



**UNIVERSIDAD NACIONAL DE LA PLATA**

**FACULTAD DE CIENCIAS EXACTAS**

**DEPARTAMENTO DE QUÍMICA**

---

**Trabajo de Tesis Doctoral:**

***Elaboración e implementación  
de una propuesta metodológica  
para la evaluación y gestión de la  
calidad del aire mediante el  
enfoque de la ciencia de datos***

---

**Tesista: Lic. Natacha Soledad Represa**

**Director en Universidad Nacional de La Plata:**

**Dr. Atilio Andrés Porta**

**Director en Universidad Politécnica de Valencia:**

**Dr. Jesús Palomar Vázquez**

**CoDirector en Universidad Politécnica de Valencia:**

**Dr. Alfonso Fernández Sarría**

**Año: 2020**



UNIVERSIDAD  
NACIONAL  
DE LA PLATA



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

# Elaboración e implementación de una propuesta metodológica para la evaluación y gestión de la calidad del aire mediante el enfoque de la ciencia de datos

Natacha Soledad Represa

Presentado ante la Universidad Nacional de La Plata y  
la Universidad Politécnica de Valencia  
bajo acuerdo de cotutela  
para acceder a los títulos de

**Doctor en Ciencias Exactas, con orientación en química**  
**Doctor en Ingeniería Geomática**

Directores

Andrés Porta (CIM – UNLP – CONICET)

Jesús Palomar-Vázquez (CIGAT – UPV)

Alfonso Fernández-Sarría (CIGAT – UPV)

Abril, 2020



*Cuando el 11 de diciembre del 2014 se firmó la Resolución N° 4815 donde me fue otorgada una Beca Interna Doctoral para realizar esta investigación, la República Argentina contaba con un Ministerio de Ciencia y Tecnología y un Ministerio de Ambiente y Desarrollo Sustentable. En el transcurso de esta tesis, ambos ministerios fueron degradados a secretarías, limitando el accionar del Estado Nacional sobre estos asuntos en el territorio argentino. El 10 de diciembre del 2019 se anunció la restitución de las dos carteras a ministerios nacionales y recuperamos la esperanza.*



*“El único héroe válido es el héroe ‘en grupo’,  
nunca el héroe individual.”*

H. G. Oesterheld<sup>1</sup>

# Agradecimientos

---

Somos producto de nuestras acciones porque nos hacemos haciendo. Pensar, desarrollar y escribir esta tesis ha sido una oportunidad única de crecimiento personal y profesional donde me he enfrentado a preguntas y situaciones que creí inabarcables. Y solo con el apoyo de muchas personas, instituciones y organizaciones fue posible este proyecto. Sin sus contribuciones gran parte de las cosas planteadas hubiesen quedado en el tintero sin salir a la luz. Infinitas gracias por los aportes y la compañía.

En especial, al **Dr. Andrés Porta** y a toda la Unidad II por invitarme a participar del **Centro de Investigaciones del Medioambiente (CIM, CONICET - UNLP)**. Emprendimos un camino riesgoso, lleno de dudas y preguntas abiertas que transitamos honrosamente. En ese cotidiano lleno de cambios, el acompañamiento de **Daniela Giuliani** y **Victoria Passucci** en el subsuelo de la **Facultad de Ciencias Exactas** de la UNLP, las visitas de **Daniela Mellado** y el apoyo de **Florencia Yanniello** fueron esenciales para sentirme segura en los caminos que tomó esta tesis. A su vez, agradezco a la **Dra. Yanina Sánchez** y al **Dr. Esteban Colman-Lerner** por las palabras justas en los momentos precisos.

Al **Geo-Environmental Cartography and Remote Sensing Group** y a los CGATos que me acompañaron durante días y noches en las oficinas de la **Escuela de Cartografía y Topografía** de la UPV. A los **Dres. Alfonso Fernández-Sarría** y **Jesús Palomar-Vázquez** por recibir abiertamente esta propuesta de trabajo y adoptarme como parte de la familia. Gracias por la enorme generosidad y por estar a disposición de mis consultas. Por ser pacientes y confiar en este trabajo que se fue gestando en diálogos interrumpidos por la distancia. Me llevo mil imágenes de Valencia, y todas las palabras quedan cortas para transmitir mi enorme agradecimiento.

Al interés de los integrantes de **“Valencia per l’Aire”** del grupo Mesura, al **Dr. César Ferri Ramírez** y a **Lidia Contrera** del **Departamento de Sistemas Informáticos y Computación (UPV)** y al **Dr. Ferrán Ballester** del **Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP)** por escuchar nuestras ideas y brindar su apoyo. No es fácil pensar en un territorio que no es el propio. Quisiera que sepan que hice eco de sus preocupaciones y adopté la problemática con mi más profundo compromiso.

A la **Dra. Gabriela Abril** y especialmente a la **Mg. María Fernanda García Ferreyra** con quienes hemos mantenido una conversación sostenida sobre la calidad del aire en Argentina a lo largo de estos años. Charlas repletas de calidez y cercanía, compartiendo ideas y pensando proyectos. Enormemente agradecida que este proceso me haya llevado a conocerlas.

---

<sup>1</sup> Héctor Germán Oesterheld. Escritor y guionista argentino víctima del terrorismo de Estado. Padre de la historieta argentina moderna. Autor de *El Eternauta*.

A la **Comisión Nacional de Actividades Espaciales (CONAE)** y al equipo de excelentes profesionales que se reúnen en el **Instituto de Altos Estudios Espaciales Mario Gulich (CONAE, UNC - CONICET)** que me enseñaron a programar en sus aulas.

A la **Dra. Claudia Natenzon** que me recibió en su casa para discutir sobre los estudios de vulnerabilidad social realizados en Argentina y me facilitó bibliografía. Al **Dr. Lucas Bali** del **Instituto de Investigaciones Científicas y Técnicas para la Defensa (CITEDEF)** que me aconsejó al comienzo de esta tesis con una generosidad inmensa. Son en esos primeros pasos cuando más apoyo se necesita y agradezco enormemente la amabilidad que supieron tener conmigo.

A la banda del **Laboratorio de Investigación en Nuevas tecnologías Informáticas de la Facultad de Informática (LINTI)** de la UNLP que me hospedaron los últimos meses de la realización de mi tesis y me ofrecieron soporte técnico cada vez que me hizo falta. En particular, a **Agustín Candia** por su compañía y apoyo diario.

A mis colegas, graduados de la **Lic. en química y tecnología ambiental** de la UNLP, con quienes compartimos actividades, espacios de reflexión y asambleas. A ellos que me recuerdan la importancia de hacer ciencia para el beneficio de nuestro pueblo.

A las chicas de **R-Ladies Buenos Aires** y al grupo **Podemos Aprender** que fueron espacios de contención y acompañamiento a la distancia. A **Mauricio Cap** por las fugazzetas de Guerrín y las conversaciones por calle Corrientes.

A mis amigas, que me alegran la vida.

A mis viejos, **Hugo Represa** y **Silvia Polenta**, de quienes aprendo a ser una adulta responsable.

# Publicaciones

Compendio de artículos publicados durante la elaboración de esta tesis doctoral

---

- **Review “Data Mining Paradigm in the Study of Air Quality”.** Represa NS., Fernández-Sarría A., Porta A., Palomar Vázquez J. (2019). *Environmental Processes*, 1-21, DOI: 10.1007/s40710-019-00407-5
- **“Daily concentrations of PM<sub>2.5</sub> in the Valencian Community using Random Forest for the period 2008-2018”.** Represa NS., Porta A., Palomar Vázquez J., Fernández-Sarría A. (2019). *Proceedings* 2019, 19, 13; DOI: 10.3390/proceedings2019019013
- **“Risk analysis of technological hazards: simulation of scenarios and application of a local vulnerability index”.** Sanchez Y., Represa NS., Mellado D., Balbi K., Acquesta AD., Colman Lerner JE., Porta A. (2018) *Journal of Hazardous Materials*, vol. 352, p. 101-110.
- **“Estudio de la vulnerabilidad social en Argentina mediante el uso de SIG. Construcción de un índice de aplicación local”.** Represa NS., Sanchez Y., Porta A. (2018) *Anuario do Instituto de Geociencias*, 41(2), 351-357. ISSN 0101-9759; e-ISSN 1982-3908
- **“Solutions for SmartCities: proposal of a monitoring system of air quality based on a LoRaWAN network with low-cost sensors”.** Candia A., Represa NS., Giuliani D., Luengo MA., Porta A., Marrone LA. *II Congreso Argentino de Ciencias de la Informática y Desarrollos de Investigación IEEE CACIDI*. 28-30 noviembre, 2018
- **“Assessment of satellite aerosol optical depth to estimate particulate matter distribution in Valencia city”.** Represa NS., Fernández-Sarría A., Porta A., Palomar Vázquez J. *International Geoscience and Remote Sensing Symposium, IGARSS* 2018.
- **“Study of the monthly correlation between AOD obtained from MODIS images and particulate matter for a coastal city”.** Represa NS., Palomar Vázquez J., Porta A., Fernández-Sarría A. *IEEE Biennial Congress of Argentina (ARGENCON 2018)*.
- **“Evaluación del Producto Satelital del Sensor OMI/AURA para el Análisis de Dióxido de Nitrógeno en la Provincia de Buenos Aires”.** Represa S., Ojeda S., Abril G., García Ferreyra MF. *IEEE Biennial Congress of Argentina (ARGENCON 2018)*.
- **“Enfoque metodológico para el análisis de riesgos de atmósferas peligrosas en entornos a establecimientos que manipulan HAZMAT”.** Represa SN.; Mellado D., Balbi KB., Acquesta AD., Porta A., Sanchez EY. *Libro de Resúmenes extendidos del Congreso*



*Colombiano y VI Conferencia Internacional de Calidad del Aire y Salud Pública, CASAP 2017.*

- **“Characterization of BTEX emission sources in industrial and urban areas using Principal Components Analysis and Nonparametric Regression Methods”**  
Mellado D.; Bali JL.; Giuliani D.; Colman Lerner JE.; Represa NS.; Jacovkis PM.; Sanchez EY. *Libro de Resúmenes extendidos del Congreso Colombiano y VI Conferencia Internacional de Calidad del Aire y Salud Pública, CASAP 2017.*
- **“Contaminación atmosférica en Dock Sud, Pcia. de Buenos Aires, Argentina. Análisis de su comportamiento temporal en el período 2011-2017”.** Represa SN., Abril G., Ojeda S.; García Ferreyra M F. *VI Congreso Bianual del Proyecto Integrador sobre Mitigación de la Contaminación Atmosférica (2017).*
- **“Aplicación de tecnologías de sistemas de información geográfica (SIG) para la estimación del riesgo por exposición a compuestos orgánicos volátiles (COVs)”**  
(2016) Represa S., Colman Lerner e., Sanchez Y., Porta A. Artículo completo. *III Congresso da Sociedade de Análise de Riscos Latino Americana SRA-LA.* 11-13 de mayo del 2016 – Sao Paulo, Brasil. ISBN: 978-85-7270-070-2

# Contenido

---

Agradecimientos .....	iv
Publicaciones .....	vi
Prólogo .....	14
Capítulo 1 - Introducción General .....	16
1. Objetivo general .....	18
Objetivos específicos .....	19
2. Estructura.....	19
Capítulo 2 - Marco teórico.....	20
1. La gestión de la calidad del aire.....	20
2. La complejidad del estudio de la calidad del aire .....	22
2.1 La atmósfera y su dinámica .....	22
2.2 Contaminación del aire, efectos en la salud y aspectos legales .....	25
2.3 Entre la legalidad y la seguridad ambiental .....	29
3. El riesgo a la salud .....	29
4. Estrategias para el monitoreo de la calidad del aire .....	30
4.1 Monitoreo Terrestre .....	31
4.2 Teledetección .....	32
5. Modelos de calidad del aire .....	36
5.1 Modelos numéricos determinísticos .....	37
5.2 Modelos matemáticos probabilísticos .....	39
6. Big data, ciencia de datos y minería de datos.....	40
7. El riesgo, la incertidumbre, la ciencia y la gestión .....	41
Capítulo 3 - Propuesta metodológica .....	44
Introducción.....	44
Módulo I - Etapa de diseño del análisis.....	46
I.I. Definición y caracterización el área de estudio .....	47
I. II. Determinación de los objetivos .....	49
I.III. Delimitación de las preguntas de estudio .....	49
I. IV. Adquisición de los datos .....	50
Módulo II - Etapa de procesamiento .....	56
II. I. Preparación de los datos .....	56

II. III. Selección de algoritmos de análisis y procesamiento de los datos .....	59
Módulo III - Etapa de comunicación .....	75
III. I. Análisis de los resultados .....	75
III.I. Reportes y Visualizaciones .....	75
III. II. Confección de índices .....	77
III. III. Confección de mapas de calidad del aire y mapas de riesgo .....	78
Capítulo 4 - Comunidad Valenciana, España.....	80
Introducción.....	82
Aplicación de la metodología .....	83
Módulo I - Etapa de diseño del análisis.....	83
Módulo II - Etapa de procesamiento .....	93
Módulo III - Etapa de resultados y comunicación .....	100
Capítulo 5 - Ciudad de Buenos Aires, Argentina.....	184
Introducción.....	186
Aplicación de la metodología .....	188
Módulo I - Etapa de diseño .....	188
Módulo II - Etapa de procesamiento .....	194
Módulo III - Etapa de resultados y comunicación .....	199
Capítulo 6 - La Plata, Argentina.....	232
Introducción.....	234
Aplicación de la metodología .....	235
Módulo I - Etapa de diseño .....	235
Módulo II - Etapa de procesamiento .....	241
Módulo III - Etapa de resultados y comunicación .....	242
Capítulo 7 - Conclusiones generales .....	254
Bibliografía.....	258

# Resumen

---

La exposición a contaminantes atmosféricos causa efectos negativos en la salud de la población, tanto de forma directa en el corto y el largo plazo, como de forma indirecta, a través de su incidencia en el cambio climático. El 92% de la población mundial se enfrenta diariamente a una mala calidad del aire y el 90% de las ciudades del mundo no cuentan con suficientes reglamentaciones para gestionar este tema. El estudio de los contaminantes atmosféricos ha sido abordado desde diversos enfoques, pero han sido pocos los trabajos destinados a la gestión de la calidad del aire con un enfoque metodológico integrador.

En base a esto, el objetivo de esta tesis ha sido elaborar una propuesta metodológica para la evaluación y gestión de la calidad del aire desde un enfoque de la ciencia de datos. Este enfoque contribuye en el estudio de la calidad del aire proporcionando una metodología de trabajo estructurada que organiza las etapas del proceso analítico, desde la adquisición de datos de distintas fuentes hasta la producción de información para los usuarios finales.

La gestión de la calidad del aire demanda un flujo de trabajo dinámico, donde se priorice la generación de productos y la comunicación de resultados. Teniendo esto como objetivo principal, se elaboró una metodología que facilita la integración de los datos disponibles, sistematiza las preguntas más relevantes, condensa los algoritmos empleados en los análisis descriptivos y predictivos, y permite generar mapas de riesgo contribuyendo a una rápida y mejor toma de decisiones.

Para validar esta propuesta, se aplicó la metodología en tres casos de estudio: la Comunidad Autónoma de Valencia (España) que cuenta con una extensa red de monitoreo de la calidad del aire, la Ciudad de Buenos Aires (Argentina) con tres estaciones de monitoreo en su zona central, y la ciudad de La Plata (Argentina) donde no se cuenta con un monitoreo continuo. Al contar con distinta información, se evaluó la versatilidad de la metodología para adaptarse a distintos escenarios.

Los resultados mostraron que el enfoque esquemático proporciona orden y flexibilidad en el estudio de la calidad del aire. La utilización de preguntas de investigación simplificó notablemente la selección de algoritmos para el análisis y el abordaje de la interpretación de los resultados. A su vez, permitió diferenciar las preguntas que fueron originarias del estudio de aquellas que surgieron tras el análisis de los datos. Los algoritmos seleccionados resultaron cumplieron con los objetivos propuestos.

Como gran limitación para el análisis de la calidad del aire se encontró la falta de mediciones sistemáticas. Si bien las concentraciones de algunos contaminantes se modelaron mediante el uso de variables medidas por teledetección, se precisaron de mediciones de campo para validar las simulaciones. En este sentido, el diálogo efectivo entre la ciencia y la gestión es un asunto prioritario, donde la ciencia debe proporcionar herramientas innovadoras para avanzar en el nivel de conocimiento de la atmósfera e identificar los problemas y soluciones que puedan surgir, pero es necesario dar un marco político que garantice los fondos económicos para su estudio.

# Abstract

---

Exposure to atmospheric pollutants causes adverse effects on the health of the population, both directly in the short and long term, and indirectly through their impact on climate change. 92% cent of the world's population faces poor air quality daily, and 90% of the world's cities do not have enough regulations to manage this issue. The study of atmospheric pollutants has been considered from various approaches, but there has been little work on air quality management with an integrative methodological strategy.

Therefore, the objective of this thesis was to develop a methodological framework for air quality assessment and management using a data science approach. This strategy contributes to the study of air quality by providing a structured working methodology that organizes the stages of the analytical process, from the acquisition of data coming through different sources until the presentation of information to the end-users.

Air quality management demands a dynamic pipeline that privileges the production of outputs and the reporting of results. Keeping this as a main aim, the methodology developed promotes the integration of available data, organizes the most relevant questions, summarizes the algorithms used in descriptive and predictive analysis, and creates risk maps to help make quick and better decisions.

To validate this proposal, we evaluated the methodology in three case studies: the Autonomous Community of Valencia (Spain) which has an extense air quality monitoring network, the City of Buenos Aires (Argentina) with three monitoring stations in its central zone, and the city of La Plata (Argentina) where there is no continuous monitoring. Since different information was available, the versatility of the methodology to adapted to different scenarios was assessed.

The results showed that the schematic design provides order and flexibility in the investigation of air quality. The use of research questions greatly simplified the selection of algorithms for analysis and the understanding of the results. In turn, it enabled the distinction the original questions from those that emerged after the analysis of the data. The selected algorithms were in accordance to the established goals.

A major obstacle to the determination of air quality was the lack of routine monitoring. Whereas the concentrations of some pollutants were modelled using remotely sensed variables, field measurements were required to validate the predictions. In this sense, the effective interchange between science and administration is a priority issue, where science must provide innovative tools to improve the level of knowledge of the atmosphere and to identify the problems and solutions that may arise. However, it is necessary to provide a political support to guarantee the economic funds for research.

# Resum

---

L'exposició a contaminants atmosfèrics causa efectes negatius en la salut de la població, tant de manera directa en el curt i el llarg termini, com de manera indirecta, a través de la seua incidència en el canvi climàtic. El 92% de la població mundial s'enfronta diàriament a una mala qualitat de l'aire i el 90% de les ciutats del món no compten amb suficients reglamentacions per a gestionar aquest tema. L'estudi dels contaminants atmosfèrics ha sigut abordat des de diversos enfocaments, però han sigut pocs els treballs destinats a la gestió de la qualitat de l'aire amb un enfocament metodològic integrador.

Sobre la base d'això, l'objectiu d'aquesta tesi ha sigut elaborar una proposta metodològica per a l'avaluació i gestió de la qualitat de l'aire des d'un enfocament de la ciència de dades. Aquest enfocament contribueix en l'estudi de la qualitat de l'aire proporcionant una metodologia de treball estructurada que organitza les etapes del procés analític, des de l'adquisició de dades de diferents fonts fins a la producció d'informació per als usuaris finals.

La gestió de la qualitat de l'aire demanda un flux de treball dinàmic, on es prioritze la generació de productes i la comunicació de resultats. Tenint això com a objectiu principal, es va elaborar una metodologia que facilita la integració de les dades disponibles, sistematitza les preguntes més rellevants, condensa els algorismes emprats en les anàlisis descriptives i predictives, i permet generar mapes de risc contribuint a una ràpida i millor presa de decisions.

Per a validar aquesta proposta, es va aplicar la metodologia en tres casos d'estudi: la Comunitat Autònoma de València (Espanya) que compta amb una extensa xarxa de monitoratge de la qualitat de l'aire, la Ciutat de Buenos Aires (l'Argentina) amb tres estacions de monitoratge en la seua zona central, i la ciutat de la Plata (l'Argentina) on no es compta amb un monitoratge continu. En comptar amb diferent informació, es va avaluar la versatilitat de la metodologia per a adaptar-se a diferents escenaris.

Els resultats van mostrar que l'enfocament esquemàtic proporciona ordre i flexibilitat en l'estudi de la qualitat de l'aire. La utilització de preguntes d'investigació va simplificar notablement la selecció d'algorismes per a l'anàlisi i l'abordatge de la interpretació dels resultats. Al seu torn, va permetre diferenciar les preguntes que van ser originàries de l'estudi d'aquelles que van sorgir després de l'anàlisi de les dades. Els algorismes seleccionats van resultar van complir amb els objectius proposats.

Com a gran limitació per a l'anàlisi de la qualitat de l'aire es va trobar la falta de mesuraments sistemàtics. Si bé les concentracions d'alguns contaminants es van modelar mitjançant l'ús de variables mesures per teledetecció, es van precisar de mesuraments de camp per a validar les simulacions. En aquest sentit, el diàleg efectiu entre la ciència i la gestió és un assumpte prioritari, on la ciència ha de proporcionar eines innovadores per a avançar en el nivell de coneixement de l'atmosfera i identificar els problemes i solucions que puguem sorgir, però és necessari donar un marc polític que garantisca els fons econòmics per al seu estudi.



*“Air is a shared resource.  
Research and tools to make it safe to breathe  
should be shared as well.”<sup>2</sup>*

# Prólogo

Sobre la ambición, los sueños, la ciencia y la innovación

---

La mala calidad del aire mata. Sin eufemismos ni alicientes. Una de cada ocho muertes en el mundo ocurre a causa de la exposición a contaminantes atmosféricos. Esto es, un total de 4 millones de personas por año. La contaminación atmosférica es responsable del 34% de los casos de cardiopatía isquémica, del 20% de los accidentes cerebrovasculares, del 18% de las neumopatías obstructivas crónicas, del 7% de los casos de cáncer de pulmón, y del 3% de las infecciones agudas de las vías respiratorias inferiores en menores [World Health Organization, 2016].

La mala calidad del aire mata más en los países subdesarrollados. El 87% de las muertes relacionadas con la contaminación del aire se producen en países de bajos y medianos ingresos, principalmente de Asia y África, seguidos por los países de la Región del Mediterráneo Oriental, América y Europa [Hajat et al., 2015].

Sin embargo, la producción de contaminantes no es equitativa, son los países desarrollados quienes más contaminan [World Health Organization, 2015, 2016]. Una serie de trabajos muestran el alto nivel de desigualdad internacional en las emisiones de gases de efecto invernadero per cápita y la gran importancia que tiene la desigualdad de ingresos en este aspecto [Groot, 2010; Duro & Padilla, 2006].

El aire es un bien común que no se detiene en las fronteras, esto hace que la gestión del aire resulte un asunto complejo con diversas aristas. Por un lado, la cuestión técnico-científica que limita la generación de acciones preventivas o reparatorias. El estudio de la calidad del aire requiere desarrollar distintos tipos de estrategias analíticas, donde cada escala de trabajo presenta dificultades específicas por corresponder a un sistema caótico. Por otro lado, la escasa inversión en capacitación y recursos informáticos que se destinada a personal de la administración pública en la gestión ambiental imposibilita el aprovechamiento del conocimiento científico disponible en el tema. Y sin información, no hay acción posible.

Frente a este escenario acuciante, resulta fundamental el rol de las universidades públicas de generar conocimiento científico de calidad, otorgando herramientas confiables que faciliten las tareas de gestión. Desde la **Universidad Nacional de La Plata** (Argentina) y la **Universidad**

---

<sup>2</sup> “Filthy air is a global disgrace”. Nature 561, 285 (2018). doi: 10.1038/d41586-018-06731-4



**Politécnica de Valencia** (España), hacemos este pequeño aporte al conocimiento en este campo. Esta tesis metodológica nace de la necesidad de atender un problema común, buscando optimizar el uso de la información disponible. A partir de una serie de preguntas orientativas se encaminará el abordaje para cada escenario de estudio analizando los datos disponibles; ya sea que provengan de monitoreos terrestres sistemáticos o manuales u obtenidos mediante teledetección.

El estudio de la calidad del aire y sus efectos en la salud pública resultan tópicos ampliamente estudiados. Sin embargo, aún falta parte del recorrido que separa el conocimiento científico básico de la aplicación tecnológica. Nuestro objetivo va en ese sentido: integrar distintas herramientas que se encuentran en fases iniciales de desarrollo a un grado más complejo -con más partes- pero, a la vez, más simple por estar más cerca del usuario final.

Una tesis metodológica en el estudio de la calidad del aire resulta un paso necesario para avanzar en la implementación de una solución que pueda atender complejidades diversas. En esta tesis se verán tres casos de aplicación: la Comunidad Valenciana (España) que cuenta con una extensa red de monitoreo terrestre que cubre áreas bien diferenciadas; la Ciudad Autónoma de Buenos Aires (Argentina) donde existen tres estaciones de monitoreo dentro del núcleo urbano; y la ciudad de La Plata (Argentina) donde no existen monitoreos automáticos y sólo se cuenta con las mediciones manuales del **Centro de Investigaciones del Medioambiente**. El desafío reviste en poder generar una metodología versátil que se amolde a estas diferentes situaciones y que además pueda ser adaptada a otras situaciones no contempladas en este trabajo.

Para alcanzar esta meta, se consideró fundamental incorporar distintas técnicas innovadoras. En la selección de fuentes de datos se decidió incorporar el uso de imágenes satelitales, buscando aprovechar la mejor resolución espacial y temporal disponible en la actualidad. El análisis de la información se apoyó en la ciencia de datos, la geoestadística, la minería de datos y en los conocimientos en inteligencia artificial. Finalmente, la integración de la información espacial se realizó bajo el paradigma de los sistemas de información geográfica.

Toda esta labor se realizó con un uso exclusivo de datos abiertos provistos por la comunidad científica y por organismos estatales. A su vez, se trabajó únicamente con software libre, tanto para el análisis estadístico como para la confección de las cartografías. En complemento, todo lo generado en el marco de esta tesis se pone a disposición del público, mediante el libre acceso en el sitio: <https://github.com/solrepresa>.

Por ser una metodología de trabajo fácilmente adaptable y escalable, se cuenta con generar un impacto positivo en el corto plazo. Esta tesis, además, presenta el análisis de los registros históricos de la calidad de aire para tres grandes núcleos urbanos (Valencia, Buenos Aires y La Plata), siendo insumos útiles de forma directa para las instituciones locales.

Se espera que tanto los análisis de casos particulares, como la metodología propuesta, sea utilizada como recurso en la creación de futuras políticas públicas, en el desarrollo de nuevas tecnologías o en la generación de nuevas investigaciones en salud. Construir conocimiento libre con herramientas y datos abiertos también puede resultar un elemento central en la construcción de una ciencia participativa, dando lugar a la intervención y control ciudadano en las políticas ambientales.

*Valencia, 13 de enero del 2019*  
*Sol Represa*

*“Existe un juego de negociación  
y finalmente de compromiso  
entre lo teóricamente deseable,  
lo metodológicamente posible y  
lo estratégicamente necesario.”*

*Rodrigo Marquez*

## *Capítulo 1*

# Introducción General

---

Los efectos negativos en la salud pública que causa la exposición a contaminantes atmosféricos son bien conocidos, estos pueden aparecer de forma directa en el corto o el largo plazo [Rückerlet et al., 2011; Weinmayr et al., 2009; Lippmann, 2014] o de forma indirecta, a través de su incidencia en el cambio climático [Charlson y Hofmann, 1992; Dickerson et al., 1997; Ramanathan, 1985]. Sin embargo, el 92% de la población mundial se enfrenta diariamente a una mala calidad del aire y el 90% de las ciudades del mundo no cuentan con reglamentaciones completas para gestionar este tema [World Health Organization, 2016 b].

El estudio de la calidad del aire ha sido abordado desde distintos enfoques: diseño de redes de monitoreo [Maré et al., 2015], elaboración de inventarios de emisiones [Gulia et al., 2015; Van Fan et al., 2018], confección de modelos espaciales para predecir los niveles en áreas donde no se realizan mediciones [Venkatram, 2015], estudios epidemiológicos [Mannucci et al., 2015], modelos matemáticos para la predicción en el corto plazo [Li et al., 2016]. Sin embargo, son muy pocos los trabajos destinados a la gestión de la calidad del aire con un enfoque metodológico integrador [Harrop, 2018].

Gestionar no solo requiere un conocimiento pormenorizado del fenómeno con el cual se trabaja sino, también, contar con estrategias que sirvan para optimizar los recursos disponibles, brindando la mejor solución posible en el menor tiempo de respuesta [Shaddick et al., 2018]. La gestión ambiental, por tanto, tiene como desafío transformar los datos duros, generalmente cuantitativos, en datos cualitativos, reconociendo particularidades y aportando una mirada más comprensiva (y menos explicativa) sobre el fenómeno en estudio. Esto requiere la internalización de una gran cantidad de datos para generar información de fácil interpretación que conduzca y refuerce la toma de decisiones.

La falta de información en contraposición con la gran cantidad de datos, los intereses políticos y económicos, los compromisos internacionales, la presión demográfica, la expansión de las áreas urbanas y la salud de la población son algunos de los puntos por donde pivota la toma de decisiones en la administración. A su vez, la inversión se retrasa cuando no hay planes establecidos y garantías claras de los beneficios que pueda traer el destino de recursos -humanos y económicos- en un tema que no genera controversias y, por tanto, no aparenta requerir medidas urgentes.

Como una serpiente que muerde su propia cola, la maraña de acciones desincronizadas que llevan las administraciones solo empeora el escenario socioambiental, repercutiendo aún más fuertemente en países periféricos. Si no se destinan fondos para monitorear, se desconoce el verdadero impacto de la contaminación sobre la salud de la población. Otras veces, se destinan fondos para el monitoreo, pero no se capacita a los operarios para el análisis de los datos. Se elaboran complejos informes que resultan obsoletos a los pocos meses de su publicación sin aportar conocimiento nuevo a la población. A su vez, la falta de aprovechamiento de las redes de monitoreo conlleva a que se desvíen los fondos para mantenerlas. Y con el tiempo, las redes se desmiembran, resultando imposible evaluar cómo evolucionó la calidad del aire desde el último informe realizado.

Como consecuencia, la desinformación genera miedo y desconfianza en la población. Las personas se oponen sin fundamentos precisos tanto a la generación de un carril bici (bici senda) como a la instalación de una industria. La opinión pública se dirime entre las comodidades de la ciudadanía, dificultando todo tipo de gestión conciliatoria que vele por el bien común.

Por último, la falta de información oficial repercute directamente sobre el resto de los organismos estatales. Los especialistas en salud no cuentan con información histórica para realizar estudios epidemiológicos o diseñar políticas públicas acordes a la realidad del territorio. Los funcionarios no pueden sopesar críticamente las alternativas frente a un proyecto ingenieril con impacto ambiental porque les faltan indicadores del estado de la situación.

La OMS ha titulado “el asesino invisible” a su última campaña de prevención contra los efectos de los contaminantes atmosféricos por la elevada capacidad de daño a la salud que generan y el gran desconocimiento que se les tiene. Gestionar en función de reducir los niveles de los contaminantes atmosféricos es prevenir muertes y mejorar la sobrevivencia de las personas, en particular, de los grupos de riesgo, como niños y niñas y adultos y adultas mayores.

Para conocer la calidad del aire hace falta un monitoreo sistemático y contar con modelos matemáticos que permitan inferir las concentraciones en las áreas donde no se monitorea. Sin embargo, cuando existen redes de monitoreo de la calidad de aire, pocas veces se integra esa información para brindar un servicio acabado al ciudadano. Esto conlleva a la falta de apropiación de la problemática y a una subestimación del riesgo, dificultando la ejecución de medidas preventivas o reparatorias.

Los modelos más utilizados para evaluar la calidad del aire de forma diaria son complejas simulaciones que se apoyan en las predicciones atmosféricas a escala global. Estos modelos tienen la desventaja de presentar una resolución espacial de meso escala, dificultando su uso en áreas urbanas donde los mecanismos de transporte de los contaminantes dependen, en gran parte, del uso del suelo [Xie et al., 2017]. Por otro lado, existen modelos de microescala, que por su naturaleza matemática requieren una gran cantidad de datos de entrada y una elevada capacidad de cálculo. Estos modelos se utilizan en situaciones concretas de desastres o evaluaciones de riesgo, pero no son propicios para un uso diario [Xie et al., 2017].

En ambos casos, los resultados que se obtienen siguen el mismo destino que los datos provistos por las redes de monitoreo: no se integran con la información de otras fuentes dificultando su interpretación para un público general. El usuario final suele ser un comité técnico que examina las concentraciones generadas por el modelo y evalúa el acatamiento de la norma, sin considerar los riesgos asociados independientemente de la violación de los valores legales. La falta de un método de trabajo que englobe y sintetice las diferentes técnicas de análisis de datos de

contaminantes atmosféricos, dificulta aún más la tarea de comunicar de forma intuitiva a un público no especializado.

Es por esto que en esta tesis se desarrolla una *Propuesta para la Evaluación de la Calidad del Aire (PECA)*, con el fin de brindar un conjunto de herramientas que permitan analizar la calidad del aire y generar mapas de los niveles de exposición a los contaminantes atmosféricos y del riesgo asociado mediante el aprovechamiento de la información satelital, y haciendo uso de las mediciones disponibles en cada sitio.

La selección del mapa como instrumento no es inocente. El mapa funciona como plataforma para repensar el territorio y redefinirlo. La representación espacial facilita su interpretación, y aporta una narrativa sobre los eventos que suceden, ya no de forma aislada, sino atravesados por la identificación de las áreas conocidas, las experiencias de vida, los valores y costumbres, y la propia historia personal y colectiva [Harley, 2018].

En la era de la información, resulta crucial el aprovechamiento de las oportunidades tecnológicas como son la minería de datos, la teledetección y el uso de sistemas de información geográfica para encarar esta problemática. En este sentido, esta tesis provee una guía para ayudar a las agencias ambientales y organizaciones sociales a evaluar el funcionamiento de las redes de monitoreo de calidad de aire, desarrollar y operar modelos matemáticos de calidad de aire basados en imágenes satelitales, y evaluar e informar los resultados obtenidos en función de los riesgos asociados a la salud.

Finalmente, el método propuesto se aplicará en tres casos de estudio reales para poder realizar su evaluación: la Comunidad Autónoma de Valencia (España) que cuenta con una extensa red de monitoreo de la calidad del aire, la Ciudad Autónoma de Buenos Aires (Argentina) con tres estaciones de monitoreo en su zona central; y el Partido de La Plata (Argentina) donde no se cuenta con un monitoreo continuo.

Como características generales, se buscó incorporar algoritmos con una matemática simple que no requieren grandes tiempos de procesamiento y permiten un uso intensivo de los datos disponibles. A su vez, se buscó mantener una estructura modular para poder cubrir distintas realidades, que fuese de fácil integración con los modelos en funcionamiento (no excluyente), escalable a nuevas fuentes de información, útil en cada fase de su aplicación, fácil de implementar, robusta en bibliografía científica y libre, haciendo uso de herramientas *open-source* y *open-data*.

## 1.1. Objetivo general

El objetivo general que persigue esta tesis es elaborar una propuesta metodológica para la evaluación y gestión de la calidad del aire mediante el enfoque de la ciencia de datos.

La adquisición de información de los datos masivos es clave para la construcción de instituciones más eficientes y eficaces en la toma de decisiones. Por ello, esta tesis busca aportar una metodología basada en el conocimiento científico que permita generar un flujo de trabajo dinámico el cual incluya la selección de las fuentes de datos, la sistematización de las preguntas más relevantes, la selección de mejores métodos actuales y la presentación de resultados.

## Objetivos específicos

- Identificar las principales fuentes de información para estudiar la calidad del aire.
- Determinar las preguntas más relevantes en la evaluación de la calidad del aire.
- Generar una metodología que permita la integración de múltiples fuentes de información.
- Estudiar las redes de monitoreo de la calidad del aire.
- Confeccionar mapas de concentración de contaminantes mediante el uso de imágenes satelitales.
- Aplicar un índice de calidad del aire para las zonas de estudio.
- Evaluar escenarios de riesgo por exposición a los contaminantes presentes y su impacto en la salud pública.
- Construir mapas de riesgo por exposición a contaminantes atmosféricos.
- Generar informes de la calidad del aire para los distintos sitios de aplicación de la metodología

## 1.2. Estructura

En el Capítulo 2 se encuentra el marco teórico de este estudio, donde se describen aspectos relacionadas a la problemática de la calidad del aire y a las principales tecnologías para su monitoreo. Se describe la estructura de la atmósfera y su dinámica, los procesos químicos, las fuentes principales de emisiones de contaminantes y cómo influyen las condiciones meteorológicas. Se presentan los principales estudios sobre los efectos en la salud producidos por la contaminación atmosférica y las estrategias para el monitoreo de la calidad del aire. Tras esto, se detallan brevemente los modelos de calidad del aire, determinísticos y estocásticos, y de riesgo para poner en contexto el trabajo.

La propuesta metodológica se presenta en el Capítulo 3. En este capítulo se propone una metodología sistemática de análisis de los datos de calidad de aire. Se identifican las principales preguntas de trabajo y las fuentes de información más utilizadas. Tras esto, se recopilan los algoritmos de análisis utilizados en el estudio de la calidad del aire, tanto en análisis descriptivos como predictivos. Finalmente, se resumen los mecanismos de evaluación de modelos y se presentan diferentes estrategias para generar reportes y visualizaciones.

En los Capítulos 4, 5 y 6 se aplica la metodología de trabajo en las tres áreas y se presentan los resultados obtenidos para cada caso de estudio. Como cierre, en el Capítulo 7, se hace una evaluación de la metodología propuesta y se da una discusión respecto a las ventajas y limitaciones que se han encontrado en cada caso. A su vez, en este capítulo se proponen una serie de acciones para mejorar la gestión de la calidad del aire.

*“Clean air is considered to be  
a basic requirement of human health and well-being.  
However, air pollution continues to pose a significant threat to health worldwide”*

*World Health Organization, 2005b.*

## *Capítulo 2*

# Marco teórico

---

## 1. La gestión de la calidad del aire

Las consecuencias de la acción directa de la humanidad sobre el planeta son visibles a distintas escalas [Paul & Meyer, 2001; Kalnay & Cai, 2003]. El uso extensivo del petróleo como principal fuente de energía, la producción y el consumo desmedido de productos para sostener la economía global, y la intervención sin reparos sobre los ecosistemas han provocado cambios evidentes en el planeta [McKinney, 2002; Paul & Meyer, 2001] conduciendo a la humanidad a la necesaria e inminente obligación de intervenir de forma activa para contrarrestar el daño causado [Stern, 2007; Grimm et al., 2008].

La presencia de contaminantes atmosféricos provoca una alteración en el balance radiativo de la Tierra conduciendo a un cambio en el clima a nivel global [Brasseur & Prinn, 2003; Davis, 1997]. En los últimos años, se ha demostrado que la temperatura de la superficie terrestre ha aumentado de forma sostenida poniendo en riesgo innumerables ecosistemas junto a su diversidad biológica [Adams et al., 1998; Jacobson, 2002]. Las predicciones no son optimistas. Se estima que las pérdidas de las áreas cultivadas serán entre un 10% y un 40% para los próximos 10 años, según el modelo matemático que se considere, sin contar las pérdidas producidas por los fenómenos extremos o por los efectos indirectos en la agricultura (sequías, inundaciones, plagas y patógenos y aumento de las tasas de degradación del suelo) [Adams et al., 1998].

Como respuesta, los Estados han celebrado una serie de acuerdos internacionales para controlar las emisiones de contaminantes a la atmósfera, como son la Convención Marco de las Naciones Unidas sobre el Cambio Climático (1992), el Protocolo de Kioto (1995), el Acuerdo de París (2015), entre otros. Los acuerdos internacionales son de obligatorio cumplimiento, debiendo los países firmantes alcanzar ciertos objetivos en función de sus situaciones económicas particulares [Eritja, 2007]. Este interés internacional promueve acciones específicas locales, aunque muchas veces estas medidas no son prioritarias para los gobiernos [Reichert, 2003].

Gozar de un ambiente sano y equilibrado es un derecho humano y debemos velar por su protección [Gibson, 1990; Hancock, 2003]. En particular, es bien sabido que la mala calidad del aire afecta a la salud. En octubre de 2013, la agencia especializada en cáncer de la OMS, la Agencia Internacional de Investigación sobre el Cáncer (IARC), clasificó la contaminación del aire como un agente carcinogénico para los seres humanos, considerándola un grave problema de salud

ambiental. Basado en la revisión independiente de más de mil estudios epidemiológicos de gran envergadura, la evaluación de la IARC mostró un aumento del riesgo a padecer cáncer de pulmón frente a niveles crecientes de contaminantes atmosféricos [Loomis et al., 2013].

La gestión del territorio está estrechamente ligada a la calidad del aire [Davis, 1997; West, 2013]. Por la alta densidad de actividades antrópicas contaminantes en las ciudades, la polución del aire es, principalmente, un problema urbano [Monks et al., 2009]. A medida que la población se concentra en las ciudades [WHO, 2015], mayor es la población expuesta. Mientras más del 50% de la población mundial se concentra en asentamientos urbanos, el 80% de estos presenta niveles nocivos de contaminantes atmosféricos [World Health Organization, 2019; Osseiran & Chriscaden, 2016].

Si bien todas las personas se ven afectadas en algún modo por la mala calidad del aire, la exposición a los contaminantes atmosféricos no es equitativa. Es la población más empobrecida quien soporta la mayor carga de contaminantes [World Health Organization, 2016; Mitchell et al., 2015], tanto en el vecindario donde viven como en los lugares de trabajo [Evans & Kantrowitz, 2002; Deguen & Zmirou-Navier, 2010; Næsset al, 2007; Wang et al., 2016]. Como contrapunto, los sectores desfavorecidos generan una menor proporción de emisiones por persona y presentan una menor posibilidad de movilidad, ya que se encuentran más limitados para alejarse de la contaminación, por ejemplo, mudándose de domicilio [Mitchell et al., 2015; World Health Organization, 2016 b].

A su vez, hay grupos específicos de la población, como ancianos y ancianas, niños y niñas, y personas con afecciones respiratorias o cardiovasculares preexistentes, que tienen una mayor probabilidad de presentar enfermedades por los contaminantes del aire [Deguen & Zmirou-Navier, 2010; Peled, 2011; Wang et al., 2016; World Health Organization, 2016 b; Pope 3rd, 2000], lo que aumenta la desigualdad en materia de salud.

La gestión de la calidad del aire busca la adecuación de los niveles de contaminación atmosférica, cualesquiera que sean las causas que la produzcan, a fin de garantizar que no se generen molestias graves o daños inmediatos o diferidos, para las personas y para los bienes de cualquier naturaleza. Gestionar la calidad del aire es, por tanto, gestionar el riesgo [World Health Organization, 2016 b]. Esto es, sopesar las opciones, analizar las alternativas, evaluar las pruebas y las incertidumbres, y tener en cuenta las opiniones de los grupos de interés y las partes interesadas en la formulación de políticas [Prüss-Üstün et al., 2016; Longhurst et al., 2006].

Sin embargo, nada de esto es posible sin información precisa y confiable que dé cuenta del estado de la calidad del aire. No se puede gestionar sin conocer los niveles de los contaminantes en la atmósfera y sin contar con herramientas que permitan evaluar la evolución, espacial y temporal de la contaminación frente a los cambios económicos y sociales que se dan en el territorio. Lamentablemente, los países más afectados por la contaminación del aire presentan el monitoreo de la calidad del aire más limitado [Petkova et al., 2013; Sulemana, 2012; Baldasano et al., 2003; Sammarco et al., 2017]. Según World Health Organization [2019], el 97% de las ciudades de los países de ingresos bajos y medios con más de 100.000 habitantes no cumplen las directrices de la OMS sobre la calidad del aire.

Un factor que podría explicar la baja inversión en el monitoreo de la calidad del aire es que las mediciones de contaminantes sólo tienen valor por el conocimiento que aportan a la creación e implementación de políticas públicas que generen mejoras para la sociedad [Zhang et al., 2012]. La falta de habilidades técnicas para el tratamiento de grandes volúmenes de datos, el

desconocimiento de algoritmos específicos para el análisis de las medidas o presentar problemas para sintetizar la información y elaborar informes son algunas de las dificultades que se pueden presentar en la administración pública [Sammarco et al., 2017; Amegah & Agyei-Mensah, 2017]. Esto no solo limita la generación de información sino también la posibilidad de comprender el potencial que estos datos tienen para la prevención, respuesta y mitigación del riesgo asociado a la exposición de una mala calidad del aire.

## 2. La complejidad del estudio de la calidad del aire

Procesos químicos y físicos que ocurren en la superficie de la Tierra pueden alterar la composición de la atmósfera. Las emisiones, el transporte, la vida media de los contaminantes y el destino de las sustancias tienen un fuerte impacto sobre las concentraciones de los compuestos químicos presentes en el aire y sobre la calidad del aire que respiramos [Wang et al., 2016]. A su vez, estas sustancias sufren transformaciones químicas por reacciones entre ellas o por la acción de la luz solar, generando nuevos compuestos. La calidad del aire es, por tanto, un fenómeno complejo que depende de múltiples factores, naturales y antrópicos [Mayer, 1999].

La complejidad de un fenómeno o sistema radica en dos dimensiones: la variedad de los eventos inherentes al problema bajo análisis y su aleatoriedad, es decir, la incertidumbre acerca del modo en que se desenvolverá en el futuro [Downey, 2018]. Los fenómenos o sistemas complejos se caracterizan por la confluencia de múltiples procesos cuyas interrelaciones constituyen la estructura de un sistema que funciona como una totalidad organizada [Ortegon, 2007; Soler, 2017].

En esta sección se desarrollarán algunas cuestiones descriptivas respecto a la dinámica y composición de la atmósfera, para facilitar la comprensión y lectura de este trabajo de investigación. No se pretende realizar un análisis exhaustivo, por el contrario, solo se presentarán los principales fenómenos que aportan al entendimiento del sistema.

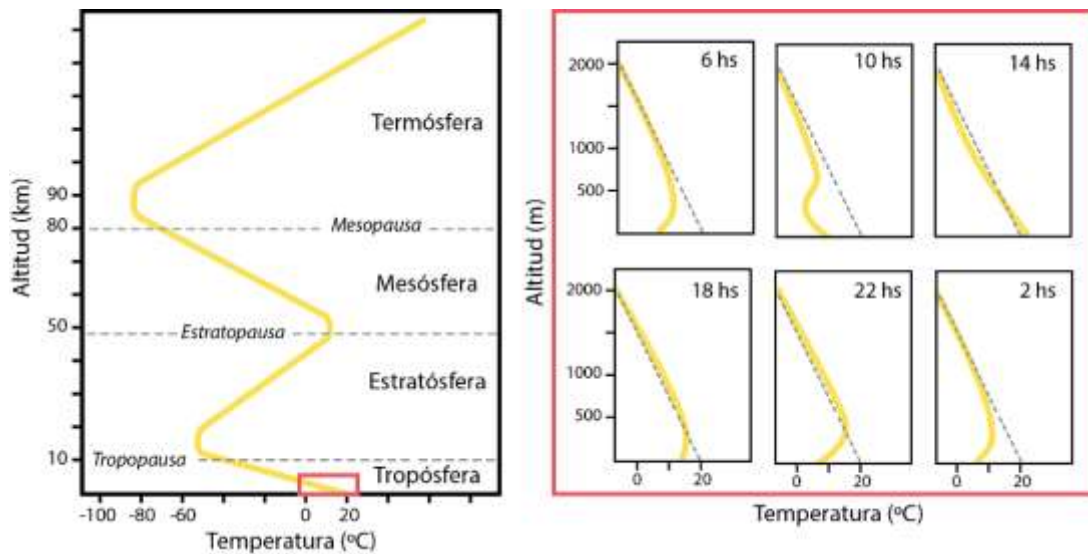
### 2.1 La atmósfera y su dinámica

La atmósfera es la capa gaseosa que rodea nuestro planeta. La estimación de su ancho oscila entre los 50-80 km, insignificante frente a los 12 mil km que presenta el diámetro de la Tierra. Sin embargo, tiene una gran importancia en los ciclos biogeoquímicos, donde funciona como sumidero de distintos elementos, como el C y el N [Arya, 1999].

