



ICADE – ICAI, Universidad Pontificia de Comillas

ANÁLISIS MULTIFACTORIAL DE LA CONTAMINACIÓN AMBIENTAL EN LA COMUNIDAD FORAL DE NAVARRA MEDIANTE TÉCNICAS DE BIG DATA.

Autor: Gregorio Egea Díaz

Tutor: Manuel Alejandro Betancourt Odio

MADRID | Noviembre 2021

Agradecimientos

Me gustaría expresar el más sincero agradecimiento a mi director de este proyecto, Alejandro Betancourt Odio por darme la oportunidad de llevar a cabo este trabajo y por la gran atención y orientación durante el desarrollo de este. Ayudándome a pensar de manera crítica y a plasmar las ideas con belleza.

Índice

Índice de tablas.....	4
Índice de Figuras	5
Índice de ejemplos.....	6
Resumen.....	7
Abstract	8
Introducción	9
Capítulo 1. Actualidad del Tema.....	14
Capítulo 2. Metodología.....	17
Capítulo 3. Aplicación de la Metodología.....	24
Conclusiones	44
Referencias Bibliográficas	46
Anexos.....	48
Anexo 1. Código desarrollado en Matlab y RStudio.....	48
Anexo 2. Scripts y funciones del trabajo.....	57

Índice de tablas

Tabla 1. <i>Lista de contaminantes del estudio.</i>	19
Tabla 2. <i>Resultado de la regresión por contaminantes (Parte 1/2). Se muestra el promedio de los coeficientes estimados para cada una de las variables independientes, así como su p-valor para su nivel de significación.</i>	28

Tabla 3. <i>Resultado de la regresión por contaminantes (Parte 2/2). Se muestra el promedio de los coeficientes estimados para cada una de las variables independientes, así como su p-valor para su nivel de significación.</i>	29
Tabla 4. <i>Impacto de los factores del estudio sobre la partícula PM2.5.</i>	34
Tabla 5. <i>Impacto de los factores sobre las concentraciones de BEN y TOL en la estación de Iturrama.</i>	42
Tabla 6. <i>Distancias más cortas entre las estaciones de contaminación y meteorológicas</i>	55

Índice de Figuras

<i>Figura 1.</i> Localización de estaciones de contaminación y detalle Pamplona	18
<i>Figura 2.</i> Localización de las estaciones meteorológicas.....	20
<i>Figura 3.</i> Fases del proyecto.....	24
<i>Figura 4.</i> Impacto promedio de los factores.	30
<i>Figura 5.</i> Efecto de los factores climatológicos sobre la concentración de PM10.	32
<i>Figura 6.</i> Efecto de los días no laborables sobre la partícula PM10.....	33
<i>Figura 7.</i> Impacto del Lockdown, Combustión Biomasa y Polvo del Sahara sobre el PM10.....	34
<i>Figura 8.</i> Impacto de las condiciones climatológicas sobre la concentración de NO2.....	35
<i>Figura 9.</i> Impacto de los días no laborables sobre la concentración de NO2.	35
<i>Figura 10.</i> Impacto del Lockdown, la combustión de biomasa y el polvo del Sahara en la concentración de NO2.....	36
<i>Figura 11.</i> Impacto de los factores medioambientales sobre la partícula NO.	37
<i>Figura 12.</i> Impacto de los días no laborables sobre la concentración de NO.	37
<i>Figura 13.</i> Impacto del confinamiento sobre la concentración de NO.	37
<i>Figura 14.</i> Impacto de los factores climatológicos sobre la concentración de NOx.....	38
<i>Figura 15.</i> Impacto de los días no laborables sobre la concentración de NOx.....	38
<i>Figura 16.</i> Impacto del Lockdown, la Combustión de Biomasa y el Polvo del Sahara sobre la concentración de NOx.....	39

<i>Figura 17.</i> Impacto de la temperatura sobre la concentración de O3.	39
<i>Figura 18.</i> Efecto del viento y el confinamiento en la concentración de O3.....	40
<i>Figura 19.</i> Efecto de la combustión de biomasa y el polvo del Sahara sobre la concentración de O3.	41
<i>Figura 20.</i> Efecto del Lockdown y de la Combustión de Biomasa sobre el SO2.....	41
<i>Figura 21.</i> Localización polígono industrial en Pamplona.	42
<i>Figura 22.</i> Esquema PREPARAR_DATOS_NAVARRA.	49

Índice de ejemplos

<i>Ejemplo 1.</i> Muestra de la estructura : <i>ficheros_importar</i>	50
<i>Ejemplo 2.</i> Resultado de almacenar tablas en la estructura usando <i>readtable()</i>	51
<i>Ejemplo 3.</i> Muestra de la estructura con las tablas limpias.	52
<i>Ejemplo 4.</i> Formato Original.	53
<i>Ejemplo 5.</i> Formato Madrid.	53
<i>Ejemplo 6.</i> Tabla resultada de la regresión.....	56

Resumen

Este trabajo realiza un análisis multifactorial del impacto de los contaminantes, SO₂, CO, NO, NO₂, PM_{2.5}, PM₁₀, NO_x, O₃, TOL, BEN y MXY, en la Comunidad Foral de Navarra, España. El estudio realizado está soportado en modelos de regresión múltiple y en el análisis espacial de datos.

Los resultados muestran claras disminuciones en las concentraciones de los contaminantes en escenarios de baja actividad económica para la mayoría de los contaminantes, como, entre los que destacan los días no laborables (reducción entre un 20 y 35% respecto los días laborables) o durante el periodo de confinamiento (38% de reducción en las concentraciones de los contaminantes de media). La excepción en este comportamiento generalizado la tiene el Ozono troposférico, ya que es el único contaminante que experimenta aumentos en su concentración en estos periodos (más del 30%). También obtenemos de los resultados que las condiciones meteorológicas adversas como la lluvia o el viento ayudan a depurar el ambiente mejorando la calidad del aire.

El análisis de los datos diarios en las 12 estaciones de medición de contaminación ubicada en la Comunidad foral de Navarra, entre los años 2015-2021 ha permitido también establecer diferencias a nivel local en cuanto a los comportamientos de las partículas contaminantes.

Palabras clave: Contaminación, Clima, Calidad del aire, Confinamiento.

Abstract

This project carries out a multifactorial analysis of the impact of the pollutants, SO₂, CO, NO, NO₂, PM_{2.5}, PM₁₀, NO_x, O₃, TOL, BEN and MXY, in the Comunidad Foral de Navarra, Spain. The study carried out is supported by multiple regression models and spatial data analysis.

The results show clear decreases in the concentrations of most of the pollutants in scenarios of low economic activity, among which stand out non-working days (reduction between 20 and 35% compared to working days) or during the period of confinement (38% reduction in pollutant concentrations on average). The exception to this generalized behavior is the tropospheric ozone, since it is the only pollutant that experiences an increase in its concentration in these periods (more than 30%). We also obtain from the results that adverse weather conditions such as rain or wind help to purify the environment, improving air quality.

The analysis of the daily data in the 12 pollution measurement stations located in the Autonomous Community of Navarra, between the years 2015-2021 has also made it possible to establish differences at the local level in terms of the behavior of the concentration of the polluting particles.

Keywords: Pollution, Climate, Air quality, Lockdown.

Introducción

La contaminación atmosférica es uno de los mayores peligros para la salud pública en todo el mundo debido a su impacto en la salud pública e individual. La exposición a los contaminantes atmosféricos está asociada con desencadenantes de enfermedades crónicas como pueden ser asma, insuficiencia pulmonar, enfermedades cardiovasculares.

Se calcula que debido a la contaminación atmosférica cada año mueren siete millones de personas en todo el mundo (WHO, 2020). Además, está asociada al cambio climático y al calentamiento global del planeta, que en última instancia afectan la salud y el bienestar de las personas a través de problemas de seguridad alimentaria, daños a la vegetación, inundaciones y las migraciones forzadas.

Las actividades humanas constituyen un motor esencial en las consecuencias negativas que ocasionan los actuales niveles de contaminación a nivel mundial ya que éstas han sido capaces de deteriorar la calidad de bienes básicos tan imprescindibles para la vida en el planeta como lo es el agua y el oxígeno que se respira.

En Europa, se presta especial atención a ciertos grupos de contaminantes no sólo por ser particularmente nocivos para la salud de las personas, sino también por encontrarse en grandes concentraciones y, por ende, rebasar los límites de la normativa impuesta para la preservación de la calidad del aire. En los casos del PM₁₀ y del PM_{2.5}, por ejemplo, pese a que su presencia se ha visto significativamente reducida en muchas de las urbes europeas en los últimos años 10 años, alrededor del 85% de los habitantes de ciudades en Europa inhala un aire que sobrepasa los niveles de estas partículas fijados como no saludables por la OMS. Las principales fuentes son el tráfico, la industria, los trabajos de construcción o la quema de biomasa para la generación de electricidad (Carceller, 2019). Además de estas fuentes producida por la actividad humana, destaca una fuente natural: las nubes de Polvo del Sahara (Querol, 2008), cuyo impacto resulta especialmente

relevante en los países del sur de Europa por la proximidad con el desierto de Marruecos, como es el caso de España.

Otros contaminantes como lo constituyen los Óxidos de Nitrógeno, y en especial, el dióxido de Nitrógeno (NO_2) ensombrecen el bienestar de alrededor del 7% de la población de Europa, ya que respiran este contaminante en cantidades perjudiciales para la salud. La causa principal de sus emisiones, pese a que existen tecnologías alternativas o que son capaces de reducirlas, son la gran cantidad de vehículos de transporte (tanto de personas como de mercancías) que utilizan motores combustión, especialmente los vehículos de combustión que utilizan diésel como combustible, aunque también procede de las emisiones y residuos de las diversas industrias y gases emitidos en procesos para la generación de energía eléctrica (Carceller, 2019).

España, país que se enmarca en la Comunidad Europea, no escapa de la problemática de la contaminación ambiental, existiendo una preocupación por la calidad del aire y por los efectos en el número de muertes prematuras relacionadas con la calidad del aire (Nunes et al. 2021). El aumento del número de vehículos y sobre todo de “la dieselización” durante los años los años 1990-2000, han contribuido a que la disminución de los niveles de contaminación por efecto de las partículas contaminantes que están directamente relacionadas por el tráfico terrestre, aéreo y marítimo, sea complejo. Por otro lado, dada la situación geográfica de la península dentro del globo terráqueo, hace que las condiciones climatológicas, caracterizado por un clima mediterráneo, estimule la presencia cada año de millones de turistas que si bien tiene un efecto positivo en la economía del país tiene un efecto negativo en los niveles de contaminación ya en sí mismo promueve el uso de vehículos contaminantes.

Otro elemento para destacar y que contribuye a las variaciones en los niveles de contaminación en España son las variables climáticas. Por ejemplo, los bajos niveles de precipitación que se han manifestado en muchas localidades españolas no facilita de forma rápida la eliminación partículas contaminante presente en la atmósfera, favoreciendo la continua re-suspensión de estas (Querol, 2008).

Como se ha expuesto, salvo excepciones, parece ser que el factor humano es el causante de gran parte de la contaminación ambiental y esto se pone de manifiesto en “el experimento de mayor escala de la historia” (metáfora con la que algunos expertos se refieren a la pandemia de la COVID-19 vivida en todo el mundo en el año 2020). Dicho “experimento” ha demostrado cómo en periodos de baja actividad industrial y desplazamiento de personas hay una menor presencia de contaminantes en el aire si se comparan con periodos similares, pero con intensa actividad económica, pero en años anteriores (Watts & Kommenda, 2020).

En el caso puntual de la Comunidad Foral de Navarra, medio objeto de nuestra investigación, aunque es pionera en la puesta en marcha de industria generadoras de energía renovables, persisten niveles de contaminación que en algún espacio temporal rebasa lo establecido por la OMS. Es por ello por lo que surge el interés de estudiar a nivel local las causas de las variaciones en los niveles de contaminación de once partículas contaminantes. Considerando todo lo descrito, surge la siguiente **pregunta de investigación**:

¿Cuáles son las principales variables que determinan los actuales niveles de contaminación en la Comunidad Foral de Navarra?

La **hipótesis de investigación** que pudiera dar respuesta a esta pregunta está dada por:

La actividad económica, industrial y el turismo como una actividad esencial dentro de la comunidad podrían generando los actuales niveles de contaminación de la región.

A partir de la posible respuesta planteada al problema de investigación el **objetivo general** de este trabajo es: Analizar el comportamiento de los niveles de contaminación de las distintas partículas contaminantes en la Comunidad Foral de Navarra entre los años 2015-2021 y determinar qué factores influyen en la calidad del aire.

Para el logro de este objetivo general se han establecido un conjunto de **objetivos específicos**:

1. Extraer datos de contaminación horaria para las 12 estaciones de calidad del aire situadas en la comunidad foral de navarra.
2. Extraer datos de variables climáticas disponible en la agencia española de meteorología (AEMET).
3. Extraer datos de la intrusión del Polvo del Sahara, disponible en los datos estadísticos ofrecidos por el Ministerio de Transición Ecológica y el Reto Demográfico.
4. Tratar los datos extraídos de forma automatizada que serán programados en MATLAB.
5. Construir modelos de regresión que expliquen la influencia de variables climáticas y otras que miden comportamiento de la actividad humana económica sobre los niveles de contaminación de las partículas contaminantes SO₂, CO, NO, NO₂, PM_{2.5}, PM₁₀, NO_x, O₃, TOL, BEN y MXY.
6. Analizar a nivel espacial los resultados obtenidos en los modelos de regresión mediante el uso de la herramienta informática PowerBI.

Para lograr validar las hipótesis y dar cumplimiento a los objetivos planteados, el trabajo se ha articulado en una introducción, tres capítulos y la exposición de las conclusiones obtenidas del análisis de los resultados.

En el capítulo 1, se dedica de forma resumida una panorámica de la actualidad del tema tratado en esta investigación, poniendo de relieve el efecto de las partículas contaminantes sobre la salud humana y qué líneas de actuación existen hoy en día para el control de las emisiones.

En el capítulo 2 se describe de forma exhaustiva la Metodología a seguir para consecución del objetivo general planteado. Seguidamente, en el capítulo 3 se aplicarán, a partir de los datos extraídos, la Metodología propuesta en el apartado dos, detallando los pasos seguidos en la elaboración del trabajo, así como la presentación de los resultados obtenidos.

Finalmente y como resumen de los resultados obtenidos se exponen las conclusiones y recomendaciones del estudio. Un apartado de anexos posterior a las conclusiones permitirá al

lector valorar y comprender con una mejor dimensión la profundidad del trabajo realizado en esta investigación. En esta parte se detallarán el desarrollo códigos máquina realizado en MATLAB.

Capítulo 1. Actualidad del Tema

La contaminación ambiental global se considera, sin duda, un problema de salud pública internacional. Las consideraciones sociales, económicas y legislativas, así como los estilos de vida, están relacionadas con este gran problema. Concretamente, la contaminación atmosférica es una de las principales amenazas para la salud pública en todo el mundo y es responsable de aproximadamente 9 millones de muertes al año (Manisalidis, Stavropoulou, Stavropoulos, & Bezirtzoglou, 2020).

El efecto de una partícula en concreto sobre la salud de una persona depende de tres factores fundamentales: el caudal de aire que respira esa persona, es decir, qué volumen de aire es procesado por el sistema respiratorio del sujeto por unidad de tiempo, la duración que la persona está expuesta a ese aire contaminado y finalmente la concentración de esa partícula en el aire. Es decir, la unidad de masa del contaminante por cada unidad de volumen de aire del ambiente. Si hacemos un análisis de unidades, podríamos decir que el producto de estos tres factores constituye la masa de agente contaminante que una persona inhala para una duración determinada.

Si no se desea alterar los hábitos de la sociedad para protegerles de la exposición a los contaminantes, el único de los tres factores susceptible de ser controlado es la concentración de partículas contaminantes en el aire inhalable. Respecto a los agentes contaminantes concretos, en Europa y como mencionado anteriormente se presta especial atención a tres tipos de partículas, con foco en las ciudades europeas más desarrolladas: En primer lugar, las partículas en suspensión (PM_{10} o $PM_{2.5}$ según su tamaño, menores de $10\ \mu m$ de diámetro para las PM_{10} y menores de $2.5\ \mu m$ de diámetro para las $PM_{2.5}$). En segundo lugar, los óxidos de nitrógeno NO_x , pero en especial el NO_2 y el NO ; y finalmente el Ozono de la troposfera o O_3 (Querol, 2008).

La nocividad de las partículas en suspensión PM_{10} y $PM_{2.5}$ se debe a que hacen disminuir la función pulmonar e interfiere en los mecanismos de defensa de los pulmones entre otros efectos

del corto plazo. En el largo plazo, pueden aumentar el riesgo de producir cáncer para las personas en la edad adulta. Por otro lado, los óxidos de nitrógeno pueden ser causantes de episodios de asma e interferir con el desarrollo de la capacidad pulmonar (Oyarzún, 2010).

Finalmente, el Ozono, popularmente está concebido de forma errónea como un gas que nos protege ante la radiación ultravioleta, no obstante, esto sólo ocurre con el Ozono en la estratosfera, el que se encuentra a decenas de kilómetros de altitud. Cuando se encuentra en la troposfera (i.e. a nivel de la tierra) es un contaminante en el punto de mira por estar asociado a irritaciones en las vías respiratorias, nariz u ojos siempre que esté en parámetros aceptables. Cuando el O₃ troposférico se da en grandes concentraciones y las personas se someten a largos periodos de exposición, se asocia con mortalidad por enfermedades cardíacas y respiratorias (Zhang, Wei, & Fang, 2019).

