure to outdoor air pollution at coarser administrative geographies in England and Wales. En: *Statistical Methods in the Atmospheric Sciences, Fourth edition.*. Elsevier. 23-75.
- Munir, S. (2017). Analysing Temporal Trends in the Ratios of  $PM_{2,5}/PM_{10}$  in the UK. *Aerosol and Air Quality Research*, 17: 34-48.
- NASA (2011). Saharan Dust off Europe. En: <https://earthobservatory.nasa.gov/images/50066/saharan-dust-off-europe>
- Parlamento Europeo y Consejo de la Unión Europea (2008). 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. *Diario Oficial de la Unión Europea*, 152: 1-44.
- Pope, R. J., Arnold, S. R., Chipperfield, M. P., Latter, B. G., Siddans, R. y Kerridge, B. J. (2018). Widespread changes in UK air quality observed from space. *Atmospheric Science Letters*, 19: e817.
- Rovira, J., Domingo, J. L. y Schuhmacher, M. (2020). Air quality, health impacts and burden of disease due to air pollution (PM10, PM2.5, NO2 and O3): Application of AirQ+ model to the Camp de Tarragona County (Catalonia, Spain). *Science of the total environment*, 703.
- Sachs, L. (1984). Statistical Decision Techniques. En: *Applied Statistics. A Handbook of Techniques, Second Edition.*. Springer. 23-194.
- Shaughnessy, W. J., Venigalla, M. M. y Trump, D. (2015). Health effects of ambient levels of respirable particulate matter (PM) on healthy, young-adult population. *Atmospheric environment*, 123: 102-111.
- Shi, Y., Tsuneto, M., Yamaguchi, Y., Z, A., Li, Z. y Gu, X. (2018). Long-term trends and spatial patterns of  $PM_{2,5}$ -induced premature mortality in South and Southeast Asia from 1999 to 2014. *Science of the total environment*, 631-632: 1504-1514.
- Stirling, E. L., Pope, R. J., Graham, A. M., Chipperfield, M. P. y Arnold, S. R. (2020). Quantifying the transboundary contribution of nitrogen oxides to UK air quality. *Atmospheric Science Letters*, 21: e955.
- Wallace, J.M y Hobbs, P.V. (2006). Atmospheric Chemistry. En: *Atmospheric Science. An Introductory Survey, Second Edition.*. Academic Press. 153-207.
- WHO (2006). Air quality guidelines. Global update 2005. Particulate matter, ozone, nitrogen dioxide and sulfur dioxide. *World Health Organization*.
- WHO (2006). Guías de calidad del aire de la OMS relativas al material particulado, el ozono, el dióxido de nitrógeno y el dióxido de azufre. Actualización mundial 2005. Resumen de evaluación de los riesgos. *World Health Organization*.
- Wilks, D.S (2019). Empirical Distributions and Exploratory Data Analysis. En: *Statistical Methods in the Atmospheric Sciences, Fourth Edition.*. Elsevier. 23-75.