La atmósfera es una mezcla de gases y partículas que se mantienen cercanas a la superficie terrestre, principalmente, por la atracción de la fuerza gravitatoria que ejerce el planeta. Si observamos las funciones de estado que gobiernan el comportamiento de esta masa de aire, podemos ver que la presión disminuye a medida que nos alejamos de la Tierra, al igual que el número de partículas por unidad de volumen [Seinfeld & Pandis, 2016]. De esta forma, la atmósfera resulta un sistema físico continuo, donde la cantidad de materia tiende a disminuir a medida que nos alejamos de la superficie terrestre [Arya, 1999].

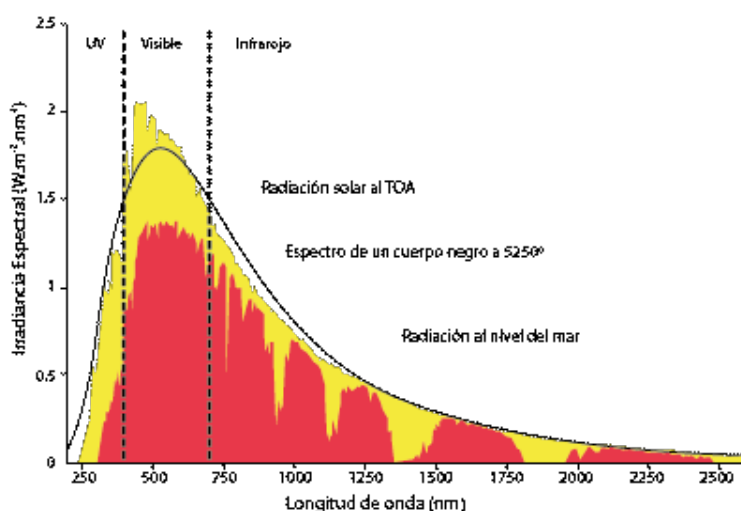


Por su parte, el gradiente de temperatura vertical muestra un comportamiento particular, vinculado con los fenómenos físicoquímicos que suceden a distintas alturas [Finlayson-Pitts & Pitts, 1999]. La capa más baja de la atmósfera, la tropósfera, se caracteriza por el descenso de la temperatura con la altura (**Figura 1**). Esta disminución continúa hasta alcanzar la estratósfera (capa ubicada entre los 45 a 55 km de altitud) donde se invierte el perfil térmico y la temperatura aumenta con la altitud. La estructura térmica vertical de la estratósfera es el resultado de la absorción de la radiación solar ultravioleta [Finlayson-Pitts & Pitts, 1999] que provoca reacciones de fotodisociación y fotoionización, principalmente.



**Figura 1.** Perfil térmico de la atmósfera y de la tropósfera a distintas horas del día.

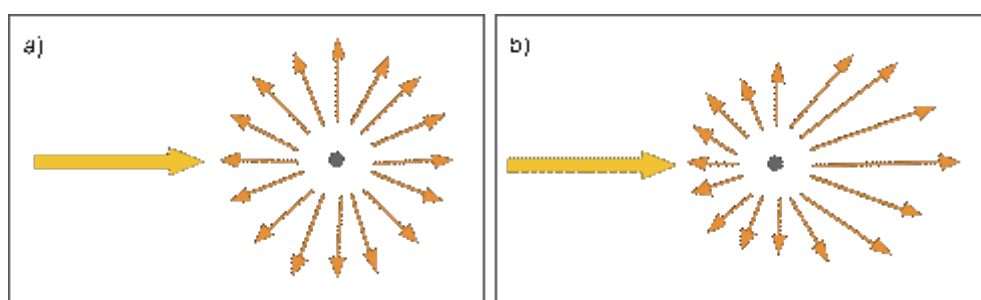
La atmósfera, por tanto, es un sistema abierto que mantiene un intercambio radiativo con la Tierra y el espacio. El Sol se comporta prácticamente como un cuerpo negro, emitiendo energía a longitudes de onda entre los 250 y 2500 nm. En la **Figura 2**, se puede apreciar el espectro de emisión de un cuerpo negro a 5250° (línea continua negra) y el espectro de emisión del Sol en amarillo, los cuales se solapan en gran medida. También, se puede apreciar la irradiancia medida al nivel del mar en color rojo, es decir, la energía que llega a la superficie terrestre por unidad de área. Esta diferencia entre la potencia radiativa al tope de atmósfera (TOA, del inglés *top of atmosphere*) y la potencia radiativa a la altura del mar se debe a la interacción entre la energía proveniente del sol y las sustancias presentes en la atmósfera.



**Figura 2.** Alveolo atmosférico. Espectro de emisión de un cuerpo negro a  $5250^{\circ}$  (línea continua negra), radiación solar al TOA (amarillo), radiación a nivel del mar (rojo) [Modificado de Seinfeld & Pandis, 2016]

Cuando la radiación del sol pasa a través de la atmósfera, la energía incidente se modifica tanto en intensidad como en su distribución espectral. Por un lado, esto se debe a la absorción de la luz por parte de las moléculas presentes, provocando fotodisociación, reordenamientos intramoleculares, fotoisomerización, fotodimerización, entre otras reacciones posibles [Seinfeld & Pandis, 2016]. Y también se observan fenómenos de dispersión de la radiación solar, dispersión de Mie y/o Rayleigh, dependientes de la relación del tamaño de las partículas con la longitud de onda incidente [Seinfeld & Pandis, 2016].

La presencia de partículas en la atmósfera provoca la dispersión de la luz solar (**Figura 3**). Cuando la dispersión está generada por moléculas o aerosoles menores que la longitud de onda incidente se conoce como dispersión de Rayleigh, donde se altera la frecuencia y la dirección del haz de luz (favoreciendo las ondas cortas). Mientras que la dispersión de Mie ocurre cuando la dispersión es provocada por aerosoles de tamaño próximos o mayores a la longitud de onda de la radiación incidente, provocando un aumento de la intensidad de la radiación (más brillo). La dispersión producida por las partículas es el proceso de atenuación de la radiación solar más importante, seguido, en orden, por la absorción de partículas, la absorción de gases y la dispersión de gases [Jacobson, 2005].



**Figura 3.** Fenómeno de dispersión de a) Rayleigh y b) Mie

La sección inferior de la tropósfera se conoce como la capa límite atmosférica que, por estar en contacto con la superficie terrestre, se encuentra fuertemente influenciada por los efectos superficiales, como la transferencia radiativa, los efectos de la fricción por arrastre, la evapotranspiración y las modificaciones surgidas por la rugosidad del terreno [Garratt, 1994]. Para nuestro estudio en la calidad del aire nos centraremos en esta capa, la cual se caracteriza por un flujo turbulento con una alta capacidad de mezcla.

La capa límite atmosférica presenta cambios a lo largo del día debido a los procesos promovidos por el calentamiento solar de la superficie terrestre. Durante el día, el calentamiento del suelo provoca un aumento de temperatura de las capas más bajas de aire, las cuales se expanden disminuyendo su densidad, provocando el movimiento convectivo ascendente del aire. Al ascender, los volúmenes de aire se enfrían hasta alcanzar un equilibrio isocinético. Este fenómeno provoca una mezcla de los componentes troposféricos en un área vertical denominada “capa de mezcla”.

Con la puesta del sol la capa de mezcla desaparece abruptamente debido a la baja capacidad calorífica del suelo. Por las noches, la superficie terrestre es fría en comparación con las capas altas atmosféricas, provocando una inversión del gradiente térmico que se ve durante el día. Esto crea una zona estable restringiendo el transporte de los contaminantes hasta alturas más elevadas, que se traduce en un descenso de la altura de la capa límite [Mahlman, 1997].

Todos estos movimientos y transformaciones hacen que las concentraciones de los componentes de la atmósfera no sean constantes en el tiempo ni uniformes en el espacio. Sin embargo, los principales componentes químicos que pueden hallarse en la tropósfera se mantienen relativamente constantes, como el N<sub>2</sub>, O<sub>2</sub>, Ar, variando solo aquellos que aparecerán en nivel de trazas (ppm, partes por millón o ppb, partes por billón).

Ya sean naturales o artificiales, las sustancias emitidas a la atmósfera pueden reaccionar modificando su estado físico (gas, líquido o sólido) y/o químico. Los productos de transformación pueden diferir de los compuestos originales en sus propiedades químicas, toxicidad u otras características. Incluso, estas nuevas sustancias pueden ser retiradas de la atmósfera por fenómenos muy diferentes al de sus precursores [Seinfeld & Pandis, 2016]. Estas particularidades dificultan el estudio cualitativo y cuantitativo de la calidad del aire, requiriendo de un conjunto de técnicas diversas para su evaluación y pronóstico.

## 2.2 Contaminación del aire, efectos en la salud y aspectos legales

La tropósfera se encuentra fuertemente influenciada por la acción humana [Garratt, 1994]. Por su cercanía, la emisión de sustancias químicas a la atmósfera altera su composición, provocando un aumento de las concentraciones de los compuestos que se encuentran naturalmente presentes o la incorporación de sustancias exógenas. Este corrimiento del estado “puro” o “natural” es el principal indicador de una situación de contaminación.

Si bien la presencia de nuevas sustancias químicas hace de esta la situación más peligrosa para el medio ambiente, una importante área de la química se especializa actualmente en este campo reduciendo el riesgo. Tras el daño provocado por los clorofluorocarbonos a finales del siglo pasado, se cuenta con un mayor estado de consciencia respecto al efecto negativo que puede

provocar la liberación masiva de sustancias químicas, mejorando los controles necesarios para su utilización [Bostrom et al., 1994].

Como estrategia para monitorear la calidad del aire a nivel global, se ha planteado el control de algunas especies químicas que naturalmente están presentes en la tropósfera y que resultan buenos indicadores de contaminación cuando se encuentran a niveles elevados [World Health Organization, 2005; Krzyzanowski & Cohen, 2008]. Estos son los llamados contaminantes criterio: PM, NO, NO<sub>2</sub>, CO, O<sub>3</sub> y SO<sub>2</sub>. Un exceso de estos da cuenta de una situación de riesgo para la población y al ambiente [Krzyzanowski & Cohen, 2008]. En el **Cuadro 1** se detallan algunas características de estos contaminantes.

**Cuadro 1.** Contaminantes criterio. Características principales, fuentes de origen y efectos a la salud

### Ozono, O<sub>3</sub>

El O<sub>3</sub> constituye menos del 0,001% v/v del aire, pero es uno de los componentes más importantes de la atmósfera por su alta reactividad. Mientras que el O<sub>3</sub> presente en la estratósfera protege la biósfera absorbiendo la radiación ultravioleta de alta energía, en la tropósfera tiene un impacto negativo en la salud humana, el clima, la vegetación y los materiales [Pires et al., 2008]. Su presencia en la tropósfera se debe mayoritariamente al ciclo de reacciones fotoquímicas que incluyen los compuestos orgánicos volátiles (COVs) y los óxidos de nitrógeno (NO<sub>x</sub>).

El O<sub>3</sub> o sus precursores pueden ser transportados por grandes distancias, afectando a zonas alejadas de la fuente. Los efectos sobre la salud, debidos a una exposición excesiva al O<sub>3</sub>, afectan esencialmente al aparato respiratorio y a los ojos [Brunelli et al., 2007].

### Óxidos de nitrógeno, NO<sub>x</sub> (NO + NO<sub>2</sub>)

Los óxidos de nitrógeno (NO<sub>x</sub>), el óxido nítrico (NO) y el dióxido de nitrógeno (NO<sub>2</sub>), son un grupo de gases altamente reactivos relacionados con los procesos de quema de los combustibles fósiles. Fuentes naturales son los procesos biológicos anaeróbicos en el suelo y el agua, la actividad volcánica y la destrucción fotoquímica de los compuestos de nitrógeno en la atmósfera superior [Pires et al., 2008; Brasseur & Prinn, 2003].

Mientras que el NO es la forma primaria de producción, el NO<sub>2</sub> representa un contaminante secundario que deriva de la oxidación del NO en la tropósfera. Asimismo, constituye un contaminante intermedio para la producción de otros contaminantes como el O<sub>3</sub>, el ácido nítrico y el ácido nitroso, con un papel fundamental en la formación del smog fotoquímico [Chameides & Walker, 1973, Pires et al., 2008b].

Distintos estudios epidemiológicos han informado de asociaciones entre el NO<sub>x</sub> y los efectos adversos para la salud, como accidentes cerebrovasculares e infartos de miocardio [Hashim et al., 2004; Kalabokas et al., 2002; World Health Organization, 2000].

## Dióxido de azufre, SO<sub>2</sub>

El SO<sub>2</sub> se libera de los procesos de combustión que utilizan combustibles fósiles en los que el azufre está presente como impureza o de la industria metalúrgica. Las fuentes antrópicas más importantes del SO<sub>2</sub> son los equipos de calefacción domésticos y la contribución del tráfico vehicular (especialmente en el caso de los vehículos con motor diesel). Como fuentes naturales podemos identificar el sulfuro de dimetilo (DMS) proveniente del mar, que se oxida para formar SO<sub>2</sub> [Pires et al., 2008], y las erupciones volcánicas. El SO<sub>2</sub> puede oxidarse para formar ácido sulfúrico, el cual es un precursor de la formación de aerosoles [De Leeuw et al., 2011].

El SO<sub>2</sub> en altas concentraciones es un irritante del sistema respiratorio, provocando limitaciones en el flujo de aire. En algunos estudios el SO<sub>2</sub>, los sulfatos y los aerosoles ácidos se han asociado con un aumento de las visitas de emergencia y hospitalizaciones por asma [Kampa & Castanas, 2008].

## Material particulado, PM<sub>10</sub> y PM<sub>2.5</sub>

PM es el término genérico utilizado para los aerosoles atmosféricos, una mezcla compleja y variable de partículas suspendidas en el aire que varían en tamaño y composición, y que son producidas por una amplia variedad de actividades naturales y antropogénicas [Kampa & Castanas, 2008]. Los aerosoles troposféricos pueden ser introducidos en la tropósfera desde la superficie o pueden formarse mediante reacciones entre los componentes gaseosos de la atmósfera [Seinfeld & Pandis, 2016].

El tamaño de las partículas puede variar. Las partículas finas y ultrafinas (PM < 2.5 µm) son capaces de llegar a los alvéolos pulmonares y al torrente sanguíneo tras ser inhaladas. Se ha demostrado que existe una relación directa entre la exposición a las partículas en suspensión y el aumento de la mortalidad diaria y anual [World Health Organization, 2018]. Es por esto que la Agencia Internacional para la Investigación del Cáncer (IARC, del inglés *International Agency for Research on Cancer*) concluyó en 2014 que la exposición a PM es cancerígena para los seres humanos (IARC Grupo 1) causando cáncer de pulmón [Hamra et al., 2014].

Una gran cantidad de actividades antrópicas generan aerosoles, siendo las más peligrosas las emisiones de los vehículos, la combustión del carbón y una variedad de procesos industriales, como la metalurgia y la fundición, que generan partículas con metales pesados absorbidos.

Si bien los aerosoles son un componente menor de la atmósfera, presentan un efecto importante en el balance radiativo del sistema climático, comparable con los gases de efecto invernadero [Stocker et al., 2013, Dickerson et al., 1997].

## Monóxido de carbono, CO

El CO proviene principalmente del transporte vehicular, y en menor medida de las calderas de vapor, los procesos industriales, la eliminación de residuos sólidos y otras fuentes diversas [Levy, 2015].

Su mecanismo de acción en el cuerpo humano está bien identificado. Cuando se inspira, el CO se difunde rápidamente a través de la membrana capilar alveolar y se une a la hemoglobina, formando carboxihemoglobina (COHb) [Smithline et al., 2003], lo que provoca un deterioro inmediato en la entrega de oxígeno a los tejidos provocando hipoxia tisular [Gorman et al., 2003].

En este sentido, se habla de una buena calidad del aire cuando las concentraciones de los contaminantes son seguras para la población, estableciéndose márgenes de confianza para delimitar las cotas máximas aceptables. Estos valores surgen de una extensa bibliografía en estudios epidemiológicos que sirvieron para determinar las concentraciones mínimas a la cual la salud de las personas no se ve afectada por estos compuestos [Kampa & Castanas, 2008; World Health Organization, 200b]. En la **Tabla 1** puede observarse los niveles guías actualmente recomendados OMS para las distintas especies químicas.

**Tabla 1.** Niveles guías recomendados para cada contaminante.

<b>Contaminante</b>	<b>Período temporal</b>	<b>Concentración</b>
<i>PM<sub>2,5</sub></i>	1 año	10 µg.m <sup>-3</sup>
	24 h (percentil 99)	25 µg.m <sup>-3</sup>
<i>PM<sub>10</sub></i>	1 año	20 µg.m <sup>-3</sup>
	24 h (percentil 99)	50 µg.m <sup>-3</sup>
<i>Ozono, O<sub>3</sub></i>	8 h, máximo diario	100 µg.m <sup>-3</sup>
<i>Dióxido de nitrógeno, NO<sub>2</sub></i>	1 año	40 µg.m <sup>-3</sup>
	1 hora	200 µg.m <sup>-3</sup>
<i>Dióxido de sulfuro, SO<sub>2</sub></i>	24 horas	20 µg.m <sup>-3</sup>
	10 minuto	500 µg.m <sup>-3</sup>
<i>Monóxido de carbono, CO</i>	24 horas	6 mg.m <sup>-3</sup> (5 ppm)
	1 hora	30 mg.m <sup>-3</sup> (25 ppm)

Las investigaciones han mostrado que para algunos contaminantes existe un valor umbral a partir del cual se observan efectos a la salud. Sin embargo, otros contaminantes, como el PM, son nocivos desde pequeñas cantidades debiéndose fijar los valores de riesgo aceptable [World Health Organization, 2005].

## 2.3 Entre la legalidad y la seguridad ambiental

Los parámetros de calidad del aire establecidos por la OMS sirven como niveles guías para garantizar la protección de la población, pero estos no son de cumplimiento obligatorio por los Estados. En cada territorio se deben sancionar leyes que reglamenten los límites legales de los contaminantes en el aire. Estas leyes no siempre existen y cuando existen, no siempre se encuentran conforme los valores recomendados por la OMS [Vahlsing & Smith, 2012].

Las normativas son una expresión del contrato social entre el Estado y la ciudadanía, donde se busca ofrecer igual protección a todas las personas [Wang et al., 2016]. En su generación se pone en tensión distintos intereses políticos, sociales y económicos que, incluso en ciertos casos, pueden generar escenarios de conflicto, donde accionar sobre los agentes causantes de la contaminación implica perjudicar económicamente al sistema productivo [Amann et al., 2004], afectando de forma indirecta a la población que se busca proteger.

Nuevamente, este análisis no puede ser acotado a un campo de estudio o insensible a las distintas realidades que se presentan. En el territorio se produce la síntesis de todas las variables socioeconómicas y ambientales que determinan las desigualdades, generando complejos escenarios de vulnerabilidad social [Adger et al., 2007]. El territorio pasa a ser un determinante de la salud de la población, donde garantizar la seguridad implica, entre otras cosas, proveer un ambiente sano para el ejercicio y la recreación de la comunidad [Mitchell et al., 2015].

Pese a las dudas que pueden aparecer en la casta política, distintos estudios han demostrado que la intervención estatal en materia de calidad de aire genera beneficios económicos a largo plazo [Olsthoorn et al., 1999; Winiwarter & Klimont, 2011; Voorhees, 2005]. Es por ello que frente a los peligros que corre la población y el ambiente resulta importante obtener una buena caracterización del riesgo para tomar medidas basadas en el conocimiento.

## 3. El riesgo a la salud

En un esquema simplificado, la noción de riesgo puede definirse por dos aspectos bien diferenciados. Por un lado, el peligro que es la probabilidad de que ocurra un evento extremo, el cual viene determinado por las características de la amenaza o agente accionante. Por el otro, la vulnerabilidad que da cuenta del grado de susceptibilidad que tiene la población de ser afectada por la amenaza [Pizarro, 2001].

Para analizar el riesgo es necesario determinar la frecuencia con la que se producen eventos catastróficos y la magnitud de sus consecuencias [Kasperson et al., 1988]. La evaluación del riesgo es, por tanto, una construcción social, histórico-cultural, generada a partir de las percepciones del daño ocurrido [Natenzon, 2003].

Desde esta perspectiva, el nivel del riesgo no está sólo determinado por cuestiones externas, sino que se encuentra compuesto por múltiples factores sociales e individuales que son definidos por los sujetos y por la valoración que estos hacen del daño [Natenzon, 2003]. Por ejemplo, será mayor la repercusión de las consecuencias negativas en grupos con carencias económicas puesto que tienen un acceso desigual al sistema de salud [O'Neill et al., 2003].

La evaluación del riesgo es un mecanismo necesario para la gestión, pues permite discernir las áreas prioritarias a ser atendidas, facilitando la toma de decisiones y la definición de políticas públicas. A partir de la evaluación del riesgo es posible sopesar las amenazas asociadas a una actividad permitiendo seleccionar una estrategia de actuación que modifique los niveles de riesgo a los que están sometidos los individuos o la población [Chalvatzaki et al., 2019].

Como se vio anteriormente, la contaminación del aire se ha relacionado con una amplia gama de efectos sobre la salud, entre ellos enfermedades pulmonares, problemas cardíacos, asma y cáncer de pulmón [Brunekreef & Holgate, 2002]. Los estudios epidemiológicos en este campo brindan evidencias directas sobre la magnitud de los impactos a la salud debidos a la contaminación atmosférica [Hemminki & Pershagen, 1994], estimándose que ocasionan el 1.4% de la carga global de muertes, las cuales ocurren, mayoritariamente, en sujetos menores de 60 años [Ostro & World Health Organization, 2004].

Un elemento central para la cuantificación de los impactos en la salud son los modelos de concentración-respuesta (CRF, del inglés *concentration-response function*) [Héroux et al., 2015]. Las CRF dan cuenta de la relación entre los niveles de exposición a una sustancia y sus efectos en los sujetos de estudio, y sirven para conocer el daño que puede ocasionar una cierta concentración.

Como la contaminación del aire es una mezcla compleja, resulta muy difícil atribuir los efectos a un único componente [Peled, 2011]. Por tanto, resulta conveniente considerar que existen contaminantes indicadores sustitutivos de toda la mezcla. En particular, se ha demostrado que  $PM_{10}$  y  $PM_{2.5}$  son indicadores sólidos del riesgo asociado a la exposición de la mala calidad del aire [Ostro & World Health Organization, 2004; Levy et al., 2002].

Un modelo CRF general tiene la forma:

$$E(Y) = e^{(\beta \cdot X_i + \text{covariables})}$$

donde el valor esperado del parámetro de salud Y (por ejemplo, recuento de mortalidad) se modela como una función exponencial de las variables explicativas, siendo  $X_i$  la concentración ambiental del contaminante  $i$  y  $\beta$  es el factor pendiente o el log-riesgo relativo asociado con un cambio unitario en la concentración ambiente medida [Yeh & Small, 2002].

Es importante notar que la mayoría de los estudios de este tipo se han desarrollado en Europa y América del Norte y la composición química del aire y los sistemas de salud pueden ser muy diferentes en otros lugares, pudiendo variar las CRFs [World Health Organization, 2016]. Y como se mencionó anteriormente, la localización es una cuestión que considerar muy importante en el momento de analizar el riesgo, debido a la influencia que ejerce la posición geográfica de diversos componentes de la vida social en el territorio [O'Neill et al., 2003; Bernard et al., 2007].

## 4. Estrategias para el monitoreo de la calidad del aire

Es necesario realizar mediciones de forma periódica para conocer los niveles de los contaminantes a los cuales estamos expuestos dada la alta variabilidad de los contaminantes en la atmósfera. Se



define como monitoreo atmosférico a todas las metodologías diseñadas para muestrear, analizar y procesar en forma continua las concentraciones de sustancias o de contaminantes presentes en el aire en un lugar establecido y durante un tiempo determinado [Finlayson-Pitts & Pitts, 1999].

Las redes de monitoreo están formadas por un conjunto de estaciones y son diseñadas con el propósito principal de obtener información primaria que permita implementar medidas de protección de la salud pública. La importancia del monitoreo atmosférico radica en el poder de la información recolectada para generar conocimiento que facilite la gestión de la calidad del aire.

La base de datos que se genera aporta información útil para la realización de estudios científicos, como pueden ser estudiar las reacciones químicas atmosféricas, identificar los tipos específicos de contaminantes presentes en el ambiente, conocer las principales fuentes de emisiones y/o evaluar modelos de dispersión de contaminantes. A su vez, aporta información para interpretar estudios epidemiológicos que relacionen daños en la salud de la población y el ambiente con las concentraciones de los contaminantes.

Con esta información, podemos determinar el cumplimiento de las normas y los criterios legales, formular nuevos estándares de calidad de aire, llevar a cabo evaluaciones de tendencias a largo plazo, implementar estrategias de control, establecer políticas de desarrollo y medir los efectos de las medidas de control en la calidad del aire. Finalmente, brinda la posibilidad de otorgar información al público acerca de la calidad del aire, y proporcionar información de fuentes y del riesgo a la salud.

A continuación, se presentan distintas tecnologías empleadas ampliamente en el monitoreo de la calidad del aire.

## 4.1 Monitoreo Terrestre

El objetivo del monitoreo terrestre de calidad del aire es conocer las concentraciones máxicas de los contaminantes a la cual están expuestas las personas (concentración de inmisión), a partir de la obtención de muestras representativas y confiables del aire que se respira. Para esto se desarrolla un proceso de toma de mediciones de los componentes atmosféricos a nivel del suelo en distintos sitios y en diferentes momentos del año.

El monitoreo puede realizarse mediante distintas técnicas, las cuales se pueden agrupar en técnicas de muestreo discreto o continuo, dependiendo si las mismas se encuentran automatizadas o no [PNUMA/OMS 2002]. Los métodos continuos implican la captación y análisis del contaminante en el punto de muestreo, de forma continua y automática, mientras que los métodos discretos o discontinuos suponen la captación del contaminante en el punto de muestreo, el transporte del contaminante captado al laboratorio y el posterior análisis.

Los muestreadores pasivos y activos corresponden a metodología discretas, implicando un tratamiento de la muestra en el laboratorio. Mientras que los analizadores automáticos y los sensores *low-cost* son metodologías continuas, proporcionando mediciones en tiempo real.

Los métodos para medir los contaminantes atmosféricos cubren los más amplios rangos de costos y características de funcionamiento. En la **Tabla 2** figuran las ventajas y desventajas de estas técnicas. Cualquiera sea el método elegido es preciso contar con un mantenimiento periódico que incluya la calibración del equipamiento.

Resulta importante resaltar que estas metodologías de monitoreo no son excluyentes. En un plan de gestión, las metodologías discretas y continuas se complementan para obtener un mejor detalle de la variación de los niveles de los contaminantes en el territorio. Por lo general, los monitoreos discretos se realizan estacional o anualmente para, por ejemplo, obtener información de zonas que no se monitorean de forma continua o contar con una caracterización más precisa de la situación.

**Tabla 2.** Resumen de ventajas e inconvenientes de distintas metodologías de medición junto a sus métodos

Metodología	Método	Ventajas	Inconvenientes
<i>Discreta</i>	Muestreadores pasivos	* Muy sencillos. * Útiles para estudios de base.	* No útiles para algunos contaminantes. * Dan medias semanales o mensuales.
	Muestreadores activos	* Fácil de operar. * Operación segura. * Datos históricos. * Coste moderado	* Medias diarias. * Trabajo intensivo. * Requiere análisis en el laboratorio.
<i>Continua</i>	Analizadores Automáticos	* Comprobados. * Datos horarios. * Información en tiempo real.	* Muy costosos.
	Sensores Low-cost	* Datos cada 20 segundos. * Información en tiempo real. * Muy económicos	* No estandarizados * Aún no comprobados.

Muchos de los países industrializados cuentan con una amplia red de estaciones de monitoreo que cubren la mayor parte de las zonas rurales y urbanas proporcionando mediciones continuas de los contaminantes. Sin embargo, esto es económicamente inviable en muchos países, requiriendo otros métodos, como las mediciones satelitales, para estimar la concentración de los contaminantes [Garaga et al., 2018].

En el cruce de las diferentes tecnologías es posible obtener una red de monitoreo confiable con buena cantidad de información a un costo menor. Por ejemplo, integrando información de modelos de calidad de aire, información meteorológica, datos de monitoreo discreto, una red de sensores continuos y mediciones realizadas por teledetección.

## 4.2 Teledetección

La teledetección siempre ha estado a la vanguardia de la adquisición automatizada de información y ha producido tradicionalmente conjuntos de datos de gran tamaño [Lillesand et al., 2015]. La función principal del monitoreo de la calidad del aire por teledetección es ayudar a vigilar los niveles de los contaminantes atmosféricos de manera sistemática, coherente y frecuente,

reforzando el monitoreo terrestre y el desarrollo de modelos matemáticos espaciales [Wentz et al., 2014].

La teledetección atmosférica se define como la medición de gases, sólidos y líquidos presentes en la atmósfera en tiempo real mediante el uso de instrumentos ópticos de observación a distancia [Simonds et al., 1994]. Los instrumentos de teledetección ópticos utilizan energía lumínica para detectar y medir los contaminantes.

Podemos diferenciar las técnicas de teledetección en activa o pasiva, dependiendo de la fuente que se utilice. En la teledetección activa, un radar o un láser emite una señal electromagnética que interactúa con la materia, para luego evaluar los cambios producidos en la señal emitida. Por su parte, la teledetección pasiva aprovecha la energía solar, centrándose en la fracción de la energía que recibe del Sol.

En ambos casos, el funcionamiento se basa en fenómenos fisicoquímicos a nivel molecular, donde los componentes de la atmósfera interactúan con la luz, mediante la absorción o la dispersión del haz incidente [Simonds et al., 1994]. Las transiciones energéticas ocasionadas por la interacción entre la materia y la luz son características de la sustancia, y pueden ser utilizadas para identificar y cuantificar las especies químicas presentes.

Cuando un fotón o cuanto de energía colisiona con un átomo o molécula, si la energía del fotón es igual a la diferencia energética entre el nivel energético ocupado (estado basal) y el nivel excitado, entonces el fotón es absorbido por la partícula. Este nuevo estado energético es inestable y, tras un breve tiempo, la energía extra se perderá mediante alguno de los mecanismos de disipación, siendo posible la emisión de un fotón. Los espectros de absorción y emisión son característicos de las sustancias y, si bien son de escala microscópicas, pueden ser registrados por instrumentos ópticos lejanos.

En la **Tabla 3** se resumen las regiones ópticas donde se utiliza la teledetección atmosférica junto con las especies químicas que comúnmente se monitorean. La relación entre la intensidad de la absorción y la concentración viene expresada por la Ley de Bouguer-Beer-Lambert:

$$A_{\lambda} = -\log(T) = a_{\lambda} \cdot c \cdot l$$

donde  $A_{\lambda}$  y  $a_{\lambda}$  son la absorbancia y la absortividad a la longitud de onda  $\lambda$ , respectivamente;  $T$  es la transmitancia;  $c$  es la concentración de la muestra y  $l$  es la longitud del camino del haz de luz.

Los fenómenos de absorción-emisión no son los únicos empleados para identificar las especies químicas presentes en la atmósfera, también la dispersión puede ser utilizada, en particular para detectar la presencia de aerosoles [Simonds et al., 1994].

Una gran parte de las investigaciones en calidad de aire por teledetección se centra en el uso de láseres y en la fotometría solar. La red AERONET (*AERosol Robotic NETwork*) es una extensa red global compuesta por fotómetros solares [Holben et al., 1998]. Las redes EARLINET (*European Aerosol Research LiDAR Network*) [Bösenberg & Matthias, 2003], MPLNET (*Micro-Pulse LiDAR Network*) [Welton et al., 2006], AD-Net (*Asian Dust and Aerosol LiDAR Observation Network*), REALM (*Regional East Aerosol LiDAR Mesonet*) [Sugimoto et al., 2014] presentan sistemas del tipo LiDAR (*Light Detection and Ranging*). A las redes antes mencionadas se le suman las distintas campañas de vuelo con sensores LiDAR, para registrar desde el aire la composición de la atmósfera a lo largo del planeta, como son las campañas ATom [Wofsy et al., 2018], ORACLES [Pi et al., 2017] o AEROCLO-SA [Formenti et al., 2019].

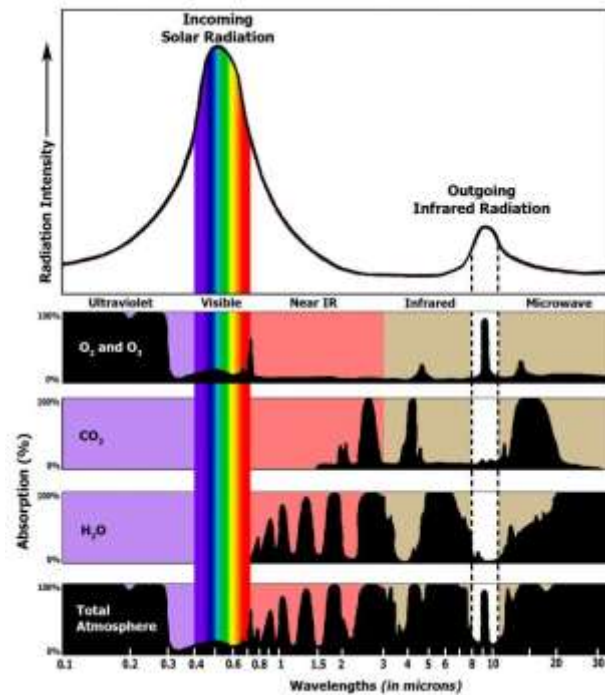
A diferencia de estos métodos, el monitoreo satelital proporciona una extensa cobertura, pudiendo monitorear amplias zonas del mundo e incluso áreas inaccesibles [Engel-Cox et al., 2004]. Mediante la captura de la irradiancia en distintos rangos del espectro electromagnético (en particular, en el UV, el visible y el IR cercano) los satélites brindan datos físicos para deducir mediante la aplicación de determinados algoritmos las concentraciones de ciertas especies químicas en la columna atmosférica para extensas áreas.

**Tabla 3.** Regiones ópticas y técnicas de teledetección [Modificado de Simonds et al., 1994]

Región óptica	Rango de longitud de onda	Principios de detección	Compuestos	Comentarios
<i>UV</i>	200-400 nm	Absorción	SO <sub>2</sub> , H <sub>2</sub> S, OH, NO <sub>x</sub> , Hg, Cl <sub>2</sub> , compuestos aromáticos	Interferencias del O <sub>2</sub>
<i>Visible</i>	400-700 nm	Absorción	SO <sub>2</sub> , NO <sub>2</sub> , O <sub>3</sub> , H <sub>2</sub> O	Poco compuestos absorben en esta región
<i>IR cercano</i>	700-2500 nm	Absorción	CO <sub>2</sub> , CO, H <sub>2</sub> O, O <sub>2</sub>	No para especies traza debido a su débil absorción
<i>IR medio</i>	2500-25000 nm	Absorción, Emisión, Fluorescencia	COVs, gases	Interferencias del H <sub>2</sub> O y CO <sub>2</sub>
<i>IR lejano</i>	0.025 - 0.5 mm	Absorción, Emisión	H <sub>2</sub> O, O <sub>3</sub> , N <sub>2</sub> O, NO <sub>2</sub>	No se utiliza por las fuertes interferencias por la absorción del H <sub>2</sub> O

Los sensores empleados en el estudio de la calidad de aire tienen una buena resolución radiométrica, permitiendo identificar los componentes de la atmósfera, y un período de revisita diario o de pocos días, pero una resolución espacial moderada que los hacen no aplicables a estudios de micro-escala (< 1 km). Pese a esto, su uso brinda información confiable de fácil integración para el modelado de la calidad del aire en las ciudades [Stafoggia et al., 2017; Just et al., 2015; Hoek et al., 2015, Young et al., 2016].

La **Figura 4** muestra de forma esquemática los espectros de absorción para el O<sub>2</sub>, O<sub>3</sub>, CO<sub>2</sub> y H<sub>2</sub>O y su superposición para conformar el espectro atmosférico. Esta figura también nos muestra las secciones del espectro electromagnético más propicias para operar con teledetección. Las técnicas de espectroscopia desde satélites trabajan a rangos de longitudes de onda precisos para cada uno de los gases en estudio.



**Figura 4.** Espectro de absorción de los gases de efecto invernadero [Adaptado de Climate Science Investigations, 2019]

El proceso de recuperación (o de determinación de la concentración) de un gas en la columna atmosférica requiere tres pasos principales. En primer lugar, se mide la radiancia a la longitud de onda de interés mediante espectroscopia de absorción óptica diferencial (DOAS, del inglés *differential optical absorption spectroscopy*) para obtener la concentración del gas en la columna inclinada por la Ley de Beer-Lambert-Bouguer [Richter y Wagner, 2011]. Luego, si se desea obtener la concentración en la columna troposférica, se elimina la señal estratosférica [Boersma et al., 2011; Bucsela et al., 2013]. Para ello, se generan los perfiles verticales de los gases utilizando modelos de transporte químico. Finalmente, la concentración de la columna inclinada troposférica (SCD, del inglés *slant column density*) debe convertirse a la concentración de la columna vertical (VCD, del inglés *vertical column density*) mediante el uso de un factor de masa en aire (AMF, del inglés *air mass factor*) según la ecuación:

$$VCD = SCD / AMF$$

Dependiendo del algoritmo específico, las interferencias producidas por las nubes pueden ser ignoradas o corregidas [Laughner et al., 2016].

Los métodos de recuperación de la concentración de aerosoles son distintos al de los gases ya que se basan en el fenómeno de dispersión de la luz. En el espectro del UV cercano, la reflectancia atmosférica está regida fuertemente por la dispersión de Rayleigh y esta relación puede utilizarse para estimar los aerosoles atmosféricos [Torres et al., 2007]. El parámetro utilizado para cuantificar los aerosoles es AOD (del inglés, *Aerosol Optical Depth*), un número adimensional que da cuenta de la cantidad de partículas presentes.

Los algoritmos de recuperación evalúan la dispersión provocada por las partículas atmosféricas que dirigen energía desde afuera del campo de vista del sensor hacia el sensor. Como estas fuentes de luminosidad no vienen del objeto, actúan como una máscara de luminosidad uniforme, causando que áreas oscuras aparezcan con mayor luminosidad y viceversa. Estos algoritmos requieren información sobre el albedo de dispersión única de aerosoles (SSA, del inglés *single scattering albedo*) para inferir el nivel de AOD. Ejemplos de algoritmos de recuperación son el *Dark Target Land and Ocean* [Teillet & Fedosejevs, 1995], el *Deep Blue* [Hsu et al., 2013] o el *MAIAC* [Lyapustin et al., 2011].

La presencia de nubes dificulta la recuperación de AOD. La contribución a la reflectancia total en el TOA por parte de la dispersión provocada por los aerosoles es muy pequeña en comparación con la reflectancia generada por las nubes. Por tanto, la contaminación por cirros o nubes de tamaño de subpíxel provoca una subestimación del SSA aumentando el error en su recuperación [Torres et al., 1998]. Como salida a esto, se utiliza la capacidad de absorción de los aerosoles en el UV permitiendo ser caracterizados incluso sobre fondos muy brillantes [Torres et al. 1998].

## 5. Modelos de calidad del aire

Los datos generados en el monitoreo de la calidad del aire, por sí mismos, resultan insuficientes para establecer una representación detallada del comportamiento espacial y temporal de los contaminantes. Mediante algoritmos matemáticos es posible predecir los valores en puntos distantes, del cual no se disponen de medidas, mediante la información recogida en otros sitios. Los modelos matemáticos son representaciones conceptuales que permiten representar el estado de los contaminantes y ampliar la cobertura espacial y/o temporal de las observaciones [Beaulant et al., 2008; Janssen et al., 2008]. Los modelos pueden utilizarse como información suplementaria o como fuente exclusiva de información para la evaluación, de esta forma se permite reducir el número de estaciones de vigilancia en una zona [Jacobson, 2005].

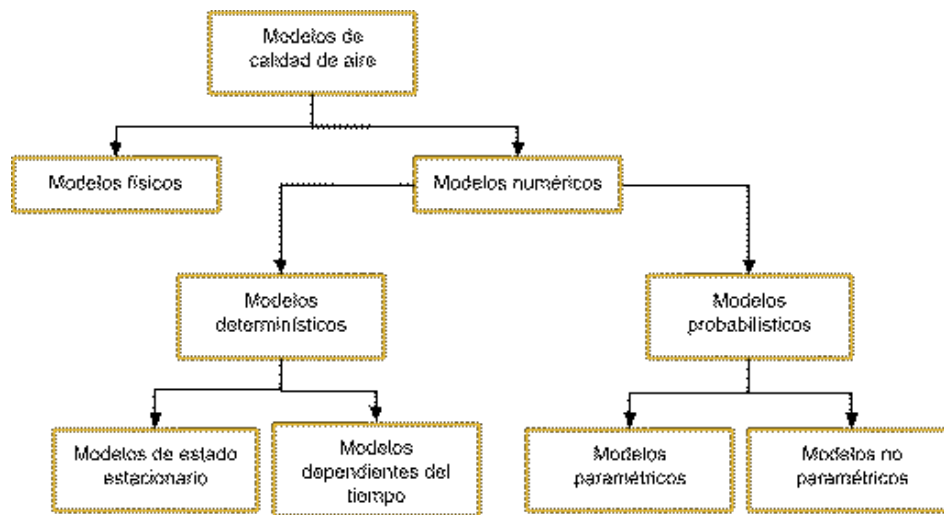
Un modelo es un sistema abstracto en el que los elementos que intervienen se conocen como variables, y las relaciones entre ellas vienen expresadas mediante conjuntos de ecuaciones. De esta forma, el uso de modelos permite entender mejor los fenómenos, comprender las propiedades físicas, químicas, dinámicas y radiativas de la contaminación del aire y la meteorología. A su vez, brinda información sobre los procesos que tienen lugar en una determinada zona y cómo estos interaccionan entre sí [Jacobson, 2005].

El empleo de modelos en la gestión facilita la tomar decisiones, y permite optimizar los recursos, ensayar hipótesis o probar modificaciones que serían imposibles de otra manera. Esto hace que se utilicen en la confección de planes de mejora o para evaluar si se superan los valores límites u objetivos en ciertas condiciones. También, son herramientas de ayuda para la preparación de propuestas de adopción de medidas, planes ó programas de reducción en el caso de superación de las concentraciones límites establecidas.

Finalmente, la modelización es utilizada en la predicción de la calidad del aire. Las simulaciones permiten estudiar el comportamiento futuro del sistema, evaluar su evolución con el tiempo o frente a un cambio de las condiciones climáticas. Otorgan la posibilidad de generar predicciones a corto plazo sobre la evolución de la situación y con ellos, realizar avisos a la población o desplegar protocolos de acción.

Los modelos de calidad de aire se pueden clasificar en 2 grandes grupos (**Figura 5**). Por un lado, encontramos los modelos físicos que son representaciones a escala de laboratorio del transporte atmosférico en una región. Por ejemplo, en un modelo de túnel de viento se coloca una maqueta que representa el área de estudio dentro del túnel de viento y se evalúa la fluidodinámica de los eventos que pudieran ocurrir. Por otro lado, los modelos numéricos que comprenden los modelos deterministas y los modelos estocásticos. En la siguiente subsección se presentan estos modelos con más detalle.

Cualquiera sea el modelo elegido, es importante señalar que los resultados que se obtienen con un modelo nunca pueden ser mejores que la calidad de los datos utilizados y que los mismos estarán atados a las hipótesis y supuestos que han sido adoptados para su realización [Janssen et al., 2008].



**Figura 5.** Familia de modelos empleados en el estudio de la calidad del aire

## 5.1 Modelos numéricos determinísticos

La modelización mediante modelos numéricos deterministas es el procedimiento que se realiza sobre un fenómeno, proceso o sistema con el fin de identificar las teorías y variables que lo definen. Los modelos determinísticos son considerados una herramienta indispensable en el estudio de la calidad del aire por incorporar un marco teórico y conceptual en su elaboración.

Estos modelos se basan en las descripciones fisicomatemáticas de la atmósfera y muestran una gran aplicación en la predicción de la dinámica atmosférica. Las ecuaciones gobernantes son las Ecuaciones Primitivas (**Tabla 4**), un conjunto de ecuaciones diferenciales parciales, no lineales, imposibles de resolver con exactitud.

Siendo el aire un fluido, para un elemento de volumen de dimensiones  $\Delta x$ ,  $\Delta y$ ,  $\Delta z$ , la Ecuación de continuidad describe la velocidad de la variación de la densidad; la Ecuación de Navier-Stokes o Ecuación de movimiento describe la aceleración de un pequeño elemento de volumen por las fuerzas que actúan sobre este; y la Ecuación de la energía total describe la conservación de la energía como una sumatoria de la energía térmica, cinética y potencial [Bird et al., 1992].

**Tabla 4. Ecuaciones primitivas.** Conjunto de ecuaciones que gobiernan la atmósfera: la Ecuación de continuidad, la Ecuación de Navier-Stokes y la ecuación de energía total. Donde  $\frac{D}{Dt}$  es la derivada sustancial o material:  $\frac{D(*)}{Dt} = \frac{\partial(*)}{\partial t} + (v \cdot \nabla)(*)$ ,  $\nabla^2$  es el operador laplaciano:  $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ ,  $\rho$  es la densidad,  $t$  es el tiempo,  $v$  es la velocidad local del fluido,  $p$  es la presión estática,  $\nabla$  es el operador gradiente,  $g$  es la aceleración de la gravedad,  $\mu$  es la viscosidad,  $\hat{U}$  la energía interna por unidad de masa,  $q$  es el vector densidad de flujo de calor,  $\hat{\Phi}$  la energía potencial por unidad de masa,  $\tau$  es el tensor de esfuerzo cortante.

*Ecuación de Continuidad*

$$\frac{D\rho}{Dt} = -\rho (\nabla \cdot v)$$

*Ecuación de Navier-Stokes*

$$\rho \frac{Dv}{Dt} = -\nabla p + \mu \nabla^2 v + \rho g$$

*Ecuación de Energía Total*

$$\rho \frac{D(\hat{U} + \hat{\Phi} + \frac{1}{2}v^2)}{Dt} = -(\nabla \cdot q) - (\nabla \cdot pv) - (\nabla \cdot [\tau \cdot v])$$

Debido a la naturaleza caótica de las Ecuaciones primitivas resulta prácticamente imposible predecir con certeza absoluta el estado de la atmósfera en un cierto momento [Lorenz, 1964]. Es por esto que, a partir, de diferentes tipos de métodos numéricos, considerando sólo los fenómenos más relevantes e imponiendo condiciones iniciales y de contorno, se generan soluciones aproximadas para las ecuaciones que facilitan el estudio de la atmósfera [Lamb y Seinfeld, 1973].