En Europa comenzó a concienciarse del problema de la contaminación y sus efectos nocivos en la salud de las personas sobre los años 60, especialmente tras el incidente de Londres de 1952, donde una densa niebla causada por gases contaminantes disparó los niveles de mortalidad y fue considerada catástrofe de primer nivel (Logan, 1953).

Es entonces cuando comenzaron las primeras medidas contra la emisión de estos gases como, por ejemplo, controlando el carbón utilizado en las calefacciones de la época. Hoy en día existe un interés social mayor que en aquella época y se plantean unos objetivos y controles mucho más rigurosos que entonces (Querol, 2008).

Por esta razón, gracias a las medidas que se han ido tomando al respecto y pese al crecimiento de la población mundial, al crecimiento de la concentración de esta población en núcleos urbanos y al enorme desarrollo de la tecnología, el consumo y el transporte, se ha conseguido reducir en Europa en los últimos 15 años las concentraciones de gases nocivos como el dióxido de azufre o el monóxido de carbono. Concretamente “contaminantes como PM₁₀ y PM_{2.5} han visto reducidas al 50 % sus concentraciones urbanas [...] Sin embargo, nos queda aún mucho camino por recorrer

para llegar a respirar un aire que pueda ser considerado como aceptable según los criterios de la OMS” (Carceller, 2019).

Poniendo el foco de atención en España, se sabe que es uno de los países que más afectado se verá por los efectos del cambio climático y emisiones de gases contaminantes en el futuro, además de ser el país europeo que más gases efecto invernadero ha emitido en las últimas dos décadas (Lazaro, Gonzalez, & Escribano, 2019).

No obstante, hay indicios que muestran que hoy en día la sociedad española vive en una época donde cada vez es más creciente la conciencia de los devastadores efectos futuros del cambio climático. En un estudio del Real Instituto Elcano “más del 90% de los entrevistados está de acuerdo con que se dedique parte de los Presupuestos Generales del Estado a compensar los daños causados por el cambio climático” (Lazaro, Gonzalez, & Escribano, 2019). España, además de tener una conciencia ambiental significativa similar a países como Suecia o Noruega parece que ha comenzado a atacar el problema mediante la regulación (zonas de circulación restringida, pago por uso de las autovías, impuestos sobre combustibles fósiles, etc.).

Concretamente en 2019, se presenta por primera vez ante la Comisión Europea un programa con el objetivo de reducir las emisiones de contaminantes atmosféricos para 2030 conteniendo 57 medidas que afectan a varios sectores (transporte, generación de energía, industria, comercio, etc.). Entre otros, se pretende reducir en un 50% las partículas en suspensión o un 66% los óxidos de nitrógeno (Ministerio para la transición ecológica, 2019).

Se ha mostrado como los protocolos contra la contaminación del aire tienen efecto, también se observa que existen y están en marcha los compromisos de cara al futuro para mejorar aún más la calidad del aire. Con una cooperación mundial y un compromiso individual, se podrá forjar un futuro en el que la sociedad pueda seguir desarrollando la tecnología y su actividad económica y social de manera sostenible y respetuosa con el medio ambiente.

Capítulo 2. Metodología

En este capítulo se revisa en una primera parte el conjunto de datos que han sido tratados en el proyecto. En una segunda parte, se formula el problema matemático que permitirá construir el modelo de regresión de manera formal.

2.1. Datos del estudio

Se acotará el perímetro del estudio a la comunidad Foral de Navarra. Navarra es una región española situada al norte de la península ibérica, concretamente es una de las diecisiete comunidades autónomas de España, limita con los pirineos, de ahí su geografía en parte montañosa, aunque la mitad sur se trata de una zona llana debido al valle del río Ebro.

Respecto a la climatología, tomando datos desde 2013 hasta 2021, Navarra tiene una temperatura media durante todo el año de entorno a los 14°C, con una temperatura máxima diaria de cerca de 20°C en promedio y una mínima diaria de 8°C en promedio. De media, las precipitaciones diarias son de en torno a 2 mm, pudiéndose multiplicar esta cantidad por más de 10 veces en los días de más lluvia (AEMET).

Atendiendo a la demografía, la comunidad Foral de Navarra tiene 654.214 habitantes de los cuales unos 200.000 habitan en la capital, Pamplona (INE, 2019). Las estaciones estudiadas están repartidas por su geografía con una mayor presencia en la capital y en otras localidades de menor tamaño como, por ejemplo, Tudela (~36.000 habitantes), Alsasua (~7.500), Sangüesa (~5.000) o Leitza (~3.000). Esto nos permite estudiar el impacto por contaminante en regiones distintas con características demográficas diversas y, por ende, actividades económicas y de transporte presumiblemente diversas también.

El total de estaciones de medición de contaminación son 12, de Norte a Sur: Lesaka, Leitza, Alsasua, Rotxapea, Plaza de la Cruz, Felisa Munarriz, Iturrama, Sangüesa, Olite, Funes, Tudela y Tudela II. Están repartidas por la geografía de la comunidad autónoma con una mayor

concentración en torno a la capital, Pamplona. El detalle de la localización se muestra en la figura de a continuación:

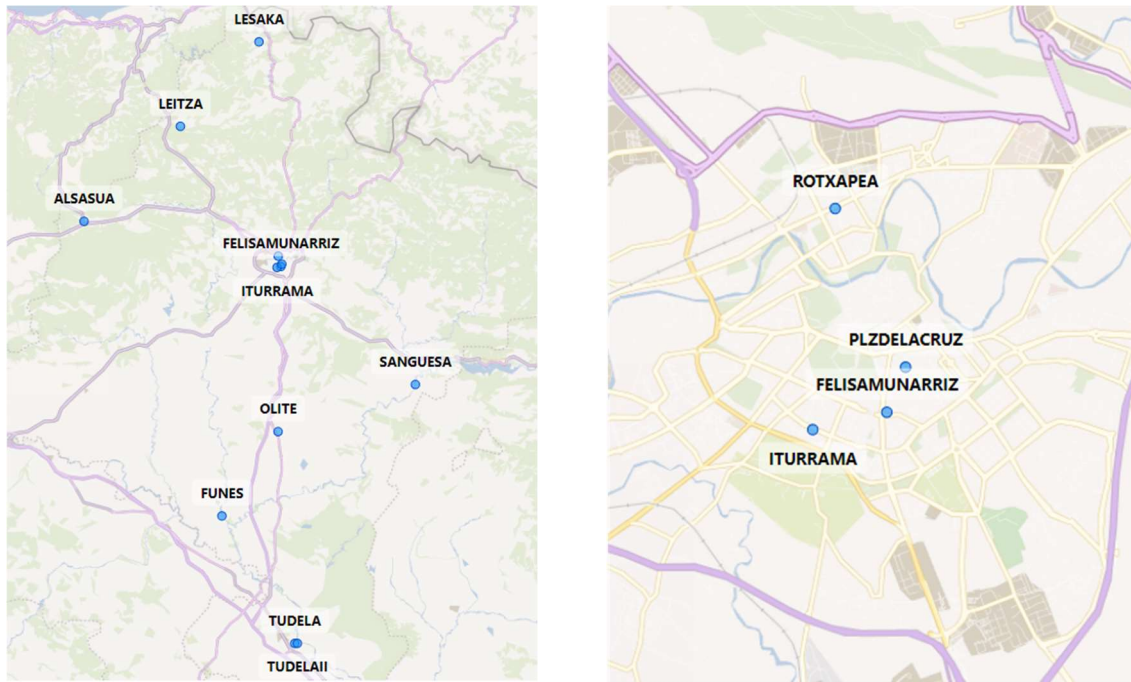


Figura 1. Localización de estaciones de contaminación y detalle Pamplona
Fuente: Elaboración propia con PowerBI.

Las partículas contaminantes medidas por el conjunto de estaciones de contaminación mostradas y se tratarán en este estudio, se listan en la siguiente tabla:

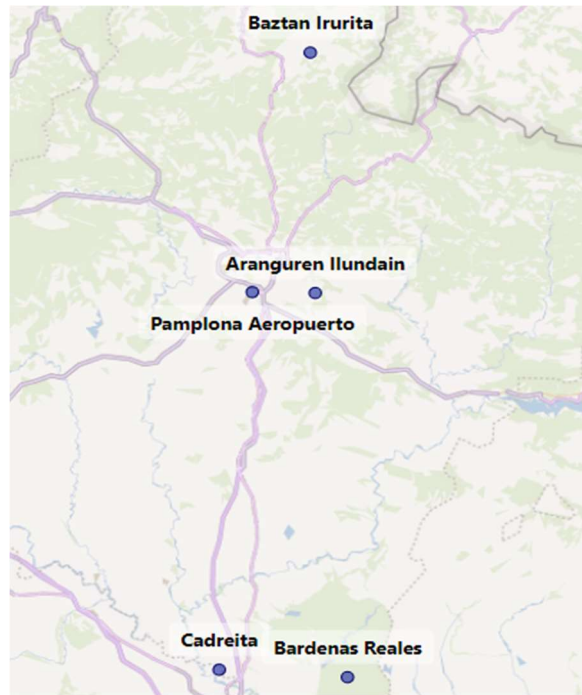
Tabla 1.

Lista de contaminantes del estudio.

Código	Nombre	Fórmula	Ud.
1	Dióxido de Azufre	SO ₂	µg/m ³
6	Monóxido de Carbono	CO	mg/m ³
7	Monóxido de Nitrógeno	NO	µg/m ³
8	Dióxido de Nitrógeno	NO ₂	µg/m ³
9	Partículas < 2.5 µm	PM _{2.5}	µg/m ³
10	Partículas < 10 µm	PM ₁₀	µg/m ³
12	Óxidos de Nitrógeno	NO _x	µg/m ³
14	Ozono	O ₃	µg/m ³
20	Tolueno	TOL	µg/m ³
30	Benceno	BEN	µg/m ³
37	Meta xileno	MXY	µg/m ³

Fuente: Elaboración propia

Por otro lado, las estaciones de medición de los parámetros del clima son Baztan Irurita, Aranguren Ilundain, Pamplona Aeropuerto, Cadreita y Bárdenas Reales, y están repartidas por el territorio como muestra la figura:



*Figura 2. Localización de las estaciones meteorológicas.
Fuente: elaboración propia con PowerBI*

Estas estaciones medirán precipitaciones a nivel diario, velocidad media del viento y temperatura media. También nos dan información sobre la orografía, con una altura media de 355m sobre el nivel del mar. Siendo la altura máxima Aranguren Ilundain con 572 m y la mínima Baztan Irurita con 183 m.

Para relacionarlas entre ellas (contaminación y clima) se utiliza un procedimiento descrito en el capítulo de Anexos y código desarrollado.

2.1. Modelo Económico

- Formulación del problema matemático:

Se conocen el número de estaciones de medición que describen los niveles diarios de contaminación a nivel local para cada una de las partículas contaminantes, así como los factores que pudieran influir en la variación de dichos niveles de contaminación.

Se desea construir un modelo econométrico que permita evaluar el impacto de dichos factores, como pueden ser la presencia del polvo Sahariano, sobre los niveles de contaminación a nivel local. Para ello ha sido necesario definir los siguientes conjuntos e índices:

- P : Conjunto de partículas contaminantes medidas en la Comunidad Foral de Navarra. El índice p hace referencia a cada una de las partículas del conjunto P tal que $p \in P$.
- E : Conjunto que contiene las estaciones de medición de calidad del aire de la Comunidad Foral de Navarra. El índice e describe cada una de las estaciones de este conjunto tal que $e \in E$.
- E_p : Subconjunto del conjunto E utilizado para describir el conjunto de estaciones de E con capacidad para medir cada una de las partículas p , tal que $E_p \subseteq E$.

Una vez definidos estos conjuntos e índices, definimos el modelo de multi regresión lineal utilizado en el análisis:

$$Z = \ln(C_{pe}) = \beta_{0p} + \beta_{Xpe} * X_{pe} + \beta_{Ype} * Y_{pe} + \varepsilon_{pe} \quad p \in P, e \in E_p$$

Donde C representa la concentración media diaria en la estación $e \in E_p$ para la partícula $p \in P$. Por otro lado, Z representa el logaritmo neperiano de la variable dependiente C . Finalmente, β_{0pe} actúa como término independiente del modelo y ε_{pe} recoge los residuos del modelo.

El vector X está compuesto por el conjunto de factores del estudio que se representan mediante variables de tipo lineal, para el caso concreto del estudio se trata de los factores meteorológicos, distinguimos los siguientes:

- T_{media} → Temperatura media medidas en grados Celsius [°C].
- $V_{vientoMed}$ → Velocidad media del viento medida en metros por segundo [m/s].
- $Prec$ → Precipitaciones medidas en milímetros [mm].

Los coeficientes que acompañan a las variables dependientes del vector X , es decir, las β_{Xpe} son los impactos marginales de dichas variables sobre la variable dependiente C_{pe} (e.g. un

incremento en una unidad de la variable $Prec$ supondrá un incremento de $\beta_{Prec_{pe}}$ % en la concentración de la partícula p medida en la estación e : C_{pe}).

Por otro lado, el vector Y se compone de las variables del estudio de carácter dicotómico, esto es, aquellas variables que únicamente toman valor 1 o 0. Podemos distinguir las siguientes:

- *Lockdown* → Periodo de confinamiento doméstico como variable dicotómica que toma valor 1 si el día corresponde con un día del confinamiento. Los días comprendidos en este variable son los correspondientes a la llamada “Primera Ola” de la fase del virus en España, entre los días 14 de Marzo y 20 de Junio de 2020, un total de 99 días. El factor Lockdown pretende estudiar el impacto vivido durante la pandemia en los días que fue efectivo el estado de alarma y se paró por completo la movilidad, y la industria (no esencial). Este parámetro en cierta forma puede entenderse como el impacto en la calidad del aire y como dañamos al medio ambiente por el efecto de la actividad humana cotidiana (desplazamientos, actividades de ocio o actividades profesionales).
- *Saturday, Sunday, Holiday* → Sábado, Domingo y festivo como variable dicotómica. Toman valor 1 cuando el día en cuestión es Sábado, Domingo o festivo respectivamente y valor 0 en el caso contrario. Se han determinado estas tres variables ya que podemos suponer que estos días se reduce el tráfico de personas en vehículo privado por toda la región y también disminuye la actividad industrial.
- *SaharanDust* → Día con presencia de nube de polvo del Sahara como variable dicotómica. Esta variable toma valor 1 los días anunciados por la autoridad competente como días donde se detecta polvo Sahariano en la zona norte de España. El polvo del Sahara es un fenómeno que ocurre cuando grandes rachas de viento provenientes del continente Africano arrastran arena de las extensiones desérticas de Marruecos y quedan flotando en el ambiente pudiendo provocar problemas respiratorios en los habitantes de las regiones afectadas.
- *CombBiomasa* → Día con Combustión Biomasa como variable dicotómica. Esta variable toma valor 1 los días que se combustionó biomasa en la generación de energía eléctrica

en la región norte de España. Este parámetro se modela en este estudio ya que es un tipo de recurso que se clasifica como renovable al no consumir un recurso finito, se trata de utilizar residuos como combustibles para la generación de energía, evitando así la quema de, por ejemplo, carbón, reduciendo emisiones de algunos gases contaminantes como el CO₂. No obstante, este método no es tan limpio como puede parecer, quemando residuos también se emiten partículas contaminantes al ambiente que pueden afectar en mayor o menor medida a la salud humana.

Los coeficientes que acompañan a las variables del vector Y , β_{Ype} representan el impacto marginal en las concentraciones en el caso de que suceda o no el evento que recoge cada una de las variables. (e.g. si la variable Lockdown toma valor 1, la concentración de la partícula p medida en la estación e sufrirá una variación de $\beta_{\text{Lockdown}pe}$ %).

Capítulo 3. Aplicación de la Metodología

En este capítulo se exponen la aplicación de la metodología propuesta en el capítulo 2 en la Comunidad Foral de Navarra. En una primera parte, se detalla el proceso de recogida y, tratamiento de los datos y pasos seguidos en el análisis, mientras que en una segunda parte trata el análisis de los resultados, por factor y por partícula contaminante.

3.1. Recogida y tratamiento de los datos

El proceso de recopilación y tratamiento de datos ha seguido 5 fases en las que ha sido necesario el uso del software MATLAB por la manejabilidad que ofrece tratando grandes volúmenes de datos (ya que los datos se encuentran a nivel horario para cada partícula y estación). Siguiendo elaboración de este proyecto y que se muestran en la siguiente figura:

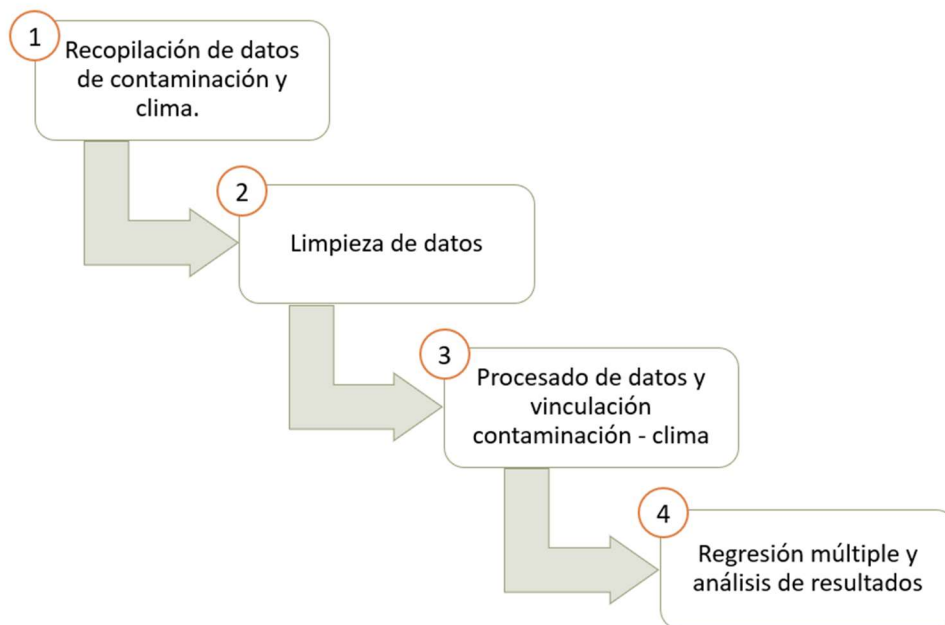


Figura 3. Fases del proyecto.