Los modelos Gaussianos son los modelos más sencillos, ya que brindan soluciones para el estado estacionario [El-Harbawi, 2013]. Los modelos Gaussianos describen la distribución tridimensional de una pluma de contaminación (con distribución Gaussiana, de allí su nombre) bajo condiciones meteorológicas y de emisión estacionarias [Leelőssy et al., 2014]. Un ejemplo del uso de modelos Gaussianos se observa en el software científico AERMOD [Cimorelli et al., 2005].

Los modelos que muestran dependencia con el tiempo se dividen en función de la forma de analizar la mecánica en medios continuos. La descripción de Lagrange consiste en hacer un seguimiento de las partículas, mientras que la descripción de Euler consiste en medir lo que sucede en puntos fijos del espacio [Collett & Oduyemi, 1997]. Ambas descripciones son equivalentes, salvo que en ciertos análisis puede resultar una más simple que la otra.

Los modelos Lagrangianos son modelos de celda fija, donde el enfoque matemático es seguir la trayectoria de una parcela de aire. Los modelos Lagrangianos tienen una capacidad más limitada que los modelos Eulerianos de incorporar procesos químicos. Estos modelos se centran en los procesos de transporte por advección, aunque también se pueden incorporar los procesos convectivos de la atmósfera [Collett & Oduyemi, 1997; Leelőssy et al., 2014]. Un ejemplo de aplicación de modelos Lagrangianos es el modelo GEOS-5 de circulación general atmosférica (AGCM, del inglés *atmospheric general circulation model*) [Lin, 2004].

Los modelos Eulerianos apuntan a incluir los procesos atmosféricos relacionados a la dispersión de contaminantes y a su transformación química. Existen distintos grados de complejidad en los modelos Eulerianos en términos de la representación de procesos, desde modelos que consideran sólo la dispersión, hasta modelos que incluyen procesos fotoquímicos. Estos modelos discretizan



el espacio en forma de grilla tridimensional e integran las ecuaciones matemáticas en el tiempo para cada una de las celdas, generando la información de la evolución temporal de las concentraciones en todas las dimensiones [Collett & Oduyemi, 1997]. Ejemplos de modelos Euleriano son el WRF-Chem [Skamarock & Klemp, 2008], CAMx [ENVIRON, 2014] y CMAQ [Binkowski & Roselle, 2003].

Un tercer tipo de modelos son los modelos combinados o tipo “puff”. Estos modelos calculan la dispersión de los contaminantes provenientes de una emisión instantánea (llamada “puff”) a lo largo de su recorrido [Collett & Oduyemi, 1997]. En caso de ser una emisión continua se simulan múltiples puff sucesivos. Su aproximación matemática consiste en estimar la dispersión en forma Gaussiana en cada punto de la trayectoria. Como ejemplo de modelo tipo puff se encuentra el CALPUFF [Scire et al., 2000].

La modelización determinista ha realizado grandes contribuciones al estudio de la atmósfera. Sin embargo, presenta una serie de desventajas a considerar. Si bien estos modelos tienen su fuerte en el planteamiento teórico, la resolución espacial limita la comprensión de los fenómenos que suceden en escalas más pequeñas que la malla del modelo [El-Harbawi, 2013] y se encuentran limitados en la incorporación de procesos relevantes en escalas temporales mayores [van Donkelaar et al., 2010].

La aplicación de modelos deterministas requiere de un amplio conocimiento de las fuentes de emisiones (inventario de emisiones), la naturaleza del contaminante, las características meteorológicas y los efectos del relieve [Engel-Cox et al., 2013; El-Harbawi, 2013]. Con una dificultad adicional para el modelado de contaminantes secundarios, donde se debe considerar las condiciones de contorno del modelo y el periodo de simulación [Jacobson, 2005].

A su vez, muchas veces se cuenta con un alto grado de incertidumbre de las condiciones iniciales y muchos de los resultados no pueden ser verificados [Wang et al., 2017 a]. Aunque en ciertos casos es posible corroborar los resultados de las simulaciones con observaciones, estas confirmaciones son parciales. Finalmente, estos modelos demandan un alto grado de experticia y computación de alto desempeño [Jiang et al., 2017], por lo que su implementación debe analizarse caso a caso.

## 5.2 Modelos matemáticos probabilísticos

Un enfoque alternativo para abordar los sistemas caóticos es basarse en la probabilidad de ocurrencia de los eventos [Zeng et al., 1993]. El modelado probabilístico, también llamado empírico o estocástico, se basa en el estudio de los eventos históricos, a fin de determinar la frecuencia de los sucesos y la importancia que presentan algunas variables para predecir este comportamiento [Watson et al., 1988]. A su vez, estos métodos permiten tener una estimación de la incertidumbre, de tal manera que el usuario final puede evaluar con más información el valor de la predicción generada.

Existen muchas variantes de modelos estocásticos que pueden agruparse en dos grandes grupos: los modelos paramétricos y no paramétricos. Los modelos paramétricos asumen que un conjunto finito de parámetros captura toda la información relevante de los datos. Ejemplo de esto son los modelos lineales (ML), los modelos lineales generalizados (GLM) o los modelos generalizados mixtos (GLMM), donde se toma como hipótesis que los errores del modelo tienen la forma de una

función de distribución de probabilidad conocida [Westerlund et al., 2014; Hasenfratz et al., 2015; Biancofiore et al., 2017].

Por su parte, los modelos no paramétricos consideran que la distribución de los datos no puede ser definida en términos de un conjunto finito de parámetros. Esto hace que sean más flexibles a modelar variables que presenten un comportamiento complejo, como la heteroscedasticidad, la tendencia temporal y/o los comportamientos cíclicos. Los métodos más usados en el aprendizaje automático (en inglés, *machine learning*) son esencialmente no paramétricos, como las redes neuronales, máquinas de vectores de soporte (SVM, *support vector machine*), Random Forest, entre otros. Su aplicación requiere de una gran cantidad de datos históricos, haciendo que la aplicabilidad de estos algoritmos se encuentre limitada a ciertos conjuntos de datos [Jiang et al., 2017].

En el estudio de la calidad del aire se han observado un amplio uso de modelos estocásticos para la predicción temporal o espacial. Los enfoques más innovadores incluyen la regresión paso a paso (en inglés, *stepwise regression*) y el análisis de la transformada Wavelet [Russo et al., 2015], redes neuronales artificiales [Sekar et al., 2015; Csépe et al., 2014; Gong & Ordieres-Meré, 2016], lógica difusa [Gong & Ordieres-Meré, 2017; Dincer & Akkus, 2018; Jiang et al., 2017], y los modelos ocultos de Markov [Gómez-Losada, 2017; Domanska et al., 2016].

Una amplia tendencia es el desarrollo de modelos multivariados, como los modelos de regresión de uso del suelo (LUR, según sus siglas del inglés: *land use regression*). Los modelos LUR utilizan un conjunto de variables explicativas para modelar las concentraciones de contaminación. Para eso se utilizan como variables predictoras información de uso del suelo, datos sobre el tráfico y otras variables físicas y ambientales relacionadas con la contaminación atmosférica [Westerlund et al., 2014, Shi et al., 2017 b, Yang et al., 2017].

En la literatura científica, los modelos LUR se pueden encontrar basados en Modelos Lineales Mixtos Generalizados (GLMMs) [Yang et al., 2018 a], Modelos Aditivos Generalizados (GAMs) [Hasenfratz et al., 2015; Hastie, 2017], algoritmos de *machine learning*. El potencial de estos modelos es predecir la concentración de un contaminante en sitios que no son monitoreados, resultandos precisos para el análisis de concentraciones de exposición [Beelen *et al.*, 2013]. También presentan una gran aplicación en el campo de la epidemiología, con modelos que relacionan los niveles de contaminantes con los efectos sobre la salud [Lin et al., 2016; Wu et al., 2016].

## 6. Big data, ciencia de datos y minería de datos

Las mediciones automáticas de contaminantes atmosféricos, ya sea mediante monitoreo terrestre o teledetección, sumado a los múltiples modelos disponibles para evaluar las concentraciones en el tiempo y el espacio, brindan una enorme cantidad de datos inimaginables hace medio siglo atrás. La masividad de datos transforma nuestra comprensión de los sistemas ambientales y, por tanto, genera nuevos paradigmas en la ciencia que se realiza [Kitchin, 2014].

*Big Data* (BD) es el nombre acuñado para hablar de enormes volúmenes de datos que se crean en tiempo real, diversos en variedad, estructurados y no estructurados en naturaleza, de alcance exhaustivo, desarrollados para captar poblaciones o sistemas enteros, con una resolución espacial de mallado fino, de naturaleza relacional, flexibles en estructura, extensibles a nuevas variables y

escalables rápidamente en tamaño [Kitchin, 2014; Sivarajah et al., 2017]. El reto que presenta su abordaje es hacer frente a la abundancia, la variedad, el desorden y el dinamismo, y el hecho de que gran parte de estos datos responden a preguntas anacrónicas o son un producto de otras actividades o experiencias [Han et al., 2011; Hastie et al., 2009]. A su vez, el BD presenta como desafío la adopción de un marco epistemológico donde quepan la complejidad del fenómeno en estudio y las herramientas para su abordaje [Kitchin, 2014].

La ciencia de datos o ciencia guiada por los datos, en contraste con las actuales tendencias al empirismo, plantean un abordaje diferente del diseño deductivo tradicional. Su abordaje busca generar nuevas hipótesis a partir de los datos [Kelling et al., 2009]. Esto es posible al incorporar un modo inductivo en el diseño de la investigación que permite identificar nuevas preguntas relevantes de ser examinadas y probadas.

El enfoque de la ciencia de datos contribuye en el estudio de la calidad del aire proporcionando una metodología de trabajo estructurada que organiza y simplifica los pasos, desde la adquisición de datos hasta la producción de información para los usuarios finales. La estrategia epistemológica de la ciencia de datos plantea un proceso de descubrimiento del conocimiento guiado por la teoría existente que dirige el proceso [Bellinger et al., 2017; Chen et al., 2017]. Como tal, la forma en que se generan o se vuelven a utilizar los datos es regida por suposiciones sustentadas por los conocimientos teóricos y prácticos. Del mismo modo, la forma en que los datos se procesan, gestionan y analizan viene dada por suposiciones en cuanto a las técnicas de análisis de datos que podrían proporcionar información significativa [Han et al., 2011].

Se conoce como minería de datos (*Data Mining*) al conjunto de herramientas utilizadas para extraer (o producir) información de grandes volúmenes de datos a través de la automatización. Para facilitar el trabajo con BD se han elaborado distintos modelos de procesos de minería de datos. Un modelo de procesos puede definirse como el conjunto de actividades y tareas marco para realizar el trabajo, incluyendo insumos y productos en cada tarea [Pressman, 2005]. El objetivo final de un modelo de procesos es hacer el proceso manejable, repetible y cuantificable. Ejemplos de estos son el pionero modelo KDD (del inglés, *knowledge discovery in databases*) desarrollado por Fayyad et al. [1996 a, b, c] y el ampliamente usado modelo CRISP-DM [Chapman et al., 2000] para el desarrollo de modelos de minería de datos.

Mientras que los modelos de procesos de minería de datos definen qué hacer, las metodologías definen cómo hacerlo. Una metodología es una parte constitutiva de un modelo de proceso, en la que, junto a las tareas, los insumos y los productos, se debe especificar la forma en que las mismas se llevan a cabo [Yang & Wu, 2006; Mariscal et al., 2010]. En [Mariscal et al., 2010] se presentan una serie de fases que se repiten en las distintas metodologías de minería de datos, estas son: la especificación del problema, la preparación de datos, el modelado, la evaluación del modelo y la producción de la información (o nuevas hipótesis) como un avance del conocimiento en el campo de aplicación [Sivarajah et al., 2017; Shah et al., 2015].

## 7. El riesgo, la incertidumbre, la ciencia y la gestión

Como se mencionó hasta aquí, conocer la calidad del aire no es una tarea sencilla. Requiere de una exhaustiva recolección de datos, del análisis e interpretación de los resultados y de la

integración de la información obtenida con los antecedentes y la normativa en el tema. A su vez, en muchas ocasiones, se precisa el trabajo en colaboración con distintas instituciones, para obtener la información necesaria. Las dificultades comienzan cuando alguna de estas partes falla.

Mientras en las ciudades más pobladas de los países industrializados se sostiene un monitoreo continuo de la calidad del aire, la situación es distinta en países periféricos donde no se cuenta con información de la calidad del aire o el monitoreo es eventual [Silva y Quiroz, 2003]. Como la adquisición de un equipo certificado para el monitoreo de la calidad del aire resulta altamente costosa en términos de instalación y de operación, muchas veces la gestión de la calidad del aire se limita a esporádicos monitoreos discretos en sitios específicos que solo aportan una imagen estática de la situación. La falta de datos a lo largo del tiempo y espacialmente incompleta complejiza la gestión de la calidad del aire, incluso comprometiendo la credibilidad de las mediciones oficiales.

Para saldar esta carencia se pueden utilizar modelos matemáticos que, como se vio, brindan un pronóstico de lo que sucede en el territorio que no se monitorea. Sin embargo, los modelos son representaciones de la realidad, y aunque se realizan esfuerzos para representar el mundo tal cual es, cada modelo requiere de suposiciones teóricas que definen sus propias limitaciones. A su vez, los resultados brindados obedecen a la calidad de los datos recopilados y la falta de datos obliga a hacer aproximaciones reduciendo la precisión de los resultados.

De esta forma, queda planteada una interesante paradoja entre la necesidad de información para gestionar y la incertidumbre sobre la calidad de la información con la que se cuenta. Hacer en la administración, por tanto, conlleva un riesgo, donde el riesgo se encuentra en el borde entre el conocimiento y la ignorancia [Sørensen, 2018]. Según la teoría social del riesgo, el riesgo es parte de la toma de decisiones. Seleccionar una posible acción implica asumir las eventuales contingencias que dicha acción pueda traer aparejadas [Cohen, 2007]. Y si bien el riesgo refiere a un cálculo de probabilidades (inmerso en un contexto sociocultural), es imposible determinar con certeza absoluta el futuro [Beck & Beck-Gernsheim, 2002].

Por esta razón, las discusiones acerca del riesgo implican reflexionar sobre el conocimiento científico, su finalidad, las limitaciones que presenta, la condición irreductible de la incertidumbre y el carácter reduccionista de las soluciones. Para ello, se requiere dejar atrás el concepto de verdad para adoptar el concepto de control o de gestión, alejándonos de la necesidad de obtener predicciones precisas, para pasar a la necesidad de controlar los futuros eventos, aceptando la inexactitud cuantitativa, a favor de una precisión cualitativa que convenga mejorar [Funtowicz & Ravetz, 2000].

La gestión, por lo tanto, demanda un flujo de trabajo dinámico, donde se priorice la producción de resultados y la obtención de productos. En este sentido, la elaboración de metodologías y sistemas de información que faciliten la selección de las fuentes de datos, sistematicen las preguntas más relevantes y condensen los mejores métodos actuales contribuyen a una más rápida y mejor toma de decisiones.

En base a lo aquí expuesto, esta tesis doctoral retoma las experiencias desarrolladas en la temática y propone una metodología para el estudio y la evaluación de la calidad del aire incorporando el enfoque sistémico de la ciencia de datos. Con su uso en este campo se espera contribuir a una mejor administración de los datos disponibles para un mayor beneficio de la sociedad.



*“Cada método esclarece lo que le conviene  
y solo una constelación de métodos puede  
captar el silencio que persiste entre  
cada lengua que pregunta.”*

*De Sousa, 2009*

### *Capítulo 3*

# **Propuesta metodológica para el estudio de la calidad del aire**

---

## **Introducción**

En el capítulo anterior se presentaron los distintos conceptos teóricos que acompañan el estudio y la gestión de la calidad del aire, quedando enmarcados dentro de un contexto socioeconómico que lo contienen, limitan y enriquecen. A lo largo de este capítulo se detalla y justifica una propuesta metodológica para el estudio de la calidad del aire, la cual se describe en términos de procedimientos prácticos que facilitan la adquisición, proceso, interpretación y comunicación de la información.

Como se ha mencionado en la Introducción General de esta tesis, cada escenario de aplicación plantea dificultades particulares y cuenta con limitadas herramientas para su análisis. El desafío que esta tesis aborda consiste en generar una propuesta de trabajo que sintetice los avances elaborados en los estudios anteriores y, a su vez, facilite la recolección, manejo y generación de nuevo conocimiento

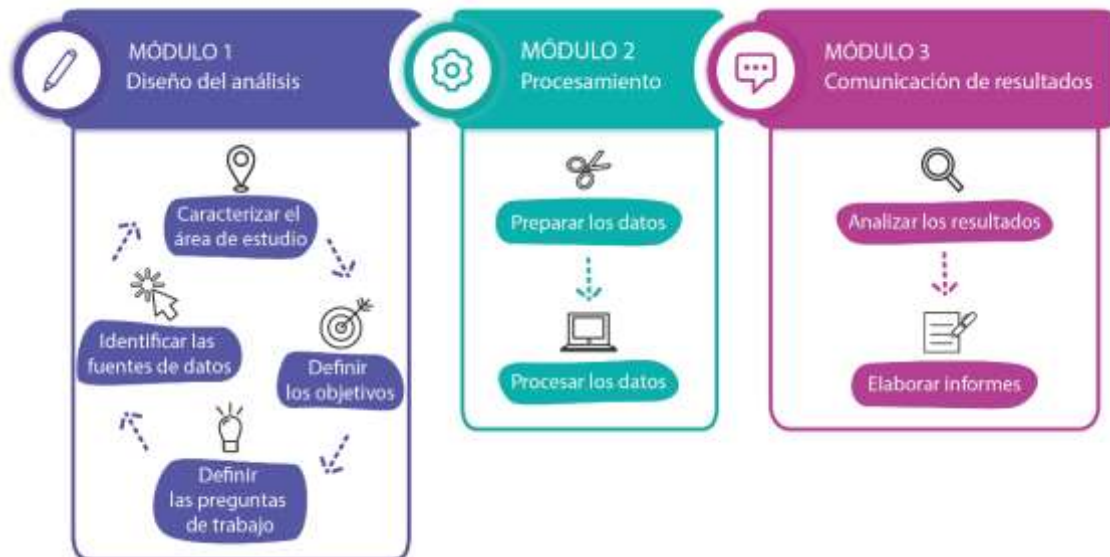
El marco de trabajo que aquí se esboza, y en los capítulos siguientes se implementará, resulta una guía de trabajo de fácil aplicación mediante la cual se puede obtener un análisis completo de la calidad de aire en función de los datos disponibles. Para simplificar su lectura, se presenta un flujo de trabajo que ordena y sintetiza los pasos necesarios en la evaluación de la calidad del aire. Este esquema de trabajo es un punto de partida, el cual no debe ser tomado en forma rígida o absoluta, sino como una propuesta que esquematiza los elementos a considerar en el abordaje de la problemática.

La generación del flujo de trabajo no fue arbitraria, por el contrario, se generó a partir de una serie de reglas de diseño. Como características generales, se pensó en que la metodología resulte de fácil integración con modelos en funcionamiento, no resultando excluyente con tecnologías anteriores y que fuera adaptable a los nuevos avances en el tema. Se buscó integrar algoritmos que fueran robustos, novedosos y fácilmente aplicables, sin requerir grandes tiempos de procesamiento para su mantenimiento, y que hicieran un uso intensivo de los datos disponibles aprovechando las fuentes mundiales de información.

Un flujo de trabajo plantea una secuencia de operaciones que se relacionan entre sí, y que se realizan de forma recursiva para obtener un producto o resultado. El flujo de trabajo ocupa un lugar central en la automatización de los procesos, ya que las tareas se estructuran en función de la dependencia de los datos y de estructuras de control. En nuestro caso, el flujo de trabajo describe los aspectos del proceso que son relevantes y que pueden ser gestionados de forma independientes y que requieren habilidades específicas para ser llevados a cabo. A estas secciones, que pueden pensarse de forma autónoma en el proceso, las llamaremos módulos operativos [Georgakopoulos et al., 1995; Russell et al., 2005].

La posibilidad de estandarizar el proceso de análisis de los datos de calidad de aire en módulos ofrece una gran cantidad de ventajas. Como primer aspecto, nos brinda la oportunidad de integrar procesos y capacidades nuevas con las preexistentes. Nos permite reconocer la organización previa dando valor a otras experiencias y saberes. Por otra parte, es posible aumentar la especificidad de la respuesta al agregar nuevos recursos o datos. Como tercer aspecto, brinda la posibilidad de adaptar la metodología al usuario mediante el desarrollo de ligeros cambios en función de las necesidades del destinatario final [Wang et al., 2011]. Un último aspecto es que el proceso por módulos permite actualizar las versiones sin perder la funcionalidad. Y de igual manera, modificar o reemplazar por partes los componentes de los módulos cuando la tecnología en la cual se basan evolucione, pero sin sacrificar el proceso completo [Tansley & Tolle, 2009].

En la **Figura 6** se sintetiza la metodología de trabajo elaborada. Se pensó en una metodología por etapas, donde cada módulo cuenta con una serie de tareas a desarrollar.



**Figura 6.** Metodología propuesta para el abordaje del estudio de la calidad del aire mediante el enfoque de ciencia de datos

El primer módulo hace foco en el conocimiento del área de estudio, la recopilación de estudios previos y datos disponibles, el análisis de la normativa, el planteamiento de un objetivo y la selección de las preguntas de estudio. Esta etapa limita los alcances del análisis que se utilicen en

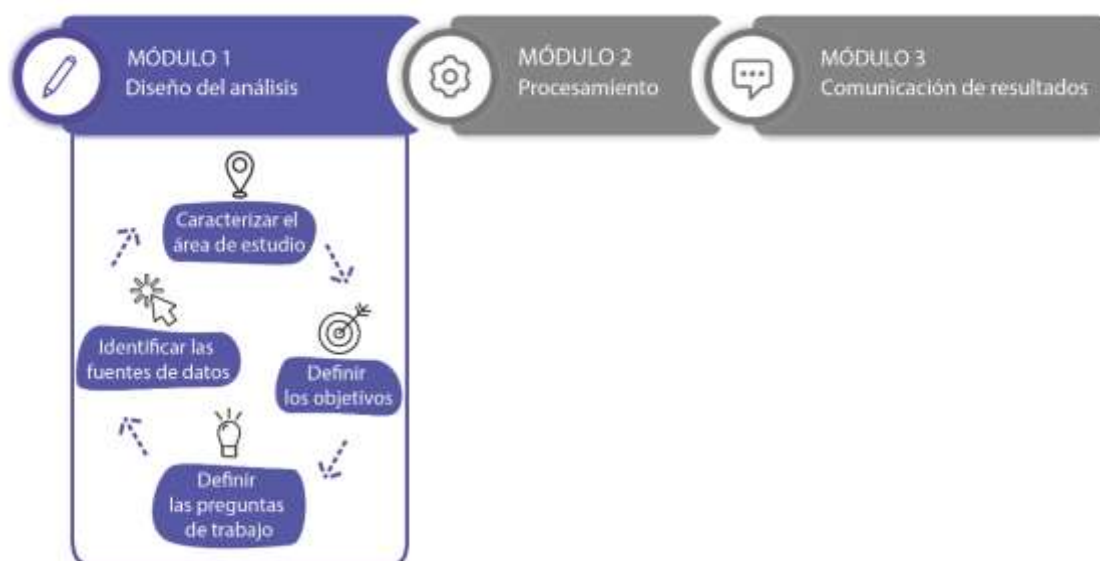
el siguiente módulo. Para su ejecución, se requiere un conocimiento experto en gestión y calidad del aire.

El segundo módulo abarca el pretratamiento y procesamiento de los datos. Esta etapa comienza una vez seleccionadas las preguntas a responder los datos con los que se cuenta. Si bien hay cuestiones generales en la preparación de los datos, existen cuestiones específicas que requieren ser consideradas en función de los algoritmos de análisis que se seleccionen. Como veremos más adelante, esta sección requiere principalmente de habilidades en el análisis e ingeniería de datos.

Finalmente, el tercer módulo se centra en el análisis de los resultados y su comunicación. Esta etapa implica la visualización de los resultados y la generación de informes. La comunicación de resultados es una parte fundamental en la gestión de la calidad del aire ya que incorpora a la población brindando la posibilidad de mejorar el manejo del riesgo. Este módulo requiere de habilidades en la comunicación científica.

A continuación, se expondrá con más detalle cada uno de estos módulos detallándose las tareas que los componen.

## Módulo I - Etapa de diseño del análisis



El primer módulo tiene como objetivo el diseño del análisis que se utilizó para la evaluación de la calidad del aire. Esto incluye una secuencia de pasos iterativos que ordena las acciones necesarias para definir las preguntas de trabajo de forma tal que orienten el análisis a realizarse en el módulo siguiente.

La etapa de diseño se la pensó como una etapa dinámica y, por tanto, resulta necesario que sea una etapa recursiva de mejora continua. Para mantener nuestro sistema actualizado con los avances científicos y tecnológicos fue importante realizar revisiones periódicas, de los objetivos y de las fuentes de información disponibles y de las preguntas que se podrían abordar.



## I.I. Definición y caracterización el área de estudio

La extensión del área de estudio, o dominio territorial, se definió en función de la magnitud del proyecto que se buscó abordar, así como de la presencia de población susceptible de ser afectada. Cada escenario de análisis presentó características particulares. Sin embargo, por las leyes físicas y químicas que gobiernan los fenómenos estudiados, hay cuestiones comunes que se repitieron en cada sitio de estudio. Identificar todas estas cuestiones facilitó la interpretación de la situación y, por tanto, el análisis. La caracterización generada nos aportó información clave para contextualizar los resultados.

En este sentido, identificamos las características que nos interesaba conocer: la demografía, las actividades económicas principales (uso de suelo), la meteorología de la región, la geografía local, el grado de urbanismo o la cubierta del suelo, la presencia de eventos climáticos característicos y el marco normativo en el que se circunscribe la problemática. Estos elementos se obtuvieron de la literatura científica, la documentación de las agencias ambientales y de los medios de comunicación, como también mediante visitas oculares y entrevistas.

### Relevamiento de fuentes de emisión

La identificación de las fuentes principales de emisión fue un paso necesario en la caracterización del área de estudio. Conocer las fuentes presentes y los niveles de emisiones, aportó información relevante para identificar las contribuciones que cada una de ellas realizaba al estado general de la calidad del aire. En este sentido, la evaluación ambiental se enfocó en los impactos que cada proyecto o actividad produjo sobre el ambiente. Esto se analizó considerando, entre otros parámetros, la magnitud y duración de las emisiones, especialmente cuando se trataron de actividades temporales y de impacto acotado.

Para esto, una herramienta muy útil son los inventarios de emisiones. Sin embargo, la confección de un inventario de emisiones es una ardua tarea de numeración de las fuentes y estimación de las emisiones de todos los contaminantes relevantes, y en el caso de contaminantes secundarios, también se deben considerar sus precursores [Pacyna & Graedel, 1995; Harrop, 2018]. Si bien se encuentran a disposición inventarios de emisiones a nivel global para distintos contaminantes, no siempre se cuenta con uno a escala local.

En nuestro caso, para fines evaluativos del área, no se precisó un inventario de emisiones de alta complejidad. Como alternativa, se identificaron las actividades económicas presentes en la zona, junto a las fuentes puntuales y fuentes de área, las fuentes vehiculares y las fuentes naturales, prestando especial interés a parques industriales, centrales eléctricas, áreas urbanas, entre otras fuentes de emisión de contaminantes. Esto permitió generar un conocimiento sobre el área de estudio que facilitó la interpretación de las mediciones de los contaminantes.

### Condiciones meteorológicas

Como se expuso en el marco teórico, las condiciones meteorológicas del clima influyen en el transporte de los contaminantes afectando sus concentraciones. Analizar este aspecto nos brinda información respecto a los ciclos anuales, permitiéndonos predecir los comportamientos entre las estaciones o los fenómenos meteorológicos que puedan provocar eventos extremos para proponer medidas o políticas proactivas.

Por esto, nos interesó conocer la velocidad media del viento y su dirección más frecuente, y los cambios que estas dos variables tuvieron a lo largo del día y de las estaciones. Como información

extra, también resultó conveniente analizar la presión atmosférica, el porcentaje de humedad relativa y los períodos de lluvia, al igual que el nivel de la radiación solar y las variaciones de la altura de la capa límite, cuando se contó con la información.

Estos datos suelen ser generados y provistos por los Servicio Meteorológicos oficiales, los cuales siguen estrictas normas de calidad para el emplazamiento de las estaciones. Sin embargo, en las estaciones de monitoreo de calidad de aire también es habitual que se registren algunos parámetros meteorológicos como la humedad y la temperatura. A diferencia de las estaciones meteorológicas que se emplazan a 10 m de altura como estándar para evitar la turbulencia del terreno, en las estaciones de calidad del aire no existe un criterio mundialmente establecido. Esta diferencia en la forma de registrar las condiciones meteorológicas comprometió la posibilidad de usar estos datos de forma alternada.

### Geografía del área

Definir la escala de trabajo y cómo esta se circunscribe en la geografía local nos permitió comprender mejor los flujos de los contaminantes sobre el relieve. La geografía limita o favorece el transporte de contaminantes. A microescala ocurre lo mismo con, por ejemplo, la presencia de edificios altos que funcionan como cañones urbanos o la presencia de áreas costeras o extensas áreas verdes que aportan humedad al sistema.

Para conocer la composición del terreno, una fuente importante de información geográfica han sido los Servicios Geográficos nacionales, los cuales tienen como tarea la generación de mapas con distintos tipos de objetos espaciales. Mediante su estudio se identificaron los flujos de vientos continentales y marinos, y los efectos del relieve en la circulación de los contaminantes. Para el análisis a microescala, resultó de gran utilidad la inspección ocular mediante visitas a los sitios de estudio.

### Eventos particulares

Otra cuestión que se tuvo en cuenta fue el análisis de eventos únicos o recurrentes que pudieron afectar la calidad del aire. Estos eventos suelen ser de origen naturales, como la bruma de mar o la influencia de un volcán, o antrópicos como la quema de pastizales o la suspensión de suelo por la actividad agrícola, y se manifiestan como un aumento de las concentraciones de algunas especies químicas en épocas puntuales del año o como hechos aislados.

### Marco normativo

Para la elaboración de este trabajo resultó imprescindible conocer el marco normativo del área de estudio. La calidad del aire es un problema de salud público y, por lo tanto, se encuentra normada en una gran cantidad de países. Por ello, fue necesario conocer las leyes vigentes y los planes de gestión aprobados para cada región.

En los casos donde se contaba con una red de monitoreo, fue conveniente analizar las circunstancias que llevaron a su implementación. Se buscaron antecedentes de informes y proyectos de gestión. Esto fue de gran utilidad puesto que en algunos casos las redes estaban asociadas a planes de mitigación con objetivos puntuales, los cuales brindaban una caracterización profunda de la problemática en el área.

## I. II. Determinación de los objetivos

El establecimiento de objetivos representó un punto de partida para la evaluación de la calidad del aire. Los objetivos pueden limitarse a una caracterización de la calidad del aire en un tiempo único en un sitio puntual, o abarcar un período de tiempo y área mayor (por ejemplo, el estudio de la evolución temporal de un contaminante a lo largo de los años o sobre una región). También, es posible querer contar con información de la red de monitoreo para mejorar la interpretación de las mediciones u obtener un pronóstico de las concentraciones en el corto plazo.

Los objetivos que fueron identificados como los más frecuentemente planteados en el estudio de la calidad de aire son:

- Conocer aspectos del monitoreo terrestre
- Determinar la calidad del aire en los sitios de medición
- Evaluar la evolución temporal de los contaminantes
- Evaluar la distribución espacial de los contaminantes

A continuación, se presenta un desglose de las implicancias de estos objetivos mediante la formulación de preguntas de estudio que direccionen el análisis. En función de los objetivos que se seleccionaron para cada caso se establecieron las preguntas de estudio que guiaron el trabajo, ya que estas se encuentran relacionadas con los algoritmos que se incluyeron o descartaron.

## I.III. Delimitación de las preguntas de estudio

En esta tesis se tomó el enfoque de ciencia de datos como un proceso de análisis basado en responder preguntas mediante datos. En base a esto, se confeccionaron una serie de preguntas orientativas del análisis. Estas preguntas representaron un conjunto específico de consultas, en términos de entidades y relaciones, que se hicieron a la base de datos que generamos para cada uno de los casos de estudio. La elaboración de estas preguntas correspondió al estudio exhaustivo de la temática y a la revisión de aspectos mencionados en informes técnicos.

Las preguntas que presentamos (**Cuadro 2**) sirven de base para un abordaje organizado del estudio de la calidad del aire. Primero, permiten discernir sobre qué datos debemos recopilar en función del alcance de la pregunta y luego, sobre qué tipos de análisis se deben o pueden realizar en función de la información disponible. Distintas preguntas requieren diferentes tipos de datos, lo que marca la diferencia en la recolección de los datos.

Sin embargo, en la práctica no siempre se contó con toda la información necesaria. Esto llevó a que se deban rever las preguntas a abordar, con el fin de obtener un conjunto final de preguntas definitiva para cada caso de estudio.

**Cuadro 2.** Preguntas orientadas para el análisis de la calidad del aire**Conocer aspectos del monitoreo terrestre**

- ¿Qué diseño sigue la red de monitoreo?
- ¿Cuáles son las técnicas utilizadas para el monitoreo?
- ¿Cuál es la calidad de la información recolectada por la red de monitoreo?
- ¿Hay redundancia entre los datos recolectados por las estaciones?
- ¿Cuáles son las características del área de emplazamiento de las estaciones de monitoreo?
- ¿Cómo influyen los vientos locales sobre los niveles registrados en las estaciones?
- ¿De qué sector del viento llegan las concentraciones mayores?

**Determinar la calidad del aire en los sitios de medición**

- ¿Cuáles son los niveles más frecuentes de los contaminantes?
- ¿Las concentraciones de los contaminantes excedieron los límites legales?
- ¿Los valores medidos de los contaminantes ponen en riesgo a la población?

**Evaluar la evolución temporal de los contaminantes**

- ¿A cuál hora del día es mayor la concentración de los contaminantes?
- ¿Las concentraciones máximas varían a lo largo de la semana?
- ¿En cuál época del año ocurren los episodios de máximas concentraciones?
- ¿Cómo han cambiado los niveles de los contaminantes con los años?

**Estudiar la distribución espacial de los contaminantes**

- ¿Dónde ocurren las máximas concentraciones?
- ¿Se observa un patrón espacial de las máximas concentraciones?
- ¿Cómo es la calidad del aire donde no se mide?
- ¿La calidad del aire en la región representa un riesgo para la población?

## I. IV. Adquisición de los datos

La etapa de adquisición de datos consiste en identificar y obtener los datos requeridos para el análisis. Los datos pueden provenir de distintas fuentes, por ejemplo, de estaciones de monitoreo oficiales, imágenes satelitales o sensores de low-cost [Candia et al., 2018]. Sin embargo, cada fuente de datos presenta sus propias limitaciones. El objetivo de este paso es la identificación de las fuentes de datos disponibles y reconocer sus posibles aplicaciones [Limaye et al., 2018].

Como punto de partida se buscaron trabajos previos en el sitio de estudio para analizar las fuentes de datos consultadas con anterioridad. Se revisó la literatura académica, los informes gubernamentales y las producciones de organizaciones sociales. Esto incluyó también la revisión de la literatura epidemiológica en la región. En este punto resultó relevante evaluar la disponibilidad de los datos cuantitativos [Limaye et al., 2018].

Una dificultad que se ha encontrado fue la falta de disponibilidad de la información. En los casos analizados, no siempre se contó con información de libre acceso. Esta falta de datos limitó las conclusiones que pudimos obtener. A su vez, la calidad de los datos que utilizamos marcó el alcance de nuestro trabajo. La calidad de los datos ha tenido una gran importancia en las conclusiones de nuestra evaluación, otorgando una mayor incertidumbre en algunos resultados.

A continuación, se presenta una breve reseña de fuentes de información con las que se trabajó en esta tesis:

### Datos de monitoreo terrestre

La principal fuente de información de la calidad del aire suelen ser los datos de monitoreo terrestre continuo brindados por organismos oficiales. Sin embargo, como se mencionó con anterioridad no siempre se cuenta con ellos y puede resultar difícil el acceso a los registros cuando estos existen.

Como alternativa, existen comunidades virtuales que buscan garantizar el acceso a los datos de calidad de aire. Por ejemplo, la comunidad detrás el sitio web Open Air Quality ([www.openaq.org](http://www.openaq.org)), brindan una plataforma para compartir los datos abiertos de calidad del aire. Con un espíritu similar, el open-hardware permite el aporte de la comunidad para generar mediciones a bajo costo. A su vez, la participación directa de la ciudadanía en la vigilancia de los contaminantes garantiza una mayor sensibilidad en la problemática [Goodchild, 2007]. Esta es aún una tecnología en fase inicial pero que cuenta con buenas perspectivas de crecimiento [Candia et al., 2018].

### Teledetección y datos de monitoreo satelital

La teledetección aporta una gran cantidad de información de forma periódica con alcance global. Dependiendo del producto satelital que se utilice, se puede contar con imágenes con pocos días de diferencia o en tiempo casi real. El uso de datos satelitales puede ayudar a los profesionales de la calidad del aire a determinar la distribución espacial de las especies químicas e identificar la ubicación de las concentraciones máximas.

La teledetección satelital aplicada al estudio de la calidad del aire ha tenido un fuerte desarrollo en los últimos años. Desde 1995 se cuenta con una serie continua de instrumentos que miden los niveles de los componentes atmosféricos. En la **Tabla 5** se detallan los satélites y sensores más relevantes en el estudio de la calidad del aire para cada especie química.

Como puede verse en la **Tabla 5**, las primeras mediciones fueron realizadas por el sensor GOME en el satélite ERS-2 de la ESA, que funcionó entre 1995 y 2011. Manteniendo una continuidad en la investigación, en 2006 fue lanzado el instrumento GOME-2 a bordo del satélite MetOp-A de EUMETSAT, aún en servicio. Un segundo GOME-2 está operando desde 2012 en el satélite MetOp-B. La resolución espacial de estos sensores es baja, pensando su aplicación para estudios globales.

Tabla 5. Sensores más utilizados para la determinación de contaminantes criterio.

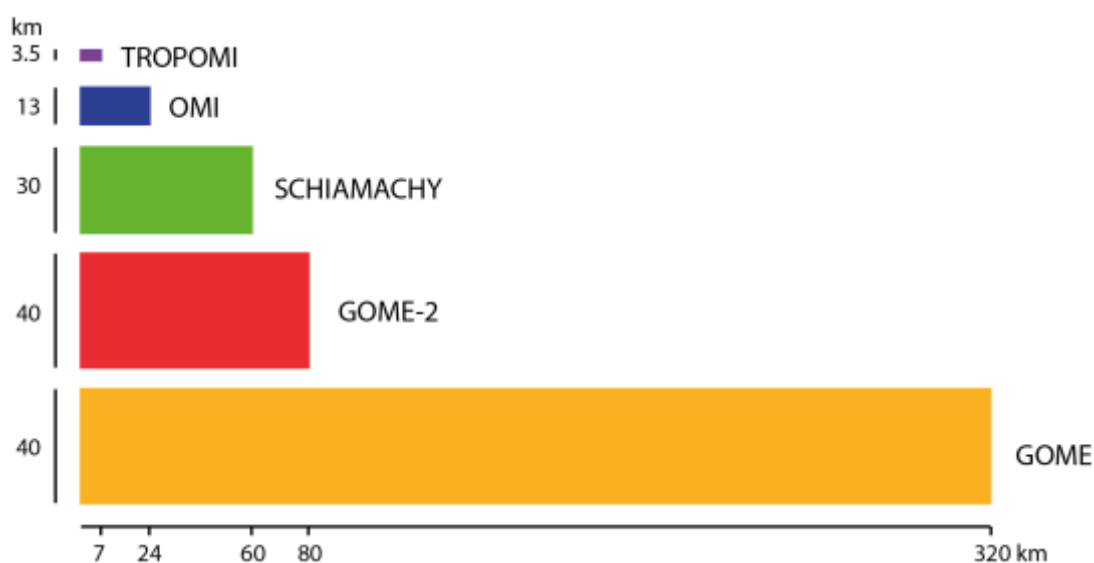
<b>Especies químicas</b>	<b>Sensor</b>	<b>Plataforma</b>	<b>Agencia</b>	<b>Período operacional</b>	<b>Resolución (km)</b>	<b>Revisita (días)</b>
<b>Aerosoles (AOD)</b>	MODIS	TERRA	NASA	2000 -	1 x 1	1-2
	MODIS	AQUA	NASA	2002 -	1 x 1	1-2
	OMI	AURA	NASA	2004 -	13 x 24	1
	MISR	TERRA	NASA	2000 -	17.6 x 17.6	9
	GOME-2	METOP-A	ESA	2006 -	40 x 80	1.5
	MSI	Sentinel-2	ESA	2015 -	10 m	5
	OLCI	Sentinel-3	ESA	2016 -	300 x 300m	3
	SCIAMACHY	ENVISAT	ESA	2002 - 2012	30 x 60	6
	GOME	ERS-2	ESA	1995 - 2003	40 x 320	3
	TROPOMI	Sentinel-5P	ESA	2017-	3.5 x 7	1
	OLI	Landsat-8	NASA	2013-	30m	16
<b>NO<sub>2</sub></b>	OMI	Aura	NASA	2004 -	13 x 24	1
	GOME-2	MEtOp-A-B-C	ESA	2007 - 2016	40 x 80	1.5
	SCHIAMACHY	ENVISAT	ESA	2004 - 2012	30 x 60	6
	GOME	ERS-2	ESA	1995 - 2003	40 x 320	3
	TROPOMI	Sentinel-5P	ESA	2017-	3.5 x 7	1
<b>O<sub>3</sub></b>	OMI	Aura	NASA	2004-	24 x13	1
	TES	Aura	NASA	2004 -	8 x 5	1
	SCHIAMACHY	ENVISAT	ESA	2004 - 2012	30 x 60	6
	JASI	MetOp	ESA	2006-	12 x 12	0.5
	GOME-2	MetOp	ESA	2006-	80 x 40	1
	GOME	ERS-2	ESA	1995 - 2003	40 x 320	3
	TROPOMI	Sentinel-5P	ESA	2017-	3.5 x 7	1
<b>SO<sub>2</sub></b>	OMPS	Suomi NPP	NASA	2014 -	3 x 250	4
	CrIS	Suomi NPP	NASA	2014 -	14 x 2	4
	OMI	Aura	NASA	2004 -	24 x13	1
	MLS	Aura	NASA	2004 -	3 x 300	1
	GOME-2	MetOp-A-B-C	ESA	2007 - 2016	80 x 40	1
	IASI	MetOp-A-B-C	ESA	2007 - 2016	40 x 80	1.5

	SCHIAMACHY	ENVISAT	ESA	2004 - 2012	30 x 60	6
	AIRS	Aqua	NASA	2002 - 2016	50 x 50	1
	MLS	Aura	NASA	2004 -	3 x 300	1
	MOPITT	Terra	NASA	2000-	22 x 22	3
<b>CO</b>	SCHIAMACHY	ENVISAT	ESA	2004 - 2012	30 x 60	6
	AIRS	Aqua	NASA	2002 - 2016	50 x 50	1
	IASI	MetOp-A-B-C	ESA	2007 - 2016	40 x 80	1.5
	TROPOMI	Sentinel-5P	ESA	2017-	3.5 x 7	1

Las medidas de las misiones MetOp fueron complementadas por el sensor SCIAMACHY a bordo de la misión Envisat de la ESA, que funcionó entre 2002 y 2012, y el sensor OMI a bordo de la nave espacial AURA de la NASA, en funcionamiento desde 2004. Como últimas adquisiciones, entre 2015 y la actualidad fueron puestos en órbita nuevos instrumentos de vigilancia con una mayor resolución, los satélites Sentinel-2, Sentinel-3 y Sentinel-5P de la ESA. Estos sensores mejoran la resolución espacial, mostrando una tendencia a evolucionar hacia la aplicación en estudios de escala urbana.

Los productos de AOD se han desarrollado utilizando datos del TOMS a una resolución de 1° [Torres et al., 2002], POLDER a una resolución de 1/6° [Leroy et al., 1997], MISR a una resolución de 17°.6 km [Dinner et al., 1999], MERIS a 10 km de resolución [Mei et al., 2017], VIIRS a 6 km de resolución [Jackson et al., 2013], Geoestacionario Himawari-8 a 0,05° de resolución [Kikuchi et al., 2018], y MODIS a 10 km, 3 km [Kikuchi et al., 2018; Gupta et al., 2018] y 1 km de resolución [Lyapustin et al., 2018]. Los satélites Landsat-8 y Sentinel-2 también brindan la posibilidad de ser utilizados para recuperar los niveles de AOD con una muy buena resolución (30 m y 10 m, respectivamente) abriéndose nuevas posibilidades de estudio a futuro [Li et al., 2019].

Para adquirir de manera efectiva mediciones de los contaminantes gaseosos, se han implementado una amplia cantidad de instrumentos satelitales a bordo de distintos satélites, tales como el GOME, el SCIAMACHY, el GOME-2 y el OMI. Sin embargo, OMI ha tenido la más amplia aplicación en estudios locales dada su mayor resolución espacial y temporal [Krotkov et al., 2016; Levelt et al., 2018]. TROPOMI, a bordo de Sentinel-5P, se encuentra en funcionamiento desde 2017 y mejora a OMI en su resolución espacial (**Figura 7**).



**Figura 7.** Comparación de la resolución espacial de los sensores más usados en el estudio de gases atmosféricos

Como principal dificultad, el trabajo con imágenes satelitales requiere del entrenamiento especializado de recursos humanos en el desarrollo de habilidades específicas. A su vez, los procesamientos matemáticos requieren de una alta capacidad de cálculo y el almacenamiento de los datos consumen mucha memoria, lo cual conduce a la necesidad de obtener recursos tecnológicos específicos e integrar nuevos saberes técnicos, como son los aportes de la ingeniería de datos o la geomática [Wang et al., 2011].

En esta tesis se trabajó con los productos de AOD del sensor MODIS, ya que han sido ampliamente utilizados en el estudio de aerosoles en la atmósfera por su extenso registro temporal a escala global [Ichoku et al., 2004]. Los archivos se descargaron como paquetes de ficheros que contenían escenas de datos (múltiples matrices numéricas). Una vez adquiridos, los productos satelitales se visualizaron y analizaron mediante programas informáticos desarrollados para esta tesis. El análisis y procesamiento se realizó en todos los casos utilizando el lenguaje de programación R [R Core Team, 2018].

### Otras variables de interés

Ya sea para contextualizar las mediciones de calidad de aire, para explicar los fenómenos extremos o para predecir los niveles de los componentes atmosféricos, resulta de suma utilidad contar con otras variables espaciales. A continuación, se presentan las fuentes de información accesorias utilizadas en este trabajo de investigación.

#### *Datos meteorológicos*

Los datos meteorológicos pueden ser datos observados o generados por un modelo numérico. Por lo general, se cuenta con los datos oficiales provistos por el servicio meteorológico. Sin embargo, estos datos son tomados en altura, y no suelen ser recomendables para la modelación del transporte de los contaminantes en superficie a microescala.



Es conveniente que los datos meteorológicos en superficie sean tomados en las mismas estaciones de monitoreo para que tengan un correlato con las concentraciones registradas de los contaminantes. Esto permite tener una noción de la influencia del viento sobre las mediciones y brinda información sobre la influencia de las fuentes de contaminantes en las mediciones, pero esto no siempre sucede así. En los casos donde no se cuenta con datos meteorológicos medidos *in situ* se pueden utilizar datos modelados. Entre los modelos numéricos de pronóstico meteorológico más conocidos se encuentra el WRF (del inglés, *Weather Research and Forecast Model*).

#### *Datos de reanálisis atmosférico*

Los datos de reanálisis proporcionan valores para las variables atmosféricas en una grilla continua que es representativa de la totalidad de la atmósfera. Los reanálisis se elaboran sobre los resultados de un modelo meteorológico al cual se le incluyen las mediciones históricas realizadas, mediante un proceso de asimilación de datos. Por lo tanto, el reanálisis resulta mucho más robusto que el modelo inicial [Deep et al., 2011].

En esta tesis se utilizaron los datos de reanálisis del modelo MERRA-2. El MERRA-2 es un modelo de reanálisis atmosférico que utiliza el *Goddard Earth Observing System, Version 5 (GEOS-5)* [Gelaro et al., 2017]. El GEOS-5 consiste en un grupo de componentes que permiten modelar distintas variables y parámetros relacionados con la atmósfera como, por ejemplo, la circulación general de la atmósfera o el transporte químico atmosférico.

Como ventaja, el MERRA-2 simula cinco especies de PM (carbono negro, carbono orgánico, sal marina, polvo, sulfato) basándose en el modelo de aerosol GOCART (del inglés, *Goddard Chemistry Aerosol Radiation and Transport*) [Chin, 2005]. GOCART es un modelo global de química y transporte que simula los principales tipos de aerosoles troposféricos: sulfato y sus precursores, OC, BC, polvo y sal marina. El modelo GOCART se describe en detalle en varias publicaciones [por ejemplo, Chin et al., 2000, 2002, 2007, 2009; Ginoux et al., 2001]. Brevemente, los procesos químicos en el modelo incluyen reacciones en fase gaseosa y líquida que convierten los precursores de sulfato (sulfuro de dimetilo o DMS y SO<sub>2</sub>) en sulfato. Los procesos físicos incluyen la emisión de aerosoles, advección, convección, así como la deposición húmeda y seca. Las emisiones antropogénicas provienen de los inventarios globales disponibles y las emisiones naturales se calculan en función de las condiciones meteorológicas [Chin et al., 2009].

#### *Elevación de terreno*

Como vimos anteriormente, las características geográficas tienen un rol importante en el transporte de los contaminantes. Para conocer las características del terreno se puede utilizar los datos provistos por los modelos de elevación de terreno (MET). Un MET es una representación en forma de grilla numérica del terreno en relación con un sistema de referencia arbitrario. Este dato toma relevancia cuando analizamos un área con una geografía marcada.

#### *Clasificación de uso de suelo*

La clasificación del suelo según su uso es una forma de ordenamiento sistemático que proporciona información básica para la formulación y aplicación de planes de gestión. Si bien su empleo es mayoritariamente agrícola, también se utiliza en otros estudios territoriales como los urbanistas o de gestión de riesgos.

En esta tesis, se utilizó para el estudio de caso de Valencia el sistema de clasificación “*Corine Land Cover 2012*” a escala 1:100.000, generado para el año 2012. Este mapa pertenece al proyecto Europeo Corine Land Cover, dirigido por la Agencia Europea de Medio Ambiente.

## Módulo II - Etapa de procesamiento



### II. I. Preparación de los datos

El propósito de este paso fue asegurar la calidad de los datos que se han seleccionado. Como se mencionó anteriormente, el hecho de que los datos no presenten anomalías y estén libres de inconsistencias es un prerrequisito para el éxito de los análisis. Por otro lado, cuanto más y mejor se conozcan las limitaciones de los datos, más fácil será interpretar los resultados.

La preparación de datos estuvo formada por distintas técnicas como son 1) limpieza de ruido, 2) imputación de valores perdidos o faltantes, 3) integración y 4) la transformación de datos y su normalización [Fayyad et al., 1996]. Algunos análisis requieren de forma obligatoria la aplicación de todas estas técnicas en la preparación de los datos, ya que sin esto los resultados obtenidos de los algoritmos estadísticos ofrecerían resultados espurios [García et al., 2016]. Sin embargo, resulta conveniente ser conservadores en cuanto a las modificaciones que se realicen de los datos de trabajo, siendo preferible cambiar las técnicas antes que aplicar transformaciones o realizar imputaciones.

#### Limpieza de ruido

El análisis del ruido en un set de datos es un paso muy importante en la preparación de los datos. El ruido es un error aleatorio o una variación azarosa de una variable. En consecuencia, las variables afectadas por el ruido tendrán valores que quedan fuera de los rangos esperados para esas variables, y por ello son considerados valores atípicos. Los valores atípicos pueden representar una oportunidad para comenzar una investigación, pues indican eventos particulares o, simplemente, ser datos incorrectos producto de errores en el registro.

En casos donde el error de entrada es la inclusión de un carácter no válido, este debe ser corregido alterando la base de datos y registrando la modificación en los metadatos. Otro tipo de valores atípicos es el que se genera cuando los mensajes de error del instrumental aparecen como números. Por ejemplo, la repetición del valor 99999 puede ser un mensaje de falla del equipo. En estos casos, debe ser computado como dato faltante.

Una situación de particular conflicto se genera con el registro de los valores por debajo del límite de detección. A menudo los valores medidos por debajo del límite de detección del equipo se registran como cero. Esto no es correcto y resulta muy importante corregir este error en nuestro set de datos puesto que puede efectuar un sesgo en los análisis paramétricos. Para corregirlo, se recomienda adoptar el criterio sugerido por la [EPA, 2006] de sustituir los valores iguales a cero por un tercio de la concentración mínima medida en esa estación dentro del período de estudio.

Hasta aquí se presentaron situaciones particulares que fueron habituales en el abordaje de los datos generados por el monitoreo automático de la calidad de aire, y su solución es de sencilla aplicación. Sin embargo, algunas técnicas de medición requieren de soluciones más complejas para identificar el ruido, eliminarlo o corregirlo.

La aplicación de límites relacionados con la desviación estándar asociada con los datos (por ejemplo, gráficos de control) no puede utilizarse porque tienen como requisito la normalidad de los datos [Martínez et al., 2014]. En su lugar, este problema requiere otro enfoque que generalmente se basa en métodos más complejos, como el uso de las técnicas de Markov Chain Monte Carlo (MCMC) [Gilks, 1995; Eastoe, 2009; Dominici et al., 2002]. Dada su complejidad, son pocos los trabajos que utilizan esta técnica, ya que resulta más conveniente adoptar algoritmos robustos o no paramétricos en el análisis de forma que no se vean afectados por la presencia de ruido o valores anómalos.

## Imputación de valores faltantes

Al trabajar con datos provenientes de sistemas de detección automática fue usual observar valores faltantes entre los registros. Es decir, nos encontramos una muestra que no presentaba ningún valor asignado para cierta variable o atributo en un determinado tiempo.

Este problema puede ser atendido de distintas formas que van desde ignorar la existencia de estos valores faltantes, rellenar manualmente, utilizar estadísticas simples o modelar los datos faltantes para obtener nuevos valores. En función del porcentaje de datos faltantes y de nuestro objetivo deberemos elegir un método u otro. Sin embargo, hay que recordar que ninguna técnica es perfecta y que hay que tener cuidado de no introducir más ruido al modelar los datos faltantes [Fernández et al., 2018].

Por ejemplo, la Directiva 2008/50/EC de la UE recomienda que en el análisis de las series de tiempo de las concentraciones de los contaminantes no imputar los datos faltantes, puesto que se corre el riesgo de obtener medidas resúmenes de datos modelados y no de mediciones reales. En casos donde hay un porcentaje de datos faltantes elevado se recomienda descartar el día o el mes, dependiendo del tipo de análisis estadístico que se realice [Directive 2008/50/EC].

Por otra parte, cuando se busca generar mapas de calidad de aire a partir de imágenes satelitales, es aconsejable rellenar las áreas de la escena donde no se haya podido recuperar el valor de AOD a través de datos modelados [Goldberg et al., 2019]. En este caso, el modelado contribuye a la

obtención de una información más completa sobre el territorio para la detección de zonas conflictivas [Stafoggia et al., 2019].

## Integración

Las técnicas de reducción de datos se orientan a obtener una representación reducida de los datos originales, manteniendo en la mayor medida posible la integridad de la información existente en los datos. La aplicación de técnicas de integración no se considera estrictamente obligatoria, pero su utilización suele facilitar la interpretación de los datos [Desarkar & Das, 2018; Gong & Ordieres Meré, 2016]. Por ejemplo, cuando trabajamos con series de tiempo en algunos casos resultó conveniente integrar los datos horarios de forma diaria o mensual [García et al., 2016]. Los promedios anuales se evitaron, pues ocultan los efectos de la variación climática interanual que tiene una gran importancia en las concentraciones de los contaminantes [Zhang et al., 2018].

Aunque la media suele ser la función de agregación más utilizada, la mediana presenta un mejor poder descriptivo cuando los datos no siguen el comportamiento normal. Por esto último, algunos países han preferido utilizar el método del percentil para evaluar la calidad del aire [Salako & Hopke, 2012]. Dado que la convención en las publicaciones científicas es utilizar la media para comparar distintos escenarios, se ha buscado comunicar ambos parámetros estadísticos cuando esto no resultaba confuso para el lector.

Al decidir la forma de integración de nuestros datos fue importante atender a las unidades de integración que se encuentran expresadas en la normativa, para generar set de datos comparables con estas. Por ejemplo, para los niveles de  $O_3$  se suele trabajar con datos integrados cada 8 hs, mientras que para  $SO_2$ ,  $PM_{2.5}$  y  $PM_{10}$  en 24hs [World Health Organization, 2006].

## Transformación de los datos

Es común en otras disciplinas donde se trabaja con series de tiempo aplicar técnicas de transformación de los datos para estabilizar la varianza y normalizar la distribución de los datos. Las transformaciones polinómicas más utilizadas en datos de calidad del aire es la familia de transformación de Box-Cox [Martínez et al., 2014].

Sin embargo, es una mejor opción no transformar los datos *a priori*, sino utilizar modelos matemáticos que se adapten a la distribución estadística que presentan los datos originales. Un ejemplo de estos modelos son los modelos lineales generalizados (GLM) o los modelos lineales generalizados mixtos (GLMM) que pueden aplicarse a distribuciones distintas de la normal como son la distribución de Poisson, binomial, binomial negativa, gamma [Zuur et al., 2009 a; Zuur et al., 2009 b]. Y en caso de que, teniendo esto en cuenta, la variación de los datos no pueda ser explicada por el modelo, sería conveniente utilizar un método no paramétrico [Fan & Yao, 2008].

## II.II. Generación de los metadatos

Dado que trabajamos con un conjunto de datos que representan variables ambientales de distintas fuentes resultó conveniente confeccionar metadatos asociados a cada variable o conjunto

de datos. Los metadatos nos permiten recordar y compartir la información recolectada aportando calidad a la información que los datos albergan. Los metadatos no sólo proporcionan una definición de los datos o de la variable, también brindan información sobre los tipos de datos, la fuente original, el formato, las unidades, el proveedor y otras características que tienen que ver con la definición de la variable [Limaye et al., 2018].

## **II. III. Selección de algoritmos de análisis y procesamiento de los datos**

Una vez adquiridos y pre-procesados los datos se dio paso a la fase de análisis y procesamiento. Esta etapa del estudio fue guiada por las preguntas que se desprendieron de los objetivos definidos, de esta forma el ejercicio de la minería de datos confluyó en conclusiones relevantes para el desarrollo de informes y planes de acción.

En la **Tabla 6**, se detallan los objetivos específicos que surgieron de los objetivos generales y las preguntas planteadas. Como se verá a continuación, cada uno de estos objetivos específicos ha sido abordado en la bibliografía por diferentes técnicas, las cuales, en muchos casos, constituyen campos enteros de estudio.

**Tabla 6.** Dimensiones del análisis de la calidad del aire: objetivos, preguntas de estudio y técnicas.

<b>Objetivos generales</b>	<b>Preguntas</b>		<b>Objetivos específicos</b>
<i>Caracterizar aspectos del monitoreo terrestre</i>	¿Qué diseño sigue la red de monitoreo?	I	Analizar el patrón de la distribución espacial de las estaciones de monitoreo.
	¿Cuáles son las técnicas utilizadas para el monitoreo?	II	Describir la instrumentación utilizada por parámetro y analizar la calidad de los datos obtenidos por las estaciones de monitoreo.
	¿Cómo es la calidad de la información recolectada por la red de monitoreo?		
	¿Hay redundancia de información entre las estaciones?	III	Evaluar la redundancia entre las estaciones de monitoreo mediante un análisis de correlación y análisis de componentes principales. Identificar las estaciones que tienen igual comportamiento mediante análisis clúster jerárquico.
	¿Cuáles son las características del área de emplazamiento de las estaciones de monitoreo?	IV	Analizar el área de influencia de las estaciones de monitoreo
	¿Cómo influyen los vientos locales sobre los niveles registrados en las estaciones?	V	Analizar los vientos en las estaciones (rosa de vientos)
	¿De qué sector del viento llegan las concentraciones mayores?	VI	Analizar los aportes de las fuentes mediante rosa de contaminación, mapa de fuentes de emisiones y modelos de retrotrayectoria.
<i>Determinar la calidad del aire en los sitios de medición</i>	¿Cuáles son los niveles más frecuentes de los contaminantes?	VII	Realizar un análisis exploratorio. Análisis descriptivo con respecto a los niveles guías. Estudio de la frecuencia y análisis de la frecuencia acumulada de las concentraciones. Análisis del comportamiento diario, mensual, anual.
	¿La concentración de los contaminantes excede los límites legales?		
	¿Los contaminantes medidos ponen en riesgo a la población?		
<i>Conocer la evolución temporal de los contaminantes</i>	¿A cuál hora del día es mayor la concentración de los contaminantes?	VIII	Evaluación de la tendencia temporal a largo plazo
	¿Las concentraciones máximas varían a lo largo de la semana?		
	¿En cuál época del año ocurren los episodios de máximas concentraciones?		
<i>Evaluar la distribución espacial de los contaminantes</i>	¿Cómo es la calidad del aire donde no se mide?	IX	Generar un modelado espacial con los datos terrestres. Confeccionar mapas de calidad de aire. Analizar comportamiento espacial y temporal de las variables desde satélite.
	¿Dónde ocurren las máximas concentraciones?		
	¿Se observa un patrón espacial de las concentraciones?		
	¿La calidad del aire en la región representa un riesgo para la población?	X	Generar mapas de zonas de riesgo a la salud

## I. Análisis del patrón de las estaciones

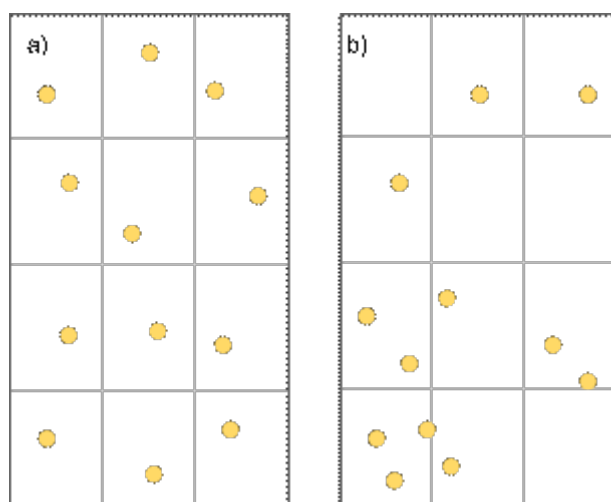
Hay distintos tipos de diseños de muestreo para una red de calidad del aire, por ejemplo, pueden obedecer a un diseño estratificado, en grilla o puede ser confeccionado a conveniencia de ciertos requerimientos, como ser denuncias ciudadanas.

El muestreo estratificado es un diseño de muestreo probabilístico en el que se divide a la región en subgrupos o estratos en función de la caracterización previa. Este tipo de diseño es muy común y suele ser conveniente cuando se desea obtener información de ciertas áreas bien diferenciadas, por ejemplo, por sus actividades productivas. Sin embargo, tiene como desventaja que suele sub-representar las áreas despobladas, dificultando la inferencia estadística a otras regiones no monitoreadas.

El muestreo en grilla no suele ser utilizado en monitoreos automáticos por el gasto que conlleva el despliegue de una gran cantidad de estaciones, pero se puede ver en muestreos eventuales, para caracterizar una zona en un momento particular. Este método tiene como ventaja que ofrece la mejor representación de toda el área de estudio, brindando suficiente información para aplicar técnicas de extrapolación espacial.

Por su parte, el muestreo por conveniencia o a criterio técnico es el más utilizado y el que mayor problema presenta -en términos estadísticos- para generar inferencias a partir de las mediciones. Cuando se tiene una distribución heterogénea de las estaciones, una zona del territorio se encuentra sobre-representada por sobre otras, generando un sesgo espacial que se traduce en los estadísticos que se calculan sin tener en cuenta esta situación.

El análisis de la distribución espacial de las estaciones aportó conocimiento respecto a la representatividad del monitoreo. Para conocer la distribución espacial de las estaciones se hizo un **análisis de patrones de puntos**. El análisis de patrones se fundamenta en la comparación entre las propiedades de una distribución teórica aleatoria (proceso puntual de Poisson) y las propiedades de la distribución observada [Dadson, 2017]. El modelo nulo es el proceso puntual de Poisson, en el cual la intensidad, número de puntos por unidad de área, se asume homogénea a lo largo de todo el espacio de estudio. Se utilizó como prueba de hipótesis para discernir entre un patrón homogéneo o heterogéneo la prueba de conteo por cuadrantes (**Figura 8**).



**Figura 8.** Esquema de una distribución espacial a) homogénea y b) heterogénea

La prueba de conteo por cuadrantes consiste en dividir el área en sub-regiones denominadas cuadrantes de área  $a_1, \dots, a_m$ . Si los conteos de puntos  $n_j$  para  $j = 1, \dots, m$  son divididos por su correspondiente área, obtenemos un simple estimado de la función de intensidad ( $\rho$ ). Así, la prueba estadística para homogeneidad, asumiendo que es un proceso puntual de Poisson, se basa en los conteos de los cuadrantes considerando como  $H_0$ : “La intensidad es homogénea” y como  $H_1$ : “Es un proceso Poisson no homogéneo”. Si  $H_0$  es verdadera, entonces los  $n_j$  son realizaciones de variables aleatorias de tipo Poisson independientes, con valores esperado  $\mu_j = \rho \cdot a_j$ . Dado que el número de puntos  $n = \sum_j n_j$  y el área total de la ventana  $a = \sum_j a_j$ , entonces la intensidad estimada es  $\hat{\rho} = \frac{n}{a}$  y el conteo esperado en el  $j$ -ésimo cuadrante es  $e_j = \hat{\rho} a_j = \frac{(n a_j)}{a}$  [Cressie, 1991].

## II. Descripción de la instrumentación y análisis de la calidad de los datos

Los datos obtenidos de los sistemas de medición automática deben representar las concentraciones reales presentes en el área de estudio. Si no se cuenta con esta certeza, los resultados obtenidos carecen de validez. Como garantía, resulta conveniente utilizar datos producidos únicamente mediante métodos de medición certificados y con instrumental comprobados y aceptados por la comunidad científica, o validados contra estos.

Otro punto crucial para asegurar la calidad de los datos se encuentra en el mantenimiento de las estaciones de monitoreo. Las calibraciones de los instrumentos, momento en donde se evalúa la trazabilidad respecto a los patrones absolutos o aceptados de metrología, deben ser periódicos y realizados por personal competente en el área. Solo con la calibración periódica del instrumental se garantiza una operación satisfactoria y durante periodos prolongados. Por esto, contar con informes de la calibración de los equipos de monitoreo es esencial para obtener mediciones exactas, precisas, comparables, reproducibles y consistentes en el tiempo [PNUMA/OMS 2002].

El problema de la calidad de los datos ha sido objeto de estudio desde varias perspectivas y por diferentes áreas, como son la estadística, la gestión o la computación [Bianchi, 2009]. Sin



embargo, la literatura sobre este tema comenzó en el contexto de los sistemas de información [Strong et al., 1997], introduciendo distintos aspectos para su evaluación.

Dado que hay varias dimensiones que hacen a la calidad de los datos resultó interesante tomar como marco internacional de referencia la norma ISO/IEC 25012. La calidad de los datos se define en ella como “el grado en que los datos satisfacen los requisitos establecidos por la organización propietaria del producto” [ISO/IEC 25012:2008]. Estos requisitos vienen dados por el usuario final de los datos. En nuestro caso, como los datos provinieron de distintas fuentes de información, se decidió trabajar con la mejor calidad que se encontró disponible de los datos.

En el caso de la gestión de la calidad del aire interesa, en particular, evaluar la completitud del set de datos y su consistencia. La completitud viene dada por la proporción de datos almacenados. Para aquella información que fue omitida, nos interesó conocer la causa de su omisión. La existencia de valores nulos influye en el análisis de los datos que se lleva a cabo, ya que al obtener estadísticas de estos se hace necesario dejar de lado los registros faltantes [Bianchi, 2009]. A modo de ejemplo, un cúmulo de valores nulos en las épocas del año donde se presentan las máximas concentraciones de un contaminante puede generar un descenso de la media.

Por su parte, el parámetro de la consistencia captura la satisfacción de reglas definidas sobre los datos. Las variables medidas siguen comportamientos que son propios del fenómeno que se estudia, en función de sus características físicas y químicas. Por ejemplo, la concentración es una unidad de medida que relaciona la masa con el volumen, siendo incorrecta la presencia de valores negativos. Si bien las inconsistencias debieron ser identificadas y corregidas en el paso I del módulo II, en este apartado nos interesa conocer cuánto afectan a la calidad del set de datos.

Como se mencionó, en algunos países existe una clara reglamentación respecto a la tolerancia que se debe tener para el porcentaje de datos faltantes, debiendo ser descartados del estudio aquellas estaciones que presenten ausencias de registros por encima del límite establecido. Por ejemplo, en la [Directiva 2008/50/CE] se establece como límite para los datos omitidos o erróneos un 15% de las medidas totales durante el período seleccionado, sobre la base del 90% de los datos capturados anualmente.

### III. Evaluación de la redundancia

La redundancia de los registros generados por distintas estaciones es una característica no deseada en una red de monitoreo, puesto que implica un costo económico para sostener el registro de las concentraciones en sitios que no aportan información nueva. En una situación ideal, cada estación de monitoreo aporta información relevante y no duplicada con las demás estaciones de monitoreo, pudiendo en conjunto generar la imagen más descriptiva de la situación de los niveles de los contaminantes y sus variaciones [Pires et al., 2008a, Pires et al., 2008b].

A continuación, se presentan una serie de procedimientos que sirvieron para identificar las estaciones que recolectaban información redundante.

### *Análisis de correlación*

El **análisis de la correlación** se utiliza para medir la fuerza de la relación entre dos variables. La correlación puede revelar la presencia potencial de un origen compartido u otra relación causal. El estudio de correlación se puede utilizar como un análisis preliminar, mediante el cual se determina si hay estaciones de control redundantes en la red. Una correlación fuerte es un indicio de redundancia.

Existen diferentes coeficientes de correlación para cuantificar la fuerza de la relación entre dos conjuntos de datos. El más utilizado es el factor de correlación de Pearson ( $r$ ), que se centra en la covarianza entre las variables. El factor de correlación de Pearson viene dado por la ecuación:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

donde  $\bar{x}$  es la media de la variable X,  $\bar{y}$  es la media de la variable Y,  $i$  es la posición de la observación y  $n$  el número total de observaciones.

Los coeficientes paramétricos, como el factor de correlación de Pearson, proporcionan una prueba estadística más potente que los coeficientes no paramétricos. Sin embargo, no son robustos respecto a los datos atípicos y generan sesgo cuando se trabaja con un conjunto de datos sin distribución normal [Yang et al., 2011, Chen et al., 2015]. Por ello, elegimos incorporar en los análisis el uso del factor de correlación de rango de Spearman ( $r_s$ ) que es un coeficiente no paramétrico, tolerante a la pérdida de normalidad y robusto con respecto a los datos extremos.

Para calcular el factor de correlación de Spearman, los datos son ordenados y reemplazados por su respectivo orden. El estadístico  $r_s$  viene dado por:

$$r_s = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n(n^2 - 1)}$$

donde  $D_i$  es la diferencia entre los pares ordenados de  $x$  e  $y$ , y  $n$  es el número de parejas de datos.

### *Análisis de PCA*

La técnica de **Análisis de Componentes Principales** (PCA) es una técnica ampliamente utilizada para la reducción de la dimensionalidad de un conjunto de datos, lo que permite explicar la mayor parte de la variabilidad utilizando menos variables [Pires et al., 2008a, Pires et al., 2008b]. En este sentido, la técnica de PCA se aplica como método de clasificación, identificando las variables que contengan un alto porcentaje de variabilidad común.

El primer paso de PCA es la selección del número de componentes principales (PCs). Comúnmente, se utiliza para este propósito el criterio de Kaiser, que selecciona PCs con valores propios superiores a 1 [Yidana et al., 2008]. Sin embargo, este criterio no suele alcanzar el 90% de la varianza original de los datos, por tanto, se propone como criterio seleccionar PCs que representen al menos el 70% de la varianza original de los datos [Mendiguchía et al., 2004, Pires et al., 2008a, Pires et al., 2008b]. Este procedimiento permite utilizar más información de los datos originales en los PCs seleccionados y, por lo tanto, aumentar la confianza de los resultados [Pires et al. 2009].

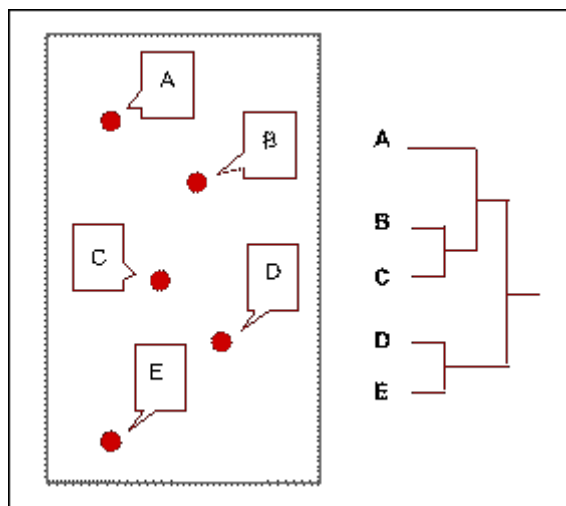
El PCA se puede realizar sobre el set de datos o sobre la matriz de correlación. Este último caso presenta grandes ventajas cuando se trabaja con una gran cantidad de datos faltantes, existe una falta de centralidad en el conjunto de datos o difiere la varianza de los datos entre las estaciones [Pires et al. 2009].

Finalmente, las cargas de los factores rotados (por ejemplo, aplicando el algoritmo de rotación Varimax) indican la influencia de cada variable en los PCs. Luego, la influencia de las variables originales de cada PC se interpreta como fuerte, moderada y débil para valores de carga absoluta  $> 0,75$ ,  $0,75-0,5$  y  $0,5-0,3$ , respectivamente [Liu et al., 2003]. Una elevada influencia sobre el mismo factor por 2 estaciones distintas es indicador de redundancia en su información.

### *Análisis de conglomerados jerárquicos*

El análisis de conglomerados es una técnica estadística multivariante que está encontrando una creciente aplicación en la investigación científica [Arora & Varshney, 2016]. El enfoque consiste en dividir un conjunto de datos en varios grupos que son distintos en términos de valores de grupo típicos de las variables. El objetivo de los análisis cluster es maximizar la varianza entre grupos y minimizar la varianza dentro del grupo [Andenberg, 1973].

Los métodos de cluster se dividen en dos familias: jerárquicos y no-jerárquicos. Para nuestro estudio, resulta conveniente adoptar un método de **aglomeración jerárquica** para agrupar las estaciones, ya que no es necesario definir el número de cluster *a priori* y permite generar una clasificación por niveles de similitud (**Figura 9**).



**Figura 9.** Esquema de clúster jerárquico en estaciones de monitoreo

Los resultados del análisis de cluster suelen estar representados por un **dendrograma** [Yidana et al., 2008]. De esta forma, la agrupación jerárquica permite determinar si cada estación de monitoreo, potencialmente redundante, tiene un sitio alternativo que pueda reemplazarla.

#### IV. Analizar el área de influencia de las estaciones de monitoreo

La calidad del aire está vinculada a las fuentes de emisión presentes en áreas vecinas a los sitios de monitoreo. A su vez, estas fuentes de emisión generan distintos contaminantes asociados a su actividad productiva. Conocer las características de las áreas circundantes a las estaciones de monitoreo nos permitió reconocer su impacto en los sitios de monitoreo.

Una herramienta útil para esto es contar con un **inventario de emisiones** del tipo *bottom-up* (del inglés, de abajo hacia arriba), en el cual se puedan localizar las principales fuentes de emisión cercanas a la estación de monitoreo [Gaitán Varón & Cárdenas Ruiz, 2017]. Sin embargo, como se mencionó anteriormente, no siempre se cuenta con ellos. Una información complementaria son los **mapas de uso de suelo** que brindan información sobre el paisaje del territorio estudiado.

Distintas agencias gubernamentales y grupos de investigación brindan mapas de uso de suelo a nivel global. Entre los proveedores más conocidos, podemos mencionar: la agencia europea Copernicus que ofrece un mapa de usos de suelo global, conocido como *Global Land Cover*, con una resolución de 100m [Büttner et al., 2004]; la NASA que brinda un mapa de cobertura del suelo (producto MCD12Q1) con una resolución de 500 m, basado en la clase dominante en el período 2001-2010 [Channan et al., 2014]; la USGS que ofrece el mapa de usos de suelo GLCC (del inglés, *Global Land Cover Characterization*) generado a partir de las mediciones anuales del sensor AVHRR.

#### V. Análisis de los vientos

El estudio de los vientos aportó conocimientos determinantes sobre la distribución y la dispersión de los contaminantes. Las condiciones climáticas influyen de forma directa sobre la generación, permanencia, movimiento y dispersión de los gases y partículas presentes en el aire. Mientras que una situación de calma y alta incidencia solar, favorecen a la acumulación de contaminantes secundarios, vientos fuertes pueden dispersar los contaminantes lejos de su fuente de emisión.

La **rosa de vientos** es una representación habitual de la frecuencia e intensidad de los vientos desde los distintos ángulos de procedencia. Los sectores o cuadrantes de la rosa de vientos muestran el origen del viento, mientras que los colores indican su intensidad. Esta representación nos permitió estudiar las velocidades y direcciones del viento más frecuentes en cada sitio de monitoreo.

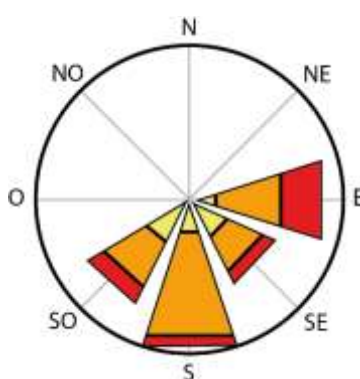
#### VI. Análisis de los aportes de las fuentes a las estaciones

Una primera aproximación de las actividades productivas que contribuyeron a las máximas concentraciones en nuestros sitios de medición se obtuvo mediante el **análisis de las rosas de contaminación** junto a un mapa de las fuentes de emisiones.

Con la misma composición gráfica que la rosa de vientos, en la rosa de contaminación, se indica la frecuencia de las direcciones de incidencia del viento y las concentraciones de los contaminantes en rangos mediante colores y su frecuencia de aparición [Vardoulakis et al., 2007; Ratto et al., 2006]. De esta forma, una misma gráfica resume 3 características claves para comprender mejor los sectores del viento desde donde suceden los aportes de los contaminantes

en el sitio de medición (**Figura 10**). Si esta información se superpone en un mapa, podemos distinguir las fuentes de emisiones dado que se posicionan siempre a barlovento del sitio de monitoreo.

Como los datos no se encuentran distribuidos linealmente, la clasificación de cuantiles no es adecuada para la selección de los intervalos de clasificación de las concentraciones. Para esto resulta conveniente utilizar el algoritmo de rupturas naturales de Jenks [Jenks, 1967]. Los cortes de clase se caracterizan porque agrupan mejor los valores similares y maximizan las diferencias entre clases. El parámetro, en nuestro caso la concentración, se divide en clases cuyos límites quedan establecidos dónde hay diferencias considerables entre los valores de los datos [Smith et al., 2015].



**Figura 10.** Ejemplo de una rosa de contaminación.

Una solución más compleja nos la brindan los **modelos de tipo receptor**. Estos modelos se utilizan para determinar las contribuciones directas de diversas fuentes de contaminación atmosférica a las concentraciones ambientales medidas en un solo receptor a escala urbana o a menor escala. Se han desarrollado diversos modelos de receptores para identificar cualitativamente y estimar cuantitativamente las contribuciones de las fuentes [Thurston y Spengler, 1985; Paatero y Tapper, 1994; Watson et al., 2001; Kim y Henry, 2000].

A una escala mayor, los **modelos de retrotrayectoria** se han utilizado generalmente en estudios de contaminación atmosférica para determinar las zonas de origen de los contaminantes medidos en un punto de muestreo. Mediante un modelo numérico se calcula el recorrido que sufrió la masa de aire en el tiempo (varios días). El modelo híbrido de Trayectoria Integrada Lagrangiana de una sola partícula (HYSPLIT, *Hybrid Single-Particle Lagrangian Integrated Trajectory*), desarrollado por el *Air Resources Laboratory* de NOAA, es uno de los modelos más utilizados para los cálculos de trayectoria y dispersión atmosférica [Stein et al., 2015]. Puede utilizarse para modelar el flujo, la dispersión y la deposición de plumas de contaminantes químicos y radiológicos.

En este trabajo nos hemos limitado a la etapa de identificar posibles fuentes mediante las rosas de contaminación, dado que resultan un excelente primer paso a los modelos más complejos. Sin embargo, es posible incorporar estas nuevas fases a futuros trabajos.

## VII. Análisis exploratorio de los datos

El **análisis exploratorio** de los datos implicó el cálculo de los principales estadísticos descriptivos (valor mínimo, media, mediana, máximo) para cada contaminante por estación de monitoreo. Con esta información fue posible identificar las estaciones que presentaron una mayor variabilidad y concentración media mayor a lo largo del período de estudio.

También resultó conveniente realizar un estudio de la frecuencia y un análisis de la frecuencia acumulada de las concentraciones para identificar las más frecuentes dentro del período. Los datos de calidad del aire presentan, por lo general, una distribución de probabilidad positiva continua, donde se observa una proporción más alta de bajas concentraciones, disminuyendo la probabilidad de ocurrencia de las concentraciones más altas. Las concentraciones de contaminantes atmosféricos suelen presentar una función de distribución gamma [Bakhtiarifar et al., 2017] o lognormal [Hasenfratz et al., 2015].

Con el fin de evaluar el cumplimiento de las normativas en calidad de aire y el riesgo a la salud, se compararon las mediciones en cada estación con la normativa local y con las recomendaciones de la OMS. En esta etapa se realizaron distintos tipos de gráficas que facilitaron el entendimiento del comportamiento de los datos. Por ejemplo, las gráficas de tipo violín que muestran la distribución de los datos y las gráficas tipo box-plot que permiten visualizar los principales percentiles.

Los **perfiles temporales** estudiados fueron (i) perfiles mensuales, (ii) semanales y (iii) por hora, y por estación de monitoreo y contaminante. Si bien la mediana es un estadístico más robusto, se utilizó por convención la media como parámetro de agregación.

## VIII. Evaluación de la tendencia temporal a largo plazo

Un estadístico relevante para la implementación de políticas públicas es la tendencia a largo plazo de las medias mensuales. La tendencia temporal da cuenta del crecimiento o decrecimiento de un parámetro a lo largo del tiempo, considerando un comportamiento lineal.

Las tendencias de los promedios mensuales de las concentraciones de los contaminantes se pueden analizar en cada estación de monitoreo utilizando la función **Theil-Sen** [Theil, 1950; Sen, 1968]. Este método prueba la significación estadística de las tendencias observadas y la magnitud de la tendencia, mediante el cálculo de las pendientes entre todos los pares de puntos ( $x$ ,  $y$ ) del conjunto de datos (siendo  $x$  el tiempo en meses e  $y$  la concentración media mensual del contaminante) y luego la estimación Theil-Sen de la pendiente que describe la mediana de todas estas pendientes [Carslaw y Ropkins, 2012].

Para los contaminantes que presentan un comportamiento cíclico se desestacionalizaron las series de datos utilizando el método STL (Desestacionalización de tendencias estacionales mediante el alisado de LOESS) para controlar las fluctuaciones [Carslaw y Ropkins, 2012].

## IX. Análisis del comportamiento espacial de la contaminación atmosférica

Desde el enfoque de la geoestadística, la concentración de un contaminante en el aire puede entenderse como una variable regionalizada, es decir, posee un comportamiento estocástico no puro. En contraposición con una variable completamente aleatoria, una variable regionalizada presenta cierto grado de autocorrelación, pero su comportamiento espacial es complejo para ser descritos por una función determinística. Esta consideración nos permite, cuando se cuenta con una gran cantidad de puntos, realizar análisis geoestadísticos como los métodos de interpolación espacial y de simulación [Deutsh y Journal, 1992]. Sin embargo, la baja densidad de puntos hace que los modelos presenten errores groseros, lo cual lleva a analizar otras soluciones.

Una mejor solución se obtiene mediante la implementación de **modelos de regresión multivariados**. En los últimos años, los *land use regression models* (LUR) han ganado protagonismo para modelar los contaminantes atmosféricos utilizando distintas técnicas de regresión como son Modelos Lineales Mixtos Generalizados (GLMMs) [Yang et al., 2018 a], Modelos Aditivos Generalizados (GAMs) [Hasenfratz et al., 2015; Hastie, 2017] y algoritmos de *machine learning* [Ni et al., 2018; Li et al., 2017; Hu et al., 2017; Huang et al., 2018; Chen et al., 2018]. A continuación, veremos en mayor detalle la construcción de modelos LUR utilizando Random Forest, un potente algoritmo de machine learning.

### *Modelos de regresión multivariados LUR*

Los **modelos LUR** combinan el monitoreo de la contaminación del aire en un pequeño número de estaciones y el desarrollo de modelos estocásticos utilizando variables predictoras obtenidas generalmente a través de teledetección y sistemas de información geográfica (SIG) [Hoek et al., 2008; Cowie et al., 2019]. El uso integrado de estos datos permite obtener una mejor comprensión de la variación espacial y temporal de los contaminantes que con cualquiera de estas fuentes por separado.

La mayoría de los estudios evalúan un gran conjunto de variables predictoras potenciales para modelar las concentraciones de los contaminantes. Los datos de predicción utilizados con mayor frecuencia incluyen: información de la cubierta del suelo, el tráfico vehicular, la distribución de la población, la elevación del terreno y las características meteorológicas [Hoek et al., 2008; de Hoogh et al., 2014; Mavko et al., 2008; Beelen et al., 2013].

A su vez, como se vio anteriormente, las imágenes satelitales aportan una buena cantidad de información relevante para el estudio de la calidad del aire, si bien su inclusión debe considerar dos grandes cuestiones. Por un lado, la resolución espacial de las imágenes debe ser buena para otorgar detalle a la escala urbana. Y por otra parte, debe ser viable la validación de los datos con mediciones locales.

La secuencia de pasos para la implementación de un modelo LUR contempla: 1) la selección del conjunto de variables predictoras para el modelo, 2) la validación de las imágenes satelitales; 3) entrenar el modelo con un set de datos de entrenamiento (80% del total de los datos); 4) evaluar el ajuste del modelo contra un set de datos de prueba (20% restante).

### Selección de las variables predictoras

En estudios previos, se ha observado que las variables predictoras seleccionadas para la incorporación en los modelos LUR difieren sustancialmente entre las distintas aplicaciones, sin mostrar un patrón de preferencia por ninguna de ellas. Esto puede estar relacionado con la disponibilidad de los datos, las características únicas del área de estudio (por ejemplo, las variaciones en el relieve) y con las decisiones arbitrarias que el personal técnico realizó en su confección. Por lo general, se espera que el modelo sea “alimentado” con más de 10 variables predictoras, y que se tenga conocimiento de las mismas de forma extensiva sobre el territorio.

### Validación de las imágenes satelitales

Las mediciones de AOD se calibraron utilizando datos de fotómetros solares provistos por la red de AERONET. Los instrumentos AERONET miden AOD usando dos protocolos diferentes. Uno de ellos deriva el AOD midiendo la radiación entrante del haz directo del sol, mientras que el segundo protocolo utiliza mediciones de radiación difusa desde varias direcciones del cielo para derivar el AOD. Estas mediciones se realizan varias veces por hora en función del ángulo del sol.

Los datos de AERONET AOD se calculan para tres niveles de calidad diferentes: Nivel 1 (datos crudos sin calibración), Nivel 1.5 (datos donde se han extraído las nubes automáticamente) y Nivel 2 (datos sin nubes, calibrados e inspeccionados manualmente). Es recomendable que los datos de AERONET que se utilicen para la validación procedan únicamente de mediciones directas de la radiación solar con calidad de Nivel 2.

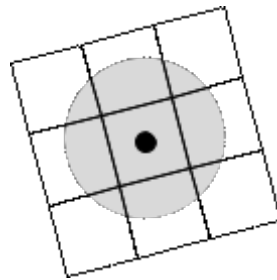
La validación de AOD se realizó por el método de interpolación cuadrática, utilizando las longitudes de onda más cercanas de AERONET a la longitud de onda empleada para las mediciones satelitales [Prasad et al., 2007]. La relación empírica entre AOD y la longitud de onda de la luz incidente viene dada por la ecuación formulada por Ångström (1929):

$$\frac{\tau_{\lambda_1}}{\tau_{\lambda_2}} = \left(\frac{\lambda_1}{\lambda_2}\right)^{-\alpha}$$

donde  $\tau$  es la profundidad óptica del aerosol a la longitud de onda ( $\lambda$ ) y  $\alpha$  se conoce como el coeficiente de Ångström, que normalmente es un indicador del tamaño de la partícula del aerosol.

La coincidencia espaciotemporal es una condición esencial para la comparación [Ichoku et al., 2002]. Por ello, para productos diarios, resulta preciso obtener la información de las horas de paso de los satélites, y utilizar las mediciones realizadas por AERONET dentro de un rango menor de 30 min de cada paso de órbita [Petrenko et al., 2012]. Como existe una inconmensurabilidad entre los datos puntuales generados por AERONET y los valores medidos por satélite que representan áreas, resulta conveniente generar un buffer en torno a la estación y promediar los valores medidos por el sensor en esa superficie para relacionarlo con los datos de AERONET [Ichoku et al., 2004]. Si bien no existe un criterio para el tamaño del buffer a utilizarse, el diámetro suele ser entre 2.5 y 3 veces el lado del píxel [NASA, 2009; Prasad et al., 2007] (**Figura 11**).





**Figura 11.** Esquema representativo de la imagen satelital, la estación de monitoreo y el área de influencia.

En el caso de las mediciones satelitales de gases, resulta mucho más compleja su validación. Si bien, no es preciso realizar validaciones contra mediciones en tierra, ya que su algoritmo de recuperación incluye modelos de dispersión atmosféricos, se puede calibrar su medición contra campañas atmosféricas realizadas por aeronaves. Sin embargo, como estos datos no se encuentran disponibles de forma extensiva, no es una práctica habitual incluir la validación de estos productos en los trabajos [Laughner et al., 2016].

### *Entrenamiento del modelo*

La bibliografía muestra un amplio uso de técnicas para el desarrollo de los modelos. Una gran cantidad de los estudios utilizan técnicas de regresión lineal estándar [Hoek et al., 2008; Henderson et al., 2007; Meng et al., 2016; Li et al., 2011], modelos aditivos generalizados [Paciorek et al., 2008] o modelos mixtos [Zhang et al., 2018; Shtein et al., 2018; Lee et al., 2011; Yap and Hashim, 2013] para desarrollar modelos de predicción. A menudo se aplican métodos de selección automática de subconjuntos hacia adelante [Schikowski et al., 2017] o hacia atrás para desarrollar un modelo parsimonioso a partir de un gran conjunto de variables predictoras que maximizan el porcentaje de variabilidad explicada ( $R^2$ ). Los trabajos más recientes aplican algoritmos de inteligencia artificial como redes neuronales [Ni et al., 2018; Li et al., 2017; Gupta & Christopher, 2009; Mao et al., 2017] o Random Forest [Zhao et al., 2019; Hu et al., 2017; Huang et al., 2018; Chen et al., 2018].

Los modelos de regresión de **Random Forest** (RF) tienen una alta precisión en la predicción, son rápidos de entrenar, relativamente robustos a los valores atípicos y difíciles de sobreentrenar [Breinman, 2001]. El algoritmo de RF se puede utilizar sin necesidad de ajustar los parámetros del algoritmo, aunque a menudo se puede obtener fácilmente un mejor modelo de clasificación mediante la optimización de pocos parámetros. Como ventaja adicional, brinda la importancia de las variables incluidas en el modelo para predecir la variable en estudio, produciendo así resultados más interpretables que otros algoritmos de *machine learning* como con las redes neuronales [Hu et al., 2017].

El modelo Random Forest está basado en la combinación de múltiples árboles de decisión [Breinman, 2001]. Para cada árbol de decisión, crea un conjunto de entrenamiento (*bootstrap*) diferente mediante el muestreo aleatorio del conjunto de datos, que contiene aproximadamente dos tercios de los datos originales (**Figura 12**).

El árbol de decisión avanza utilizando el conjunto de datos del bootstrap mediante particiones recursivas. Cada nodo del árbol representa una prueba binaria contra la variable predictiva seleccionada, del tipo  $X_i > t$  donde  $t$  representa un valor al azar dentro del dominio de la variable. Los nodos terminales contienen el número máximo especificado de ejemplos de los que se obtiene el valor predicho mediante un promedio.