Fuente: Elaboración propia

En una primera fase se recopilan los datos disponibles en la web del Gobierno de Navarra con respecto a la calidad del aire y los datos proporcionados por la AEMET para la provincia de Navarra. Ambos para el periodo de estudio 2015-2021. Para la descarga de los datos de contaminación, se debe acceder al portal de Navarra al catálogo de datos y descargar cada carpeta (una por cada año) conteniendo un fichero por cada estación en formato *.csv*. Estos ficheros correspondientes a la calidad del aire se guardan en una carpeta en el equipo local creando así la BBDD (Gobierno de Navarra).

Respecto a la descarga de los datos meteorológicos, se debe acceder a la web de la AEMET y solicitar previamente una clave API para descargar los archivos en formato *JSON*. En lugar de descargar año a año como ocurre con los datos de contaminación, en este caso es necesario descargar el periodo de estudio estación a estación (AEMET).

Finalmente, para la descarga de los datos del polvo del Sahara y de la Combustión de Biomasa ha sido necesario acceder a los informes de episodios naturales año por año proporcionados por el ministerio para la transición ecológica español (Ministerio para la transición ecológica y el reto demográfico). La descarga consiste en un informe en formato PDF de cada año de donde ha sido necesario recoger a mano los datos deseados.

La segunda fase tras la recopilación de los datos es la limpieza de datos, que consiste en realizar una serie de programas que permitan realizar tareas como: unificar los nombres de las magnitudes en torno a un mismo criterio para su fácil acceso y agregación (e.g. algunas tablas pueden hacer referencia a la magnitud del ozono como “O₃” o como “OZONO”), unificar los criterios coma-punto como separador decimal, tratamiento de datos faltantes (i.e. celdas en blanco) y deshacer las variables en neperiano entre otros ajustes menores.

Por otro lado, previo a la vinculación de los datos de contaminación y clima, ha sido necesario cierta transformación del formato de los datos: trasponer los valores de tal forma que cada línea de la base de datos represente un día para una estación de medición y para una magnitud en concreto, mientras que cada columna de la base de datos representa cada una de las 24 horas del

día. Esto se debe a que la estructura de datos original dedica cada una de las filas para una hora de un día concreto, y cada una de las columnas con las observaciones de concentraciones de las distintas partículas contaminantes que se miden en la estación en concreto. Se transforma al formato de Madrid para aprovechar sinergias y funciones recicladas y optimizar así el trabajo.

A continuación, se llevó a cabo la tercera fase: tratamiento de los datos. Esta fase ha consistido principalmente en vincular para cada hora y para cada día los datos de contaminación con los datos de climatología disponibles para poder relacionarlos posteriormente mediante una regresión. Tras esta cuarta fase obtenemos ya una tabla para cada contaminante y estación, vinculada con todas las variables independientes y factores que queremos evaluar.

En la cuarta fase, la última del proyecto, se han utilizado las tablas obtenidas para hacer una regresión multifactorial con la formulación expuesta anteriormente en el Capítulo 2 con la ayuda del software R y RStudio. Tras una primera regresión, se tratan las tablas del resultado manteniendo únicamente los factores cuyos coeficientes resultan estadísticamente significativos para así volver a ejecutar el modelo eliminando ruido y afinando los resultados.

Finalmente, con los datos obtenidos de la regresión, se ha utilizado el software de Microsoft PowerBI para visualizarlos y obtener los resultados y conclusiones separados a nivel local por cada una de las estaciones.

3.2. Resultados

En esta sección se muestran, en un primer, lugar los resultados globales de la regresión (tablas de a continuación). En un segundo lugar, se lleva a cabo un análisis del efecto de cada uno de los factores aislados del estudio en promedio y su impacto de manera general sobre el conjunto de los contaminantes. Finalmente aterriza en un análisis partícula a partícula y sus sensibilidades ante las distintas variables independientes en las diferentes regiones del estudio de manera visual.

En las tablas 2 y 3 de los resultados de las regresiones se muestran primero las β o coeficientes de la regresión estimados por el modelo acompañados por los respectivos p – valor que indican el nivel de significación de las estimaciones. Como criterio, se ha tomado que todas aquellas

estimaciones cuyos p - valor sean inferiores a 0,05 serán considerados como significativos y por tanto interpretable.

En este caso, y como se explica en la metodología, al estar la variable del modelo en logaritmo, la interpretación de los coeficientes representa la sensibilidad de la variable independiente ante los factores de forma porcentual. (e.g. si la β que acompaña al factor Lockdown para los modelos que explican la concentración del Benceno toma un valor de media -0,282 entre todas las estaciones, y además tiene un p – valor menor que 0,05; interpretamos que la concentración del Benceno decrecerá en un 28,2% de media si el día observado es un día de confinamiento).

Los resultados completos se muestran en las dos tablas de a continuación:

Tabla 2.
Resultado de la regresión por contaminantes (Parte 1/2). Se muestra el promedio de los coeficientes estimados para cada una de las variables independientes, así como su p-valor para su nivel de significación.

VARIABLES INDEPENDIENTES	BEN		CO		MXY		NO		NO2	
	Coef. Beta	p-valor	Coef. Beta	p-valor	Coef. Beta	p-valor	Coef. Beta	p-valor	Coef. Beta	p-valor
Lockdown	-0.282	0.000	-0.091	0.012	-1.188	0.000	-0.583	0.000	-0.580	0.000
Temperatura Media	-0.024	0.000	-0.003	0.065	0.002	0.568	-0.032	0.000	-0.029	0.000
Velocidad Media del Viento	-0.099	0.000	-0.025	0.000	-0.122	0.000	-0.154	0.000	-0.121	0.000
Sábado	-0.147	0.001			-0.262	0.001	-0.305	0.000	-0.202	0.000
Domingo	-0.184	0.000			-0.525	0.000	-0.335	0.007	-0.296	0.000
Festivos							-0.363	0.002	-0.304	0.000
Polvo del Sahara	0.211	0.000			0.243	0.003	-0.148	0.006	0.165	0.002
Precipitaciones			-0.006	0.002			-0.009	0.015	-0.002	0.005
Combustión de Biomasa	0.234	0.000							0.187	0.001

Fuente: Elaboración propia

Tabla 3.

Resultado de la regresión por contaminantes (Parte 2/2). Se muestra el promedio de los coeficientes estimados para cada una de las variables independientes, así como su p-valor para su nivel de significación.

Variables independientes	NOx		O3		PM10		SO2		TOL		PM2.5	
	Coef. Beta	p-valor	Coef. Beta	p-valor	Coef. Beta	p-valor	Coef. Beta	p-valor	Coef. Beta	p-valor	Coef. Beta	p-valor
Lockdown	-0.516	0.000	0.262	0.001	-0.202	0.000	-0.176	0.002	-0.466	0.000	0.218	0.005
Temperatura Media	-0.030	0.000	0.022	0.000	0.012	0.004	0.009	0.000			-0.025	0.000
Velocidad Media del Viento	-0.129	0.000	0.126	0.000	-0.045	0.000	-0.008	0.000			-0.055	0.000
Sábado	-0.207	0.000			-0.094	0.005			-0.269	0.002		
Domingo	-0.309	0.000			-0.141	0.001			-0.433	0.000		
Festivos	-0.279	0.003			-0.220	0.005						
Polvo del Sahara	0.103	0.002	0.180	0.000	0.403	0.000			0.496	0.000	0.515	0.000
Precipitaciones	-0.008	0.000	0.005	0.000	-0.012	0.000	-0.006	0.000			-0.020	0.002
Combustión de Biomasa	0.188	0.002	0.255	0.001	0.204	0.000	0.205	0.000				

Fuente: Elaboración propia

3.2.1. Análisis de los factores

Tras la regresión lineal, analizaremos primero de forma conjunta el resultado promedio del impacto de los factores sobre todas concentraciones de las partículas contaminantes del estudio, la figura que se muestra a continuación recoge un resumen de estos:

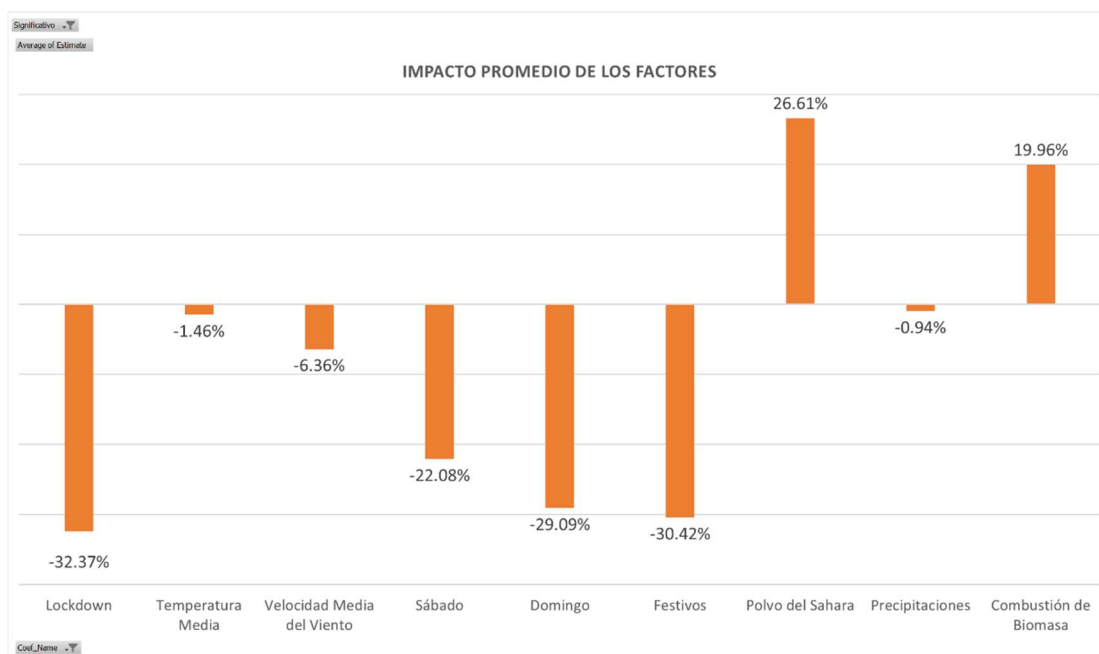


Figura 4. Impacto promedio de los factores.

Fuente: Elaboración propia

Como muestra la figura, las variables con impacto positivo en la concentración de contaminantes son la Combustión Biomasa (*cada día que se produce la combustión de biomasa la concentración de partículas contaminantes en el aire aumenta en un ~20% de media*) y el Polvo del Sahara (*los días con presencia de polvo proveniente del Sahara, la concentración de partículas contaminantes en el aire aumenta en un ~27% de media*).

Por otro lado, el resto de los factores tienen impacto negativo en la concentración de contaminantes: los días Sábado, Domingo y Festivos suponen una reducción del ~22, 29 y 30% en la concentración de partículas contaminantes de media respectivamente. Los días pertenecientes al periodo de Lockdown durante el estado de alarma vivido en España en 2020, suponen una reducción de las concentraciones de en torno al 32%, siendo este el impacto más notable de todos los factores.

Con respecto a los factores medioambientales observamos que tienen una relación inversamente proporcional a la concentración. Las precipitaciones afectan a la concentración en un -0,94% por cada mm de precipitación adicional, la Temperatura media en un -1,46% por cada grado Celsius (°C) adicional y la velocidad media del viento en un -6,36% por cada m/s adicional.

○ 3.2.1. *Análisis de los contaminantes*

En esta segunda parte, se analizan y presentan los resultados del estudio a nivel partícula de aquellas que han resultado tener un impacto estadísticamente significativo en al menos alguno de los factores del estudio:

▪ **Partícula contaminante PM₁₀**

La Temperatura tiene un efecto sobre la concentración del PM₁₀ directamente proporcional de media 1.39%. Teniendo en cuenta que esta sensibilidad hace referencia a cada incremento de grado centígrado, los días más calurosos, donde se registran temperaturas de más de 40°C (AEMET), suponen una diferencia de aproximadamente 20°C con la temperatura media anual, es decir, la concentración de PM₁₀ previsiblemente incrementará en ~28%.

El efecto de los otros dos factores climatológicos sobre la concentración de las partículas en suspensión PM₁₀ tienen una relación inversamente proporcional. La velocidad media del viento con una sensibilidad media de -5,3% por cada m/s adicional y las precipitaciones con un -1,36% por cada mm. La distribución por la comunidad Foral de Navarra se muestra en las figuras de a continuación:

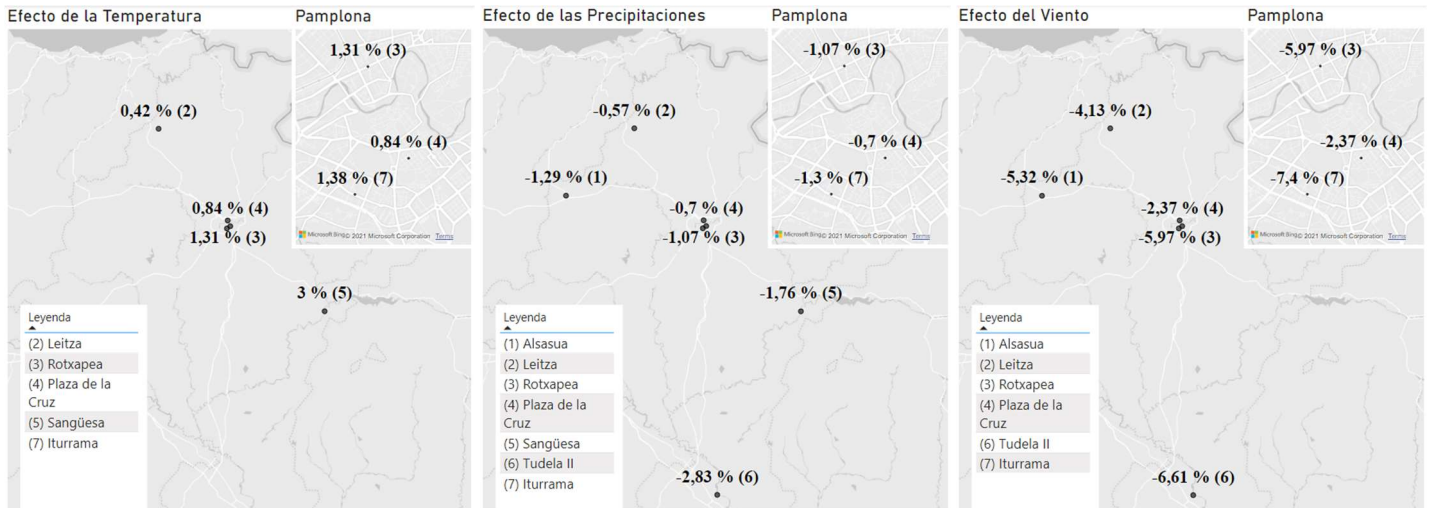


Figura 5. Efecto de los factores climatológicos sobre la concentración de PM10.

Fuente: Elaboración propia con PowerBI

Respecto al efecto de los sábados, domingos y festivos sobre la concentración de esta partícula, obtenemos un resultado esperado, disminuye en aproximadamente 9, 16 y 22 % respectivamente, aunque no de manera tan generalizada como en el caso de las condiciones meteorológicas. La drástica reducción del transporte estos días concretos es presumiblemente la causa de esta disminución en la concentración. También se puede apreciar como esta reducción es más drástica en Tudela, donde hay menos población que en la ciudad de Pamplona, esto se podría explicar por el hecho de que en las zonas más urbanizadas seguirá habiendo transporte los días no laborables por tener más centros de ocio y, sobre todo, más población. La distribución de los impactos por estación se muestra a continuación:

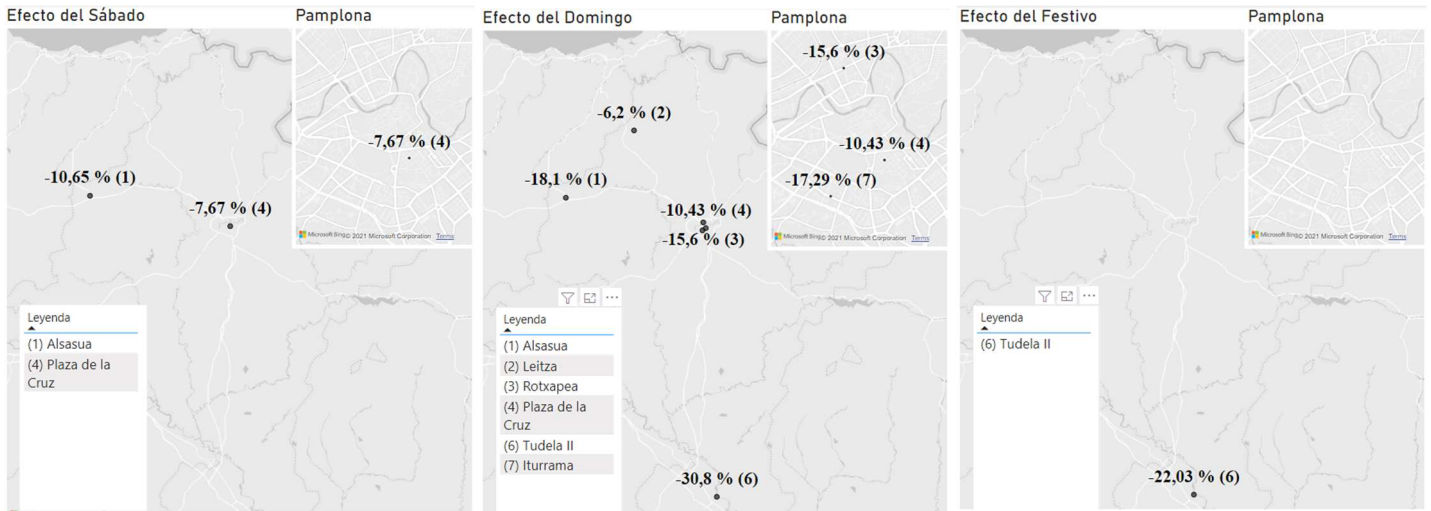


Figura 6. Efecto de los días no laborables sobre la partícula PM10.