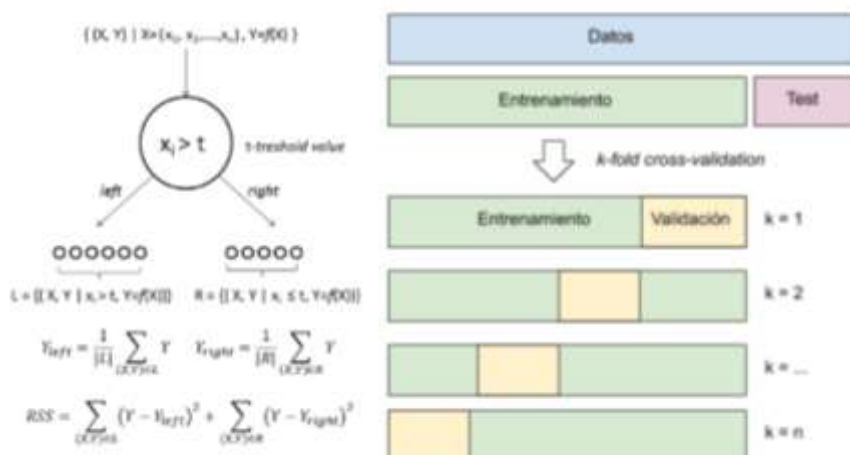


Figura 12. Esquema de funcionamiento de un árbol de decisión para el problema de regresión y la división de los datos de entrada en el algoritmo k-fold cross-validation

Una buena forma de mejorar los modelos predictivos es mediante la **validación cruzada** [Rodríguez et al., 2010]. En la validación cruzada (*k-fold cross-validation*), la muestra original se divide aleatoriamente en  $k$  submuestras de igual tamaño. Mientras una de las submuestras se conserva para la validación del modelo, las submuestras  $k-1$  restantes se utilizan como datos de entrenamiento del modelo. El proceso de entrenamiento se repite  $k$  veces. Cada una de las submuestras se utiliza una única vez tanto en el entrenamiento como en la validación (**Figura 12**).

### Evaluación el ajuste del modelo

Evaluar la bondad de predicción de un modelo es el último paso antes de su implementación. La evaluación es, en general, un procedimiento complejo [Chang y Hanna, 2004] que comprende diferentes etapas (evaluación científica, verificación de códigos, validación de modelos, análisis de sensibilidad, etc.). Los modelos aplicados para la evaluación reglamentaria de la calidad del aire se evalúan comúnmente sobre la base de comparaciones con las observaciones.

Este proceso de evaluación se conoce como *análisis de desempeño estadístico*, ya que mediante distintos indicadores estadísticos y el análisis gráfico se determina la capacidad del modelo para reproducir las concentraciones medidas. Aunque la comparación entre las concentraciones modeladas y las observadas no puede dar una visión completa de las propiedades del modelo, se considera un buen primer paso en la evaluación del desempeño del modelo [Derwent et al., 2010; Irwin et al., 2008].

Existen distintos parámetros para evaluar un modelo pero, en general, se comparan los valores obtenidos del modelo con un conjunto de mediciones no utilizados para su construcción. A continuación, mencionaremos los parámetros utilizados en esta tesis (**Tabla 7**).

El **error cuadrático medio (RMSE)**, el **error absoluto medio (MAE)** y el **error porcentual absoluto medio (MAPE)** vienen dados por las fórmulas presentadas en la **Tabla 7**, donde  $n$  es el número de pares de datos,  $e$  es el valor estimado y  $m$  es el valor medido. Estas medidas pueden servir como parámetros de la exactitud del modelo, pues miden la distancia al promedio.

Por otra parte, el **sesgo (BIAS)** nos proporciona información sobre la tendencia del modelo, a sobreestimar o subestimar un variable, otorgando una medida de la fiabilidad del modelo. Debido a que BIAS carece del término en valor absoluto o al cuadrado, se permite la cancelación de errores individuales positivos y negativos, poniendo en evidencia solamente el exceso y, por tanto, cuantifica el error sistemático del modelo.

**Tabla 7.** Parámetros estadísticos para evaluar la bondad de ajuste de un modelo. Donde  $n$  es el número de pares de datos,  $e$  es el valor estimado y  $m$  es el valor medido.

Descripción	Símbolo	Fórmula
<i>BIAS</i>	BIAS	$\frac{1}{n} \sum_{i=1}^n (e_i - m_i)$
<i>Error cuadrático medio</i>	RMSE	$\sqrt{\frac{1}{n} \sum_{i=1}^n (e_i - m_i)^2}$
<i>Error absoluto medio</i>	MAE	$\frac{1}{n} \sum_{i=1}^n  e_i - m_i $
<i>Error porcentual absoluto medio</i>	MAPE	$\frac{100}{n} \sum_{i=1}^n \frac{ e_i - m_i }{m_i}$

## X. Mapas de riesgo

Como se mencionó en el marco teórico, para calcular la mortalidad provocada por la contaminación del aire se han confeccionado distintas funciones de riesgos relativos (RR), a partir de un gran número de trabajos en distintas ciudades y con tiempos promedios de estudio diferentes [Ostro & World Health Organization, 2004; Pope et al., 2002].

El RR es la probabilidad de que se produzcan efectos sobre la salud (por ejemplo, mortalidad por todas las causas, mortalidad por cáncer de pulmón) en una población expuesta a un nivel de contaminación atmosférica superior al que se considera de fondo [Ostro & World Health Organization, 2004],  $10 \mu\text{g}\cdot\text{m}^{-3}$  para  $\text{PM}_{10}$  y  $3 \mu\text{g}\cdot\text{m}^{-3}$  para  $\text{PM}_{2.5}$ ). En la **Tabla 8** se pueden ver las funciones de RR y los valores obtenidos para  $\beta$  en función de la métrica de exposición y del tamaño de PM evaluado.

**Tabla 8.** Métrica de exposición y funciones de riesgo relativa recomendados. Editado de [Ostro & World Health Organization, 2004]. Donde X es la concentración media anual medida o modelado de PM ( $\mu\text{g}\cdot\text{m}^{-3}$ ),  $X_0$  es la concentración base para una población sin exposición ( $3 \mu\text{g}\cdot\text{m}^{-3}$  para  $\text{PM}_{2.5}$  y  $10 \mu\text{g}\cdot\text{m}^{-3}$  para  $\text{PM}_{10}$ ) y  $\beta$  es el factor de pendiente.

Métrica de exposición	Función de riesgo relativo	Coefficiente $\beta$ (95% de IC)	Subgrupos
<i>Todas las causas de muerte por exposición a corto plazo de <math>\text{PM}_{10}</math></i>	$RR = e^{\beta (X - X_0)}$	0.0008 (0.0006, 0.0010)	Todas las edades
<i>Mortalidad respiratoria por exposición a corto plazo de <math>\text{PM}_{10}</math></i>	$RR = e^{\beta (X - X_0)}$	0.0016 (0.0003, 0.0030)	Edad < 5 años
<i>Mortalidad cardiopulmonar por exposición a largo plazo de <math>\text{PM}_{2.5}</math></i>	$RR = \left(\frac{X + 1}{X_0 + 1}\right)^\beta$	0.155 (0.056, 0.025)	Edad > 30 años
<i>Cáncer de pulmón por exposición a largo plazo de <math>\text{PM}_{2.5}</math></i>	$RR = \left(\frac{X + 1}{X_0 + 1}\right)^\beta$	0.232 (0.086, 0.379)	Edad > 30 años

Un RR mayor de 1 indica un aumento respecto a una población control. De esta forma, el RR tiene una interpretación intuitiva, un RR igual a 2 expresa que el riesgo es el doble al de un grupo control.

Una vez determinado el RR, se puede conocer la fracción atribuible de muertes (FAM) que es la proporción de muertes por una dada enfermedad que podrían evitarse si los niveles del contaminante se reducen:

$$FAM = \frac{(RR - 1)}{RR}$$

## Módulo III - Etapa de comunicación de los resultados



### III. I. Análisis de los resultados

Muchas veces el análisis de los resultados obtenidos hizo que se vuelva a la fase de preprocesamiento para obtener más datos o más atributos. Para que el proceso confluya, fue esencial tener el conjunto de datos preprocesados, los metadatos correspondientes elaborados y los algoritmos elegidos. Esta etapa final del proceso ha sido crucial en nuestro trabajo y por ello se le ha brindado especial detalle en cada uno de los casos estudiados.

### III.II. Reportes y Visualizaciones

Como última etapa se elaboraron informes y visualizaciones para resaltar los resultados obtenidos en los pasos anteriores.

La comunicación de los resultados es el mayor desafío del proceso, ya que requiere mostrar grandes conjuntos de datos de diferentes tipos y múltiples fuentes a personal técnico o no experto que no siempre conocen la complejidad de los datos o de los análisis realizados [Yeganeh et al., 2017]. Es, por tanto, necesario que los reportes sean comprensibles para un público amplio y autodescriptivos. Esta es la única parte del proceso analítico que llega a la audiencia. Y que incluso cuando las conclusiones pueden resultar evidentes, el público también juega un papel central en la interpretación de los resultados.

No es aconsejable que el mensaje suene apocalíptico, transmitiendo una amenaza a la cual no se puede hacer frente. Por el contrario, el mensaje es más efectivo cuando la percepción del peligro no excede la percepción de los recursos o de las capacidades para enfrentarlo [Brulle, R. J. (2010)].

La mala presentación de las conclusiones puede oscurecer cuestiones clave y sesgar las percepciones del público sobre la calidad del aire [Ma et al., 2012]. La comunicación efectiva requirió la identificación de puntos específicos del análisis que se desearon mostrar. En este

sentido, fue necesario dejar de lado los gráficos exploratorios elaborados para comprender las relaciones más profundas del conjunto de datos. Asimismo, se descartaron las visualizaciones diseñadas exclusivamente para el análisis exploratorio [Knafllic, 2015].

La representación visual que se seleccione debe proporcionar una imagen clara de los patrones y relaciones descubiertos facilitando la interpretación y comprensión de la información por parte de la audiencia. Es aconsejable que la presentación de los datos pueda ser presentada en forma de narración y con una presentación visual adecuada [Ma et al., 2012]. A continuación, se presentan algunos ejemplos de las visualizaciones más utilizadas que ya fueron introducidas en el texto.

El gráfico más utilizado para visualizar las relaciones entre distintas variables es el dendrograma que se ha visto en la **Figura 9**. Este tipo de representación es particularmente útil para presentar la disposición de los clusters producidos por las técnicas de agrupación jerárquica. Cuando se desea mostrar la agrupación de las estaciones de monitoreo, puede ser conveniente presentarla en un mapa para su representación en el espacio [Xie et al., 2018; Wang et al., 2018].

Las series temporales se utilizan para visualizar el patrón de tiempo. Es común agregar datos para mostrar el comportamiento durante la semana y los meses. También es útil mostrar la serie temporal completa con la tendencia del período analizado [Elangasinghe et al., 2014]. Para permitir comparaciones diarias, mensuales y anuales de la variabilidad de la contaminación del aire es conveniente emplear una tabla de calendario para una fácil visualización [Leung et al., 2018].

Las rosas de contaminación representan los niveles de contaminación observados por cada estación de monitoreo según la dirección del viento (**Figura 10**). El modelo de la rosa de la contaminación se basa en la simple idea de clasificar las mediciones de concentración obtenidas en un sitio determinado para un período determinado según la dirección en la que sopla el viento en el momento de la medición [Duboue, 1978]. Esta gráfica proporciona una noción relativa de la ubicación de las fuentes de contaminación.

Los mapas son una forma eficaz de visualizar y comunicar la variación espacial de las concentraciones de contaminantes. Los mapas de calor se utilizan para identificar las áreas de mayor concentración [Austin et al., 2013]. Sin embargo, si no se selecciona una escala de colores apropiada, se vuelven confusos. Por tanto, resultó conveniente utilizar una paleta monocromática para mostrar las zonas de excedencia y una multicolor para mostrar las diferencias en el territorio.

Finalmente, para reforzar la narrativa de la presentación de los resultados, se compararon las concentraciones medidas o previstas con los estándares de calidad del aire. Dado que, en general, existen diferencias entre las normas de calidad del aire y las recomendaciones de la Organización Mundial de la Salud (OMS), fue conveniente compararlas con ambos valores. Las normas de calidad del aire a menudo están condicionadas por cuestiones de política económica y tardan mucho tiempo en actualizarse, mientras que las recomendaciones de la OMS se basan en investigaciones científicas que respaldan estos valores y se revisan cuando se reúnen evidencias suficientes [Zhang et al., 2018].

## Confeción de índices

Con el fin de facilitar la interpretación del público, la comunicación de información sobre la calidad del aire al público y a los responsables de la formulación de políticas se realiza, en gran medida, mediante el uso de indicadores e índices [Zhang et al., 2012].

Mientras que un indicador es un parámetro estadístico que se crea a partir de los datos científicos con el fin de obtener y compartir información para que se puedan tomar decisiones o establecer políticas [Hsu et al., 2013], un índice es una agregación de indicadores, que a veces es simplificado para que sea más fácil de comunicar al público que múltiples valores. Los índices suelen ser adimensionales y pueden proporcionar consistencia a un valor comunicado para múltiples contaminantes del aire con diferentes unidades de medida (por ejemplo, de 0 a 100).

Para la comunicación pública de las previsiones o en tiempo real, se suele utilizar un índice de calidad del aire (AQI, por sus siglas en inglés: *air quality index*), donde se establecen múltiples rangos de severidad, dependiendo del nivel de respuesta requerida por los diferentes grupos afectados. Por lo general, el AQI convierte una concentración de contaminante en un número en una escala conocida, y los niveles de contaminantes se clasifican en "muy alto", "alto", "moderado" y "bajo" [Sadat et al., 2015, Engel-Cox et al., 2013]. Los puntos de ruptura en la categorización varían de un contaminante a otro [Mintz, 2012]. Una variante puede observarse en [Olvera-García et al., 2016] donde se aplicó un análisis difuso para generar ponderaciones AQI de contaminantes con mayores efectos en la salud.

Informar mediante un AQI presenta tres grandes ventajas. En primer lugar, garantiza el cuidado de las personas mediante consejos que les permitan adaptar su comportamiento y/o tomar medidas de mitigación para evitar los efectos en la salud. En segundo lugar, sirven para condensar los datos complejos facilitando la toma de decisiones. Una tercera razón es llamar la atención del público en la problemática de la contaminación atmosférica, generando un mayor nivel de sensibilización reportando la calidad del aire de forma diaria [Van den Elshout et al., 2014].

A pesar de las ventajas, los AQI presentan ciertas limitaciones en su empleo dado que los factores que intervienen entre la calidad del aire y los efectos en la salud son múltiples y diversos, interactúan de maneras aún poco conocida, y su naturaleza exacta es difícil de cuantificar.

Además, el valor del índice reportado se refiere a un momento cuando los efectos en la salud ocurren a diferentes tiempos de exposición. Por ejemplo, un AQI basado en los efectos a corto plazo para la salud puede indicar "niveles buenos" de calidad del aire, implícitamente señalando que la calidad del aire no es un problema, y al mismo tiempo, superar los valores límite anuales.

A su vez, no todos los contaminantes perjudiciales para la salud se encuentran regulados y controlados, y por tanto, se excluyen comúnmente de los índices. Otro inconveniente es el hecho que pocos índices se basan en datos epidemiológicos locales careciendo de especificidad [Stieb et al. 2008]. Estas cuestiones obligan a tomar con reparo el uso de un AQI.

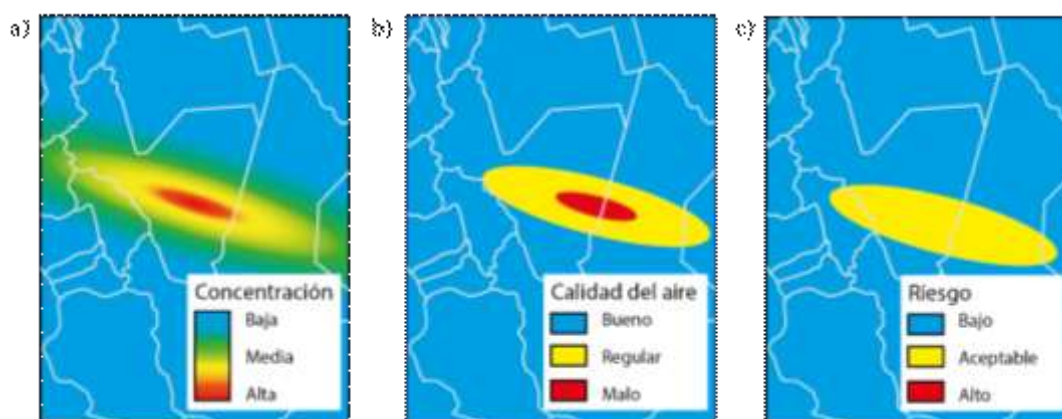
En esta tesis se trabajó con el AQI propuesto en [World Air Quality Index project, 2015] ya que se basa en las recomendaciones de la OMS para su confección [Krzyzanowski & Cohen, 2008], estableciendo una escala cromática para cada uno de los niveles (**Tabla 9**).

**Tabla 9.** AQI de la calidad del aire [World Air Quality Index project, 2015].

Categoría	Nivel	PM <sub>2.5</sub>	PM <sub>10</sub>
<i>Buena</i>	Nivel guía de calidad del aire	0 - 10	0 - 20
<i>Moderada</i>	Objetivo intermedio - 1	10 - 15	20 - 30
<i>Dañina para personas sensibles</i>	Objetivo intermedio - 2	15 - 25	30 - 50
<i>Dañina</i>	Objetivo intermedio - 3	25 - 35	50 - 70
<i>Muy dañina</i>	Por encima del objetivo	35 - 53	70 - 105
<i>Peligrosa</i>	Significativamente por encima del objetivo	53 - ∞	105 - ∞

### Confección de mapas de calidad del aire y mapas de riesgo

Como un último recurso, los mapas de calidad de aire y de riesgo son elementos ampliamente utilizados en el diseño y aplicación de políticas territoriales (**Figura 13**). La visualización mediante mapas facilitó la identificación de patrones espaciales, permitiendo delimitar las áreas que requieren una mayor intervención.



**Figura 13.** Esquema de a) mapa de concentración, b) mapa de calidad de aire y c) mapa de riesgo

Mientras los mapas de concentración muestran los niveles de los contaminantes en el territorio, los mapas de calidad de aire presentan una categorización de las concentraciones a la cual se expone la población y se la relativiza en función de los niveles recomendados por la OMS [Hasenfratz et al., 2015]. De igual forma, el riesgo por exposición a los contaminantes se calcula sobre los valores de concentración del mapa [Soares & Pereira, 2007], tal como se vio anteriormente cuando se presentó el cálculo del riesgo relativo.



Estos mapas se pueden confeccionar de forma diaria, mensual o anual dependiendo su finalidad [Saisana et al., 2004]. Si bien para pronósticos conviene trabajar con mapas diarios por el alcance de las predicciones, con fines epidemiológicos resulta conveniente contar con mapas mensuales o anuales, acorde al mínimo nivel de agregación de los datos sanitarios.

Por último, los mapas de riesgo son productos cartográficos que se pueden visualizar dentro de un Sistema de Información Geográfico (SIG), permitiendo su integración con otras capas de información en función de la aplicación que se desee dar, como pueden ser las fuentes de contaminación o los servicios de emergencias disponibles.

*Caso de aplicación I*

## **Comunidad Valenciana, España**

---



## Introducción

En los capítulos anteriores se presentaron el marco teórico, desde donde se enfoca esta tesis, y una propuesta metodológica de trabajo para el abordaje del estudio de la calidad del aire. En los próximos capítulos se presentarán distintos casos de aplicación para evaluar la propuesta metodológica realizada.

Como primer sitio de estudio se seleccionó la Comunidad Autónoma de Valencia, España. España se sitúa al sur de Europa ocupando gran parte de la península ibérica y una serie de islas e islotes en el mar Mediterráneo y en el océano Atlántico. La forma de organización interna del país es mediante comunidades autónomas, distribuyéndose el territorio español en 17 comunidades autónomas. La Comunidad Valenciana, sobre la costa este, incluye un total de 542 municipios con 5 millones de habitantes, distribuidos en una superficie de 23 mil km<sup>2</sup> (densidad poblacional 219.1 hab/km<sup>2</sup>)

España se encuentra enmarcada dentro de la UE, la cual presenta precisas directivas respecto a los límites legales de los contaminantes atmosféricos y a los procedimientos administrativos requeridos para la evaluación de la calidad del aire. Esta estructura legal le otorga a España elementos jurídicos que aportan a la construcción de una administración eficiente, además de recursos económicos para llevar adelante evaluaciones y proyectos de mejora de la calidad del aire.

En general, España presenta niveles moderados de contaminantes atmosféricos en comparación con otros países de Europa [EEA, 2017] con superaciones recurrentes en áreas urbanas como Madrid, Barcelona y Granada [Ministerio para la Transición Ecológica, 2018]. Además de los efectos de la contaminación por el crecimiento urbano y el desarrollo industrial [Aldabe et al., 2011], se presentan incrementos de los niveles de aerosoles de forma natural por las intrusiones de masas de aire provenientes de África [Díaz et al., 2017]. Las plumas de polvo procedentes de las regiones desérticas (Sahara) y semiáridas (Sahel) son transportadas por largas distancias, a lo largo del Océano Atlántico y del mar Mediterráneo [Engelstaedter & Washington, 2007], afectando a España y a una extensa parte del territorio europeo.

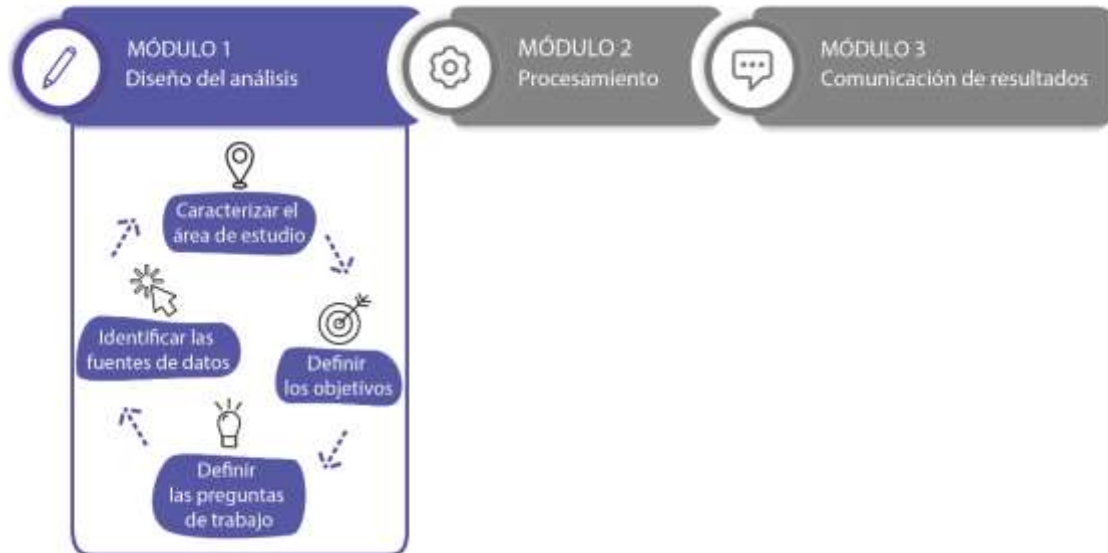
En la Comunidad Valenciana se han realizado distintos estudios sobre efectos atribuibles a la contaminación atmosférica. [Tenías et al., 1998] señalaron que las concentraciones de los contaminantes atmosféricos en Valencia se asociaban significativamente con las visitas a la sala de urgencias por asma. En [Ballester et al., 1996] se encontró una relación positiva entre la contaminación atmosférica y la mortalidad. Y en [Ballester et al., 2010], se demostró que la exposición prenatal al NO<sub>2</sub> redujo el crecimiento fetal.

En el último tiempo, distintas organizaciones sociales han manifestado una preocupación por la calidad del aire [www.valenciaperlaire.org], poniendo de manifiesto el sub-aprovechamiento de los datos recolectados y la escasa producción de información pública generada por los organismos responsables. Esto ha llevado a considerar la Comunidad Valenciana como un sitio excelente de aplicación.

A continuación, se presenta la aplicación de la metodología propuesta en los datos históricos que presenta la Comunidad Valenciana:

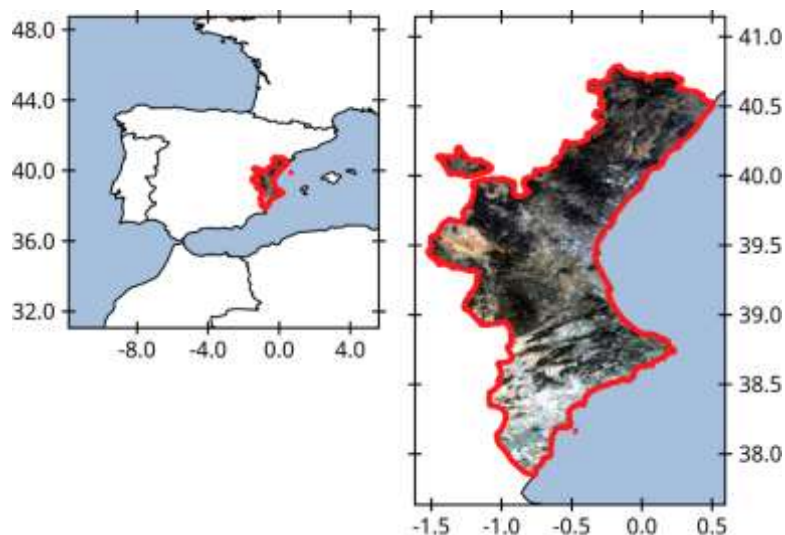
# Aplicación de la metodología

## Módulo I - Etapa de diseño del análisis



### I.I. Caracterización del área de estudio

La **Comunidad Valenciana** es una comunidad autónoma de España situada en el este de la península ibérica, en el litoral del mar Mediterráneo. Las costas de la región valenciana forman dos grandes arcos: el golfo de Valencia (Costa de Azahar) y el golfo de Alicante (Costa Blanca), ambos separados por las prolongaciones montañosas que se hunden en el mar (**Figura 14**). El dominio de estudio se definió entre los límites: W  $-1.5^\circ$ , N  $40.8^\circ$ , E  $0.7^\circ$ , S  $37.8^\circ$ .



**Figura 14.** Mapa de la Comunidad Autónoma de Valencia.

Desde una visión de conjunto, la región valenciana es la zona de contacto y final de dos grandes sistemas montañosos: el sistema Ibérico por el norte y el sistema Bético por el sur. Las cordilleras Béticas son un conjunto de sistemas montañosos que se alarga más de 600 km con orientación OSO - ENE, por el sur de la península ibérica, con una cota máxima es de 3479 msnm en el pico Mulhacén en Granada, siendo el conjunto de relieve más importante de España. El sistema Ibérico, por su parte, se extiende a lo largo de 500km por el interior de la península en dirección NO-SE, alcanzando los 2314 msnm en su pico más alto, el Moncayo.

Los sistemas montañosos tienen una gran influencia en el clima. Algunos importantes efectos en los flujos atmosféricos están provocados por las cadenas montañosas más elevadas de la región, como son el sistema Ibérico, Pirineos, Macizo Central, Alpes, Apeninos y Atlas. Estas pueden actuar como barrera que limitan el movimiento de las masas de aire, y pueden generar vórtices aerodinámicos o extensos bloqueos. A su vez, la alta concentración de centros ciclónicos en la región, y la cercanía con el vórtice polar estratosférico, con especial influencia en invierno, convierten a la región en un complejo sistema atmosférico [Belda et al., 2001].

En particular, la región de la Comunidad Valenciana se caracteriza por la irregularidad y los contrastes regionales entre zonas lluviosas y zonas áridas. En la parte nororiental, por su proximidad al mar se presentan inviernos poco rigurosos. El máximo de las precipitaciones ocurre en las áreas con mayor relieve, con un volumen promedio de 700 - 800 mm.año<sup>-1</sup>. En otoño, el efecto pantalla que ejerce el relieve favorece las precipitaciones en las zonas abiertas. La influencia de los flujos de aire húmedos de levante hace de ésta la estación más lluviosa.

Por otra parte, en el sector al noroeste se registran las temperaturas medias anuales más bajas y el volumen de precipitaciones anuales decrece gradualmente hacia el suroeste. Las temperaturas medias siguen un gradiente decreciente de interior a costa en invierno, levemente creciente en verano, por efecto de la altitud.

La evolución de las temperaturas medias mensuales sigue un cierto retraso respecto al ritmo solar. El régimen de brisas por la influencia marítima tiene como consecuencia una marcada suavización de las temperaturas máximas, sobre todo en las áreas costeras. El aumento de temperatura se produce a un ritmo lento e irregular, que culmina en julio o agosto. Los principales cambios en la temperatura se dan de abril a mayo, cuando se produce el máximo de la mayor radiación solar y la desaparición de gran parte de las penetraciones polares; y de junio a julio, cuando la incidencia solar relativa llega a su máximo [Milán et al., 1997].

### *Contexto jurídico-normativo*

España posee un complejo marco jurídico para evaluar los niveles de contaminantes atmosféricos y atender las situaciones en las que se presenta una mala calidad del aire. Su estudio data de medio siglo atrás, siendo el primer referente legislativo nacional la Ley 38/1972 de protección del medio atmosférico, sobre normas técnicas para análisis y valoración de los contaminantes presentes en el aire [Patier, 2015]. Con la entrada de España a la UE, se sancionaron distintas Leyes y Reales Decretos acordes donde se adoptaron los valores legislados y los métodos a las exigencias.

Las Directivas de la UE brindan un marco regulatorio estableciendo valores límite de concentración de contaminantes en el aire, y normas sobre las fuentes de emisión. La Directiva

2008/50/CE, que condensa las anteriores Directivas 1996/62/CE y 1999/30/CE, dispone medidas para definir y establecer objetivos de calidad del aire ambiente para evitar, prevenir o reducir los efectos nocivos para la salud humana y el medio ambiente en su conjunto. A su vez, establece métodos y criterios comunes para la gestión de la calidad del aire. Finalmente, propone reunir información y asegurar que esta se halle a disposición de la ciudadanía, y fomentar el incremento de la cooperación entre los Estados miembros para reducir la contaminación atmosférica.

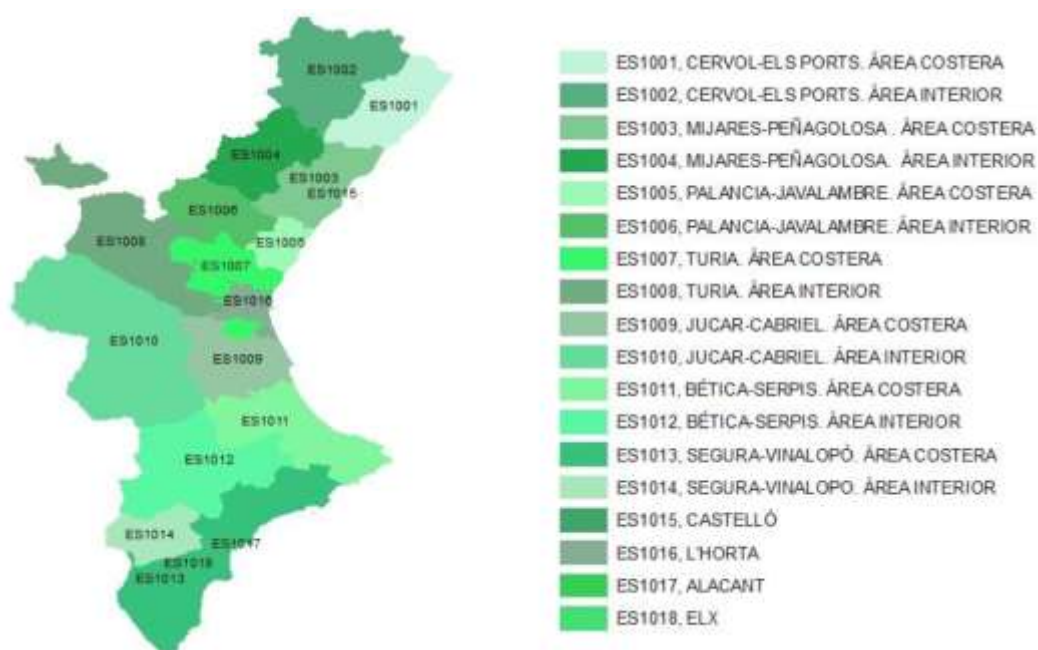
La **Tabla 10** muestra los valores legislados en Europa. Como se señaló anteriormente, la implantación de los estándares de calidad del aire se establece por criterios de costo-beneficio [Boldero & Querol, 2014], por ello pueden diferir de los valores recomendados por la OMS. Por ejemplo, la recomendación de la OMS para  $PM_{2.5}$  ( $10 \mu\text{g}\cdot\text{m}^{-3}$ ) es menor a de la mitad que el valor límite europeo ( $25 \mu\text{g}\cdot\text{m}^{-3}$ ), mientras que Estados Unidos reguló un valor límite de  $12 \mu\text{g}\cdot\text{m}^{-3}$ . Una particularidad para destacar es que la Directiva contempla un número de excedencias permitidas por año para algunos de los contaminantes.

**Tabla 10.** Estándar de calidad del aire de la Unión Europea [Directiva 2008/50/EU]

Contaminante	Concentración ( $\mu\text{g}\cdot\text{m}^{-3}$ )	Período de tiempo	Número de excedencias permitidas por año
$PM_{2.5}$	25	1 año	n/a
$SO_2$	350	1 hora	24
	125	24 horas	3
$NO_2$	200	1 hora	18
	40	1 año	n/a
$PM_{10}$	50	24 horas	35
	40	1 año	n/a
$CO$	10	Máximo diario de promedio de 8 horas	n/a
$O_3$	120	Máximo diario de promedio de 8 horas	Promedio de 25 días en 3 años

Para la gestión del territorio, la Directiva 2008/50/CE en su Artículo 23 dispone que los Estados miembros de la UE elaborarán planes de calidad del aire para las zonas y aglomeraciones donde se superen los valores límites o valores objetivos, así como los márgenes de tolerancia. Esta Directiva se complementa con la Ley Nacional 34/2007 de calidad del aire y protección de la atmósfera, que establece que los municipios deberán adoptar planes teniendo en cuenta los planes de las respectivas comunidades autónomas. Y, posteriormente, es resaltado por el Real Decreto 102/2011 que establece que las autoridades locales podrán elaborar sus propios planes considerando las propias comunidades autónomas. De esta forma, el territorio español se encuentra zonificado para un mejor monitoreo y control de la calidad del aire donde la responsabilidad del monitoreo recae en las comunidades autónomas.

En el caso de la Comunitat de València, el territorio se ha clasificado en 18 zonas donde se monitorea la calidad del aire, atendiendo a la orografía, la dinámica atmosférica, los usos del suelo, las fuentes de emisiones de contaminantes y la distribución de población (**Figura 15**). Estas zonas engloban las aglomeraciones Castelló, L'Horta, Alacant y Elx, y 14 áreas que se caracterizan por una concentración de población menor de 250 mil habitantes o con una baja densidad poblacional. En total se cuenta con 60 estaciones de monitoreo y los contaminantes que se evalúan de forma rutinaria son SO<sub>2</sub>, NO<sub>2</sub>, PM<sub>10</sub>, PM<sub>2.5</sub>, C<sub>6</sub>H<sub>6</sub>, CO, O<sub>3</sub>, Pb, metales y benzo(a)pireno, aunque no en todas las estaciones se miden los mismos contaminantes.



**Figura 15.** Zonas para la evaluación de la calidad del aire en la Comunidad de Valencia.

Las zonas presentan características heterogéneas, de las cuales, las más importantes por las problemáticas que engloban, se describe a continuación:

Las zonas **ES 1001** y **ES 1002** están caracterizadas por ser comarcas rurales de gran interés natural. En su cercanía se encuentra emplazada la Central Térmica de Andorra donde se detectaron daños forestales por emisiones de SO<sub>2</sub> atribuidas a esta central. En el año 1994, se firmó el convenio “Situación de la calidad del aire en las comarcas del norte de la Comunidad Valenciana” entre 27 Ayuntamientos de la zona, la Generalitat Valenciana, 3 grupos ecologistas y ENDESA para la creación de la Fundación Els Ports y una comisión de Vigilancia Medioambiental de la Zona els Ports-Maestrat. En este marco se definió implementar una nueva red de calidad del aire en la comarca, el seguimiento de las emisiones de la Central Térmica, y la evaluación y seguimiento de resultados de calidad del aire en la implantación de una planta de desulfuración.

La zona **ES 1003** es una zona de actividad cerámica, donde se genera el 43% de la producción de cerámica de la UE. Esta tarea cuenta con procesos de extracción de materia prima, preparación vía húmeda y prensado en seco de las baldosas cerámicas. Los contaminantes asociados con esta



actividad industrial son: PM, HF, HCl, SO<sub>2</sub>, NO<sub>x</sub>, CO<sub>2</sub>, metales pesados. En esta zona la población asociada es de 365760 habitantes. En 2008, se aprobó por Acuerdo del Consell el “*Plan de mejora de la calidad del aire de la zona ES 1003*”, con finalización de la segunda fase en 2012.

Las zonas **ES 1013** y **ES 1017** presentan la influencia de la actividad cementera por la presencia de 2 grandes plantas de fabricación de cemento a una corta distancia entre ellas. En el “*Plan de mejora de la calidad del aire de la zona ES 1013-ES 1017*”, aprobado el 5 de noviembre del 2010, se acordó el cierre anticipado de la fábrica CEMEX-I por su cercanía al territorio urbano (inferior a 1 km). En esta zona se detectaron valores críticos de SO<sub>2</sub> y PM<sub>10</sub>. Un estudio de caracterización de PM<sub>10</sub> realizado en el 2006, determinó que el 44% de las emisiones pertenecían a la industria cerámica, 23% a las emisiones de la industria cementera, 13% al tránsito, 9% a la actividad residencial y servicios. De las emisiones de SO<sub>2</sub>, el 56% se atribuyó a la industria cementera, el 33% emisiones industria cerámica, y el 7% restante al transporte y uso de maquinarias móviles.

La zona **ES 1016** es un área de tráfico intenso. Esta aglomeración agrupa a las comarcas de Valencia, L’Horta Nord, L’Horta Oest y L’Horta Sud, con un total de 25 municipios y 1 371 296 hab. Mientras que las zonas **ES 1007** y **ES 1008** se caracterizan por la influencia de las emisiones del tránsito vehicular y de las industrias [Conselleria de Agricultura, Desarrollo Rural, Emergencia Climática y Transición Ecológica, 2017a; Conselleria de Agricultura, Desarrollo Rural, Emergencia Climática y Transición Ecológica, 2017b].

Bajo el Decreto 161/2003 del Consell, se crea la Red Valenciana de Vigilancia y Control de la Contaminación Atmosférica, que tiene como función realizar un seguimiento de los niveles de los contaminantes atmosféricos más importantes en las principales áreas urbanas e industriales de la Comunidad Valenciana.

## I.II. Determinación de los objetivos

Al contar con el monitoreo extensivo de la Comunidad Valenciana, es posible realizar distintos tipos de análisis. Para este estudio nos planteamos los siguientes objetivos:

- *Conocer aspectos del monitoreo terrestre*

Descripción general de las estaciones de monitoreo. Evaluar la calidad de los datos. Caracterizar los sitios de monitoreo.

- *Determinar la calidad del aire en los sitios de medición*

Evaluar si las concentraciones de los contaminantes respetan las recomendaciones de la OMS y si se encuentran dentro de los parámetros legales.

- *Evaluar la evolución temporal de los contaminantes*

Calcular la tendencia de las concentraciones a largo plazo. Estudiar la tendencia en función de las direcciones del viento. Analizar el comportamiento diario, semanal e interanual de los contaminantes.

- *Estudiar la distribución espacial de los contaminantes*

Comparar las concentraciones medidas entre las estaciones. Analizar el comportamiento espacial del material particulado. Confeccionar mapas de material particulado mediante información

satelital. Confeccionar mapas de calidad del aire. Confeccionar mapas de riesgo por exposición a contaminantes atmosféricos y de la fracción atribuible de muertes.

### I.III. Delimitación de las preguntas de estudio

A continuación, se detallan los análisis que se pueden realizar a partir de la información existente a la fecha:

#### **Conocer aspectos del monitoreo terrestre**

- ¿Qué diseño sigue la red de monitoreo?
- ¿Cuáles son las técnicas utilizadas para el monitoreo?
- ¿Cuál es la calidad de la información recolectada por la red de monitoreo?
- ¿Hay redundancia de información entre las estaciones de monitoreo?
- ¿Cuáles son las características del área de emplazamiento de las estaciones de monitoreo?
- ¿Cómo influyen los vientos locales sobre los niveles registrados en las estaciones?
- ¿De qué sector del viento llegan las concentraciones mayores?

#### **Determinar la calidad del aire en los sitios de medición**

- ¿Cuáles son los niveles más frecuentes?
- ¿La concentración de los contaminantes excede los límites legales?
- ¿La concentración pone en riesgo a la población?

#### **Evaluar la evolución temporal de los contaminantes**

- ¿A cuál hora del día es mayor la concentración de los contaminantes?
- ¿Las concentraciones máximas varían a lo largo de la semana?
- ¿En cuál época del año ocurren los episodios de máximas concentraciones?
- ¿Cómo han cambiado los niveles de los contaminantes con los años?

#### **Estudiar la distribución espacial de los contaminantes**

- ¿Dónde ocurren las máximas concentraciones?
- ¿Se observa un patrón espacial de las máximas concentraciones?
- ¿Cómo es la calidad del aire donde no se mide?
- ¿La calidad del aire en la región representa un riesgo para la población?

### I.IV. Adquisición de los datos

#### *Monitoreo terrestre de la calidad del aire*

La *Red Valenciana de Vigilancia y Control de la Contaminación Atmosférica* cuenta con 60 estaciones permanentes y con tres unidades móviles y una Unidad de Intervención Rápida, las cuales, obtienen información sobre 11 emplazamientos repartidos en el área interior de la Comunidad Valenciana (**Tabla 11**). En la actualidad, se mantiene en funcionamiento un sistema automático y continuo de monitoreo de los siguientes contaminantes: SO<sub>2</sub>, PM<sub>10</sub>, PM<sub>2.5</sub>, PM<sub>1</sub> y metales pesados, NO, NO<sub>2</sub>, NO<sub>x</sub>, CO, O<sub>3</sub>, BTX y BaP, SH<sub>2</sub>, junto a algunos parámetros meteorológicos. Para la determinación de estos contaminantes se emplean equipos basados en los métodos oficiales recogidos en el anexo VII del R.D. 102/2011, relativo a la mejora de la calidad del aire, modificado por el Real Decreto 39/2017.

Para este trabajo, los datos se descargaron del sitio oficial (<http://www.agroambient.gva.es/va/web/calidad-ambiental>) para el período comprendido entre el 1 de enero de 2008 al 30 de septiembre de 2018.

**Tabla 11.** Estaciones de monitoreo presentes en la Comunidad Valenciana

<b>Provincia</b>	<b>Nombre</b>	<b>Estación</b>	<b>Longitud</b>	<b>Latitud</b>	<b>Altitud</b>
ALACANT	Alacant	X03014006	-0.4711	38.3586	45
ALACANT	Alacant	X03014008	-0.5067	38.3403	55
ALACANT	Alacant	X03014009	-0.5139	38.3511	20
ALACANT	Alacant	X03014010	-0.5025	38.3347	5
ALACANT	Alacant	X03014012	-0.4919	38.3369	15
ALACANT	Alacant	X03014013	-0.4967	38.3367	0
ALACANT	Alacant	X03014014	-0.505	38.3311	3
ALACANT	Alcoi	X03009006	-0.4669	38.7064	534
ALACANT	Benidorm	X03031002	-0.1464	38.5731	44
ALACANT	Elda	X03066003	-0.8064	38.445	340
ALACANT	Elx	X03065006	-0.6828	38.2422	44
ALACANT	Elx	X03065007	-0.7175	38.2592	99
ALACANT	Orihuela	X03099002	-0.8536	38.0867	32
ALACANT	Pinoso	X03105001	-1.0647	38.4517	642
ALACANT	Torreveija	X03133002	-0.69	37.9911	10
CASTELLON	Almassora	X12009007	-0.0564	39.9453	27
CASTELLON	Benicassim	X12028001	0.0728	40.0622	50
CASTELLON	Burriana	X12032001	-0.065	39.8922	37
CASTELLON	Castellò de la Plana	X12040008	-0.0572	40.0114	106
CASTELLON	Castellò de la Plana	X12040009	-0.0361	39.9592	44
CASTELLON	Castellò de la Plana	X12040010	0.0103	39.9819	10
CASTELLON	Castellò de la Plana	X12040015	-0.0264	39.9886	18
CASTELLON	Cirat	X12046001	-0.4683	40.0542	466
CASTELLON	La Pobla de Benifass	X12093004	0.0847	40.6917	1200
CASTELLON	La Vall d'Uix	X12126003	-0.2444	39.8236	250
CASTELLON	L'Alcora	X12005005	-0.2069	40.0681	160

CASTELLON	Morella	X12080007	-0.0925	40.6372	1153
CASTELLON	Onda	X12084003	-0.25	39.9628	163
CASTELLON	San Jorge	X12099001	0.3728	40.5547	181
CASTELLON	Torre Endomènech	X12120001	0.0789	40.2694	259
CASTELLON	Villafranca del Cid	X12129001	-0.2492	40.4169	1125
CASTELLON	Viver	X12140002	-0.6033	39.9306	616
CASTELLON	Zorita del Maestrazgo	X12141002	-0.1694	40.735	619
VALENCIA	Albalat dels Tarongers	X46010001	-0.3367	39.7053	320
VALENCIA	Algar de Palancia	X46028001	-0.3592	39.7822	200
VALENCIA	Alzira	X46017002	-0.4578	39.15	60
VALENCIA	Benigànim	X46062001	-0.4419	38.9375	195
VALENCIA	Buñol	X46077006	-0.7839	39.4272	400
VALENCIA	Burjassot	X46078004	-0.4178	39.5094	25
VALENCIA	Caudete de las Fuentes	X46095001	-1.2828	39.56	794
VALENCIA	Cortes de Pallás	X46099001	-0.9344	39.2608	493
VALENCIA	Gandia	X46131002	-0.1903	38.9689	22
VALENCIA	La Pobla de Vallbona	X46202002	-0.5369	39.5794	0
VALENCIA	Ontinyent	X46184002	-0.6828	38.7953	510
VALENCIA	Paterna	X46190005	-0.4608	39.5514	122
VALENCIA	Quart de Poblet	X46102002	-0.4472	39.4811	41
VALENCIA	Sagunt	X46220003	-0.2314	39.665	10
VALENCIA	Sagunt	X46220009	-0.2789	39.6847	54
VALENCIA	Sagunt	X46220010	-0.2664	39.6328	3
VALENCIA	Torrebaja	X46242001	-1.2539	40.0994	728
VALENCIA	Torrent	X46244003	-0.4825	39.4253	100
VALENCIA	Valencia	X46250030	-0.3767	39.4581	11
VALENCIA	Valencia	X46250043	-0.3694	39.4794	11
VALENCIA	Valencia	X46250046	-0.3375	39.4797	7
VALENCIA	Valencia	X46250047	-0.3428	39.4575	7
VALENCIA	Valencia	X46250048	-0.4083	39.4811	15
VALENCIA	Valencia	X46250050	-0.3964	39.4503	15
VALENCIA	Valencia	X46250051	-0.3186	39.3156	0

VALENCIA	Vilamarxant	X46256001	-0.6256	39.5647	510
VALENCIA	Villar del Arzobispo	X46258001	-0.8319	39.7081	430

### *Producto satelital de AOD - MODIS*

La profundidad óptica de aerosoles (AOD) es la integración en la columna atmosférica de la dispersión de la luz. MODIS provee información histórica de los aerosoles con una resolución moderada. El producto de datos MCD19A2 Versión 6 es un producto combinado de las mediciones realizadas por el sensor MODIS a bordo de los satélites Terra y Aqua, mediante la implementación del algoritmo *Multi-angle Implementation of Atmospheric Correction* (MAIAC) para las estimaciones de AOD. El producto es una cuadrícula diaria con una resolución de 1 km de píxel. Se puede descargar de <https://ladsweb.modaps.eosdis.nasa.gov/> en formato HDF4, y contiene distintas capas de datos científicos (SDS).