Fuente: Elaboración Propia con Power BI

Con respecto al efecto del confinamiento, el modelo muestra un impacto parecido al que mostraba con los días no laborables con una reducción en la concentración aún más acusada de entorno al 27%.

La combustión de biomasa supone un incremento significativo en todas las regiones para el contaminante PM₁₀. Esta partícula aumenta de media más de un 20% los días que se combustiona biomasa para la generación de electricidad, siendo la zona del norte de la provincia la menos afectada como se muestra en la figura.

Finalmente, los días con presencia de polvo sahariano, suponen el mayor impacto sobre esta partícula. La partícula PM₁₀ se define como componentes sólidos o líquidos que flotan en el aire, esta definición encaja precisamente con la nube de polvo del Sahara que está formada por el levantamiento de pequeñas partículas de arena que son transportadas por el viento recorriendo largas distancias. El incremento promedio de la concentración de PM₁₀ es del 46%, teniendo el impacto más acusado en Tudela con más de un 80%. También se puede apreciar una ligera tendencia decreciente en dirección Suroeste – Noreste, debida presumiblemente a la trayectoria del polvo al proceder del continente africano.

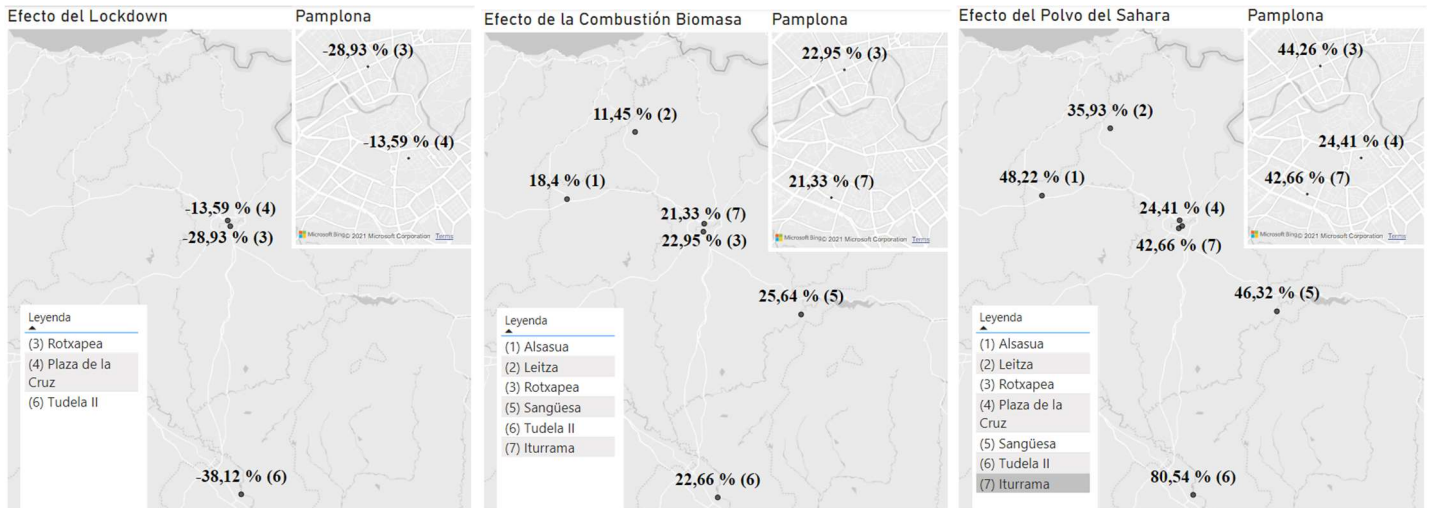


Figura 7. Impacto del Lockdown, Combustión Biomasa y Polvo del Sahara sobre el PM10

Fuente: Elaboración propia con PowerBI

▪ **Partícula contaminante PM_{2.5}.**

Para las partículas en suspensión de menor tamaño, PM_{2.5}, el modelo solo arroja sensibilidades significativas para la estación de medición de Tudela y para los factores relacionados con la climatología y con el confinamiento. Esto indica que las concentraciones de este contaminante oscilan de manera más independiente a los factores aquí propuestos. Los resultados para donde sí muestra una sensibilidad estadísticamente significativa se corresponden con las tendencias del apartado anterior de PM₁₀ se muestran en la siguiente tabla:

Tabla 4.
Impacto de los factores del estudio sobre la partícula PM_{2.5}.

Estación	Temperatura Media	Velocidad Media del Viento	Precipitaciones	Lockdown
Tudela II	-4.91%	-11.06%	-3.95%	43.65%

Fuente: Elaboración propia

▪ **Partícula Contaminante NO₂**

La temperatura media diaria y la velocidad media del viento hacen que las concentraciones de dióxido de nitrógeno se vean afectadas en $\frac{-3,26\%}{^{\circ}C}$ y $\frac{-13,88\%}{m/s}$ en promedio. Mientras que las precipitaciones parecen no diluir de manera relevante esta partícula del ambiente.

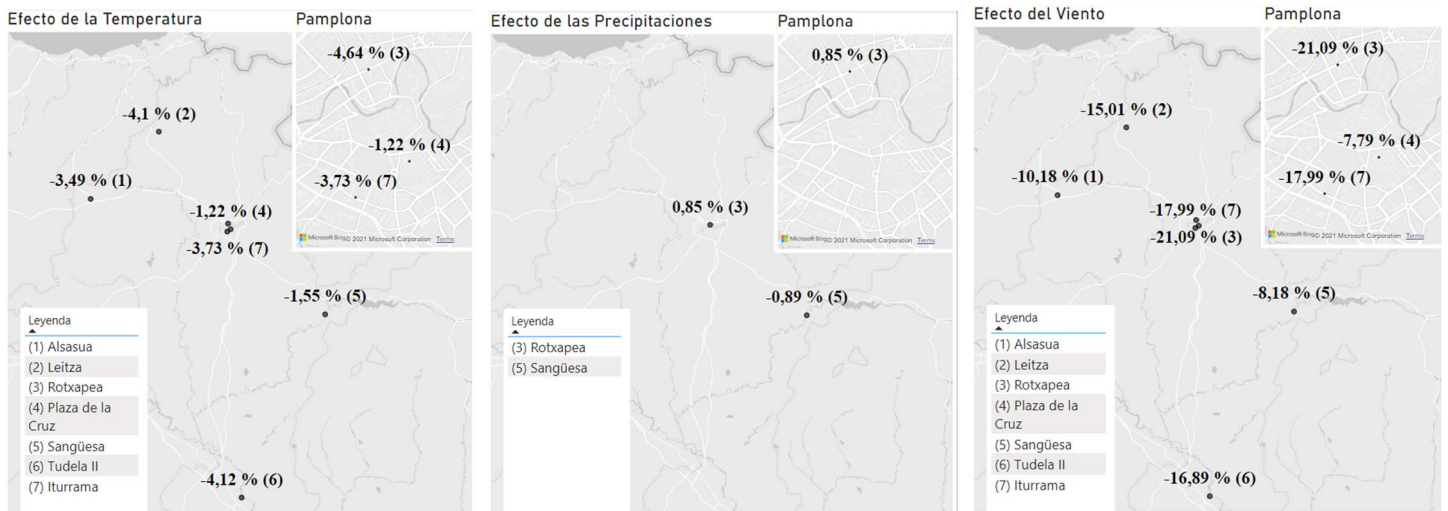


Figura 8. Impacto de las condiciones climatológicas sobre la concentración de NO₂.

Fuente: Elaboración propia con Power BI

El efecto de los días no laborables, al igual que pasaba con las partículas en suspensión es bastante relevante, con impactos de -20, -34 y -36 % para sábados, domingos y festivos respectivamente. Al ser el NO₂ un contaminante característico de las emisiones de los vehículos de combustión diésel, este efecto es el esperado por el menor flujo de personas en estos días frente a los días laborables.

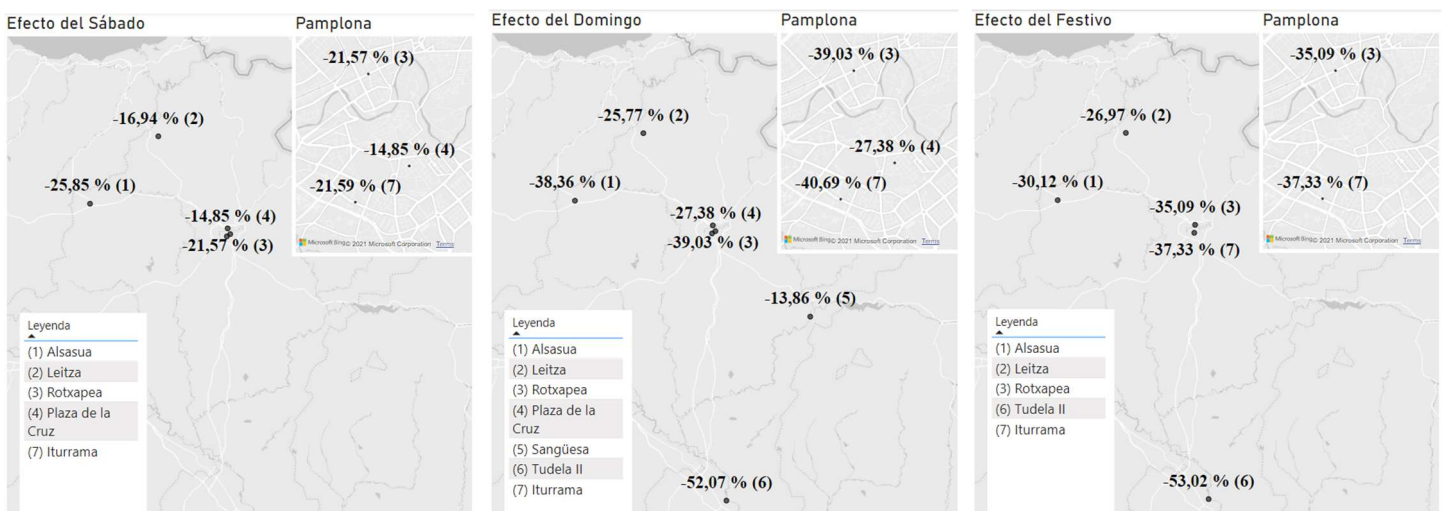


Figura 9. Impacto de los días no laborables sobre la concentración de NO₂.

Fuente: Elaboración propia con Power BI

El factor *Lockdown*, supone un descenso de la concentración de dióxido de Nitrógeno promedio de entre -30 y -71% alcanzando máximos en las zonas con más densidad de población,

en la ciudad de Pamplona. Es entendible por el transporte prácticamente nulo de las personas y el cese de la industria no esencial durante esos días.

Por otro lado, la combustión de biomasa y el polvo del Sahara hacen que aumente la presencia del NO₂ en el aire en 23 y 17% respectivamente.

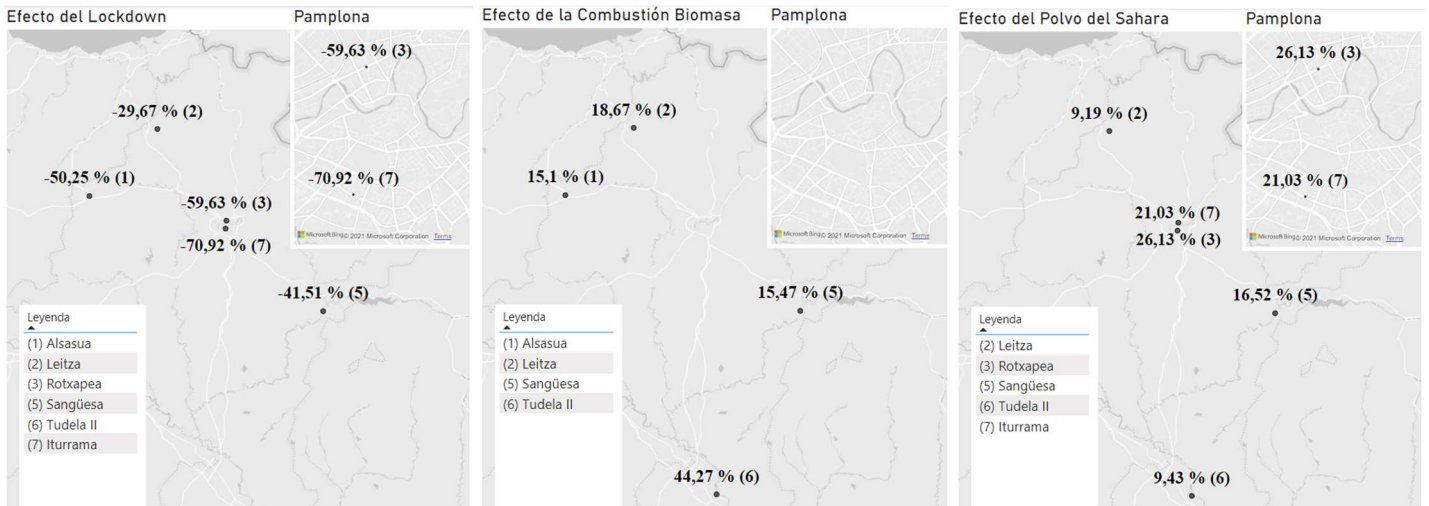


Figura 10. Impacto del Lockdown, la combustión de biomasa y el polvo del Sahara en la concentración de NO₂.

Fuente: Elaboración propia con Power BI

▪ **Partícula Contaminante NO**

Los valores para el monóxido de Nitrógeno son muy similares a los del NO₂ por proceder en gran medida de las mismas fuentes. El efecto de la temperatura es del -3.7% por °C y el del viento de -18% por m/s. Siendo el de las precipitaciones de en torno al -1% por mm.

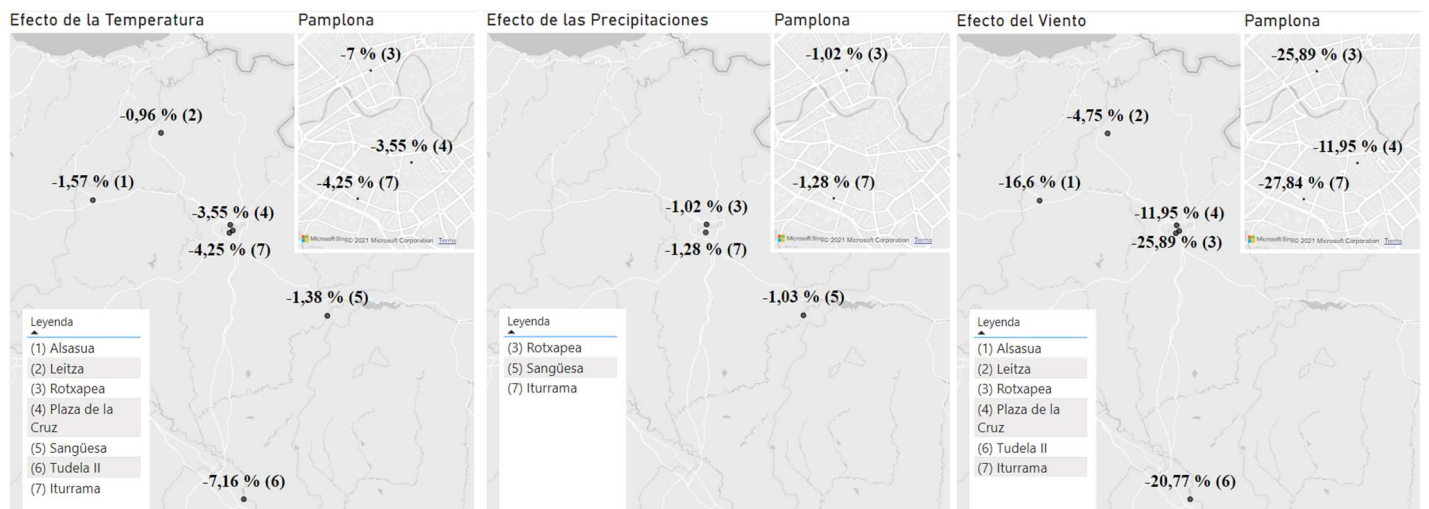


Figura 11. Impacto de los factores medioambientales sobre la partícula NO.

Fuente: Elaboración propia con Power BI

Para los días no laborables, el NO se comporta de manera muy similar al NO₂ con grandes descensos en su concentración. Si cabe, más acusado aún en los días sábado y domingo con sensibilidades de -30% y -44% de media.

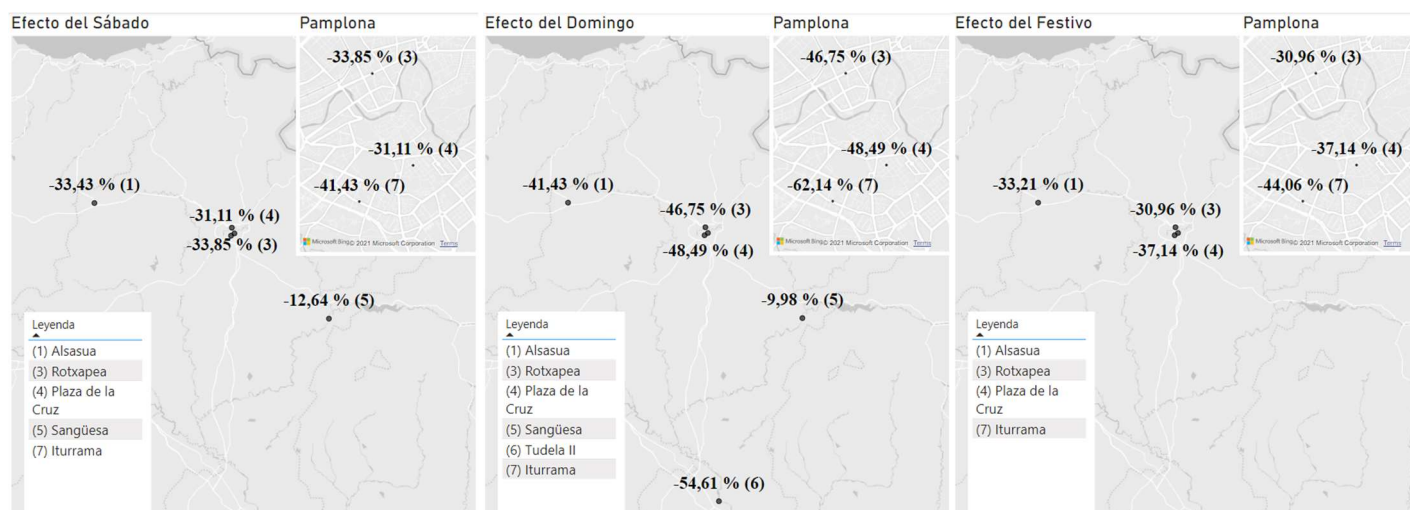


Figura 12. Impacto de los días no laborables sobre la concentración de NO.

Fuente: Elaboración propia con Power BI

A diferencia del NO₂, la concentración de NO no se muestra afectada ni por la combustión de biomasa ni por la presencia de polvo del Sahara, según el modelo. No obstante, el efecto del confinamiento altera su concentración de manera negativa por lo general, como mostrado en la figura:

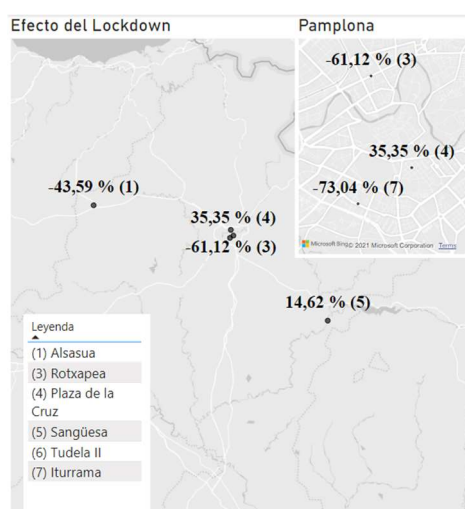


Figura 13. Impacto del confinamiento sobre la concentración de NO.

Fuente: Elaboración propia con Power BI

▪ **Partícula Contaminante NO_x**

El NO_x, es el tercer y último tipo de óxidos de Nitrógeno estudiados, y al provenir de fuentes similares a los otros dos, se pueden esperar efectos parecidos como se muestran a continuación:

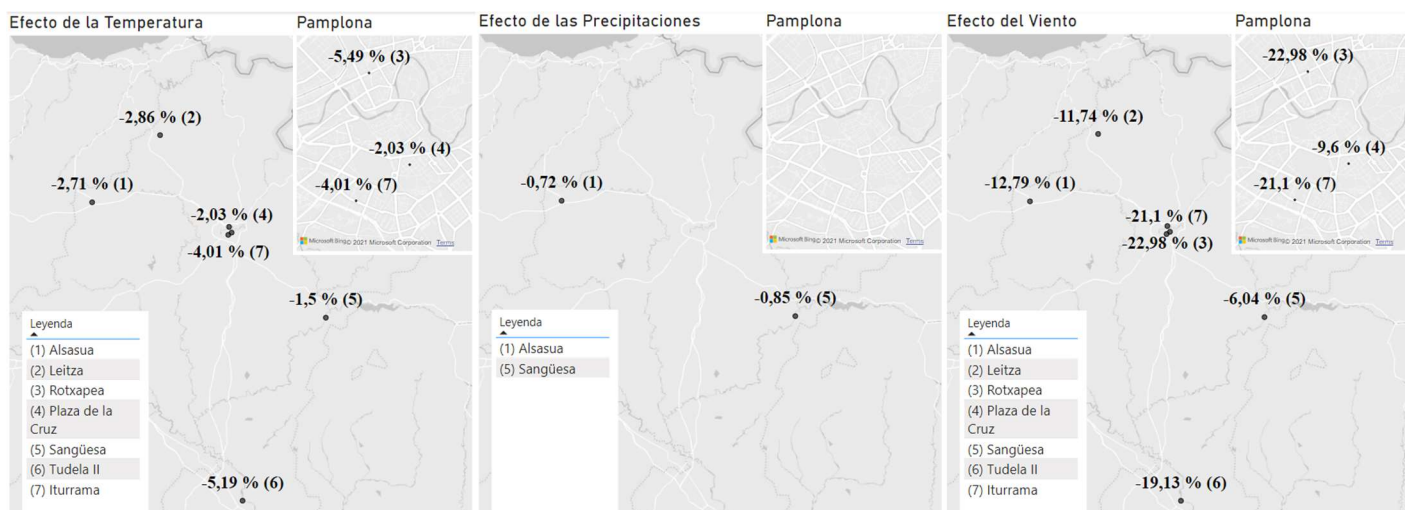


Figura 14. Impacto de los factores climatológicos sobre la concentración de NOx.

Fuente: Elaboración propia con Power BI

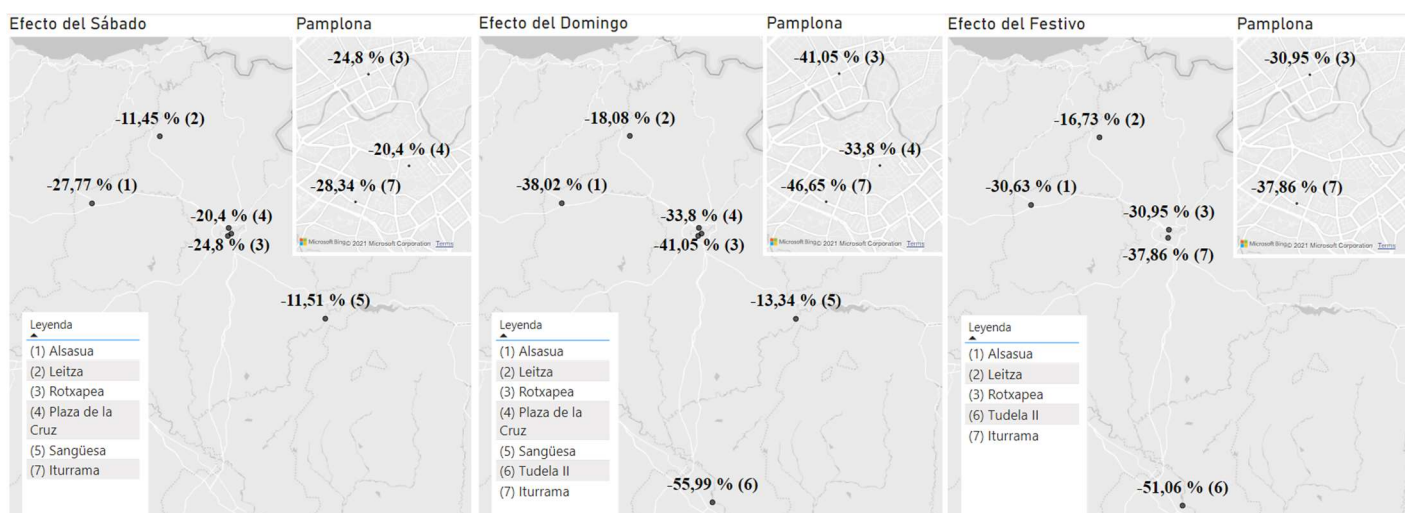


Figura 15. Impacto de los días no laborables sobre la concentración de NOx.

Fuente: Elaboración propia con Power BI

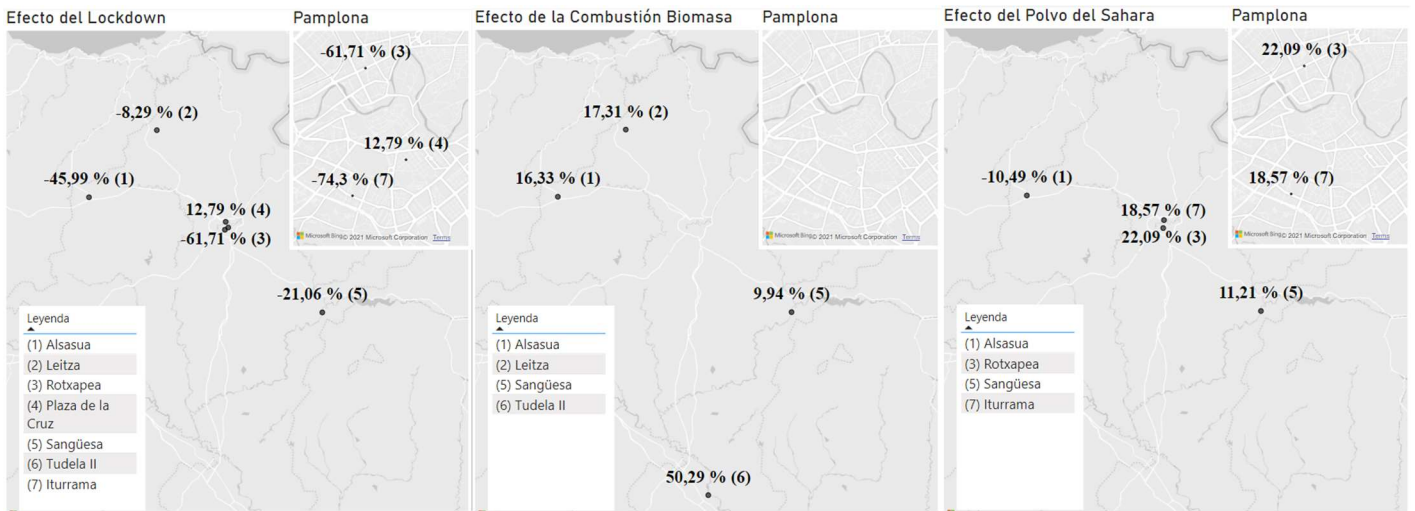


Figura 16. Impacto del Lockdown, la Combustión de Biomasa y el Polvo del Sahara sobre la concentración de NOx.

Fuente: Elaboración propia con Power BI

▪ **Partícula contaminante O₃**

El Ozono troposférico presenta un comportamiento peculiar al no provenir de una fuente directa en la mayoría de los casos. En su lugar, surge de la reacción entre otros contaminantes en presencia de radiación solar. Se puede asumir, que la relación entre luz solar y temperatura media es directamente proporcional, y por esta relación los resultados del factor temperatura tienen sentido:

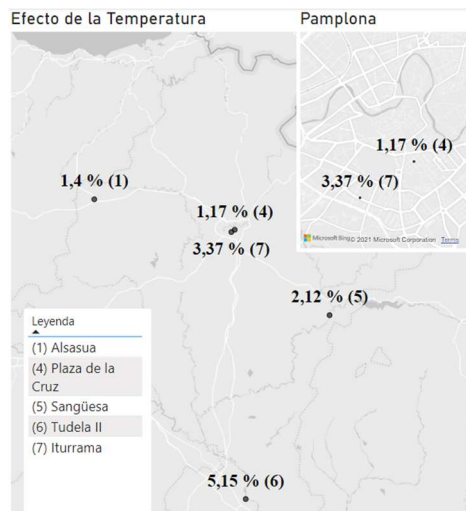


Figura 17. Impacto de la temperatura sobre la concentración de O₃.

Fuente: Elaboración propia con Power BI

El Ozono aumenta su concentración con la temperatura de manera más o menos uniforme por toda la provincia en una media de 2,64% por cada grado centígrado.

Aunque a priori, se podría pensar que al ser el ozono fruto de una reacción entre óxidos de nitrógeno y componentes orgánicos volátiles, mantendría una tendencia proporcional a la que llevan estas partículas según la presencia de los factores del estudio, no obstante, los efectos de factores como la velocidad del viento o el Lockdown resultan contra intuitivos y no son evidentes en vista del resultado obtenido

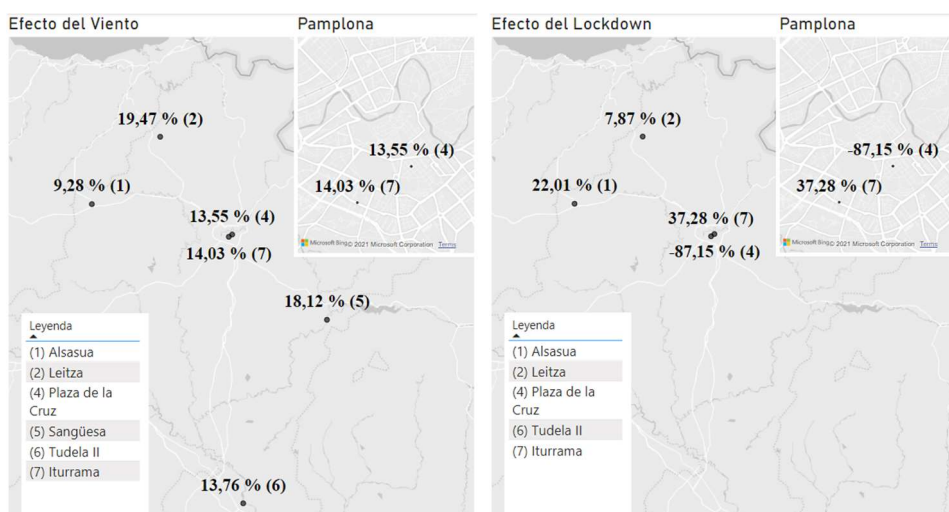


Figura 18. Efecto del viento y el confinamiento en la concentración de O3.

Fuente: Elaboración propia con Power BI

El viento, en lugar de depurar el ambiente y reducir la concentración de O₃ como hace con el resto de los contaminantes, hace que aumente cerca de un 15% de media, relativamente uniforme por todas las regiones del estudio. Puede deberse a su complejo proceso de formación o a que el viento transporte una concentración de O₃ elevada desde su procedencia. De la misma manera el Lockdown, solo hace que se reduzca la concentración de O₃ en la zona central de la ciudad de Pamplona, provocando un aumento en el resto de las zonas de estudio.

Los factores Combustión de Biomasa o Polvo del Sahara tienen un comportamiento esperado de provocar un aumento del ozono troposférico, aunque el modelo sólo arroja resultados estadísticamente significativos para tres de las estaciones de estudio:

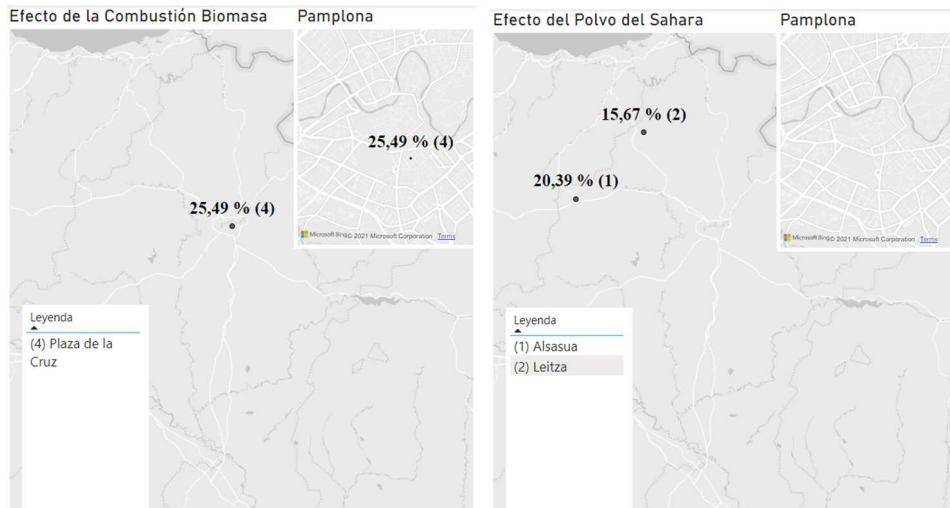


Figura 19. Efecto de la combustión de biomasa y el polvo del Sahara sobre la concentración de O3.

Fuente: Elaboración propia con Power BI

▪ **Partícula Contaminante SO₂**

El dióxido de azufre es una partícula que ha resultado independiente de los factores atmosféricos para el ámbito de estudio con el volumen y granularidad de los datos disponibles.

El modelo señala un decrecimiento generalizado en la concentración de SO₂ en los días de confinamiento y un crecimiento significativo en la estación de Sangüesa para los días de combustión de Biomasa.