### *AERONET*

Para la validación de los datos AOD de las imágenes MODIS se cuenta con las mediciones de AERONET. La estación AERONET más cercana se ubica en la ciudad de Burjassot (39°30'28" N, 0° 25' 04"W, altura 30m). El sitio se encuentra a 5 km al noroeste del centro de la ciudad de Valencia y a 10 km al oeste de la costa mediterránea. Los datos se descargaron del sitio oficial de AERONET, <https://aeronet.gsfc.nasa.gov>.

### *Reanálisis atmosférico de MERRA - 2*

Se trabajó con las salidas del reanálisis atmosférico MERRA-2 que brinda datos de las condiciones de la atmósfera y contaminantes modelados mediante la asimilación de distintos productos satelitales, incluyendo el AVHRR (*Advanced Very High Resolution Radiometer*), MISR (*Multi-angle Imaging SpectroRadiometer*), MODIS y AERONET [Gelaro et al., 2017]. Los datos se descargaron del sitio web de MERRA-2 (<https://disc.gsfc.nasa.gov/daac-bin/FTPSubset2.pl>) para el área de la Comunidad Valenciana. Como estos datos se utilizaron para los modelos multivariantes, se descargaron los sets de datos de las horas más cercanas a la hora de sobrevuelo de los satélites Aqua y Terra.

MERRA-2 simula cinco especies de PM<sub>2.5</sub> (carbono negro, carbono orgánico, sal marina, polvo, sulfato) basándose en el modelo de aerosol GOCART [Ginoux, et al., 2001], las cuales se encuentran a disposición en el producto `tavg1_2d_aer_Nx`. De estas, las variables seleccionadas fueron las concentraciones máxicas de sulfuro de dimetilo, polvo, carbón orgánico, sal marina, sulfatos y dióxido de azufre, para el intervalo de tiempo de 13.30 - 16.30 UTC.

También se trabajó con los modelos de MERRA-2 de la altura de la capa límite, flujo total de precipitación superficial, velocidad de viento y velocidad máxima de viento sobre la superficie (`tavg1_2d_flg_Nx` para el intervalo 10.30 - 13.30 UTC). También, se incorporaron las variables radiométricas como proporción del albedo y fracción de área de nube para nubes altas y bajas del producto `tavg1_2d_rad_Nx` para el intervalo 10.30 - 13.30 UTC); y datos de superficie como

presión de superficie, humedad relativa, componentes del viento (u, v), temperatura del aire (inst3\_3d\_asm\_Np para el intervalo horario 12.00 - 15.00 UTC).

### *Topografía del terreno*

Los datos de elevación del terreno del área de la Comunidad Valenciana fueron descargados del sitio web del *Consortio de Información Espacial* del CGIAR [Jarvis et al., 2008] con una resolución de 3 segundos de arco (90 m en el ecuador) y cobertura global (<http://srtm.csi.cgiar.org/>). Este producto corresponde al modelo digital de elevación creado a partir de la misión SRTM (del inglés *Shuttle Radar Topographic Mission*) del año 2000. Se trabajó con los archivos: srtm\_37\_04.tif, srtm\_36\_05.tif, srtm\_37\_05.tif y srtm\_36\_04.tif, con los cuales se confeccionó un mosaico único para el área de estudio. Este producto fue elegido sobre el EU-DEM elaborado por el CLMS puesto que tiene cobertura global y puede servir de ejemplo para futuros trabajos en otras áreas de estudio.

### *Clasificación uso de suelo*

Para la información de uso de suelo se utilizó el sistema de clasificación “Corine Land Cover 2012” a escala 1:100.000, generado para el año 2012. Este mapa pertenece al proyecto europeo *Corine Land Cover*, dirigido por la Agencia Europea de Medio Ambiente. La información se codifica en 3 niveles de detalle para su presentación oficial en Europa. Esta capa se genera a partir de la actualización de SIOSE 2011, por generalización de esta base de datos cartográfica. El proveedor fue la *Conselleria d'Habitatge, Obres Públiques i Vertebració del Territori*.

## Módulo II - Etapa de procesamiento



### II.I. Preparación de los datos

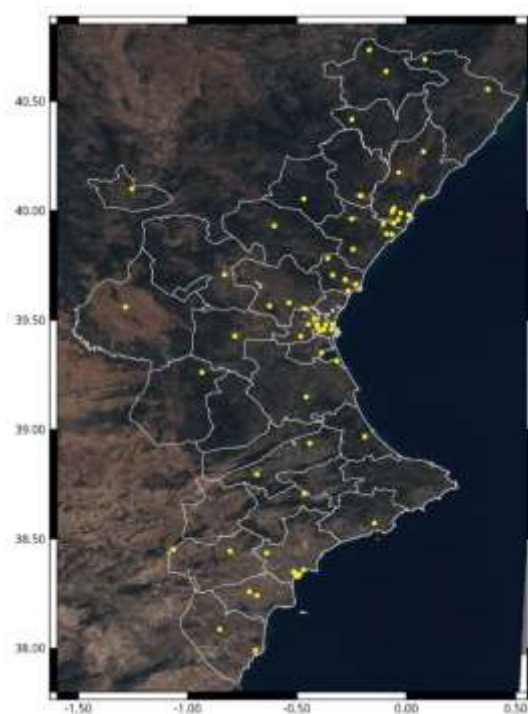
En este trabajo, previo al procesamiento de datos, se inspeccionaron los registros de las concentraciones de contaminantes y las mediciones de las variables ambientales en busca de errores en los registros. A su vez, en esta etapa se procedió a elaborar las series de tiempo agregadas, a partir de los sets de datos horarios, para obtener series diarias y mensuales. Como las concentraciones de contaminantes tienen una variabilidad considerable a lo largo del día, sólo se consideraron los días con al menos 12 mediciones horarias [Xie et al., 2018] dentro del período comprendido entre el 1 de enero de 2008 al 30 de septiembre de 2018.

De la base de datos, se corrigieron los valores por debajo del límite de detección de los equipamientos por un tercio de la concentración mínima medida en esa estación dentro del período de estudio [EPA, 2006]. A su vez, de la serie diaria se realizaron las series temporales mensuales. Los meses en los que faltaban más del 25% de los datos se descartaron de las series temporales mensuales [Directiva 2008/50/CE].

### II.II. Selección de algoritmos de análisis

#### *I. Conocer aspectos del monitoreo terrestre*

En la **Figura 16** se muestra la posición de los 60 sitios donde se monitorea la calidad del aire en la Comunidad Valenciana. Con el objetivo de obtener más información descriptiva respecto al diseño de la red de monitoreo, se analizó el patrón de distribución de los sitios mediante un análisis de patrones de puntos [Dadson, 2017].



**Figura 16.** Mapa con los 60 sitios de monitoreo de la calidad del aire en la Comunidad Valenciana

Para explorar la consistencia de la información se analizó la falta de datos mediante gráficas exploratorias.

La redundancia entre las estaciones es una característica no deseada en una red de monitoreo puesto que implica el sostener el registro de las concentraciones en sitios que no aportan información nueva del estado del aire con respecto a la información que recolectan el resto de las estaciones. Como primer paso se realizó un análisis de correlación entre las estaciones para determinar si existía redundancia. Se analizó la correlación entre las estaciones de monitoreo para cada uno de los 6 contaminantes en estudio. Elegimos el coeficiente de correlación de Spearman que es un coeficiente no paramétrico, tolerante a la pérdida de normalidad y robusto con respecto a los datos extremos.

Luego, se realizó un PCA para cada contaminante y en cada una de las distintas estaciones. Se trabajó sobre la matriz de correlación, que presenta grandes ventajas cuando se cuenta con datos faltantes, existe una falta de centralidad en el conjunto de datos y la varianza de los datos entre las estaciones difiere. El PCA se utiliza comúnmente para reducir el número de variables y evitar la multicolinealidad [Pires et al., 2008a, Pires et al., 2008b].

Finalmente, se aplicó el algoritmo de aglomeración jerárquica que permite agrupar las estaciones más similares. La aplicación del algoritmo de clasificación jerárquica se realizó sobre matrices de correlación de Pearson.

Para la identificación de potenciales fuentes de emisión en las estaciones de monitoreo dentro del área de la Generalitat Valenciana se utilizó el mapa de ocupación de suelo *Corine Land Cover 2012*. Se generaron áreas circulares de 1 km para cada estación de monitoreo y se procedió a una



inspección ocular de la tabla de atributos para determinar los principales usos. La corrección de la misma requirió unir atributos por geolocalización, separar geometrías y reproyectar a la proyección de trabajo.

Para evaluar las características meteorológicas y su influencia en las mediciones de los contaminantes, se confeccionaron rosas de vientos y rosas de contaminación en cada estación y para cada contaminante. Los intervalos utilizados como parámetros de corte en las visualizaciones cloropléticas se construyeron utilizando el método de los umbrales naturales o Jenks. El método de Jenks realiza comparaciones iterativas de las sumas de las diferencias al cuadrado entre valores observados dentro de cada clase y la media de las clases. Como los datos no mantienen una distribución uniforme, este algoritmo representa mejor el conjunto de datos que la utilización de otros métodos, como los cuantiles o el uso de intervalos constantes. La mejor clasificación se considera cuando se encuentran aquellos umbrales que minimizan la suma intra-clase de diferencias al cuadrado.

El algoritmo de Jenks se aplicó sobre una muestra al azar de 5000 datos de las 92760 observaciones para cada contaminante y para las velocidades de viento, estableciendo 5 categorías de corte para la obtención de los intervalos las cuales se mantuvieron en las rosas de vientos y rosas de contaminantes. Se utilizó el paquete *openair* para graficar las rosas de vientos y de contaminantes [Carslaw & Ropkins, 2012, Carslaw, 2013], y el paquete *classInt* para el cálculo de los intervalos mediante el algoritmo de rupturas naturales de Jenks [Bivand, 2019].

## II. Determinar la calidad del aire en los sitios de monitoreo

Para la evaluación de los niveles históricos se realizó un análisis exploratorio, donde se obtuvieron los principales estadísticos descriptivos para cada contaminante por estación de monitoreo. Con el fin de evaluar el cumplimiento de las normativas en calidad de aire y el riesgo a la salud, se compararon las mediciones en cada estación con las Directivas de la UE [Directiva 2008/50/CE] y con las recomendaciones de la OMS [World Health Organization, 2005]. Se generaron gráficas de tipo violín para mostrar la distribución de los datos junto a los límites legales y los valores recomendados. En los casos donde se produjeron superaciones de los valores legales y recomendados se diseñaron gráficos para visualizar la frecuencia de superación anual.

La descripción horaria de los niveles de los contaminantes puede ser utilizada para comprender el comportamiento de las emisiones, información útil en el funcionamiento de los modelos locales y regionales de calidad del aire. Para conocer el comportamiento de los contaminantes hemos realizado (i) perfiles mensuales, (ii) semanales y (iii) por hora por estación de monitoreo y contaminante. Se exploraron los datos utilizando gráficos de las medias horarias y mensuales de las concentraciones de los contaminantes por estación de monitoreo.

Los datos horarios se trabajaron en R [R Development Core Team, 2011] utilizando las librerías *openair* [Carslaw & Ropkins, 2012, Carslaw, 2013], *dplyr* [Wickham, 2009], *reshape2* [Wickham, 2007] y *ggplot2* [Wickham, 2016].

### III. Conocer la evolución temporal de los contaminantes

Un estadístico relevante para la implementación de políticas públicas es la tendencia a largo plazo de las medias mensuales. La tendencia temporal da cuenta del crecimiento o decrecimiento de un parámetro a lo largo del tiempo, considerando un comportamiento lineal.

Las tendencias de los promedios mensuales de las concentraciones de los contaminantes se analizaron en cada estación de monitoreo utilizando la función Theil-Sen (Theil, 1950; Sen, 1968). Para la estimación se realizó una desestacionalización. La pendiente obtenida representa la tasa de aumento o disminución anual de la concentración de cada contaminante en la estación de monitoreo analizada. Las tendencias calculadas se consideraron estadísticamente significativas con un nivel de significación de 0.05 o inferior.

El procesamiento se realizó en R, utilizando la librería *openair* (Carslaw y Ropkins, 2012, Carslaw, 2013). Se analizó la tendencia temporal para cada contaminante en cada estación, y tras esto se analizó la tendencia de cada contaminante en función de la dirección del viento en las estaciones que contaban con información de la meteorología local.

### IV. Evaluar la distribución espacial de los contaminantes

Las imágenes satelitales aportan información útil de la distribución espacial de los contaminantes. Sin embargo, no existe una relación lineal entre la variable detectada por el satélite (AOD) y las concentraciones medidas en superficie. Por ello, fue preciso integrar esta información dentro de un modelo multivariable que permitió simular las concentraciones de  $PM_{2.5}$  y  $PM_{10}$  a nivel de suelo con una resolución de 1km.

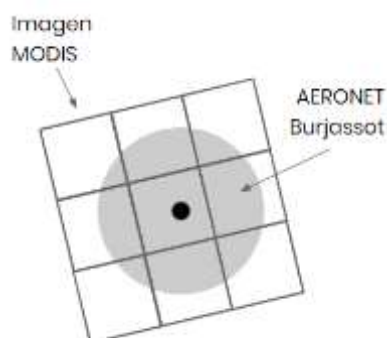
Para esto, se descargaron un total de 15704 imágenes del producto satelital MCD19A2 generadas entre el 1 de enero de 2008 y el 30 de septiembre de 2018. La descarga se realizó mediante la aplicación del comando *wget*. Los productos descargados se convirtieron del formato HDF5 a formato GEOTiff mediante el desarrollo de un código en Python 2.7 utilizando la librería *gdal*.

Como se deseaba trabajar con la máxima calidad de los valores de AOD se aplicó como máscara la capa SDS de "AOD AQ" que viene dada en el producto satelital. Esta capa se encuentra codificada como un número binario de 16 bits. Su utilización requirió la recodificación de la capa como una máscara, la cual se aplicó sobre la capa con valores de AOD. Esto se realizó en R empleando la librería *raster*.

Para la validación de los productos satelitales AOD se utilizaron los datos de AERONET de la estación Burjassot. Dado que AERONET mide a distintas longitudes de onda, pero no mide a la longitud de onda de 550 nm que es donde registra el sensor MODIS, fue necesario interpolar los datos de AERONET. Esto se realizó mediante una interpolación cuadrática de las mediciones de AERONET registradas a las longitudes de onda 440, 500 y 675 nm, para modelar el valor de AOD a 550 nm.

El producto MCD19A2 es un producto diario que resume la información de distintas órbitas de los satélites Terra y Aqua, por tanto, fue preciso extraer la información de las horas de sobrepaso de los satélites sobre el sitio de AERONET para relacionarlas con las mediciones realizadas por AERONET. La validación se realizó utilizando un promedio de las mediciones de AERONET dentro de un rango temporal de 15 y 30 min de la hora de paso de los satélites.

A su vez, para obtener los valores de MODIS en la estación Burjassot se generó un buffer de 3 y 7.5 km de diámetro entorno a la estación. Se tomaron estas dos medidas de buffer para evaluar el mejor tamaño para la validación de AOD (**Figura 17**).



**Figura 17.** Esquema representativo de la imagen satelital, la estación de monitoreo y el área de influencia para la validación de AOD MODIS-MAIAC.

Como los productos de AOD presentaron una gran cantidad de datos faltantes, se decidió realizar un llenado mediante interpolación por el método de distancia inversa ponderada (IDW) con una distancia de alisamiento de 5 km [Goldberg et al., 2019]. Para evaluar el error de la interpolación se realizó una validación cruzada de 5 veces sobre el total de los datos.

Se analizó la correlación de AOD con  $PM_{10}$  y  $PM_{2.5}$  para seleccionar la variable a modelar. Las capas de información consideradas para su inclusión en el modelo multivariante pueden verse en la **Tabla 12**. Para la generación del modelo LUR se contó con la información meteorológica extraída de MERRA-2, la clasificación de uso de suelo provista por CORINE y la elevación del terreno para el área de estudio.

**Tabla 12.** Lista completa de variables accesorias analizadas para confeccionar el modelo LUR

Código	Variable	Unidades	Resolución espacial	Resolución temporal
AOD	MAIAC AOD	-	1km x 1km	Diaria
BCSMASS	Concentración másica superficial de carbono negro	kg.m <sup>-3</sup>	0.5° x 0.625°	3 horas
DMSSMASS	Concentración másica superficial de sulfuro de dimetilo	kg.m <sup>-3</sup>	0.5° x 0.625°	3 horas
DUSMASS	Concentración másica superficial de polvo	kg.m <sup>-3</sup>	0.5° x 0.625°	3 horas
DUSMASS25	Concentración másica superficial de polvo de tamaño $PM_{2.5}$	kg.m <sup>-3</sup>	0.5° x 0.625°	3 horas

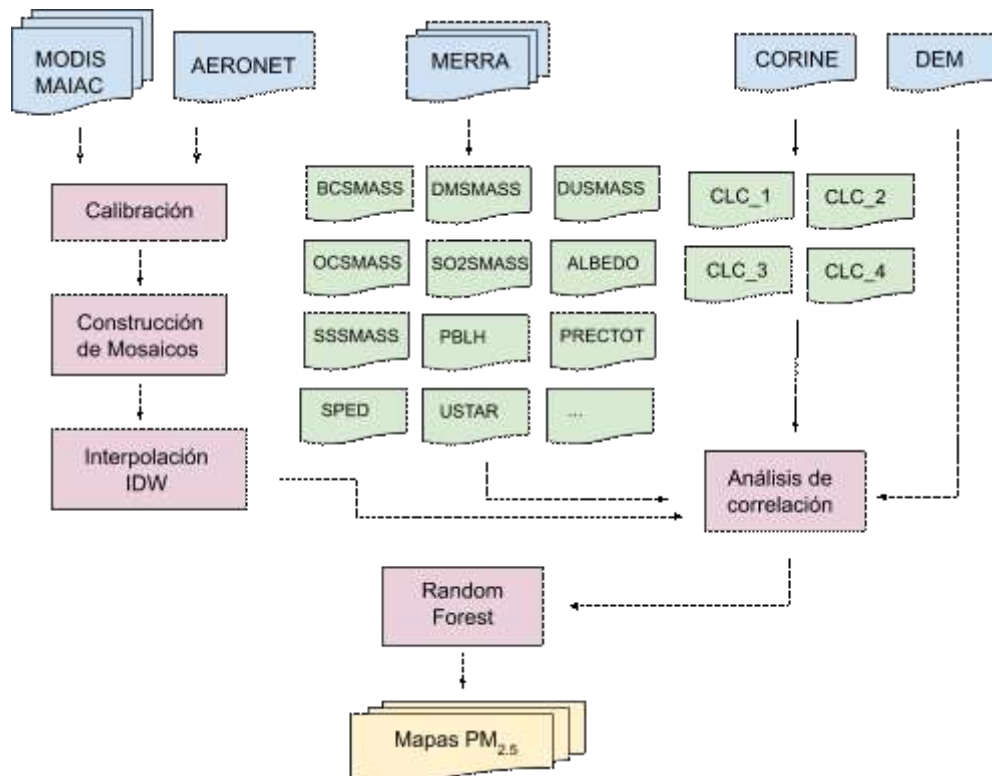
OCSMASS	Concentración másica superficial de carbono orgánico	kg.m <sup>-3</sup>	0.5° x 0.625°	3 horas
SO2SMASS	Concentración másica superficial de SO <sub>2</sub>	kg.m <sup>-3</sup>	0.5° x 0.625°	3 horas
SO4SMASS	Concentración másica superficial de SO <sub>4</sub>	kg.m <sup>-3</sup>	0.5° x 0.625°	3 horas
SSSMASS	Concentración másica superficial de sal marina	kg.m <sup>-3</sup>	0.5° x 0.625°	3 horas
SSSMASS25	Concentración másica superficial de sal marina tamaño PM <sub>2.5</sub>	kg.m <sup>-3</sup>	0.5° x 0.625°	3 horas
PBLH	Altura de la capa límite planetaria	m	0.125° x 0.125°	3 horas
PRECTOT	Precipitación superficial total	kg.m <sup>-2</sup> .s <sup>-1</sup>	0.125° x 0.125°	3 horas
SPEED	Velocidad superficial del viento	m.s <sup>-1</sup>	0.125° x 0.125°	3 horas
SPEEDMAX	Velocidad superficial máxima del viento	m.s <sup>-1</sup>	0.125° x 0.125°	3 horas
USTAR	Escala de velocidad de superficie	m.s <sup>-1</sup>	0.125° x 0.125°	3 horas
H1000	Altura a 1000 hPa	m	0.125° x 0.125°	1 hora
ALBEDO	Surface albedo	%	0.125° x 0.125°	6 horas
CLDHGH	Fracción de área cubierta por nubes bajas	%	0.125° x 0.125°	3 horas
CLDLOW	Fracción de área cubierta por nubes altas	%	0.125° x 0.125°	3 horas
PS	Presión en superficie	hPa	0.125° x 0.125°	3 horas
T	Temperatura del aire a 1000 hPa	K	0.125° x 0.125°	3 horas
RH	Humedad relativa a 1000 hPa	%	0.125° x 0.125°	3 horas
U	Componente de viento hacia el este	m.s <sup>-1</sup>	0.125° x 0.125°	3 horas
V	Componente de viento hacia el norte	m.s <sup>-1</sup>	0.125° x 0.125°	3 horas
DEM	Elevación del terreno	m	1 km x 1 km	-
	Usos de suelo		Polígono	-
CL1	Proporción de superficie artificial	%	300 m x 300 m	-
CL2	Proporción de superficie agrícola	%	300 m x 300 m	-
CL3	Proporción de superficie humedales	%	300 m x 300 m	-
CL4	Proporción de superficie agua	%	300 m x 300 m	-

En todos los casos donde fue necesario, la información de entrada al modelo se recortó a nivel Comunidad Valenciana, se proyectó a WGS84, y se adaptó a la grilla de MAIAC de 239 filas por 158 columnas con una resolución de 0.01°, mediante un remuestreo mediante el método de interpolación bilineal.

Previo a ser incluidas en el modelo, todas las variables fueron centradas y normalizadas mediante el parámetro Z. Luego, se hizo un análisis de correlación para evitar la generación de resultados espurios por información redundante.

Los modelos fueron implementados en el lenguaje R [R Core Team, 2017]. Se utilizó el paquete *RandomForest* [Liaw y Wiener, 2002] para el entrenamiento y estimación de RF, y el paquete *caret* [Kuhn et al., 2019] para la elaboración de los sets de entrenamiento y de prueba.

Como medida de calidad del ajuste del modelo se utilizó el coeficiente de determinación ( $R^2$ ), el RMSE y el MAE. Toda la metodología empleada en la confección de los mapas de concentración se resume en la **Figura 18**.



**Figura 18.** Resumen de la metodología desarrollada para confeccionar los mapas de  $PM_{2.5}$ .

#### V. Confección de mapas de calidad de aire, mapas de riesgo y mapas de FAM

Obtenidos los mapas de concentración se elaboraron una serie de cartografías temáticas para el área de la Comunidad Valenciana. Los mapas se realizaron en función de la metodología descrita en la propuesta metodológica para todas las fechas a lo largo del período de estudio. En la versión impresa se muestran solo los mapas anuales como ejemplo de los productos elaborados.

## Módulo III - Etapa de comunicación de los resultados

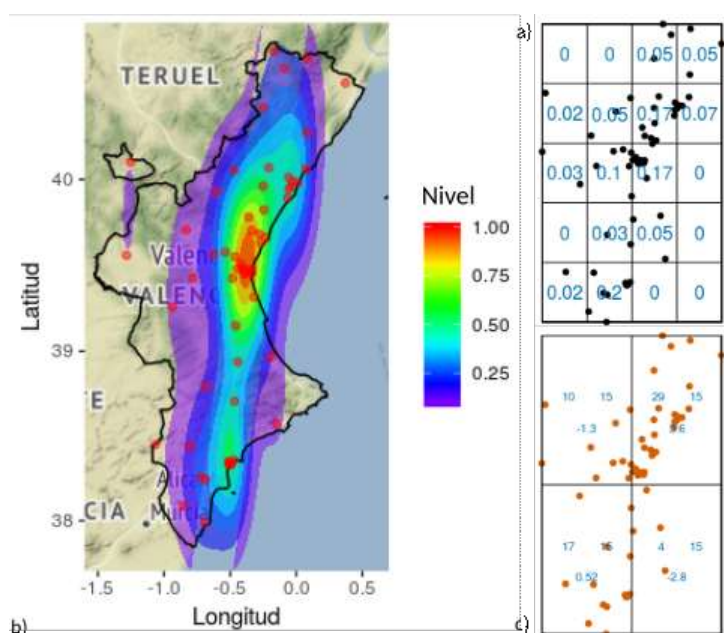


### III.I. Resultados

#### I. Análisis del patrón de las estaciones

Las estaciones de monitoreo de la calidad del aire en Valencia fueron incorporadas en el territorio en función de distintas preocupaciones. En este punto nos interesa evaluar la distribución de las estaciones, si la misma responde a un patrón uniforme a lo largo del territorio o si se presentan zonas más monitoreadas que otras.

Para ello, en la **Figura 19** se muestra la frecuencia relativa por cuadrante de las estaciones, la función de intensidad y la prueba de Chi-cuadrado.



**Figura 19.** Análisis de patrones de las estaciones en la Comunitat Valenciana. a) Función de intensidad, b) Frecuencia relativa por cuadrante; c) Prueba de Chi-cuadrado

Para evaluar si estamos frente a un proceso homogéneo de Poisson (CSR) se realizó una prueba de chi-cuadrado basada en conteos por cuadrantes (**Figura 19, c**). Como hipótesis nula los sitios se encuentran idénticamente distribuidos en el espacio y como hipótesis alternativa, existe una distribución heterogénea de las estaciones.

En base a los resultados, podemos descartar la hipótesis nula a favor de la hipótesis alternativa ( $X^2 = 23.07$ ,  $df = 3$ ,  $p\text{-value} < 0.005$ ). Tal como se puede observar en la **Figura 19 c**, las estaciones se encuentran distribuidas de forma no homogénea presentándose cuadrantes con un mayor número de estaciones.

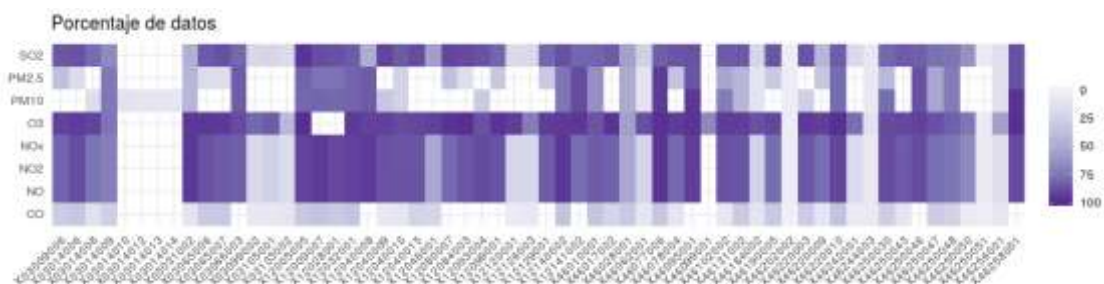
La mayor densidad de estaciones se presenta en la zona este-noreste de la Comunidad Valenciana, sobre el área donde se concentra la población (Valencia y Castellón). Una segunda concentración de estaciones se da al sur, en las cercanías de la ciudad de Alicante.

## II. Evaluación de calidad de los datos de las estaciones de monitoreo

En este trabajo se utilizaron las mediciones de la Red de Calidad de Aire de la Comunidad Valenciana para el período comprendido entre 2008 y 2018. En ese período hay un registro de concentraciones horarias en 60 puntos de la Comunidad Valenciana. En algunos de estos sitios se desarrolló un monitoreo temporal, de meses a unos pocos años, y en otros, hubo un monitoreo continuo durante los 10 años que involucra este estudio.

Las estaciones de monitoreo empleadas utilizan diferentes equipamientos, y se observan variaciones respecto a los contaminantes medidos en cada estación a lo largo del período. A su vez, se presentaron extensos períodos de registros faltantes y se tiene conocimiento de modificaciones de los sensores utilizados en el transcurso del monitoreo. Esto generó una base de datos heterogénea con un alto porcentaje de datos faltantes.

En la **Figura 20** se puede observar el porcentaje de las mediciones efectuadas para cada contaminante en cada estación de monitoreo terrestre dentro del período de tiempo en estudio. Tal como se puede ver, son pocas las estaciones que muestran una proporción superior al 70% de las mediciones para los distintos contaminantes.



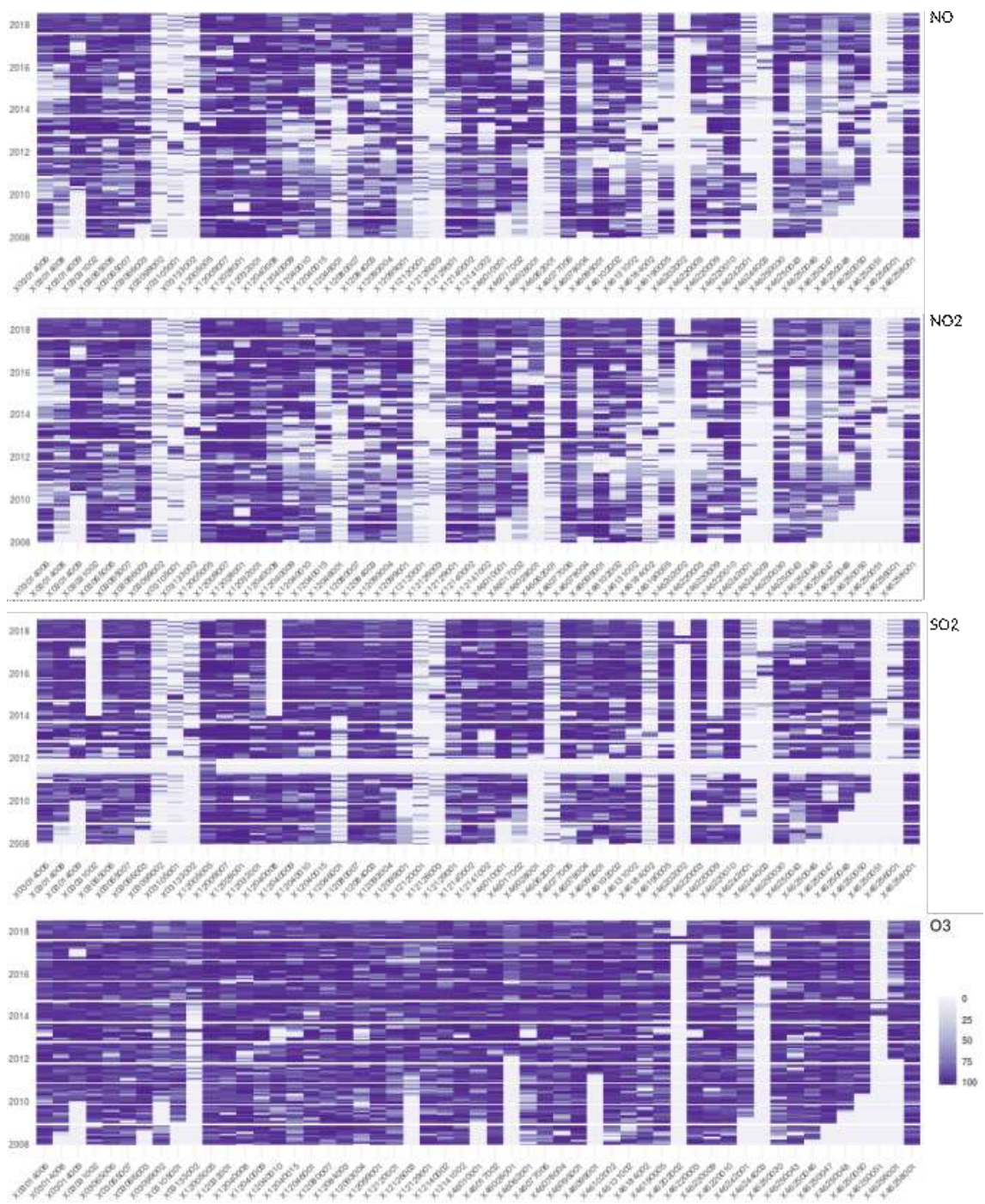
**Figura 20.** Porcentaje de datos por contaminantes para cada estación de monitoreo

La serie cronológica completa presenta 92 760 registros horarios, es decir, 127 meses de datos. Sin embargo, como se mencionó anteriormente, existe un alto porcentaje de datos faltantes. En la **Figura 21** se puede observar el % de datos existentes por mes para las mediciones de los contaminantes gaseosos en cada una de las estaciones. El CO no se muestra puesto que solo se registraron mediciones en el período 2012-2017, con extensos períodos sin mediciones, dejando solo los años 2013, 2015 y 2016 con mediciones confiables de ser analizadas. Por esto, se descartó el CO como contaminante a estudiar.

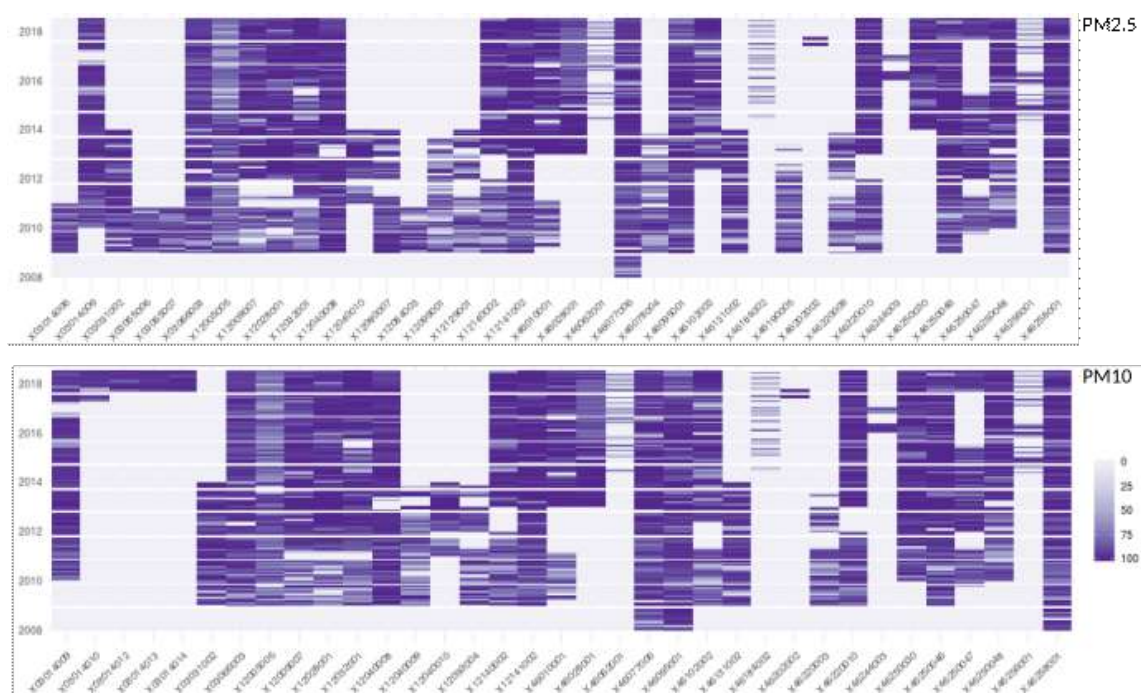
En algunos sitios solo se registran mediciones por escasos meses o se presentan altas tasas de datos faltantes a lo largo de todo el período. En otros casos, las estaciones fueron relocalizadas, faltando varios años de información. Un caso llamativo en el registro son las mediciones de SO<sub>2</sub> que muestra en la mayoría de las estaciones una discontinuidad para el año 2011. La falta de información detallada por parte de la agencia ambiental respecto al mantenimiento de las estaciones dificulta la interpretación de estas anomalías en la base de datos.

La **Figura 22** se presentan las estaciones que monitorean PM<sub>10</sub> y PM<sub>2.5</sub>. A simple vista puede verse que el número de estaciones que registran estos contaminantes es menor a los que registran gases y que en la mayoría de las estaciones se han producido discontinuidades severas en las mediciones.





**Figura 21.** Porcentaje de las mediciones mensuales de las concentraciones de los gases NO, NO<sub>2</sub>, SO<sub>2</sub> y O<sub>3</sub> para cada estación



**Figura 22.** Porcentaje de las mediciones mensuales de las concentraciones del material particulado  $PM_{2.5}$  y  $PM_{10}$  para cada estación

En vistas de la falta de continuidad en algunos sitios de monitoreo y de la alta proporción de datos faltantes en el set de datos se decidió realizar una selección de aquellas estaciones que presentaban la mayor cantidad de datos a lo largo del período de estudio seleccionado. Para elegir las estaciones se siguió el criterio establecido por la Directiva de la EU [Directiva 2008/50/CE]. Si en una estación han faltado más del 25% de las mediciones horarias de uno de los contaminantes correspondientes a un año determinado, ese año concreto se excluyó del análisis de los datos horarios.

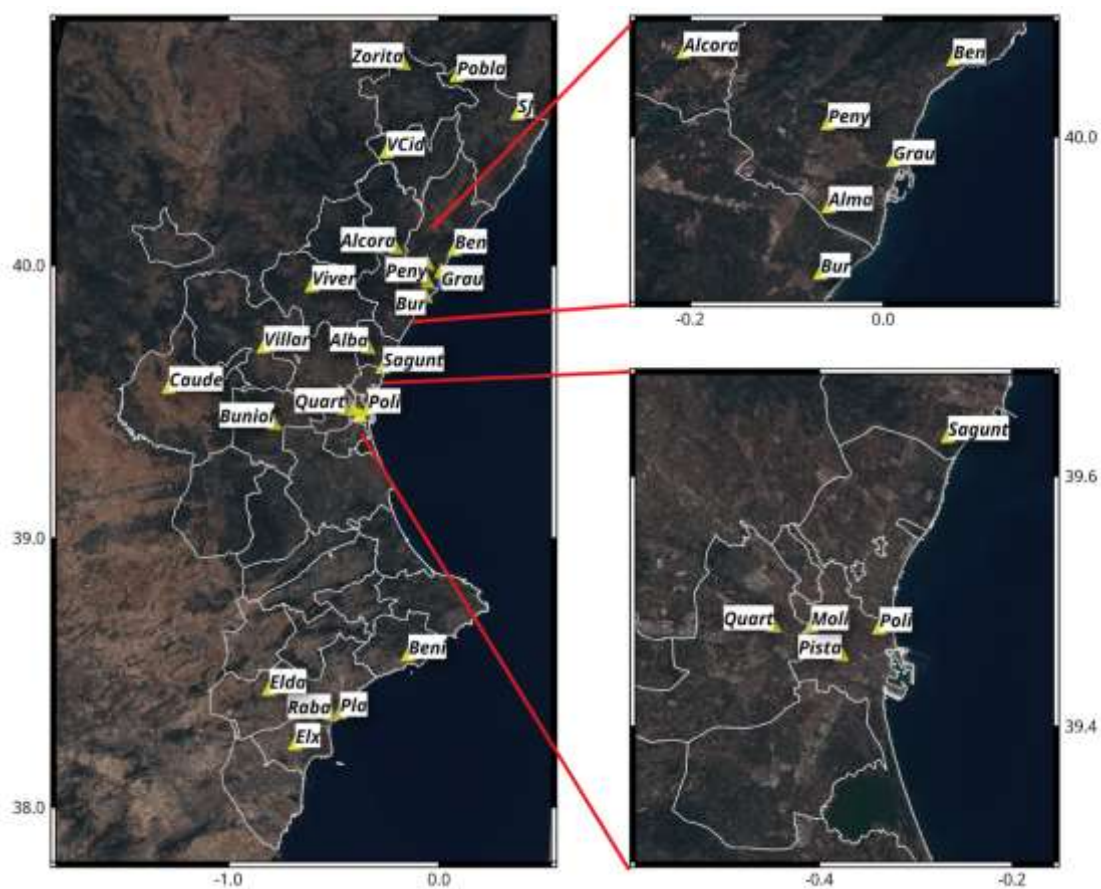
Luego, como este trabajo pretende dar cuenta de un análisis histórico de las mediciones, si se contaba con menos de 6 años de mediciones ( $2/3$  del período), la estación fue descartada del análisis. Finalmente, se decidió incluir solo aquellas estaciones que presentaban registros de al menos 3 de los contaminantes seleccionados para el estudio ( $NO$ ,  $NO_2$ ,  $O_3$ ,  $SO_2$ ,  $PM_{10}$ ,  $PM_{2.5}$ ) para poder realizar análisis cruzados entre los contaminantes.

A partir de los criterios antes mencionados, se seleccionaron 25 estaciones para trabajar en el período comprendido entre las 00:00 hs del 01 de enero del 2008 hasta las 23:00 hs 31 de julio de 2018 (**Tabla 13**).

**Tabla 13** - Estaciones incluidas en el análisis junto con su código, el nombre corto empleado y los contaminantes que se miden

<b>Nombre completo</b>	<b>Código</b>	<b>Nombre corto</b>	<b>NO</b>	<b>NO<sub>2</sub></b>	<b>O<sub>3</sub></b>	<b>SO<sub>2</sub></b>	<b>PM<sub>2.5</sub></b>	<b>PM<sub>10</sub></b>
<i>Alacant - El Pla</i>	X03014006	Pla	X	X	X	X		
<i>Alacant - Rabassa</i>	X03014009	Raba			X	X	X	X
<i>Benidorm</i>	X03031002	Beni	X	X	X			
<i>Elx - Agroalimentari</i>	X03065006	Elx	X	X	X	X		
<i>Elda - Lacy</i>	X03066003	Elda	X	X	X	X	X	X
<i>L'Alcora</i>	X12005005	Alcora	X	X	X	X	X	
<i>Almassora - C. P. Ochando</i>	X12009007	Alma	X	X		X	X	X
<i>Benicassim</i>	X12028001	Ben	X	X		X	X	X
<i>Burriana</i>	X12032001	Bur	X	X	X	X		
<i>Castellò - Penyeta</i>	X12040008	Peny	X	X	X		X	X
<i>Castello - Grau</i>	X12040010	Grau	X	X	X	X		
<i>Coratxar</i>	X12093004	Pobla	X	X	X	X		
<i>Sant Jordi</i>	X12099001	SJ	X	X	X			
<i>Vilafranca</i>	X12129001	VCid	X	X	X	X		
<i>Viver</i>	X12140002	Viver	X	X	X	X	X	X
<i>Zorita</i>	X12141002	Zorita			X	X	X	X
<i>Albalat dels Tarongers</i>	X46010001	Alba	X	X	X	X	X	
<i>Buñol - Cemex</i>	X46077006	Buniol	X	X	X	X	X	X
<i>Caudete de las Fuentes</i>	X46095001	Caude	X	X	X	X	X	X
<i>Quart de Poblet</i>	X46102002	Quart	X	X	X	X		
<i>Sagunt - CEA</i>	X46220010	Sagunt	X	X	X	X	X	X
<i>Valencia - Pista de Silla</i>	X46250030	Pista	X	X	X			X
<i>Valencia - Politecnic</i>	X46250046	Poli			X	X	X	X
<i>Valencia - Moli del Sol</i>	X46250048	Moli			X	X	X	X
<i>Villar del Arzobispo</i>	X46258001	Villar	X	X	X	X	X	X

Las ubicaciones de las estaciones pueden verse en la **Figura 23**.



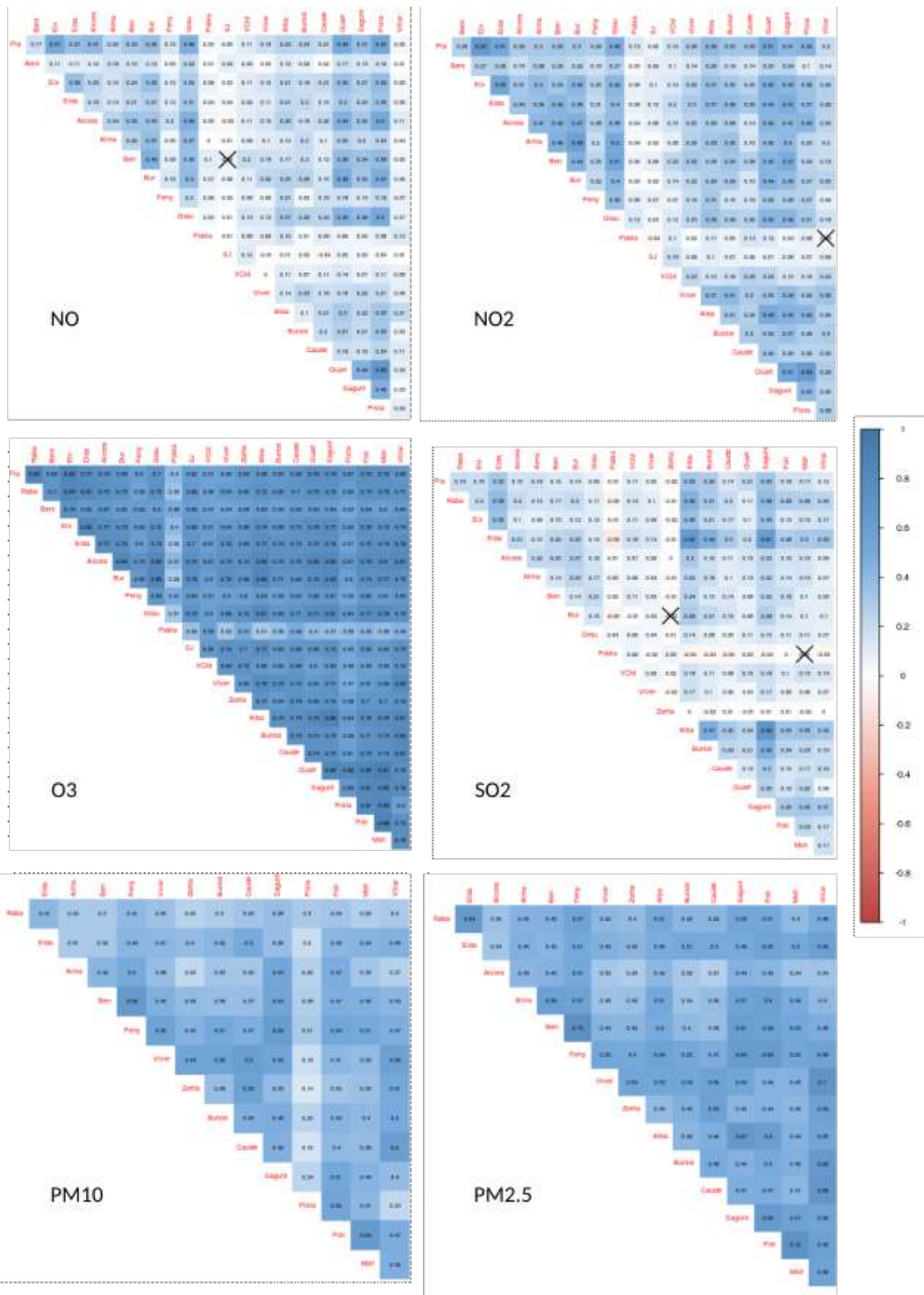
**Figura 23.** Estaciones de monitoreo que registraron una alta tasa de datos medidos para al menos 3 de los contaminantes estudiados en el período 2008-2018.

En general, no se utilizó ninguna imputación para rellenar los datos faltantes, excepto cuando así se precisó en las técnicas analíticas que *a posteriori* se aplicaron. Esta decisión se debe a que el objetivo de este trabajo es obtener la mejor caracterización del área de estudio y al incluir datos modelados sobre series de tiempo con comportamiento azaroso es posible introducir errores que sesguen los resultados de los análisis.

### III. Evaluación de la redundancia entre las estaciones

#### Análisis de correlación entre estaciones

Para conocer la redundancia entre las estaciones se realizó un análisis de correlación para cada contaminante en cada estación con los datos horarios. Como los datos presentaron una distribución estadística con un fuerte corrimiento de la normalidad, se seleccionó el coeficiente de correlación de Spearman para evaluar la fuerza de la relación. La **Figura 24** muestra los coeficientes obtenidos para cada par de estaciones.



**Figura 24.** Factor de correlación de Spearman entre pares de estaciones de monitoreo para cada contaminante ( $p < 0.05$ ). La cruz en el recuadro indica que la correlación fue no significativa.

La correlación de los contaminantes no fue uniforme para todos los contaminantes, si bien se observó que en la mayoría de los casos fue significativa. NO y NO<sub>2</sub> presentaron una correlación promedio baja de 0.18 [-0.04, 0.68] y 0.26 [-0.04, 0.65], respectivamente, indicándose entre corchetes el rango de los factores de correlación. Las estaciones SJ y Ben no mostraron correlación para NO, al igual que las estaciones Pobla y Villar para NO<sub>2</sub>.

De forma similar, SO<sub>2</sub> mostró una correlación media baja de 0.15 [-0.09, 0.70]. En particular, las estaciones Zorita y Bur presentaron las correlaciones más débiles y correlación no significativa con Pobla y Moli respectivamente.

En contraste, el O<sub>3</sub> presentó la mayor correlación media: 0.68 [0.29, 0.91], donde la estación Pobla se diferenció sutilmente del resto por mostrar un menor grado de redundancia ( $r = 0.40$ ). Los factores de correlación para material particulado también fueron elevados: 0.41 [0.14, 0.66] y 0.49 [0.27, 0.75] para PM<sub>10</sub> y PM<sub>2.5</sub>. La estación Pista se diferenció con una correlación media más baja del resto ( $r = 0.27$ ) para PM<sub>10</sub>.

Este análisis muestra correlaciones significativas ( $p < 0.05$ ) en la mayoría de los casos, siendo las más altas para el O<sub>3</sub>, PM<sub>2.5</sub> y PM<sub>10</sub>, lo que sugiere que la red de monitoreo contiene estaciones potencialmente redundantes. A su vez, se observa que las estaciones más alejadas de las zonas urbanas (Zorita, Pobla) se diferencian del resto de las estaciones.

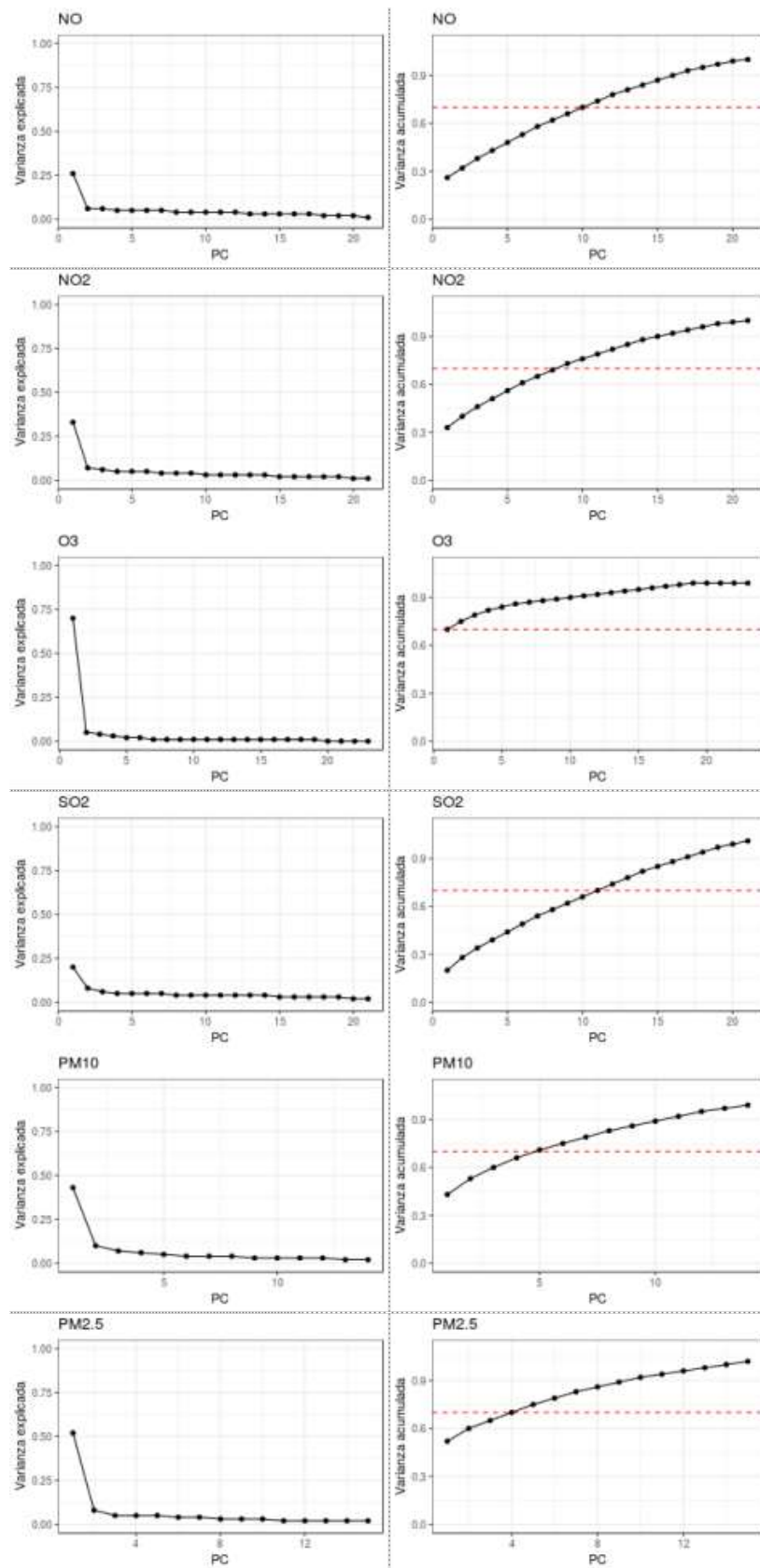
Estos resultados conllevan a continuar el análisis, en la búsqueda de identificar las estaciones de monitoreo potencialmente redundantes con un PCA, y luego confirmar la redundancia con un análisis de cluster jerárquicos.

### Análisis de componentes principales

El conjunto de datos recolectados por las estaciones de monitoreo terrestre nos brinda información respecto de la variabilidad de las concentraciones para los distintos contaminantes en el territorio. La técnica de PCA nos permite identificar la parte de esa variabilidad que es compartida entre las distintas estaciones de monitoreo.

Un PCA arroja como resultado el número de componentes principales que se deben considerar para conocer la variabilidad de la información que se encuentra en el conjunto de datos. En una situación ideal, sin redundancia, esperaríamos que existan tantos componentes como estaciones. En ese caso, cada estación estaría cubriendo un objetivo particular en la recolección de la información. Se decidió como límite de corte el 70% de la varianza acumulada [Oshima et al. 2002].

Como podemos ver en la **Figura 25**, para NO, NO<sub>2</sub>, SO<sub>2</sub>, PM<sub>10</sub> y PM<sub>2.5</sub> la proporción de la varianza explicada por el primer componente fue inferior al 50%. Esto, en principio, es positivo ya que indica que no existe una alta redundancia entre las mediciones de las estaciones de monitoreo.



**Figura 25.** Proporción de la varianza explicada y de la varianza acumulada por componente principal para cada contaminante

El 70% de la varianza de las mediciones de cada uno de estos contaminantes se explica en 10, 8, 11, 5 y 4 componentes principales para NO, NO<sub>2</sub>, SO<sub>2</sub>, PM<sub>10</sub> y PM<sub>2.5</sub> respectivamente. Por otra parte, las concentraciones de O<sub>3</sub> muestran un mayor grado de redundancia, alcanzando el 70% de explicación de la varianza con solo un componente principal.

Considerando que se trabaja con 21 o 15 estaciones, dependiendo del contaminante, el 70% de la varianza se explica con un número relativamente bajo de estaciones. En otras palabras, para la mayoría de los contaminantes, la mitad de las estaciones no están aportando información nueva y, por tanto, muestran información redundante, pudiéndose relocalizar para obtener un mejor conocimiento de los fenómenos estudiados.

Continuando el análisis, se aplicó una rotación de tipo “varimax” y se repitió el análisis para identificar las estaciones comprendidas dentro de cada componente principal. La **Figura 26** muestra el porcentaje de la varianza de las mediciones en cada estación por contaminante que queda explicado en cada uno de los componentes principales (sólo se muestran los porcentajes mayores al 40% para facilitar la lectura).

Como caso particular, para el O<sub>3</sub>, que mostró un alto nivel de explicación con tan solo un único componente, se encontró que la varianza de todas las estaciones se encontraba explicada por encima del 75% en ese componente, con excepción de la estación Peny que mostró una redundancia del 47% dentro de ese componente principal.

En algunos casos, podemos ver que la información recolectada por una estación no es posible de ser reemplazada por otra. Por ejemplo, las mediciones efectuadas para el NO en las estaciones Beni, Peny, SJ, VCid, Pobra se encuentran mayoritariamente representadas por un único componente y ninguna otra componente aporta un porcentaje alto para su comprensión. Para NO<sub>2</sub> esto sucede con las estaciones Beni, Peny y Pobra; para el SO<sub>2</sub> con las estaciones Zorita, Viver, VCid, Pobra, Elx, Caude y Alma; para el PM<sub>2.5</sub> con la estación Buniol y para PM<sub>10</sub> con la estación Raba.

En otros casos se observa que porcentajes elevados de la información recolectada por una estación se solapa dentro de un mismo componente principal con otra estación. Esto es un fuerte indicador de redundancia. Para conocer si existen agrupaciones significativas se realizó un análisis de cluster, tal como se detalla a continuación.



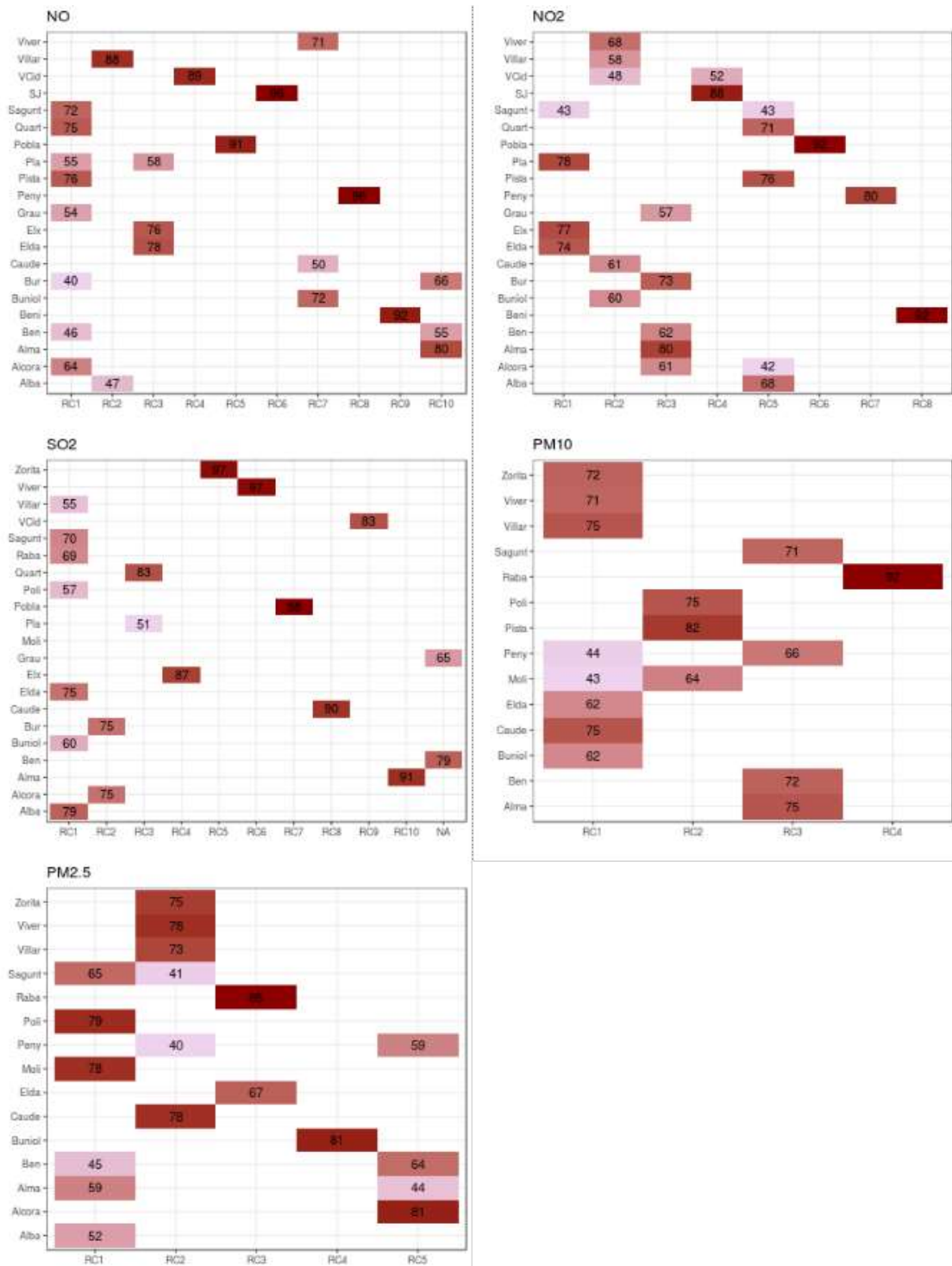
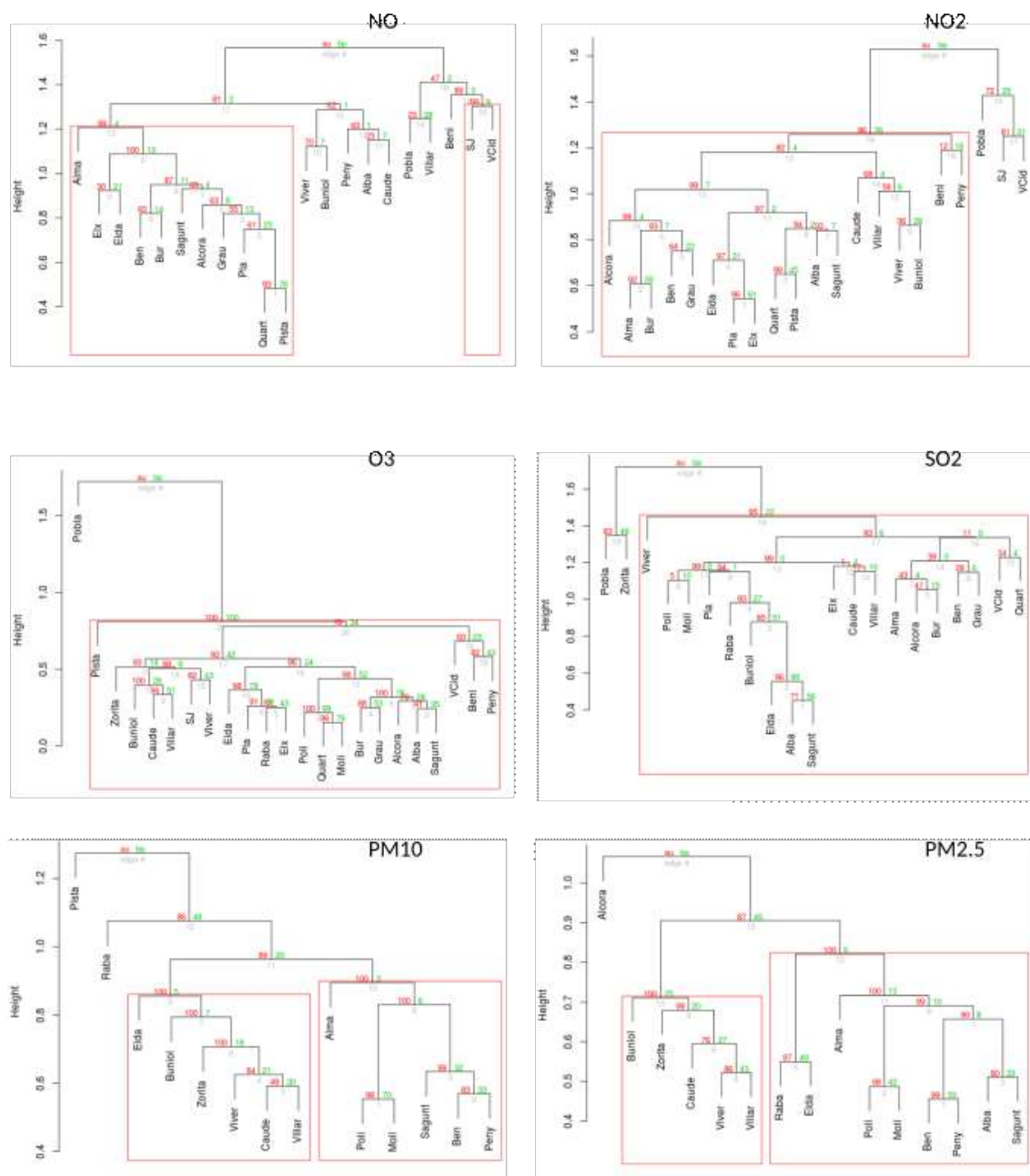


Figura 26. Porcentaje de la varianza explicada de cada una de las estaciones en cada uno de los componentes rotados (RC). Se omitieron los porcentajes inferiores al 40% para facilitar la lectura.

### Análisis clúster de las estaciones

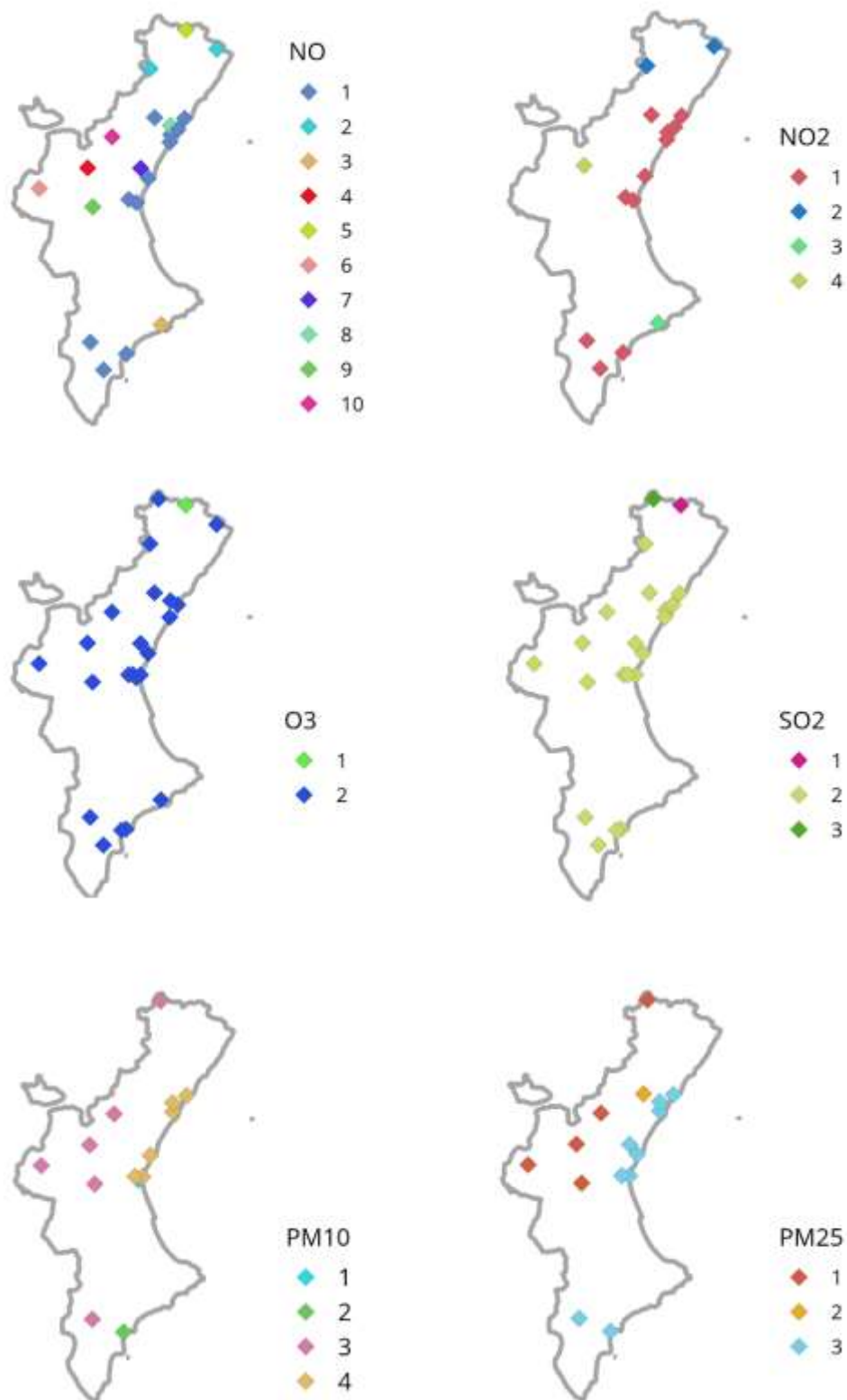
Para agrupar las estaciones se realizó un análisis cluster jerárquico, partiendo de la matriz de correlación. Se utilizó la técnica de cluster jerárquicos ya que es una técnica que no requiere supervisión para definir el número de cluster y se empleó la medida de distancia euclídea para evaluar la disimilitud entre las estaciones. A su vez, se aplicó el algoritmo desarrollado en [Shimodaira, 2004] para establecer los agrupamientos significativos ( $p < 0.05$ ).

En la **Figura 27** se observa el agrupamiento de las estaciones de monitoreo en cluster jerárquicos por cada contaminante medido. Para NO se identificaron 2 cluster, dando un total de 10 grupos diferenciados. Para el NO<sub>2</sub>, el análisis devolvió la formación de un gran cluster conteniendo 18 estaciones, generando 4 categorías diferenciadas. Para el O<sub>3</sub> se obtuvo un cluster con todas las estaciones con excepción de la estación Pobla. Algo similar sucedió con las mediciones de SO<sub>2</sub>, donde un gran cluster agrupó todas las estaciones con excepción de Pobla y Zorita. Finalmente, para PM<sub>10</sub> se formaron 2 cluster con 6 estaciones cada uno, dejando fuera las estaciones Pista y Raba. De forma similar, para PM<sub>2.5</sub> se obtuvieron 2 cluster grandes, dejando sin agrupar a la estación Alcora.



**Figura 27.** Dendrograma para cada uno de los contaminantes medidos en las estaciones de monitoreo de la Generalitat Valenciana

Para visualizar estas relaciones en el territorio, la **Figura 28** presenta un mapa con la distribución espacial de las estaciones. De esta forma, se puede apreciar más claramente un agrupamiento de las estaciones vinculadas a las zonas urbanas para el NO (Alma, Elx, Elda, Ben, Bur, Sagunt, Alcora, Grau, Pla, Quart, Pista), PM<sub>2.5</sub> (Raba, Elba, Alma, Poli, Moli, Ben, Penny, Alba y Sagunt) y PM<sub>10</sub> (Alma, Poli, Moli, Sagunt, Ben y Penny).



**Figura 28.** Mapa de ubicación de los clústeres generados para las estaciones de monitoreo.

En función de los resultados obtenidos, se evaluó la redundancia de las estaciones para todos los contaminantes. Para ello, se realizó un análisis de patrones, donde a cada cluster se le asignó un símbolo para identificar con mayor claridad los casos repetidos. La **Tabla 14** muestra el código generado para cada estación de monitoreo.

**Tabla 14.** Análisis de patrones para los clústeres generados para cada estación por contaminante. Cada letra representa un agrupamiento clúster para cada contaminante.

Estación	Contaminantes					
	NO	NO <sub>2</sub>	O <sub>3</sub>	SO <sub>2</sub>	PM <sub>10</sub>	PM <sub>2.5</sub>
Pla	A	A	A	A	-	-
Raba	-	-	A	A	C	B
Beni	C	A	A	-	-	-
Elx	A	A	A	A	-	-
Elda	A	A	A	A	A	B
Alcora	A	A	A	A	-	C
Alma	A	A	-	A	B	B
Ben	A	A	-	A	B	B
Bur	A	A	A	A	-	-
Peny	D	A	A	-	B	B
Grau	A	A	A	A	-	-
Pobla	E	B	B	B	-	-
SJ	B	C	A	-	-	-
VCid	B	D	A	A	-	-
Viver	F	A	A	A	A	A
Zorita	-	-	A	C	A	A
Alba	G	A	A	A	-	B
Buniol	H	A	A	A	A	A
Caude	I	A	A	A	A	A
Quart	A	A	A	A	-	-
Sagunt	A	A	A	A	B	B
Pista	A	A	A	-	D	-

Poli	-	-	A	A	B	B
Moli	-	-	A	A	B	B
Villar	J	A	A	A	A	A

Analizando la **Tabla 14** queda en evidencia que las estaciones Pla, Elx, Bur, Grau y Quart; Alma y Ben; y Poli y Moli pertenecen al mismo cluster para los distintos contaminantes incluidos en este estudio, siendo completamente redundante en todos los contaminantes que se miden. Por otra parte, las estaciones Raba, Beni, Elda, Alcora, Peny, Pobla, Sj, VCid, Viver Zorita, Alba, Buniol, Caude, Sagunt, Pista y Villar presentan comportamientos únicos.

#### *IV. Analizar el área de influencia de las estaciones de monitoreo*

En el monitoreo de la calidad del aire las mediciones que se realizan están fuertemente influenciadas por las actividades que se efectúan en el área donde las estaciones se encuentran emplazadas. Es por esto que, en el diseño del monitoreo, es preciso considerar que las características de la zona a estudiar sean homogéneas con el resto del territorio al cual se espera representar.

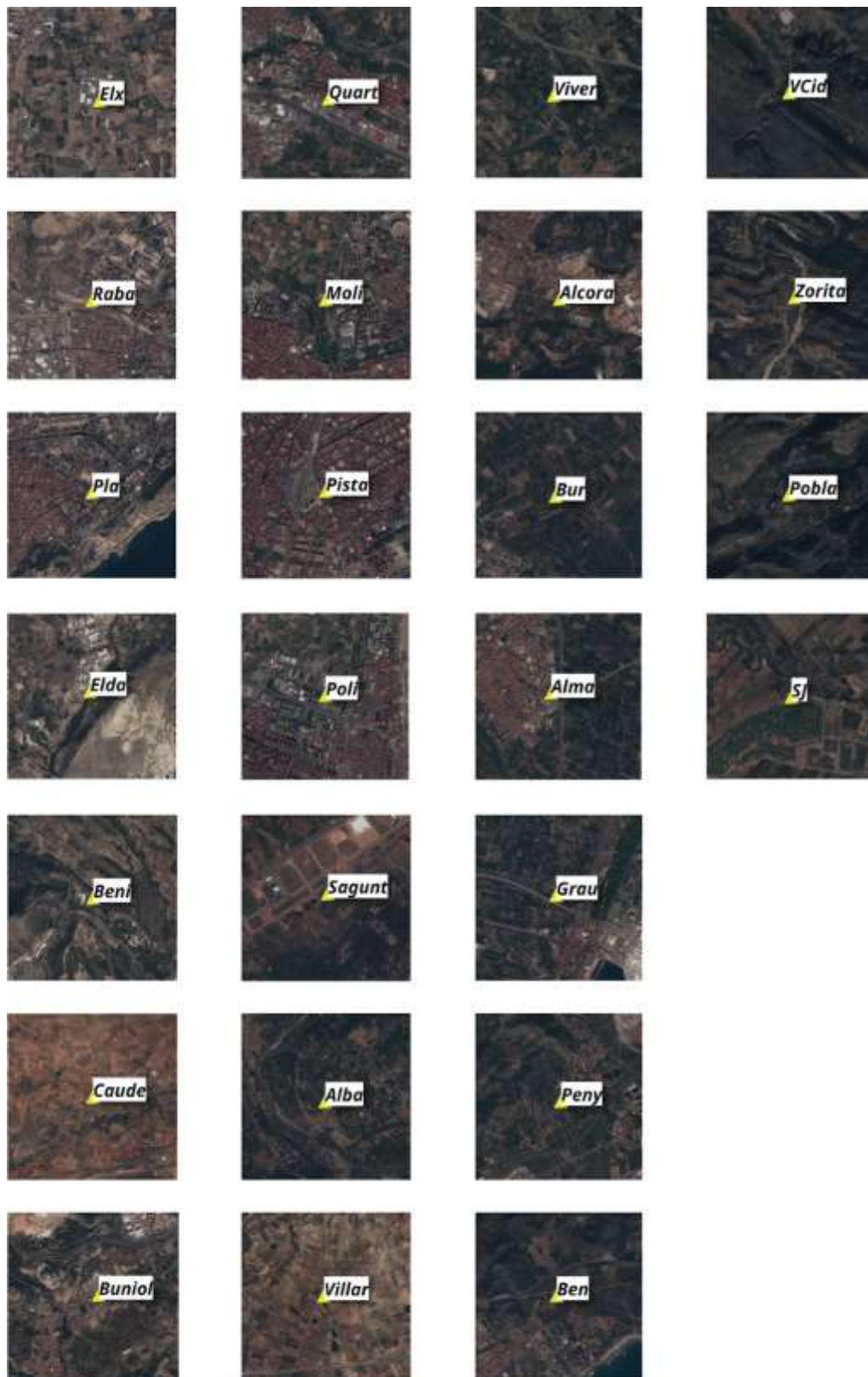
En este apartado nos interesa conocer las actividades productivas cercanas al área de influencia de las estaciones de monitoreo para contextualizar con mayor detalle el resto de los resultados. En la **Tabla 15** pueden verse las estaciones con las que se trabajó y las zonas en las cuales se ubican.

**Tabla 15.** Estaciones anidadas por zona y actividad contaminante principal.

<b>Zona</b>	<b>Estaciones</b>	<b>Actividad contaminante principal identificada</b>
ES1001	SJ	Zona de influencia de la Central térmica de Andorra
ES1002	Zorita, Pobla, VCid	
ES1013	Elx	Zona de actividades cementeras
ES1017	Raba, Pla, Elda, Beni	
ES1003	Ben, Grau, Bur, Alcora, Peny, Alma	Zona de actividades cerámicas
ES1016	Quart, Poli, Moli, Pista	Áreas de tráfico intenso
ES1007	Sagunt, Alba	Área de influencia de emisiones de tráfico e industriales
ES1008	Villar, Viver	

Las estaciones Buniol y Caude se encuentran en la zona ES1010, pudiendo ser influenciadas por la zona ES1008.

Para identificar las potenciales fuentes de emisión se utilizó la clasificación de uso de suelo CORINE. Se aplicó un buffer de 1 km entorno a la estación y se analizaron las clases de usos de suelo presentes en estas áreas. En la **Figura 29** se pueden ver los sitios de monitoreo en el territorio y en la **Figura 30** los sitios de monitoreo sobre el mapa de uso de suelo CORINE.



Ubicación de las estaciones de monitoreo de calidad de aire de la Generalitat Valenciana  
Proyección EPSG:3857 - Escala 1:9000

**Figura 29.** Vista detalle del área circundante a las estaciones de la Red de Monitoreo de la Comunidad Valenciana incluidas en este estudio.

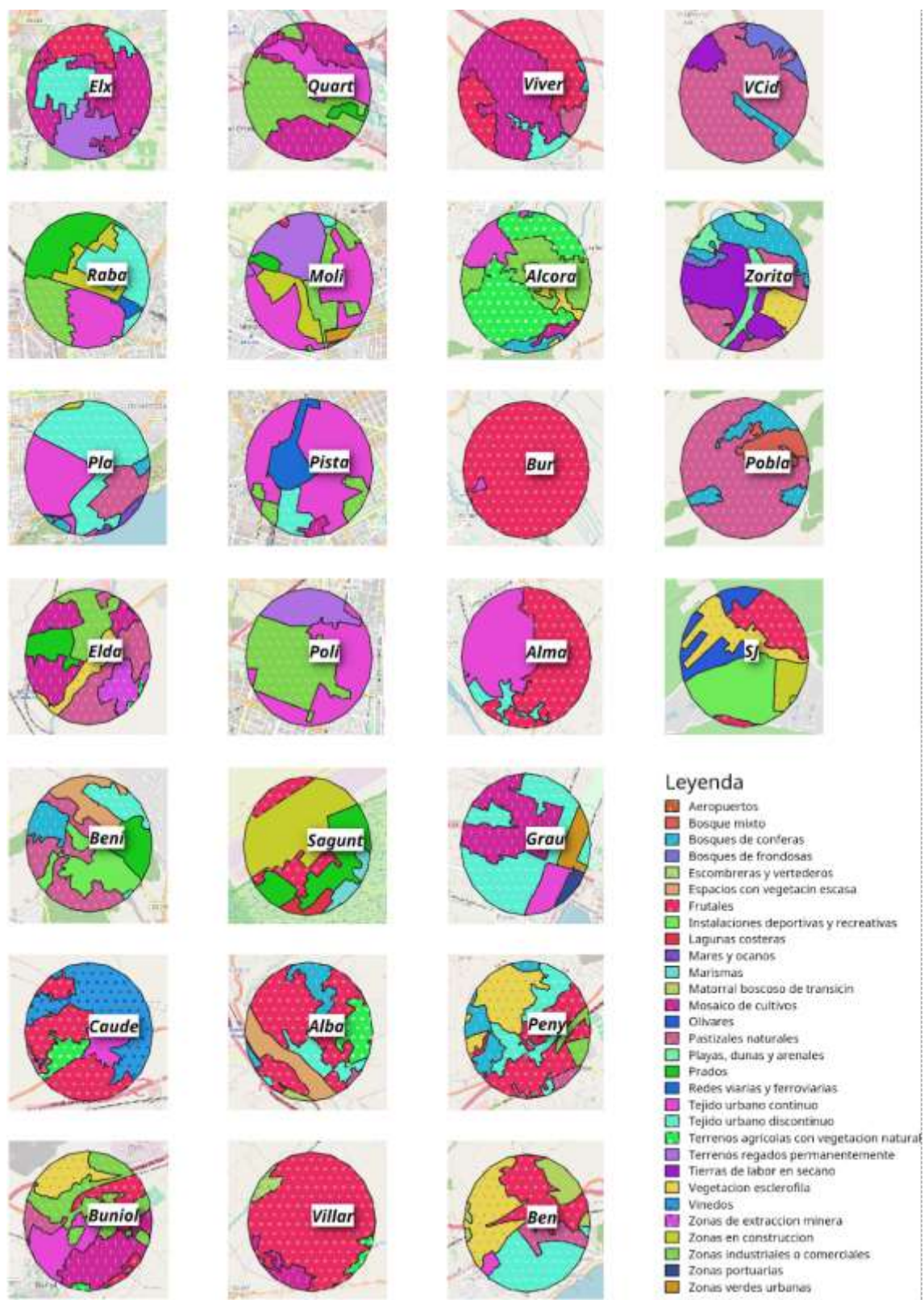


Figura 30. Vista detalle de las clases CORINE de uso de suelo en el área circundante a las estaciones de la Red de Monitoreo de la Comunidad Valenciana incluidas en este estudio.



El área de ocupación de las categorías de usos de suelo provista por CORINE fue integrada para cada estación de la Red de Monitoreo. En la **Figura 31** se muestra el porcentaje de usos de suelo que influyen en cada una de las estaciones de monitoreo.



**Figura 31.** Porcentaje de las clases CORINE en las estaciones de monitoreo.

Para más detalle, en la **Tabla 16** se presentan los porcentajes de cada categoría para cada una de las estaciones. Se tomó como punto de corte el 60% de ocupación para asignar una categoría a la estación. De esta forma, las estaciones que según categoría CORINE 2012 pueden representar a una superficie característica son para la:

- Superficie artificial (A): Pla, Raba, Grau, Quart, Pista, Poli y Moli;
- Zonas agrícolas (B): Elx, Alcora, Bur, Viver, Alba, Caude y Villar;
- Bosques y áreas seminaturales (C): Pobla, VCid y Zorita;

Mientras, las estaciones identificadas con usos mixtos fueron: Beni, Elda, Alma, Ben, Peny, SJ, Buniol y Sagunt.

Comparando estos resultados con los anteriormente presentados, se observa que el cluster Poli y Moli pertenecen al mismo uso de suelo, mientras que el cluster compuesto por las estaciones Pla, Grau, Quart, Elx y Bur se encuentra dividido entre la clase artificial y la zona agrícola. Sin embargo, ambos usos comparten las mismas fuentes de contaminantes, como puede ser el transporte automotor.

**Tabla 16.** Característica de las estaciones de monitoreo de calidad de aire de Valencia. En color las estaciones que superan el 60% en alguna de las categorías.

Estación de monitoreo	Nombre corto	Clases %					
		1	2	3	4	5	
X03014006	El Pla	Pla	80.3	17.2	2.5	-	-
X03014009	Rabassa	Raba	76.3	23.2	-	-	-
X03031002	Benidorm	Beni	34.2	12.1	53.7	-	-
X03065006	Agroalimentari	Elx	21.9	78.1	-	-	-
X03066003	Lacy	Elda	27.4	36.5	36.2	-	-
X12005005	L'Alcora	Alcora	-	86.8	13.2	-	-
X12009007	C. P. Ochando	Alma	48.5	51.5	-	-	-
X12028001	Benicassim	Ben	29.7	25.9	44.4	-	-
X12032001	Burriana	Bur	1.0	99.0	-	-	-
X12040008	Penyeta	Peny	26.8	34.3	38.8	-	-
X12040010	Grau	Grau	66.1	33.9	-	-	-
X12093004	Coratxar	Pobla	-	-	100	-	-
X12099001	Sant Jordi	SJ	42.3	38.0	19.7	-	-
X12129001	Vilafranca	VCid	0.4	6.5	93.2	-	-
X12140002	Viver	Viver	6.0	89.4	4.7	-	-
X12141002	Zorita	Zorita	-	33.8	66.2	-	-
X46010001	Albalat dels Tarongers	Alba	9.9	63.3	26.8	-	-
X46077006	Cemex	Buniol	43.4	38.8	17.9	-	-
X46095001	Caudete de las Fuentes	Caude	10.6	87.7	1.7	-	-
X46102002	Quart de Poblet	Quart	62.5	37.5	-	-	-
X46220010	CEA	Sagunt	44.3	44.9	-	6.2	4.5
X46250030	Pista de Silla	Pista	97.6	2.4	-	-	-
X46250046	Politecnic	Poli	81.5	18.5	-	-	-
X46250048	Molí del Sol	Moli	72.9	27.1	-	-	-
X46258001	Villar del Arzobispo	Villar	3.5	96.5	-	-	-

### V. Analizar los vientos locales en los sitios de monitoreo

Como se observó anteriormente, el estudio de la calidad del aire requiere analizar las variables meteorológicas que contribuyen a la dispersión de los contaminantes. En este sentido, resulta fundamental contar con información del viento en el mismo punto y momento de los registros de los contaminantes.

Los flujos del viento que llegan a las estaciones pueden sufrir variaciones respecto a los valores registrados por las estaciones meteorológicas en altura. Por ello, resulta conveniente utilizar la información recolectada en las estaciones de monitoreo que se encuentran a nivel del suelo.

De las estaciones seleccionadas para analizar en este trabajo, no todas las estaciones presentan registros de los parámetros meteorológicos. En 16 de las 25 estaciones incluidas en este trabajo se mide la dirección y la velocidad del viento *in situ*. Estas estaciones son: Raba, Beni, Elx, Elda, Alma, Bur, Peny, Grau, Pobla, SJ, VCid, Zorita, Buniol, Sagunt, Pista y Villar.

Para la construcción de las rosas de los vientos se utilizó como parámetro de corte el método de Jenks, tal como se describió anteriormente. La **Figura 32** muestra la frecuencia de los vientos en función de su ángulo de procedencia para cada una de las estaciones, los colores indican la intensidad de los vientos registrados y el ancho de la sección, su frecuencia. Puede verse que en las estaciones Alma, Bur, Elx, Peny, Raba, Sagunt, SJ y VCid existe mayor frecuencia de los vientos del sector WN. Mientras en las estaciones Beni y Grau cobra una mayor relevancia el sector WS, en la estación Zorita el WN, N y el S, en la estación Buniol y Pista el W. En la estación Villar los vientos más influyentes provienen del W y del SE.

El viento máximo para todo el período se registró en la estación Villar y alcanzó los 24.8 m.s<sup>-1</sup>. Mientras que la velocidad media máxima se registró en la estación Pobla (4.14 m.s<sup>-1</sup>). En la tabla de valoración de vientos de Beaufort, estos son vientos tranquilos, considerados bajo la denominación de ventolinas. Las estaciones con vientos medios más calmos fueron: Pista (0.83 m.s<sup>-1</sup>), Alma (0.93 m.s<sup>-1</sup>) y Raba (0.95 m.s<sup>-1</sup>), correspondiendo a estaciones urbanas y de uso mixto. Esta situación de calma y suaves ventolinas dificulta la dispersión de los contaminantes, favoreciendo escenarios de mayores concentraciones de los contaminantes en el territorio.

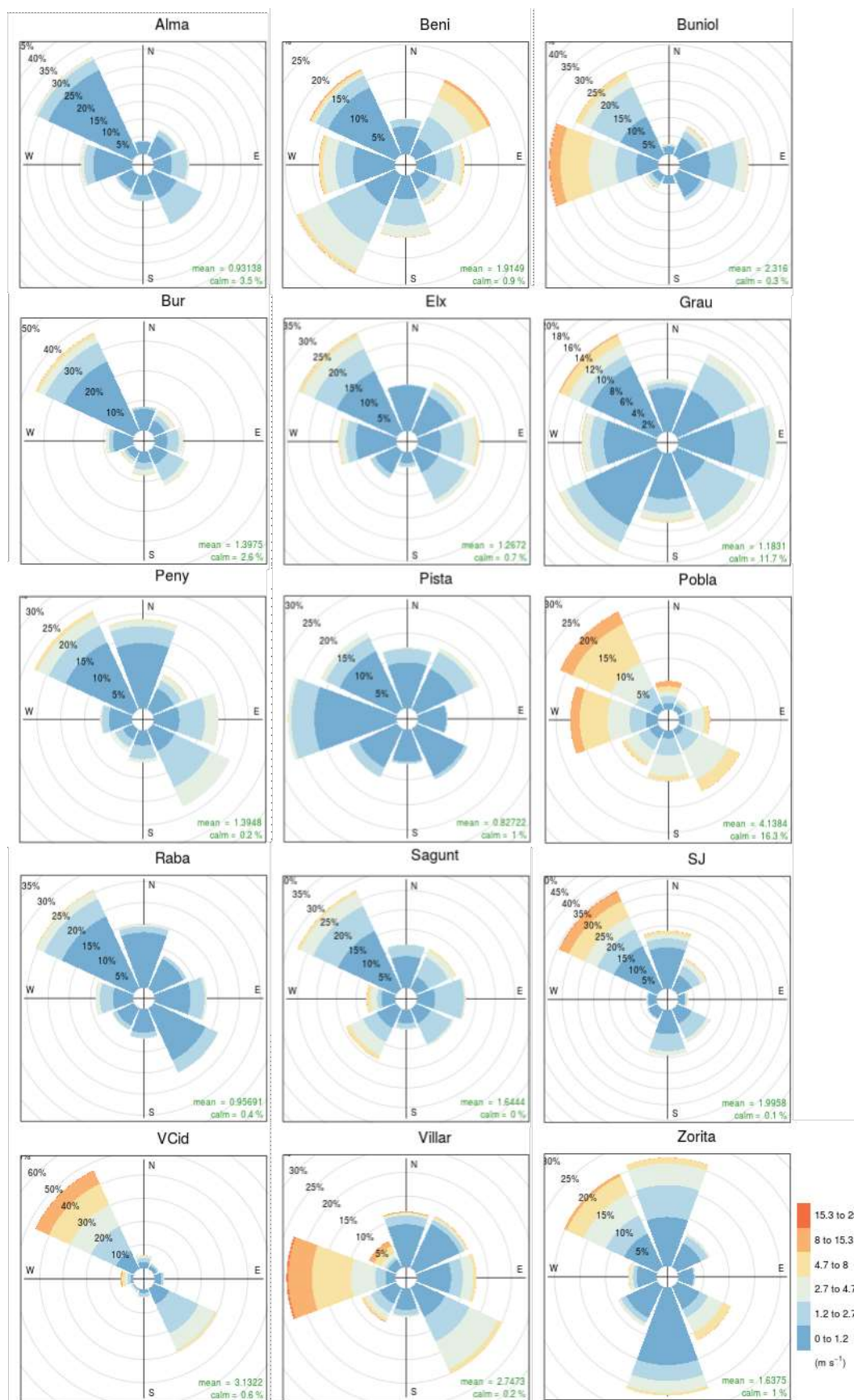


Figura 32. Rosa de vientos (m.s<sup>-1</sup>)

Como puede verse en la **Figura 33**, la dirección de proveniencia de los vientos varía a lo largo del año. Incluso, en algunas estaciones, cambia la dirección del viento más frecuente. Al modificarse el patrón de vientos y existiendo un viento fuertemente predominante, pueden arrastrarse contaminantes desde distintas fuentes hasta la estación de monitoreo a lo largo del año.

En general, los vientos son más calmos durante el verano y más intensos en el invierno (**Figura 33**). Los cambios interanuales característicos de esta región no se observan en todas las estaciones por igual. En el verano, para las estaciones Alma, Buniol, Elx, Pobla y Segunt aparece un viento predominante desde el sector E, mientras que en las estaciones Villar, VCid y Zorita se observa desde el sector SE. En las estaciones Grau y Peny se observa un corrimiento hacia el sector E desde la primavera. Estas diferencias pueden asociarse a los obstáculos que las estaciones presenten en sus cercanías.

Dado que los vientos cambian estacionalmente, variando su intensidad y dirección predominante, es posible considerar que el área de influencia considerada para las estaciones se vea modificada a lo largo del año.