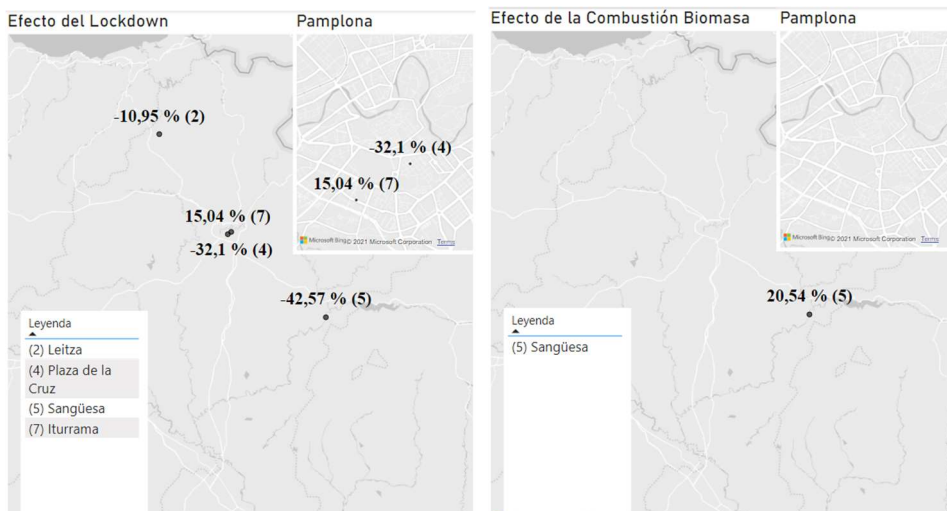


Figura 20. Efecto del Lockdown y de la Combustión de Biomasa sobre el SO2

Fuente: Elaboración propia con Power BI

▪ **Partículas Contaminantes BEN y TOL**

El conjunto de datos presenta una situación particular para las partículas de Benceno y Tolueno. Tras la regresión, observamos que sólo hay factores significativos para los dos contaminantes en la estación de medición de Iturrama y se ven afectados por los factores de sábado, domingo y Lockdown como muestra la tabla:

Tabla 5.

Impacto de los factores sobre las concentraciones de BEN y TOL en la estación de Iturrama.

Partícula	Lockdown	Sábado	Domingo
BEN	-28.16%	-14.74%	-18.37%
TOL	-46.60%	-26.87%	-43.26%
Total	-37.38%	-20.81%	-30.82%

Fuente: Elaboración propia con Power BI

Estos tres factores (Lockdown, sábado y domingo) representan episodios de baja actividad industrial, por lo que es lógico que tengan estos efectos sobre las partículas de Benceno y Tolueno que son contaminantes típicos de ciertas industrias. Para comprender por qué ocurre solo en la estación de medición de Iturrama podemos atender al siguiente mapa:



Figura 21. Localización polígono industrial en Pamplona.

Fuente: Elaboración propia con Google Maps

Observando los alrededores de la estación de medición, se podría explicar el efecto por ser la estación con más proximidad al polígono industrial de Pamplona, entre 2 y 4 km. La localización de este polígono cerca de la ciudad y sobre todo cerca de la zona universitaria pone en riesgo la salud de muchas personas y jóvenes que respiran aire con presencia de estas toxinas los días laborables.

Conclusiones

En este trabajo se ha realizado un análisis del efecto de distintos factores sobre los niveles de contaminación de distintas partículas en la Comunidad Foral de Navarra. A partir de los resultados obtenidos se han establecido las siguientes conclusiones:

1. El objetivo de la investigación ha sido logrado y la hipótesis de investigación validada ya que se ha demostrado que existe una influencia directa entre la actividad económica y la movilidad de las personas con los niveles de contaminación de determinadas partículas contaminantes.
2. En los periodos de festivos y durante el confinamiento establecido por el gobierno central de España para enfrentar la problemática del COVID-19 los niveles de los contaminantes supusieron reducciones medias en el orden del 22, 33, 34 y 38 % para los sábados, domingos, festivos y periodo de confinamiento. Con especial incidencia en las partículas de óxidos de nitrógeno, NO, NO₂ y NO_x con unos descensos promedios del orden de 58, 58 y 52 % respectivamente.
3. La Combustión de Biomasa, fruto de la generación de energía eléctrica para satisfacer la demanda energética natural de la vida de las personas, supone un 22,42% de aumento promedio en los contaminantes de la Comunidad Foral de Navarra.
4. Durante el período de confinamiento el ozono, presentó incrementos en los niveles, para todas las localidades de Navarra en el orden de un 31,44%. Este comportamiento se debe al complejo proceso de formación de este contaminante, que se vio influenciado inversamente por los cambios en los niveles de NO y NO₂.
5. Las variables climatológicas viento y lluvia, ayudan a reducir la presencia de contaminantes en el aire, pero en muy baja proporción. En este sentido los resultados muestran que se logran disminuir la presencia de los niveles de contaminación en un 7,31% por m/s y de 1,04% por cada mm respectivamente. De nuevo, a excepción del

Ozono, que experimenta aumentos del 14,70% por cada metro por segundo de velocidad media del viento y un aumento del 0,54% por cada mm de precipitación registrada. Esto demuestra que el clima en España con condiciones climáticas anticiclónicas típicas del mediterráneo, dificultan la eliminación y depuración del medio ambiente, convirtiendo a países como el nuestro en lugares con predisposición a una mala calidad del aire.

6. La presencia del polvo Sahariano debido a la situación geográfica de España respecto a la proximidad con el continente Africano impacta en la concentración de los contaminantes del aire en un 28,92% los días con presencia de este, afectando especialmente a las partículas PM_{10} y $PM_{2.5}$ con aumentos de en torno a 40 y 50 % respectivamente.
7. Entre las localidades de la Comunidad Foral de Navarra donde más destacan los niveles de contaminación son Felisa Munarriz y Olite con una media de concentración de partículas contaminantes de 34 y 21 $\mu\text{g} / \text{m}^3$. Las localidades que gozan con mejor calidad del aire son Leitza e Iturrama, ambas con concentraciones promedio de los contaminantes medidos de 13 $\mu\text{g} / \text{m}^3$.

Referencias Bibliográficas

- AEMET. (n.d.). *AEMET Open Data*. Retrieved from <https://opendata.aemet.es/centrodedescargas/inicio>
- ATSDR. (2016, 5 6). *Agencia para Sustancias Tóxicas y el Registro de Enfermedades*. Retrieved from Resúmenes de Salud Pública: https://www.atsdr.cdc.gov/es/phs/es_ph3.html#bookmark3
- Betancourt-Odio, M., Martínez-de-Ibarreta, C., Budría-Rodríguez, S., & Wirth, E. (2021). Local Analysis of Air Quality Changes in the Community of Madrid before and during the COVID-19 Induced Lockdown. *Atmosphere*.
- Carceller, X. Q. (2019). Calidad del aire urbano en Europa: retos científicos y de política ambiental. *Revista de Salud Ambiental*, 19, pp. 75-77.
- Gobierno de Navarra. (n.d.). *Catálogo de Datos*. Retrieved from Gobierno Abierto de Navarra: https://gobiernoabierto.navarra.es/es/open-data/datos/lista-catalogo?combine=CALIDAD+DEL+AIRE&term_node_tid_depth=All&tid=All&field_datos_od_formato_nid=All&field_clasif_nubetags_tid=All
- INE. (2019). *Población por sexo, municipios y edad*. INE.
- Lazaro, L., Gonzalez, C., & Escribano, G. (2019). *Los españoles ante el cambio*. Real Instituto Elcano.
- Logan, W. (1953). *Mortality in the London fog incident, 1952*. *Lancet*.
- Manisalidis, I., Stavropoulou, E., Stavropoulos, A., & Bezirtzoglou, E. (2020). *Environmental and health impacts of air pollution: a review*. *Frontiers in public health*. doi:10.3389/fpubh.2020.00014
- Ministerio para la transición ecológica. (2019). *El Gobierno aprueba el I Programa Nacional de Control de la Contaminación Atmosférica*. Madrid: Gabinete de Prensa.
- Ministerio para la transición ecológica y el reto demográfico. (n.d.). Retrieved from Histórico de Informes de Episodios Naturales: <https://www.miteco.gob.es/es/calidad-y-evaluacion-ambiental/temas/atmosfera-y-calidad-del-aire/calidad-del-aire/evaluacion-datos/fuentes-naturales/anuales.aspx>.
- Nunes, R. A., Alvim-Ferraz, M. C., Martins, F. G., Peñuelas, A. L., Durán-Grados, V., Moreno-Gutiérrez, J., ... & Sousa, S. I. (2021). Estimating the health and economic burden of shipping related air pollution in the Iberian Peninsula. *Environment International*, 156, 106763.
- Rafael A.O. Nunes, Maria C.M. Alvim-Ferraz, Fernando G. Martins, Antonio L. Peñuelas, Vanessa Durán-Grados, Juan Moreno-Gutiérrez, Jukka-Pekka Jalkanen, Hanna Hannuniemi, Sofia I.V. Sousa,
- Oyarzún, M. (2010). Contaminación aérea y sus efectos en la salud. *Revista chilena de enfermedades respiratorias* 26.1, 16-25.

- Querol, X. (2008). *Calidad del aire, partículas en suspensión y metales*. SciELO Public Health.
- Watts, J., & Kommenda, N. (2020). *Coronavirus pandemic leading to huge drop in air pollution*. The Guardian. Retrieved from <https://www.theguardian.com>
- WHO Air Pollution. Resource Document. WHO. 2020. Available online: <http://www.who.int/airpollution/en> (accessed on 21 September 2020).
- Zhang, J., Wei, Y., & Fang, Z. (2019). Ozone pollution: a major health hazard worldwide. *Frontiers in immunology*, 10, 2518.

Anexos

Anexo 1. Código desarrollado en Matlab y RStudio.

En este apartado, se detallará la lógica de programación en Matlab seguida y el código desarrollado durante la elaboración de este trabajo. Cada una de las secciones siguientes corresponde con los tres softwares utilizados en este trabajo fin de grado han sido Matlab (mayoritariamente), R y PowerBI.

Trabajo desarrollado en MATLAB

El trabajo realizado en MATLAB representa la gran mayoría del proyecto, ya que alberga las fases 1-4 detalladas en el Capítulo 3

Todo el código se engloba en un Script llamado SCRIPT_PRINCIPAL_NAVARRA.m Desde el cual se accede a un fichero Excel donde se “lee” el periodo de estudio que se desea estudiar y a continuación se ejecutan los dos ejecutables principales: PREPARAR_DATOS_NAVARRA.m y CARGAR_NAVARRA.m

- *PREPARAR_DATOS_NAVARRA.m*

Este ejecutable recoge todos los pasos necesarios para llevar a cabo las Fases 1-3. Nos apoyaremos en el siguiente esquema para guiarnos a través del código:

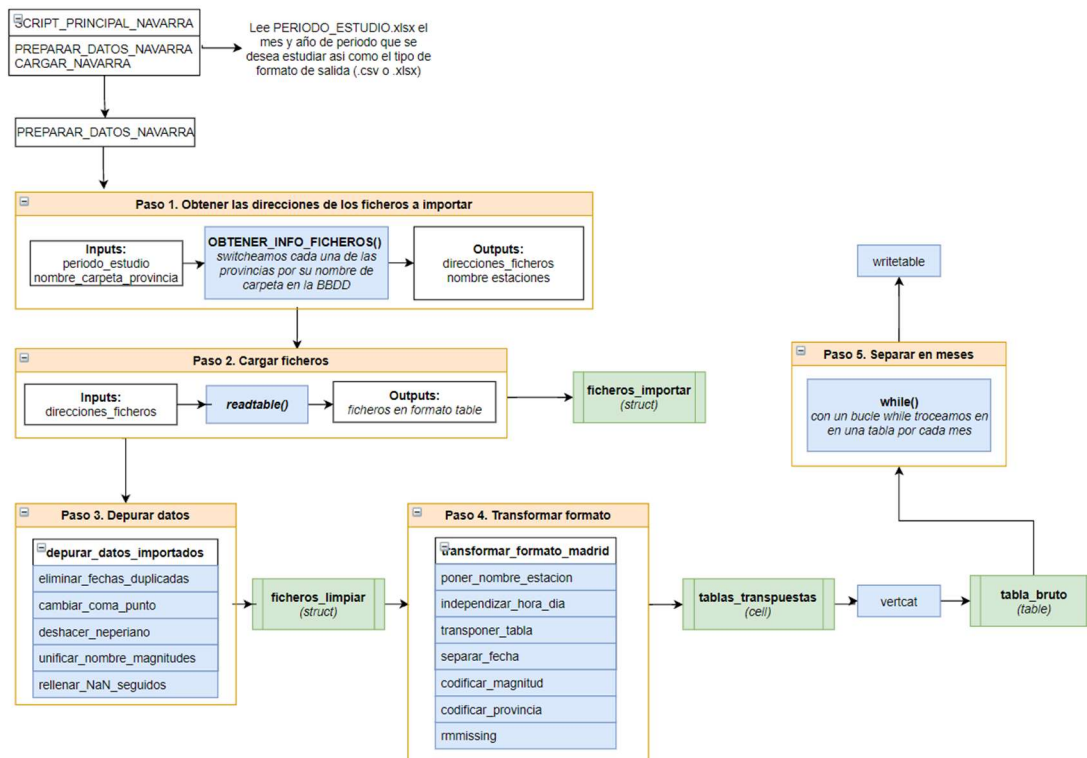


Figura 22. Esquema PREPARAR_DATOS_NAVARRA.

Fuente: Elaboración Propia

Como se aprecia en el esquema de este primer ejecutable, hay 5 secciones principales compuestas por una o más funciones cada una que van resolviendo las necesidades correspondientes en cada punto del código. Veremos paso a paso las secciones y funciones que componen este ejecutable:

➤ **Paso 1: Obtener las direcciones de los ficheros a importar**

En este primer paso recibe como inputs el periodo de estudio y el nombre de la provincia como se muestra en la carpeta de la BBDD y devuelve una estructura con las direcciones de los ficheros y el nombre de la estación correspondiente a cada fichero.

Fields	direcciones	estacion
1	'C:\Users\Gregorio\...	'Alsasua'
2	'C:\Users\Gregorio\...	'Funes'
3	'C:\Users\Gregorio\...	'Iturrama'
4	'C:\Users\Gregorio\...	'Leitza'
5	'C:\Users\Gregorio\...	'Olite'
6	'C:\Users\Gregorio\...	'Plz. de la C...
7	'C:\Users\Gregorio\...	'Rotxapea'
8	'C:\Users\Gregorio\...	'Sangüesa'
9	'C:\Users\Gregorio\...	'Tudela'

Ejemplo 1. Muestra de la estructura : `ficheros_importar`.

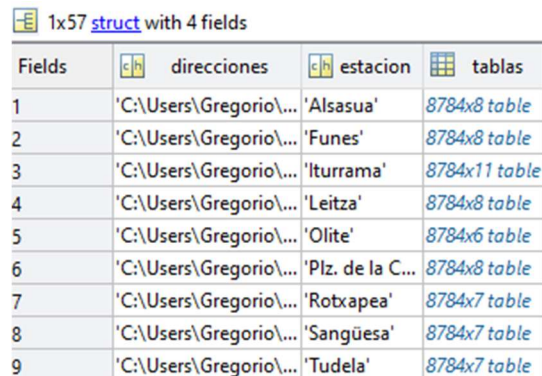
Fuente: Elaboración propia mediante el uso de MATLAB

Para hacer esto posible ha sido necesario la creación de una función: **OBTENER_INFO_FICHEROS.m** donde se ejecutan otras dos funciones:

- **OBTENER_RUTAS_FICHEROS.m** – Primero utiliza el comando `dir()` de MATLAB somos capaces de obtener una lista con todos las carpetas que se encuentran en la BBDD, correspondientes a los distintos años de datos que hay disponibles. Una vez tenemos listadas las carpetas de cada año seleccionamos únicamente las que nos interesan para el periodo de estudio y volvemos a utilizar el comando `dir()`, esta vez dentro de un *for loop*, para extraer las direcciones de todos los ficheros dentro de cada una de las carpetas. Registraremos esas direcciones en el campo “direcciones” de la estructura **ficheros_importar**.
- **OBTENER_NOMBRE_ESTACION.m** – Para obtener el nombre de la estación nos basamos en el nombre de cada fichero, que tiene la forma ‘*EstacionAÑO.csv*’ con el comando `Split()` de Matlab somos capaces de fácilmente extraer el nombre de la estación y incluirlo en el campo “estación” de la estructura **ficheros_importar**.

➤ **Paso 2: Cargar los ficheros en un nuevo campo de la estructura**

En esta sección del ejecutable no ha sido necesario crear funciones propias, utilizando la función `readtable()` de MATLAB cargamos en el campo “tablas” de la estructura **ficheros_importar** todos los ficheros cuyas direcciones hemos extraído anteriormente.



Fields	direcciones	estacion	tablas
1	'C:\Users\Gregorio\...	'Alsasua'	8784x8 table
2	'C:\Users\Gregorio\...	'Funes'	8784x8 table
3	'C:\Users\Gregorio\...	'Iturrama'	8784x11 table
4	'C:\Users\Gregorio\...	'Leitza'	8784x8 table
5	'C:\Users\Gregorio\...	'Olite'	8784x6 table
6	'C:\Users\Gregorio\...	'Plz. de la C...	8784x8 table
7	'C:\Users\Gregorio\...	'Rotxapea'	8784x7 table
8	'C:\Users\Gregorio\...	'Sangüesa'	8784x7 table
9	'C:\Users\Gregorio\...	'Tudela'	8784x7 table

Ejemplo 2. Resultado de almacenar tablas en la estructura usando `readtable()`.

Fuente: Elaboración propia mediante el uso de MATLAB

➤ **Paso 3: Depurar datos**

En esta sección se ha creado una función **DEPURAR_DATOS_IMPORTADOS.m** que contiene las siguientes subfunciones, también de creación propia:

- **eliminar_fechas_duplicadas.m** – Esta función identifica la columna de datos que contiene la fecha y hora y utilizando la función `unique()` elimina las líneas de datos con fechas duplicadas.
- **cambiar_coma_punto.m** – Esta función identifica qué valores se han grabado con coma como separador decimal y los reemplaza por punto con el comando `replace()`.
- **deshacer_neperiano.m** – Identifica por el nombre de la variable (e.g. CO_2 (\ln)) aquellas variables que estén grabadas en la BBDD con el logaritmo neperiano y les aplica la función exponencial para deshacerlo.

- **unificar_nombre_magnitudes.m** – Elimina los espacios, paréntesis y nombres de estación que contaminan el nombre de la magnitud. (e.g. original: *Iturrama_CO (ln)*, nuevo: *CO*).
- **rellenar_NAN_seguidos.m** – Con el comando *fillmissing()* se rellenan con interpolación lineal los datos faltantes hasta 6 horas seguidas, más de 6 horas se considerará contaminar la información.

El resultado de estas tablas limpias, se van grabando en cada paso en una estructura llamada **ficheros_limpiar**.

uct with 8 fields

direcciones	estacion	tablas	provincia	tablas_coma_punto	sin_neperiano	tablas_titulos_limpios	limpias
C:\Users\Greg...	'Alsasua'	8783x8 table	'NAVARRA'	8783x8 table	8783x8 table	8783x8 table	8783x8 table
C:\Users\Greg...	'Funes'	8783x8 table	'NAVARRA'	8783x8 table	8783x8 table	8783x8 table	8783x8 table
C:\Users\Greg...	'Iturrama'	8783x11 table	'NAVARRA'	8783x11 table	8783x11 table	8783x11 table	8783x11 table
C:\Users\Greg...	'Leitza'	8783x8 table	'NAVARRA'	8783x8 table	8783x8 table	8783x8 table	8783x8 table
C:\Users\Greg...	'Olite'	8783x6 table	'NAVARRA'	8783x6 table	8783x6 table	8783x6 table	8783x6 table
C:\Users\Greg...	'Plz. de la C...	8783x8 table	'NAVARRA'	8783x8 table	8783x8 table	8783x8 table	8783x8 table
C:\Users\Greg...	'Rotxapea'	8783x7 table	'NAVARRA'	8783x7 table	8783x7 table	8783x7 table	8783x7 table
C:\Users\Greg...	'Sangüesa'	8783x7 table	'NAVARRA'	8783x7 table	8783x7 table	8783x7 table	8783x7 table
C:\Users\Greg...	'Tudela'	8783x7 table	'NAVARRA'	8783x7 table	8783x7 table	8783x7 table	8783x7 table

Ejemplo 3. Muestra de la estructura con las tablas limpias.

Fuente: Elaboración propia mediante el uso de MATLAB

➤ Paso 4: Transformar formato

En este paso es necesario transformar las tablas al formato de Madrid, que consiste a grandes rasgos en una transposición, poniendo la variable hora (antes en columna) como las distintas columnas de la tabla, y las magnitudes (antes como columnas) recogidas bajo una columna llamada “magnitud”. Para ello ha sido necesario crear la función: **TRANSFORMAR_FORMATO_MADRID.m** con las siguientes subfunciones:

- **Poner_nombre_estacion.m** – Añade una columna a cada tabla con el nombre de la estación
- **Independizar_hora_dia.m** – Separa en dos columnas la fecha y hora utilizando la función *hour()* y cambia el formato de la variable “hora” a *h1, h2, h3, ... h24*.

- **Transponer_tablas.m** – utilizando el comando *unstack()*, transpone la columna hora por las columnas de las magnitudes.
- **Separar_fecha.m** - Separa la columna “fecha” en año, mes y día
- **Codificar_magnitud/provincia.m** – Cambia las magnitudes y la provincia (en este caso solo Navarra) por los códigos numéricos que se utilizan en el formato de Madrid.
- **Vertcat** – Concatena verticalmente todas las tablas en una sola denominada **tabla_bruto**.
- **rmmissing** – Elimina las filas de datos con más de 6 horas seguidas faltantes (mismo criterio que utilizado anteriormente).

A continuación, se muestran dos ejemplos del efecto de transformar los datos al formato de Madrid:

ficheros_limpiar(1).limpias								
	1 fecha	2 PM10	3 OZONO	4 NO	5 NO2	6 NOx	7 CO	8 SO2
1	01/01/2016 01:00	29	2	25.9000	27.8000	2.0648e+29	1.8221	244.6919
2	01/01/2016 02:00	28	2	14.3000	25.8000	5.7448e+20	1.8221	164.0219
3	01/01/2016 03:00	32	3	7.8000	21.9000	4.7770e+14	1.6487	121.5104
4	01/01/2016 04:00	21	2	5	22.3000	7.9167e+12	1.4918	73.6998
5	01/01/2016 05:00	18	12	4.3000	18.3000	7.2005e+10	1.3499	99.4843
6	01/01/2016 06:00	16	58	3.9000	9.8000	6.5830e+06	1.2214	148.4132
7	01/01/2016 07:00	13	59	4	7.3000	7.2942e+05	1.2214	121.5104
8	01/01/2016 08:00	12	57	4	6.8000	4.8894e+05	1.2214	164.0219
9	01/01/2016 09:00	11	61	4.1000	6.3000	2.9656e+05	1.2214	148.4132

Ejemplo 4. Formato Original.

Fuente: Elaboración propia mediante el uso de MATLAB

127770x30 table										
	1 provincia	2 estacion	3 magnitud	4 ano	5 mes	6 dia	7 h1	8 h2	9 h3	10 h
1	31	'Alsasua'	10	2016	1	1	29	28	32	
2	31	'Alsasua'	10	2016	1	2	6	9	9	
3	31	'Alsasua'	10	2016	1	3	23	11	12	
4	31	'Alsasua'	10	2016	1	4	13	10	7	
5	31	'Alsasua'	10	2016	1	5	8	8	3	
6	31	'Alsasua'	10	2016	1	6	3	2	6	
7	31	'Alsasua'	10	2016	1	7	4	1	1	
8	31	'Alsasua'	10	2016	1	8	5	4	1	
9	31	'Alsasua'	10	2016	1	9	11	9	4	

Ejemplo 5. Formato Madrid.

Fuente: Elaboración propia mediante el uso de MATLAB

➤ **Paso 5: Separar en meses**

Por último, la función **SEPARAR_EN_MESES.m** mediante un bucle *while()* va troceando la **tabla_bruto** en subtablas, de un mes cada una y se exportan en formato *.xlsx* en una carpeta en la BBDD. Estos datos servirán como input del siguiente ejecutable del proceso.

- *CARGAR_NAVARRA.m*

Este ejecutable consiste en la adaptación del código utilizado para tratar los datos de la Comunidad de Madrid utilizados en el estudio de referencia de este trabajo (Betancourt-Odio, Martínez-de-Ibarreta, Budría-Rodríguez, & Wirth, 2021). Por esta razón, se proporcionará una breve descripción de las acciones que se llevan a cabo, pero no se detallarán las funciones que incluye al no ser de elaboración propia sino recicladas. No obstante, el código está incluido al final de este capítulo para su consulta.

Como las estaciones de medición de contaminación y las estaciones de medición de clima son distintas, ha sido necesario realizar un paso previo donde se asocia cada estación de contaminación a la estación climatológica más cercana en un fichero Excel. Utilizando la fórmula de distancia entre dos puntos sobre la superficie terrestre (en km):

$$6371 * ACOS(COS(RADIANES(90 - LATITUD_1)) * COS(RADIANES(90 - LATITUD_2)) + SENO(RADIANES(90 - LATITUD_1)) * SENO(RADIANES(90 - LATITUD_2)) * COS(RADIANES(LONG_1 - LONG_2)))$$

Se obtienen todas las distancias posibles y se escogen las más cercanas:

Tabla 6.

Distancias más cortas entre las estaciones de contaminación y meteorológicas

ESTACION_CONTAMINACION	ESTACION_METEO	KILOMETROS
Alsasua	PAMPLONA_AEROPUERTO	44.473
Felisa Munarriz	PAMPLONA_AEROPUERTO	3.306
Funes	CADREITA	14.070
Iturrama	PAMPLONA_AEROPUERTO	3.090
Leitza	BAZTAN_IRURITA	30.729
Lesaka	BAZTAN_IRURITA	17.888
Olite	CADREITA	30.922
Plz. de la Cruz	PAMPLONA_AEROPUERTO	3.838
Rotxapea	PAMPLONA_AEROPUERTO	5.544
Sangüesa	ARANGUREN_ILUNDAIN	30.310
Tudela	CADREITA	18.359
Tudela II	BARDENAS_REALES	18.225

Fuente: elaboración propia.

Con esta asociación, ya es posible asociar datos clima-contaminación.

El ejecutable *CARGAR_NAVARRA.m* comienza añadiendo una columna nueva a las tablas origen conteniendo la media diaria de la magnitud en cuestión. A continuación, vincula las variables del clima con las de contaminación utilizando la fecha como nexo y la tabla de asociación de estaciones calculada anteriormente.

De igual forma, con la fecha como nexo, vincula las variables de Datos Adicionales. Finalmente reparte los datos de dos formas distintas, la primera, genera unas tablas por cada partícula contaminante, y además genera unas tablas por cada partícula y cada estación, este último tipo de tabla es el que se usará para la regresión.

- 5.2. Trabajo realizado en R y RStudio

Una vez preparados los datos, se utiliza R para hacer las regresiones logísticas correspondientes. Se ha escogido una regresión logística, donde a la variable dependiente se le ha aplicado el logaritmo neperiano, para una interpretación de los resultados más intuitiva. De esta manera, la estimación de los coeficientes β_n representarán el cambio porcentual en la concentración de la partícula ante el aumento de una unidad de la variable dependiente.

Ejemplo: Si para la regresión del CO obtenemos que el parámetro que acompaña a la variable Combustión Biomasa es estadísticamente significativo e igual a 30. Significará que la concentración de CO es previsible que aumente en un 30% los días en los que hay combustión de biomasa.

Por cada estación y partícula se lleva a cabo la regresión múltiple expuesta en el Capítulo 2.

Al ser por cada estación y partícula, se itera la regresión mediante un *for loop* donde se recorre 53 veces el código **Regresión_Partículas.R** adjunto al final de este capítulo.

Utilizamos las tablas resultado, como la del ejemplo de a continuación, de esta regresión para ver qué variables no son significativas y eliminarlas del modelo, para identificarla se ha tomado como criterio el $Pr_t < 0.01$.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.8216448529	0.025518315	-32.1982403	7.329442e-159
TEMP_MEDIA	-0.0041460987	0.001331638	-3.1135335	1.898556e-03
VELOCMED_VIENTO	-0.0218853805	0.004564597	-4.7945916	1.864580e-06
PREC	-0.0003497051	0.001436077	-0.2435141	8.076546e-01
LOCKDOWN	-0.3987919036	0.033065184	-12.0607798	1.820453e-31
SATURDAY	-0.0240799934	0.025046899	-0.9613962	3.365739e-01
SUNDAY	-0.0210262388	0.025068739	-0.8387434	4.018039e-01
HOLYDAYS	0.0512724437	0.052365072	0.9791344	3.277387e-01
SAHARAN_DUST	-0.0359138829	0.026964449	-1.3318975	1.831823e-01
COMBUSTION_BIOMASA	-0.0407154779	0.033030243	-1.2326727	2.179733e-01

Ejemplo 6. Tabla resultada de la regresión.

Fuente: Elaboración propia con RStudio

Finalmente, vuelve a ejecutar el código R únicamente con las variables que resultaron significativas de la primera regresión para reducir el ruido, mejorar así la calidad del modelo y obtener los resultados finales de nuestro estudio.

- *5.3. Trabajo realizado en PowerBI*

El resultado del paso anterior donde se realiza la regresión logística son 53 ficheros de texto plano, cada uno contiene la tabla de resultados de la regresión para cada partícula y estación. Antes de cargarlos en PowerBI se agregan todas las tablas con los resultados de la segunda regresión en un fichero Excel.

Una vez agregados los valores, se crea una página para cada uno de los 9 factores a estudiar compuesta por un diagrama de mapa (utilizando la latitud y longitud de las estaciones de medición) y asociando el coeficiente de la regresión a cada una de las ubicaciones.

Anexo 2. Scripts y funciones del trabajo.

En este apartado final se incluyen los códigos desarrollados necesarios para el tratamiento y análisis de los datos:

- Script principal navarra:

```
clear all;
close all;
clc

%% LEER PERIODO DE ESTUDIO

addpath('C:\ Users\ Gregorio\ Desktop\ YOYO25022021\ Academics\ TFG ADE\
PROYECTO\ CODIGO_NAVARRA\ funciones');

periodo_estudio=....
    readtable('C:\ Users \ Gregorio\ Desktop \ YOYO25022021\ Academics \
TFG ADE \ PROYECTO \ CODIGO_MATLAB \ NAVARRA \ PERIODO_ESTUDIO.xlsx ','
Sheet', 'periodo ');

formato_exportar=periodo_estudio.formato{:};

codigos_magnitud=...
    readtable('C: \ Users \ Gregorio \ Desktop\ YOYO25022021 \Academics
\TFG ADE \PROYECTO \CODIGO_MATLAB \NAVARRA \INDICE_MAGNITUDES.xlsx ');

codigos_provincia=...
```

```
readtable('C:\ Users \ Gregorio \ Desktop \ YOYO25022021 \ Academics \
TFG ADE \ PROYECTO \ CODIGO_MATLAB \ NAVARRA \ CODIGOS_PROVINCIA.xlsx ');
```

```
%% En esta seccion se ejecuta provincia a provincia
```

```
PREPARAR_DATOS_NAVARRA;
```

```
%CARGAR_NAVARRA();
```

- Preparar datos navarra:

```
% Primero grabamos el nombre de la provincia con el nombre que tenga en
nuestra BBDD
```

```
nombre_provincia= 'NAVARRA';
nombre_carpeta_provincia='COMUNIDAD_FORAL_NAVARRA';
```

```
%% FUNCIONES PARA LEER Y PROCESAR
```

```
%% PASO 1: Obtener las rutas, los nombres de las estaciones y los años de
los ficheros que tenemos que cargar
```

```
%-----
```

```
ficheros_importar=...
    OBTENER_INFO_FICHEROS(periodo_estudio,nombre_carpeta_provincia);
```

```
%% PASO 2: Cargar los ficheros en un nuevo campo de la estructura
```

```
%-----
```

```
tablas_importadas=...
    arrayfun(@(x)
readtable(x.direcciones,"PreserveVariableNames",true),ficheros_importar,...
    'UniformOutput',false);
```

```
[ficheros_importar.tablas]=tablas_importadas{:};
```

```
    %añadimos un campo con el nombre de la provincia
    celda_provincia=cell(length(ficheros_importar),1);
    celda_provincia(:)={nombre_provincia};
```

```
[ficheros_importar.provincia]=celda_provincia{:};
```

```
%% PASO 3: "limpiamos" las tablas PASO A PASO PARA DEPURAR
```

```
%-----
```

```
ficheros_limpiar=...
    DEPURAR_DATOS_IMPORTADOS(ficheros_importar);
```

```
%% PASO 4: Rellenar datos faltantes
```

```
%-----
```

```
ficheros_limpiar=...
    RELLENAR_NAN_SEGUIDOS(ficheros_limpiar);
```

```
%% PASO 5: Transformamos las tablas al formato de trabajo
```

```
%-----
```

```
tabla_bruto=...
    TRANSFORMAR_FORMATO_MADRID
(ficheros_limpiar,codigos_provincia,codigos_magnitud);
```

```
%algunas correcciones
```

```

tabla_bruto.estacion=replace (tabla_bruto.estacion,'Plaza','Plz. de la
Cruz');

tabla_bruto.estacion=replace (tabla_bruto.estacion, 'Rochapea','Rotxapea');

tabla_bruto.estacion=replace (tabla_bruto.estacion, 'Sanguesa','Sangüesa');

%% PASO 5: Separamos en meses en una nueva estructura
%-----

fecha_inicio=...
    datetime(periodo_estudio.ano_inicio,periodo_estudio.mes_inicio,1);
fecha_fin=...
    datetime(periodo_estudio.ano_fin,periodo_estudio.mes_fin,1);

ficheros_transformados=...

SEPARAR_EN_MESES(tabla_bruto,fecha_inicio,fecha_fin,formato_exportar,nombre
_carpeta_provincia);

%
% %% ESCRIBIMOS LOS DATOS EN FICHEROS
%
% %dependiendo del tipo de formato, se exportara en .csv o .xlsx
%
% disp(' EXPORTANDO...')
%
% if strcmp(formato_exportar,'.csv')
%
% for jj=1:length(ficheros_transformados)
%
% writetable(ficheros_transformados(jj).tabla,...
%     ficheros_transformados(jj).direccion_definitiva,'Delimiter',';');
%
% end
%
% elseif strcmp(formato_exportar,'.xlsx')
%
%     for jj=1:length(ficheros_transformados)
%
%         writetable(ficheros_transformados(jj).tabla,...
%             ficheros_transformados(jj).direccion_definitiva);
%
%     end
%
% end
%
%

```

- Cargar datos Navarra

```

function [] = CARGAR_NAVARRA()

%ESTA FUNCION LEE LOS DATOS DE LA COMUNIDAD FORAL DE NAVARRA QUE ESTAN
DENTRO DE
%LA CARPETA "COMUNIDAD_FORAL_NAVARRA" SEGÅSN EL FICHERO
%"DATOS_TRATAR"

%FUNCIÅN QUE DEVUELVE DIRECTORIOS

[directorio_datos,directorio_codigo,directorio_datos_tratados,datos_estacio
n_particula,directorio_datos_clima,directorio_leer_nuevos_datos,directorio_
tablas_unicas_contam] = DEVOLVER_DIRECTORIOS('NAVARRA');

```

```

leer_datos=strcat(directorio_codigo,'\NAVARRA_DATOS_A_LEER.txt');
variables_leer=LEER_VARIABLES_TXT(leer_datos);
cd(directorio_codigo);

[NAVARRA] = LEER_Y_JUNTAR_TODOS_FICHEROS_SEGUN_VARIABLES_NAVARRA
(variables_leer, directorio_datos, directorio_codigo );

cd(directorio_codigo);

answer=GENERAR_ALERTA_AUSENCIA_DATOS(NAVARRA);

if( strcmp( answer,'OK' ) ~=1 )

    disp('----- TRATANDO FECHAS EN LOS DATOS -----');

    NAVARRA = DEVOLVER_FECHAS_DATOS(NAVARRA);

    %NAVARRA=ASIGNAR_NOMBRE_MAGNITUD(NAVARRA,directorio_codigo);

    %disp('----- QUITANDO ACENTOS NOMBRE ESTACIONES -----');

    %NAVARRA=GENERAR_ESTACIONES_SIN_ACENTOS(NAVARRA); %DE MOMENTO NO HACE
    %FALTA PARA NAVARRA

    NAVARRA=LIMPIAR_DATOS (NAVARRA);

    NAVARRA=ANADIR_MEDIA_DIARIA(NAVARRA);

    %SELECCIONAR FECHA DE TRABAJO

    %fechas = DEVOLVER_FECHAS_DATOS_CATALUNA(NAVARRA);
    %en esta linea de arriba, la tabla NAVARRA YA ES UN TIMETABLE, no hace
    %falta, ver linea de abajo:

    fechas=NAVARRA;

    %SELECCIONAR SOLO COLUMNA DE FECHA Y ORDENARLA

    FECHAS_DATOS=sort(fechas.Time);

    [~,~,FECHAS_TRABAJO] = OBTENER_FECHAS_TRABAJO();

    answer_fechas =
    GENERAR_ALERTA_ERROR_FECHAS(FECHAS_TRABAJO,FECHAS_DATOS);

    if(strcmp(answer_fechas,'OK')~=1)

        disp('-----REPARTIENDO TABLAS POR ESTACIONES-----');

%tabla=REPARTO_POR_ESTACIONES(NAVARRA,NAVARRA.CODIEOI,NAVARRA.NOMESTACIO);
    % EL FÓRMATO ES DISTINTO ASI QUE HARÁ% UNA FUNCION NUEVA QUE
    % DEVUELVA EL MISMO RESULTADO, MIRAR SIGUIENTE LINEA:

    tabla=REPARTO_POR_ESTACIONES_V2(NAVARRA);

    fclose('all');

    %GENERA TABLAS PARA CADA PARTICULA Y ESTACION

    [TAB_ESTAC_PART,NOMB_PARTICULAS] =
    GENERAR_TABLAS_ESTAC_CONT_CATALUNA(tabla,directorio_datos_clima,directorio_
    codigo,directorio_leer_nuevos_datos);

```

%GENERAR UNA TABLA ÚNICA DE CONTAMINACION POR CADA PARTICULA (CALCULANDO LAS MEDIAS DE TODAS LAS ESTACIONES)

```
SCONT_UNICA=GENERAR_UNICA_TABLA_CONTAMINANTE(NOMB_PARTICULAS,TAB_ESTAC_PART);
```

```
    respuesta= 'Añadir'; %proviene de reciclar codigo, por eso lo falseo aqui
```

```
PASAR_STRUCT_EXCEL_TABLAUNICA(SCONT_UNICA,directorio_tablas_unicas_contam,respuesta,directorio_codigo);
```

```
PASAR_STRUCT_A_EXCEL_PARTICULA(TAB_ESTAC_PART,datos_estacion_particula,respuesta,directorio_codigo);
```

```
    disp('-----ESCRIBIENDO FICHEROS EN EXCEL-----');
```

```
PASAR_STRUCT_A_EXCEL(tabla,directorio_datos_tratados,directorio_codigo,respuesta);
```

```
    cd(directorio_codigo);
```

```
    else
```

```
        return
```

```
    end
```

```
else
```

```
    return
```

```
end
```

```
disp('-----DATOS LISTOS PARA SER USADOS-----');
```

```
end
```

- **Obtener info ficheros**

```
function [ficheros_importar]=...
```

```
    OBTENER_INFO_FICHEROS(periodo_estudio,nombre_carpeta_provincia)
```

```
%OBTENER_INFO_FICHEROS Obtiene las rutas, nombre de la estacion y año
% devuelve una estructura con los campos de ruta del fichero, el nombre
% de la estacion, el nombre de la estacion con caracteres especiales y el
% año
```

```
%PASO 1: LAS RUTAS DE LOS FICHEROS QUE QUEREMOS IMPORTAR
```

```
rutas_ficheros=...
```

```
    OBTENER_RUTAS_FICHEROS(periodo_estudio,nombre_carpeta_provincia);
```

```
%PASO 2: LOS NOMBRES DE LAS ESTACIONES DE CONTAMINACION
```

```
estaciones= OBTENER_NOMBRE_ESTACION(rutas_ficheros);
```

```
%PASO 3: CAMBIAMOS LOS CARACTERES ESPECIALES SEGÚN READTABLE
```

```
%NO HACE FALTA CON LA OPCION "PreserveVariableNames",true
```

```
%estaciones_especial= CAMBIAR_CARACTERES_ESPECIALES(estaciones);
```

```
%PASO 4: GRABAMOS TODO EN UNA ESTRUCTURA QUE DEVOLVEREMOS
```

```
ficheros_importar=cell2struct(...
```

```
    rutas_ficheros,{'direcciones'},length(rutas_ficheros));
```

```
[ficheros_importar.estacion]=estaciones(:);
```

End

- Depurar datos importados

```
function [ficheros_limpiar] = DEPURAR_DATOS_IMPORTADOS(ficheros_importar)
%DEPURAR_DATOS_IMPORTADOS limpia las tablas para poder transformarlas al
%formato de trabajo mas adelante

    ficheros_limpiar=ficheros_importar;

%~~~~~
%1. eliminamos fechas duplicadas

tablas_importadas=arrayfun(@(x) eliminar_fechas_duplicadas(x.tablas),...
    ficheros_limpiar,'UniformOutput',false);

[ficheros_limpiar.tablas]=tablas_importadas{:};

%~~~~~
%2.cambiamos las comas por puntos

tablas_coma_punto=arrayfun(@(x)...
    cambiar_coma_punto(x.tablas),...
    ficheros_limpiar,'UniformOutput',false);

    [ficheros_limpiar.tablas_coma_punto]=tablas_coma_punto{:};

%~~~~~
%3.Deshacemos las variables neperiano

tablas_sin_neperiano=...
    arrayfun(@(x) deshacer_neperiano(x.tablas_coma_punto),...
    ficheros_limpiar,'UniformOutput',false);

    [ficheros_limpiar.sin_neperiano]=tablas_sin_neperiano{:};

%~~~~~
%4. limpiamos los nombres de las columnas

tablas_titulos_limpios=...
    arrayfun(@(x) ...
    unificar_nombres_magnitudes(x.sin_neperiano,x.estacion),...
    ficheros_limpiar,'UniformOutput',false);

[ficheros_limpiar.tablas_titulos_limpios]=tablas_titulos_limpios{:};

%Aquí podemos añadir mas pasos para limpiar los datos...
%.....
%.....

[ficheros_limpiar.limpias] = tablas_titulos_limpios{:};

end
```

- Transformar datos al formato madrid

```
function [tabla_bruto] = TRANSFORMAR_FORMATO_MADRID
(ficheros_limpiar,codigos_provincia,codigos_magnitud)
%TRANSFORMAR_FORMATO_MADRID Transformamos los datos al formato de datos de
%madrid
```

```

%~~~~~
%1. ponemos una columna extra con el nombre de la estacion y otra con el
%nombre de la provincia

    tablas_con_estacion=...
    arrayfun (@(x)
poner_nombre_estacion_provincia(x.limpias,x.estacion,x.provincia),...
    ficheros_limpiar,'UniformOutput',false);

    [ficheros_limpiar.tablas_con_estacion]=tablas_con_estacion{:};

%~~~~~
%2. independizamos la hora de la fecha en dos columnas

    tablas_con_hora= arrayfun(@(x)...
    independizar_hora_dia(x.tablas_con_estacion),...
    ficheros_limpiar,'UniformOutput',false);

%~~~~~
%3. transponemos las columnas contra la columna de la hora
%para lograr el formato de madrid

    tablas_transpuestas=cellfun(@(x)
transponer_tabla(x),tablas_con_hora,...
    'UniformOutput',false);

%~~~~~
%4.independizamos la fecha en tres columnas: año, mes, dia

    tablas_transpuestas = cellfun(@(x)
separar_fecha(x),tablas_transpuestas,...
    'UniformOutput',false);

%lo junto todo en una sola tabla de para luego trocearlo
%en meses mas adelante

tabla_bruto=vertcat(tablas_transpuestas{:});

%~~~~~
%5.Codificamos la provincia, la estacion y la magnitud

tabla_bruto= CODIFICAR_PROVINCIA (tabla_bruto,codigos_provincia);
tabla_bruto= CODIFICAR_MAGNITUD (tabla_bruto, codigos_magnitud);

%-----QUITO LAS FILAS CON DEMASIADOS NaNs-----

minimo_NaN_eliminarfila=6; %<-- Variar este numero segÃn criterio

[tabla_bruto,indice_filas_eliminadas] = ...
    rmmissing(tabla_bruto,1,'MinNumMissing',minimo_NaN_eliminarfila);

mensaje_alerta=...
    sprintf('Se eliminaron %d filas de datos por contener mÃs de %d datos
faltantes
consecutivos',sum(indice_filas_eliminadas(:)==1),minimo_NaN_eliminarfila);
fig = uifigure;
uialert(fig,mensaje_alerta,'Datos Faltantes','Icon','info');

%-----

End

```

- Rellenar NaN seguidos

```

function [ficheros_limpiar] = RELLENAR_NAN_SEGUIDOS(ficheros_limpiar)
%RELLENAR_DATOS_FALTANTES Rellena los datos faltantes segun criterio
elegido

    %contar
    NaN_seguidos_rellenar=6;

    tablas_limpias =... Primero pasamos a timetable para poder usar
fillmissing
    arrayfun(@(x) table2timetable(x.limpias),ficheros_limpiar,...
    'UniformOutput',false);

    tablas_limpias=cellfun(@(x)...
fillmissing(x,'nearest','MaxGap',...
hours(NaN_seguidos_rellenar)),...
    tablas_limpias,'UniformOutput',false);

    tablas_limpias =... Volvemos al formato table
cellfun(@(x) timetable2table(x),tablas_limpias,...
    'UniformOutput',false);

    [ficheros_limpiar.limpias]=tablas_limpias{:};
end

```

- Separar en meses:

```

function [ficheros_transformados] = ...

SEPARAR_EN_MESES(tabla_bruto,fecha_inicio,fecha_fin,formato_exportar,nombre
_carpeta_provincia)
%SEPARAR_EN_MESES Separa la tabla_bruto en tablas de longitud =1 mes
%
%Metemos en una nueva estructura las tablas separadas por meses, para
%separarlas usaremos un bucle while.

ficheros_transformados = struct ;
mes=month(fecha_inicio) ;
ano=year(fecha_inicio) ;
ii=1;

while datetime(ano,mes,1) <= fecha_fin

    %campos de la estructura donde se indican mes y año
    ficheros_transformados(ii).mes = mes;
    ficheros_transformados(ii).ano = ano;

    %S es un indice que vale 1 cuando el mes y el año coinciden con el mes
    %y año en cuestion

    S=logical((tabla_bruto.mes == mes).*(tabla_bruto.ano==ano));
    ficheros_transformados(ii).tabla=tabla_bruto(S,:);

    %nombre del fichero con forma "mes_año"
    ficheros_transformados(ii).nombre_fichero=...
        [num2str(ano),'_',num2str(mes),formato_exportar];

    ficheros_transformados(ii).direccion_definitiva=...
        ['C:\Users\Gregorio\Desktop\YOYO25022021\Academics\TFG
ADE\PROYECTO\DATOS',...

        '\',nombre_carpeta_provincia,'\DATOS_TRATAR','\ ',ficheros_transformados(ii)
.nombre_fichero];

    ii=ii+1;
    mes=mes+1;

```



```

    if mes>12
        ano=ano+1;
        mes=1;
    end
end

```

```
end
```

```
end
```

- Devolver direcciones

```

function [direcciones_ficheros,estaciones,estaciones_especial,numero_ano]
=...
    devolver_direcciones(periodo_estudio,nombre_carpeta_provincia)
%devolver_direcciones Localiza carpetas con los ficheros excel
% Busca los ficheros excel según la , la provincia y los años que
% deseamos y devuelve las direcciones a los ficheros y los nombres de las
% estaciones

direccion_carpeta_provincia='D:\CODIGOS_MATLAB\CODIGO_BASE_DATOS_CALIDAD_AI
RE_ESPANA\DATOS\COMUNIDAD_FLORAL_NAVARRA\DATOS_TRATAR\';

switch nombre_carpeta_provincia
    %-----
    %- SWITCHEAMOS CADA COMUNIDAD AUTONOMA CON SU NOMBRE DE CARPETA -
    %-----
    case 'COMUNIDAD_FORAL_NAVARRA'

        %% Obtenemos las direcciones
        %Primero obtenemos las direcciones de las subcarpetas dentro de la
        %comunidad autonoma en cuestion

        subcarpetas=dir(direccion_carpeta_provincia);

subcarpetas=subcarpetas(~ismember({subcarpetas(:).name},{'.','..' }));
idx =1;
for i=1:length(subcarpetas)

    nombre_subcarpeta= split(subcarpetas(i).name, '_');

    % A continuacion solo grabaremos las carpetas de los años que nos
    % interesan en direcciones_subcarpetas.

    if length(nombre_subcarpeta)>1

        ano_subcarpeta=str2double(nombre_subcarpeta(2));

        if ano_subcarpeta>=periodo_estudio.ano_inicio && ...
            ano_subcarpeta<=periodo_estudio.ano_fin

            direcciones_subcarpetas{idx}=...
                [direccion_carpeta_provincia,'\ ',subcarpetas(i).name];

            idx=idx+1;

        end

    end

end

end

end

%Una vez tenemos las direcciones de las subcarpetas que tenemos que

```

```

%importar, grabaremos en direccion_ficheros las direcciones de
%todos los archivos .csv correspondientes a las estaciones de esa
%provincia

for j=1:length(direcciones_subcarpetas)

    ficheros{j}=dir(direcciones_subcarpetas{j});

ficheros{j}=ficheros{j}(~ismember({ficheros{j}}(:).name},{','..''}));

end

ficheros_importar=vertcat(ficheros{:});
ficheros_importar=...
    rmfield(ficheros_importar,{'date','bytes','isdir','datenum'});

%Añadimos el campo de las direcciones como el resultado de
concatenar los
%dos primeros
direcciones_ficheros=...
    arrayfun(@(x) cat(2,x.folder,'\',x.name),ficheros_importar,...
        'UniformOutput',false);
direcciones_ficheros=transpose(direcciones_ficheros);

%% Obtenemos el nombre y año de las estaciones

%obtenemos el nombre de la estacion del nombre del fichero
split_estacion_ano= ...
    arrayfun(@(x) split(x.name,'_'),ficheros_importar,...
        'UniformOutput',false);

estaciones=...
    cellfun(@(x) x{1},split_estacion_ano,'UniformOutput',false);

estaciones_especial=cell(size(estaciones));
for k=1:length(estaciones)
%-----
%corregimos de forma manual aquellos nombres con caracteres especiales

%Caracteres especiales no identificados por el readtable
caracteres_especiales=...
    {' ','\.','ü'};
caracteres_equivalentes=...
    {'_'};
expresion_regular= strjoin(caracteres_especiales,'|');

caracter_cambiar=unique(regexp(estaciones{k},expresion_regular,'match'));

%igualamos el nombre de la estacion especial con el de la original para
%posteriormente cambiarle los caracteres en el switch
estaciones_especial{k}=estaciones{k};

if ~isempty(caracter_cambiar)

    for numero_caracter=1:length(caracter_cambiar)

switch caracter_cambiar{numero_caracter}
    case ' '
        posicion_cambiar=regexp(estaciones_especial{k},' ');
        estaciones_especial{k}(posicion_cambiar+1)=...
            upper(estaciones_especial{k}(posicion_cambiar+1));

        estaciones_especial{k}=replace(estaciones_especial{k},' ','');

    otherwise

        estaciones_especial{k}=replace(estaciones_especial{k},...
            caracter_cambiar{numero_caracter},...

```

```

        caracteres_equivalentes);
end
end

end

end

%obtenemos el año del nombre del fichero
numero_ano=...
    cellfun(@(x) x{2},split_estacion_ano,'UniformOutput',false);

numero_ano=cellfun(@(x) split(x,'.csv'),numero_ano,...
    'UniformOutput',false);

numero_ano=cellfun(@(x) x(~cellfun('isempty',x)),...
    numero_ano,'UniformOutput',false);

numero_ano=...
    cellfun(@(x) str2double(x{:}),numero_ano,'UniformOutput',false);

end
end

```

Por motivos de eficiencia, orden y sostenibilidad, he incluido el resto de los códigos secundarios utilizados en el proyecto en el siguiente enlace. Los nombres de los anexos corresponden con los nombres utilizados en la memoria para su rápida consulta.

<https://www.notion.so/Anexos-TFG-Egea-D-az-Gregorio-b1117de99d054450ad17b007b715c69b>