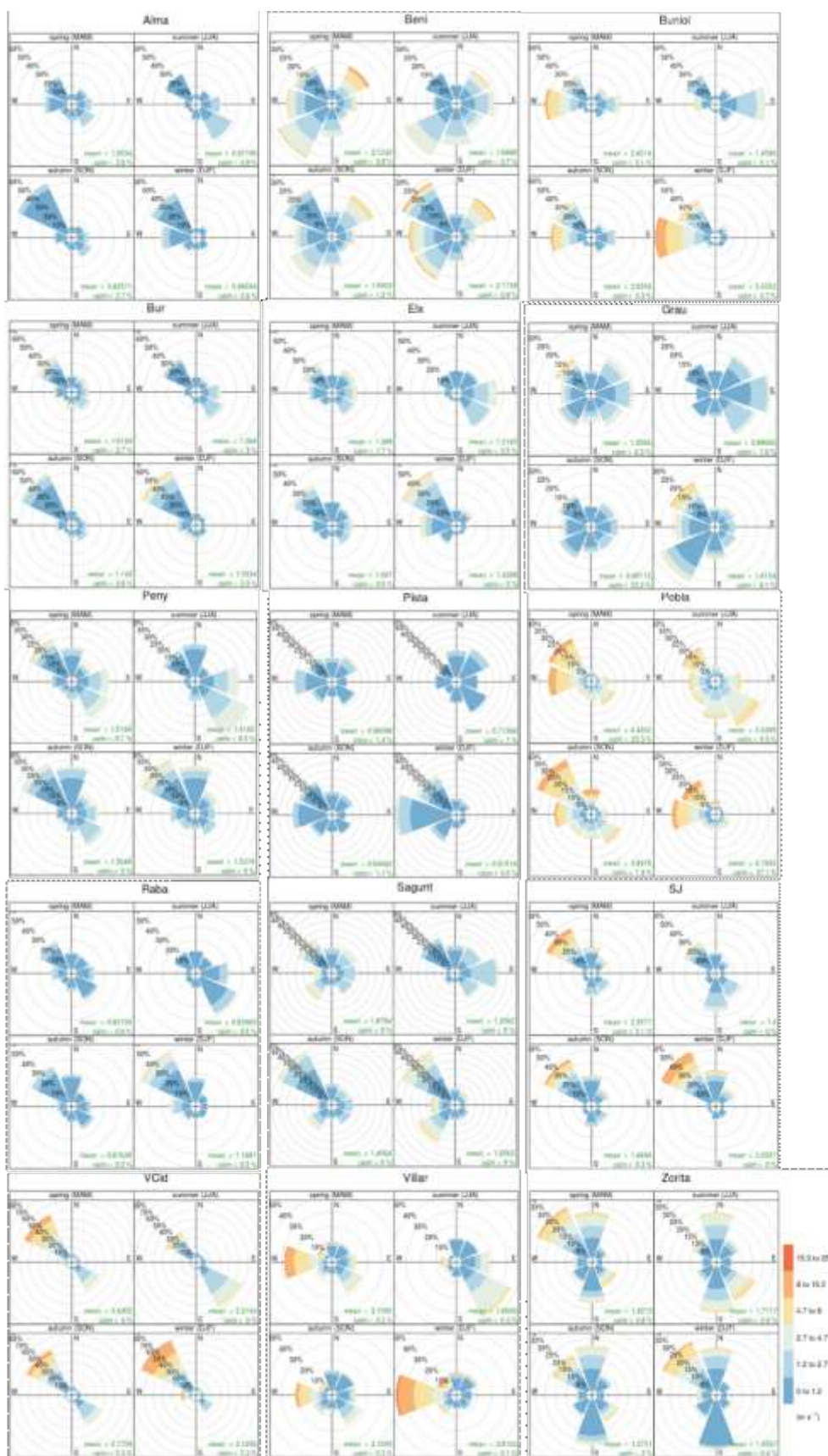
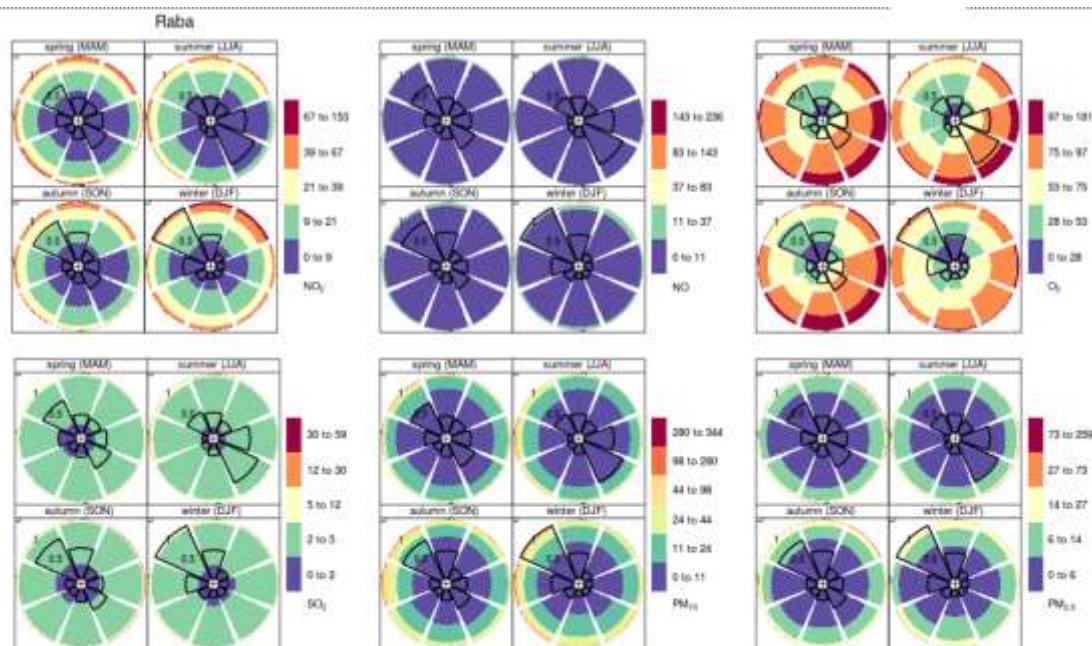


Figura 33. Rosa de los vientos por estación climática (m.s<sup>-1</sup>)

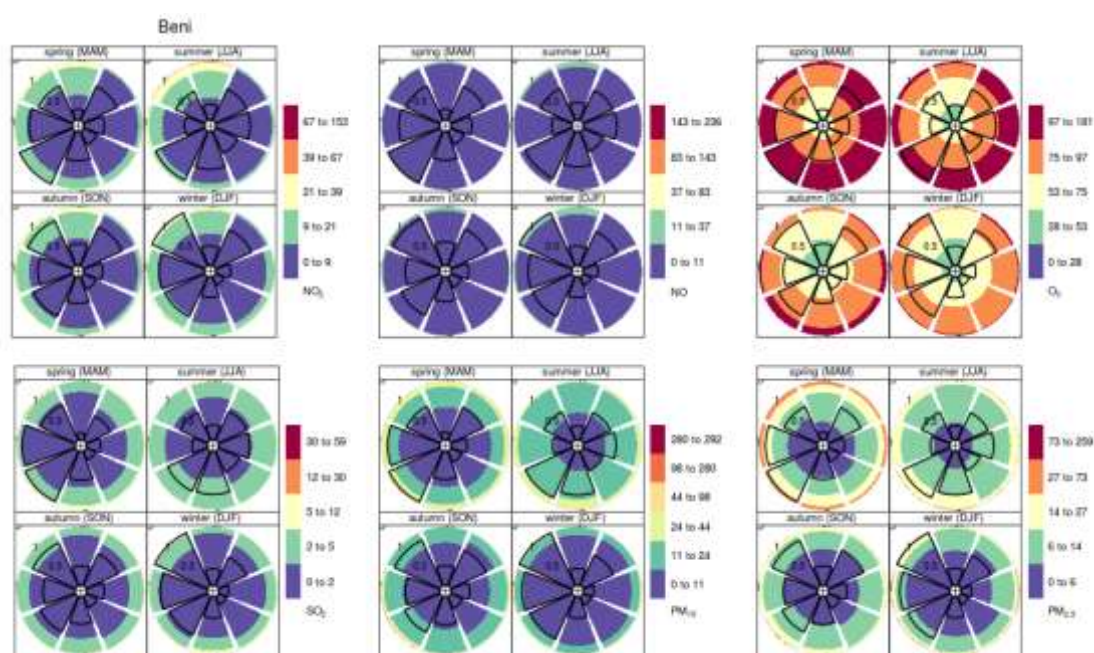
## VI. Análisis de aportes

Las rosas de contaminación nos permitieron identificar los sectores de donde llegaron los contaminantes. Para la estación Raba se observó que las mayores concentraciones del NO y NO<sub>2</sub> provinieron del N y NE, llegando con más frecuencia las concentraciones mayores en el invierno y la primavera (las estaciones con mayores vientos). Para O<sub>3</sub> se observó que los mayores aportes en todas las estaciones provinieron del E, SE y S, donde las concentraciones mayores llegaron con más frecuencia en primavera, otoño y verano. Respecto al SO<sub>2</sub>, se observó una distribución equitativa, mostrando mayores concentraciones desde el sector N, E y SE en otoño. Para PM<sub>10</sub> se recibieron aportes mayores desde el W y NW, en particular en otoño e invierno (**Figura 34**).



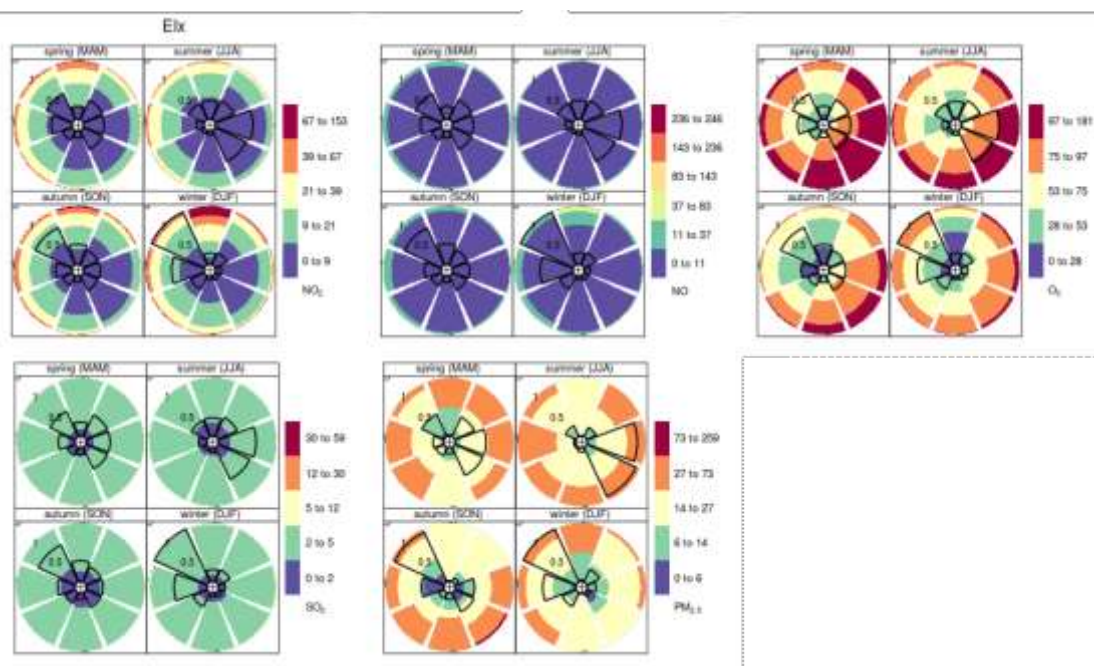
**Figura 34.** Rosas de contaminación en función de las direcciones del viento en la estación de monitoreo Raba. Los colores indican las concentraciones en µg.m<sup>-3</sup>. El tamaño de las categorías indica el % de la frecuencia del viento.

En la estación Beni, las concentraciones mayores de NO y NO<sub>2</sub> llegaron de forma descendente del N hasta el E. Las concentraciones mayores de O<sub>3</sub> provinieron desde los sectores E, NE, SE, S, SO y en menor grado del O. Estos contaminantes mostraron un mayor nivel en verano y primavera. El SO<sub>2</sub> llegó en niveles mayores desde el S. PM<sub>10</sub> y PM<sub>2.5</sub> se midieron mayormente en primavera y verano, con concentraciones mayores desde el sector O y NO (**Figura 35**).



**Figura 35.** Rosas de contaminación en función de las direcciones del viento en la estación de monitoreo Beni. Los colores indican las concentraciones en  $\mu\text{g}\cdot\text{m}^{-3}$ . El tamaño de las categorías indica el % de la frecuencia del viento.

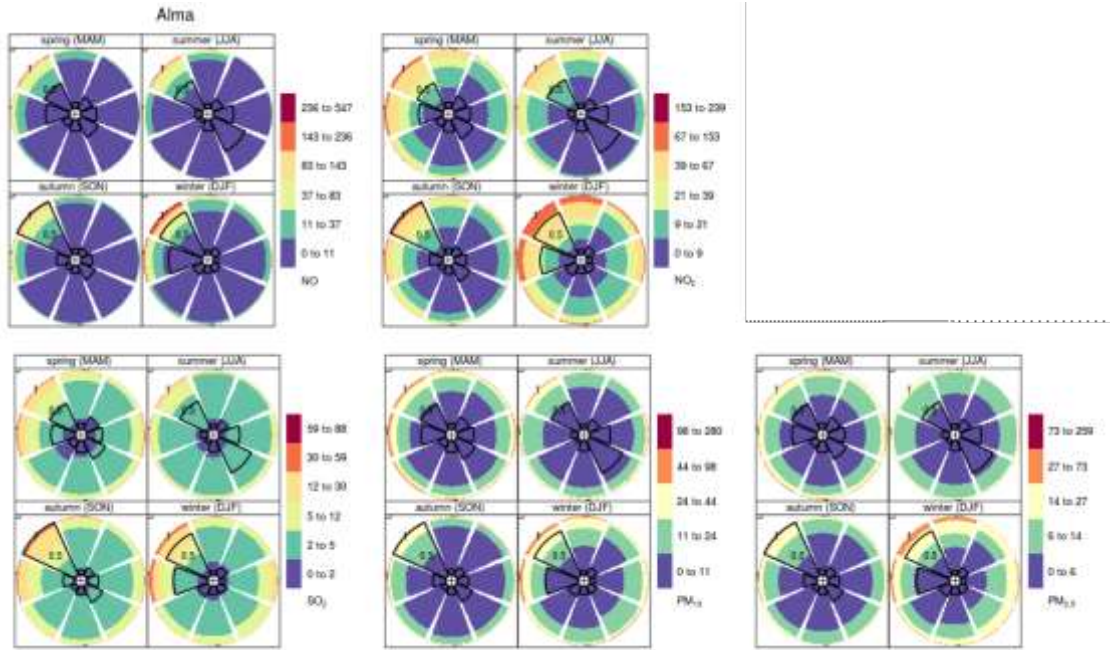
En la estación Elx las mayores concentraciones de  $\text{NO}$  y  $\text{NO}_2$  fueron en invierno del sector N. Las mayores concentraciones de  $\text{O}_3$  llegan del sector SE, E y S en las estaciones de primavera, verano y otoño, descendiendo para el invierno. El origen de  $\text{PM}_{2.5}$  varía a lo largo del año, apareciendo del N, E, SE y O en primavera, del SW en verano, de SE, y S y SW en otoño (**Figura 36**).



**Figura 36.** Rosas de contaminación en función de las direcciones del viento en la estación de monitoreo Elx. Los colores indican las concentraciones en  $\mu\text{g}\cdot\text{m}^{-3}$ . El tamaño de las categorías indica el % de la frecuencia del viento. En esta estación no se monitorea el  $\text{PM}_{10}$ .

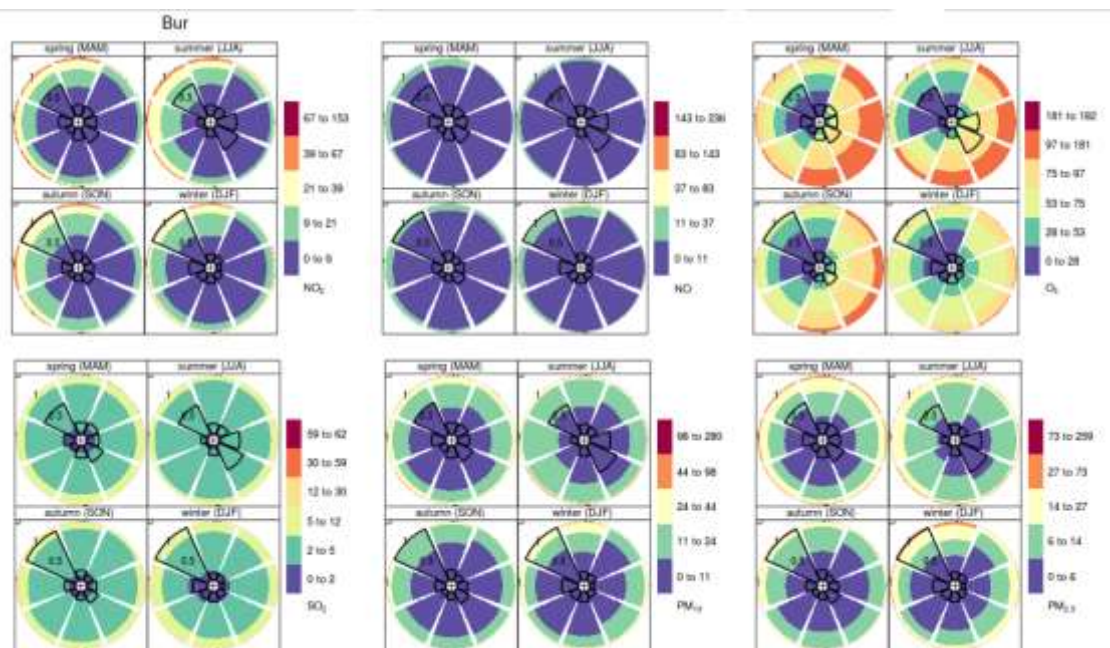


En Alma, para NO, NO<sub>2</sub>, SO<sub>2</sub>, PM<sub>10</sub> y PM<sub>2.5</sub> proviene en todas las estaciones desde NW, N y O, con el mayor aporte en invierno (**Figura 37**).



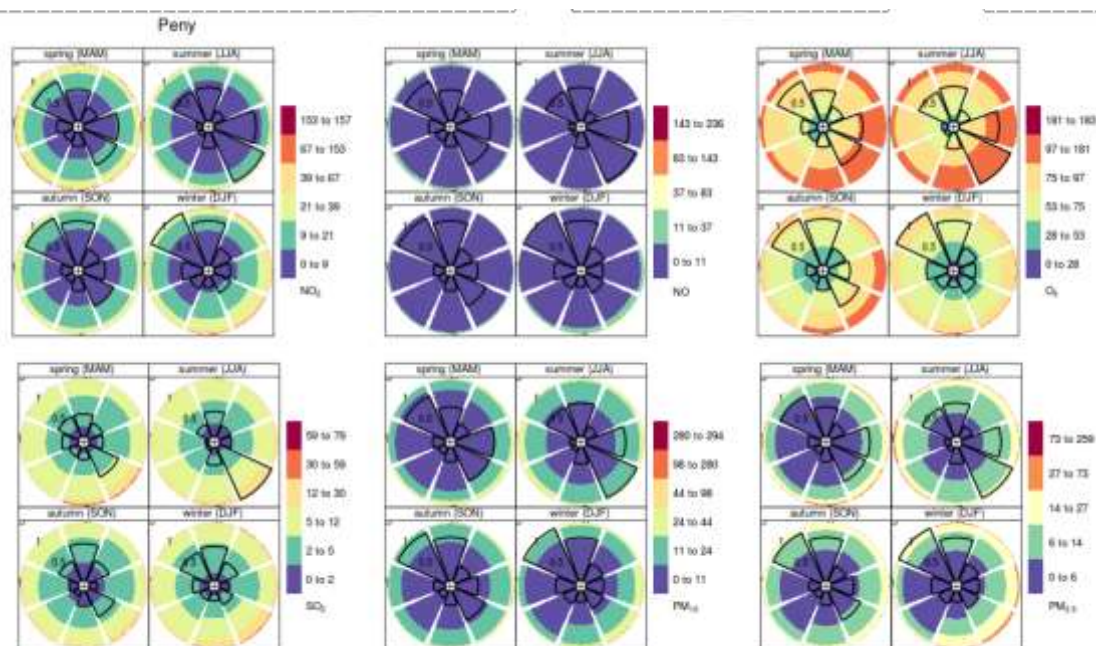
**Figura 37.** Rosas de contaminación en función de las direcciones del viento en la estación de monitoreo Alma. Los colores indican las concentraciones en  $\mu\text{g}\cdot\text{m}^{-3}$ . El tamaño de las categorías indica el % de la frecuencia del viento. En esta estación no se monitorea el O<sub>3</sub>.

En la estación Bur, el aporte mayoritario de NO, NO<sub>2</sub>, PM<sub>10</sub> y PM<sub>2.5</sub> llega del sector N al SW. El NO tiene sus niveles más altos en invierno, mientras que para el NO<sub>2</sub> llega en primavera verano. En contraposición, las mayores concentraciones llegan desde el sector NE al S (**Figura 38**).



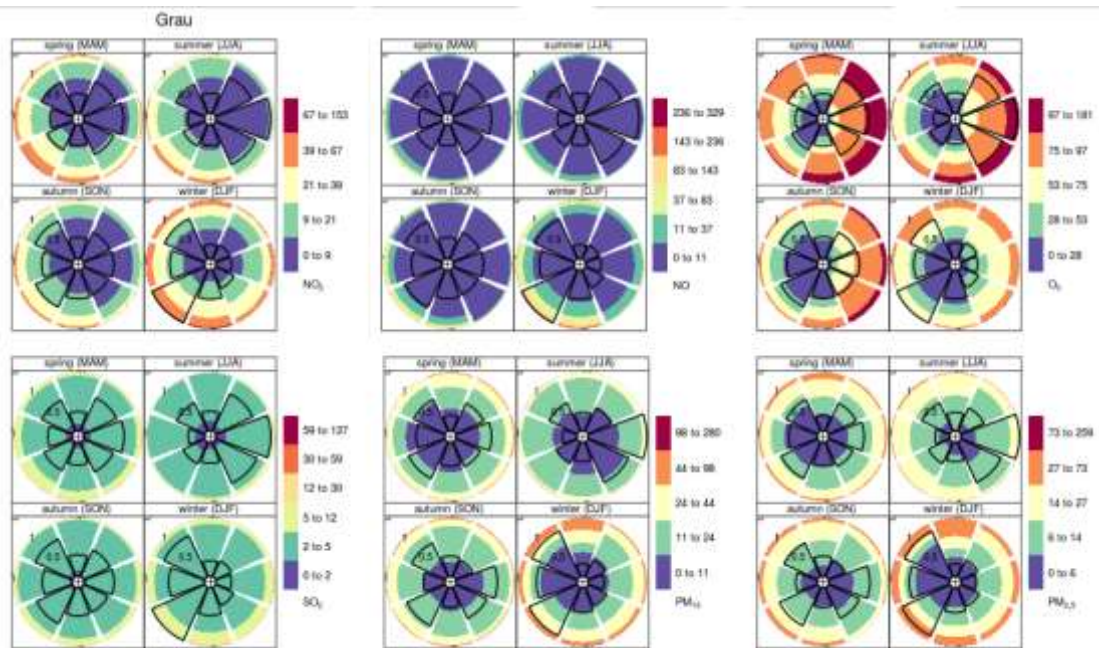
**Figura 38.** Rosas de contaminación en función de las direcciones del viento en la estación de monitoreo Bur. Los colores indican las concentraciones en  $\mu\text{g}\cdot\text{m}^{-3}$ . El tamaño de las categorías indica el % de la frecuencia del viento.

La estación Peny recibe el principal aporte desde  $\text{O}_3$ ,  $\text{PM}_{10}$  y  $\text{PM}_{2.5}$  desde el NE al S. Las mayores concentraciones de  $\text{O}_3$  suceden en verano y primavera.  $\text{NO}$  y  $\text{NO}_2$  presentan los mayores niveles en invierno y primavera desde estos mismos sectores (**Figura 39**).



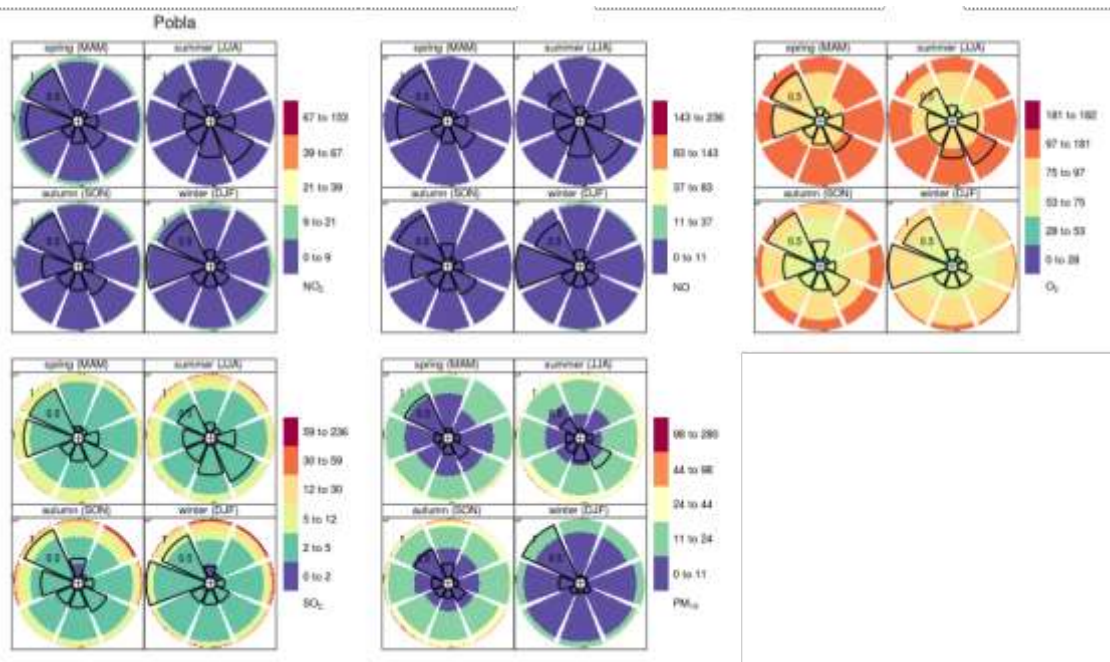
**Figura 39.** Rosas de contaminación en función de las direcciones del viento en la estación de monitoreo Peny. Los colores indican las concentraciones en  $\mu\text{g}\cdot\text{m}^{-3}$ . El tamaño de las categorías indica el % de la frecuencia del viento.

Para la estación Grau,  $\text{NO}$  y  $\text{NO}_2$  provienen de los mismos sectores con concentraciones máximas en primavera e invierno. Por su parte el  $\text{O}_3$  muestra concentraciones máximas del NE, E, SE, y S con concentraciones máximas desde el sector NE, E, SE y S, con máximos en primavera y verano.  $\text{PM}_{10}$  y  $\text{PM}_{2.5}$  proviene del sector N, NW, W y SW mayoritariamente en el invierno (**Figura 40**).



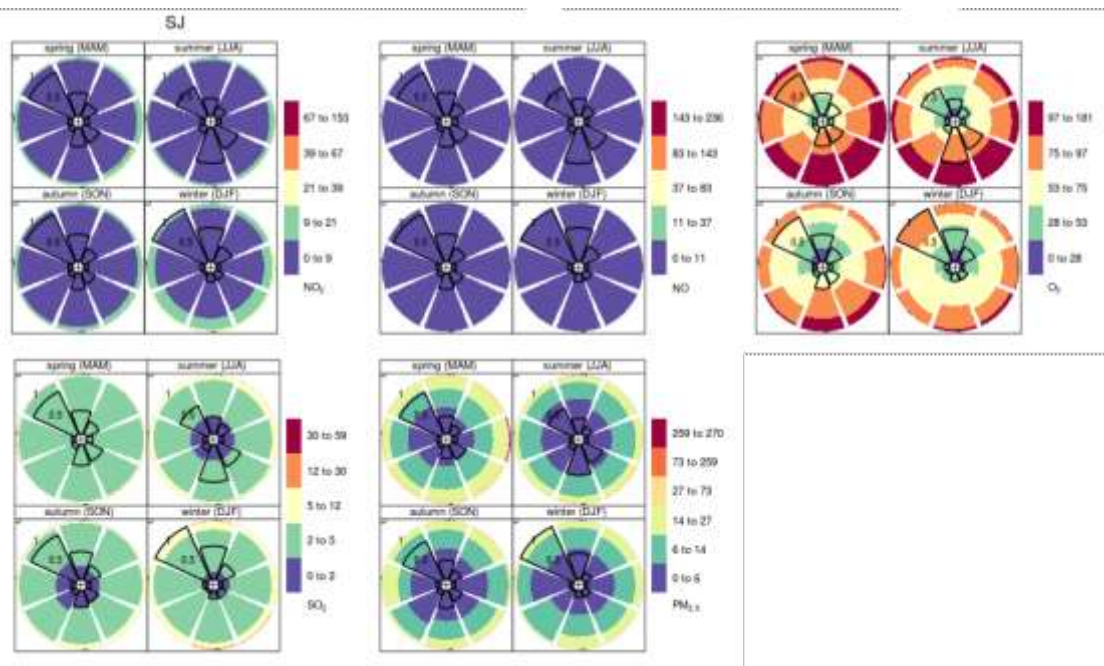
**Figura 40.** Rosas de contaminación en función de las direcciones del viento en la estación de monitoreo Grau. Los colores indican las concentraciones en  $\mu\text{g}\cdot\text{m}^{-3}$ . El tamaño de las categorías indica el % de la frecuencia del viento.

La estación Pobla presenta concentraciones bajas de  $\text{NO}$  y  $\text{NO}_2$  a lo largo de todo el año. sin embargo, las concentraciones de  $\text{O}_3$  varían, siendo mayores en diciembre y primavera.  $\text{PM}_{10}$  muestra mayores concentraciones en otoño y verano (**Figura 41**).



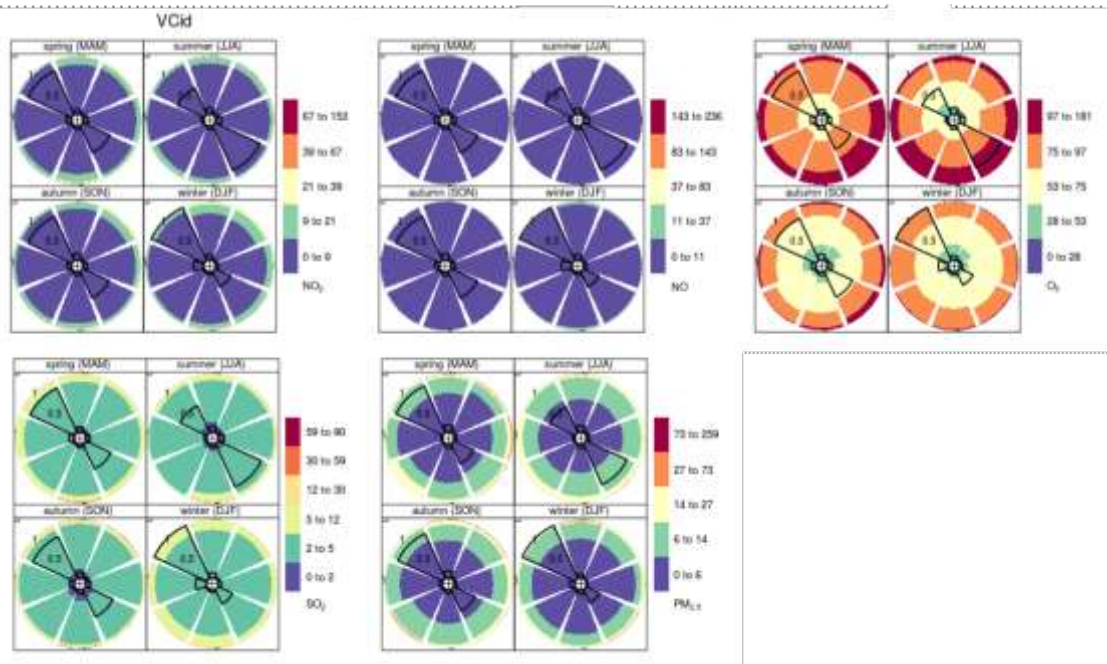
**Figura 41.** Rosas de contaminación en función de las direcciones del viento en la estación de monitoreo Pobla. Los colores indican las concentraciones en  $\mu\text{g}\cdot\text{m}^{-3}$ . El tamaño de las categorías indica el % de la frecuencia del viento. En esta estación no se monitorea el  $\text{PM}_{2.5}$ .

La estación SJ muestra valores bajos de NO y NO<sub>2</sub>, siendo mayores para este último en invierno desde el SE. Desde este mismo sector llegan las mayores concentraciones de O<sub>3</sub>, siendo marcadamente mayores en verano y primavera. Para PM<sub>2.5</sub> hubo un fuerte aporte desde el sector E en verano (**Figura 42**).



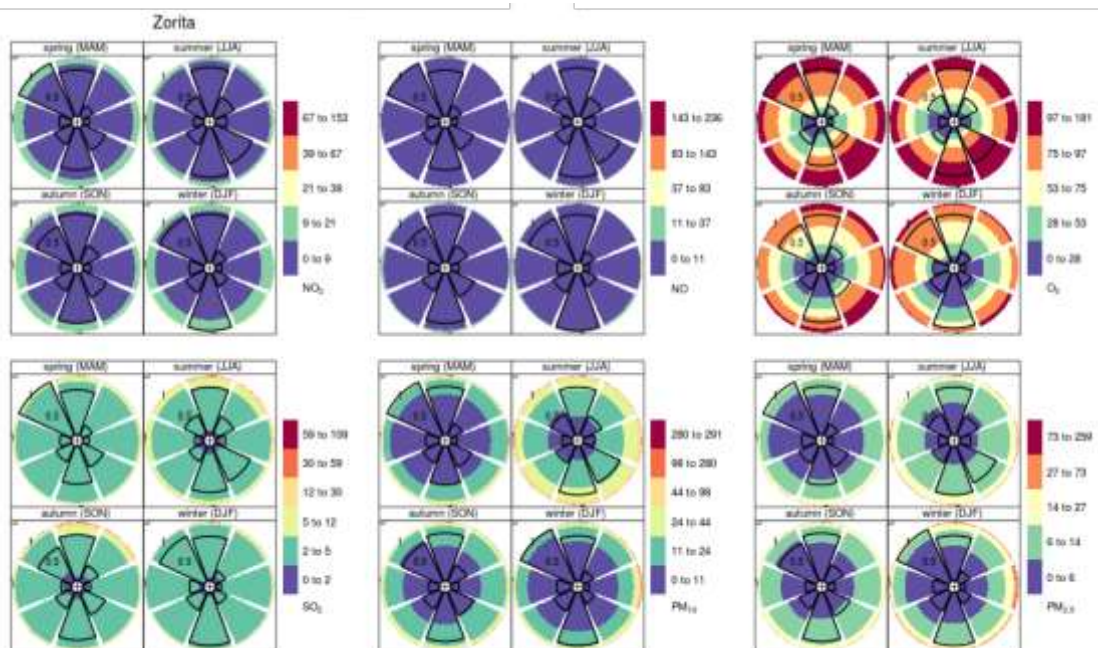
**Figura 42.** Rosas de contaminación en función de las direcciones del viento en la estación de monitoreo SJ. Los colores indican las concentraciones en µg·m<sup>-3</sup>. El tamaño de las categorías indica el % de la frecuencia del viento. En esta estación no se monitorea el PM<sub>10</sub>.

Estación VCid, NO y NO<sub>2</sub> concentraciones mayores en invierno. O<sub>3</sub> concentraciones mayores en verano y primavera, desde distintos sectores, descendiendo desde el E siguiendo el sentido del reloj. Para PM<sub>2.5</sub> se observa concentraciones altas desde el SW en otoño y desde el SE en invierno con poca frecuencia. Para SO<sub>2</sub>, concentraciones altas con poca frecuencia desde el sector N (**Figura 43**).



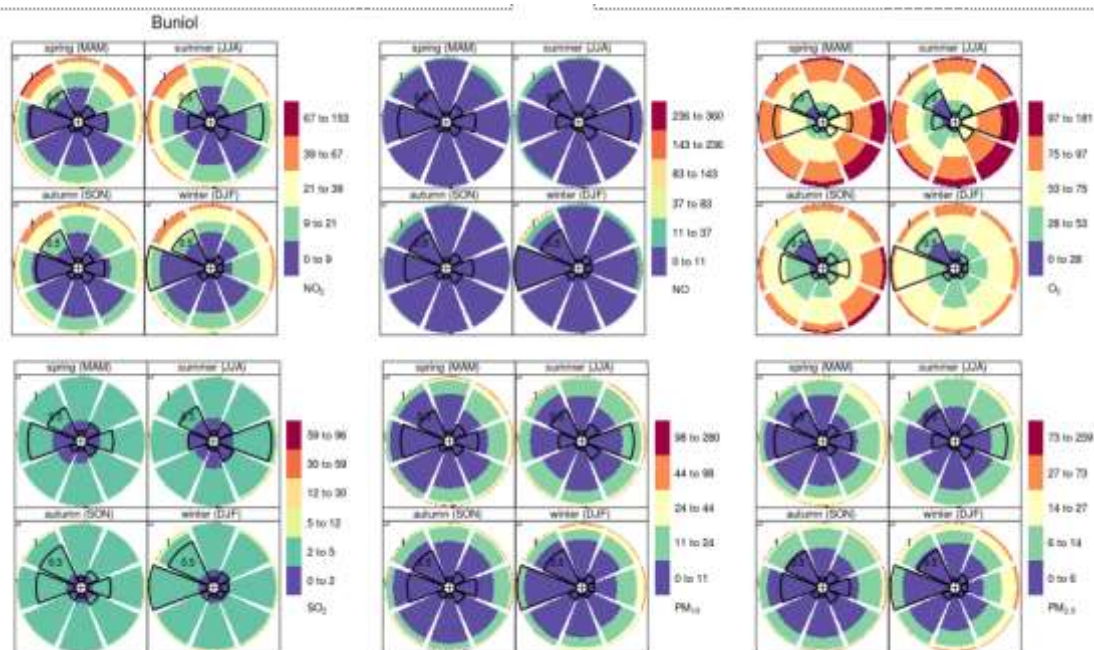
**Figura 43.** Rosas de contaminación en función de las direcciones del viento en la estación de monitoreo VCid. Los colores indican las concentraciones en  $\mu\text{g}\cdot\text{m}^{-3}$ . El tamaño de las categorías indica el % de la frecuencia del viento. En esta estación no se monitorea el  $\text{PM}_{10}$ .

Estación Zorita, altas concentraciones en primavera y verano, desde distintos sectores.  $\text{PM}_{2.5}$  y  $\text{PM}_{10}$  muestra un aumento de la frecuencia de las concentraciones elevadas en verano (**Figura 44**).



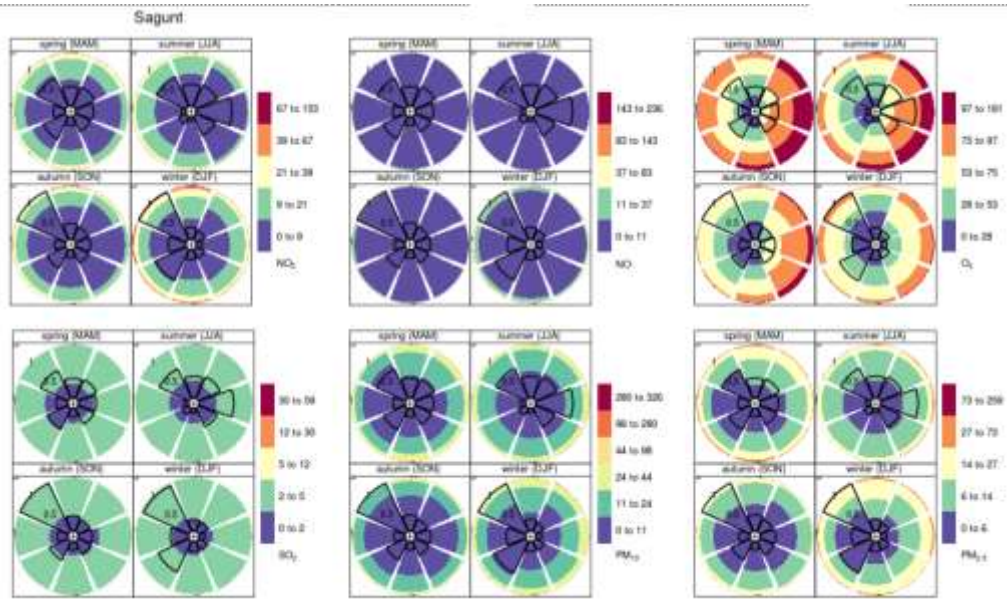
**Figura 44.** Rosas de contaminación en función de las direcciones del viento en la estación de monitoreo Zorita. Los colores indican las concentraciones en  $\mu\text{g}\cdot\text{m}^{-3}$ . El tamaño de las categorías indica el % de la frecuencia del viento.

La estación Buniol recibe el mayor aporte de NO y NO<sub>2</sub> desde el sector NW. Ocurriendo en invierno para el NO y en primavera para el NO<sub>2</sub>. El sector que más aporte O<sub>3</sub> es el E, SE y S, siendo mayor en verano y primavera. Esta diferencia puede indicar una fuente de emisión extra de O<sub>3</sub> respecto al NO<sub>2</sub>. PM<sub>10</sub> y PM<sub>2.5</sub> provienen mayoritariamente de los sectores del N- NE y S, con mayores concentraciones y con mayor frecuencia en invierno (**Figura 45**).



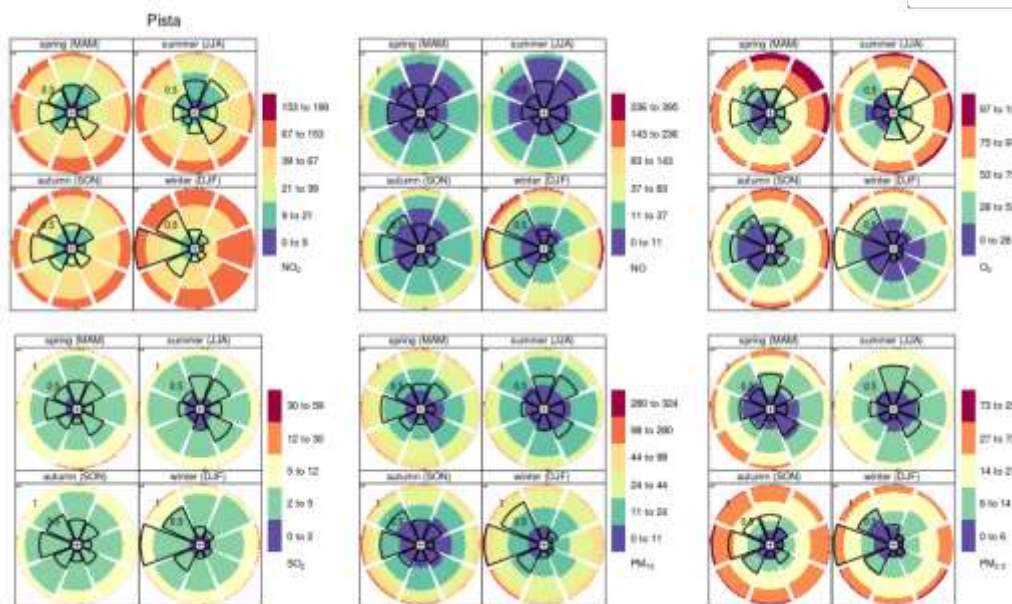
**Figura 45.** Rosas de contaminación en función de las direcciones del viento en la estación de monitoreo Buniol. Los colores indican las concentraciones en  $\mu\text{g}\cdot\text{m}^{-3}$ . El tamaño de las categorías indica el % de la frecuencia del viento.

En la estación Sagunt, el NO y NO<sub>2</sub> presenta una mayor frecuencia de concentraciones altas en invierno desde el N, W, SW y S. Las concentraciones mayores de O<sub>3</sub> provienen del NE, E, SE y S, con niveles más altos en primavera, verano y otoño (**Figura 46**).



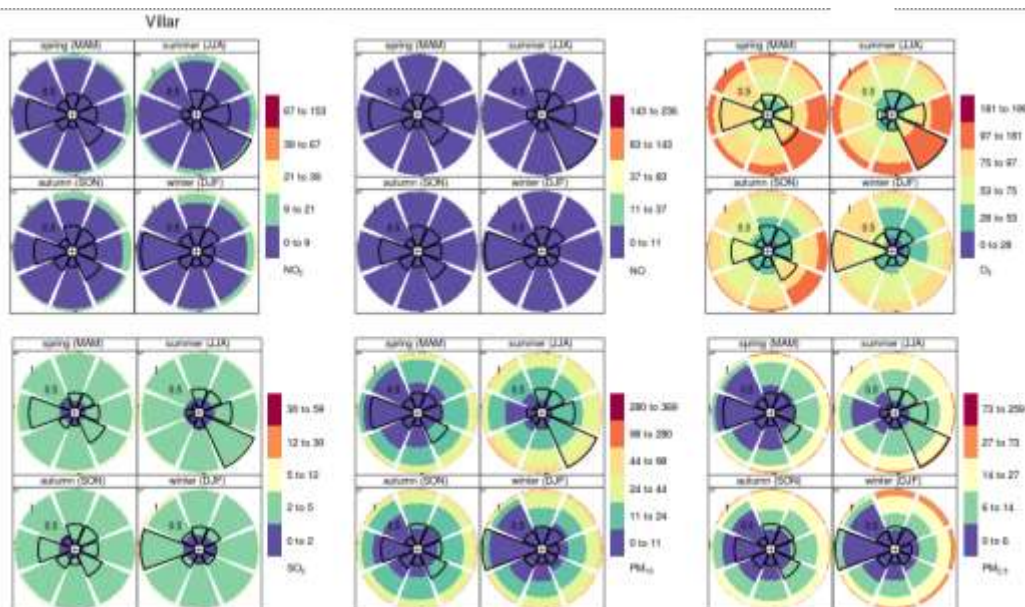
**Figura 46.** Rosas de contaminación en función de las direcciones del viento en la estación de monitoreo Sagunt. Los colores indican las concentraciones en µg.m<sup>-3</sup>. El tamaño de las categorías indica el % de la frecuencia del viento.

A la estación Pista llegan altos niveles de NO desde el sector E, N, NW, W y SW principalmente durante el invierno. Esto no es así con el NO<sub>2</sub> que mantiene una distribución más equitativa entre las distintas direcciones de viento. Para el O<sub>3</sub> se observa un fuerte aporte desde el sector N, NE, E, SE y S en especial en primavera y verano. Las concentraciones más elevadas de PM<sub>2.5</sub> y PM<sub>10</sub> ocurren en otoño - invierno desde el W, SW, S, SE y E (**Figura 47**).



**Figura 47.** Rosas de contaminación en función de las direcciones del viento en la estación de monitoreo Pista. Los colores indican las concentraciones en µg.m<sup>-3</sup>. El tamaño de las categorías indica el % de la frecuencia del viento.

En la estación Villar las concentraciones de NO<sub>2</sub> mayores provienen del primer cuadrante y menos el segundo. Nuevamente se ve un desfase con el O<sub>3</sub>, donde el sector que más aporta es el segundo cuadrante, sobre todo en primavera y verano. Durante todo el año los sectores que menos aportan PM<sub>10</sub> y PM<sub>2.5</sub> son el W y el NW (**Figura 48**).



**Figura 48.** Rosas de contaminación en función de las direcciones del viento en la estación de monitoreo Villar. Los colores indican las concentraciones en  $\mu\text{g}\cdot\text{m}^{-3}$ . El tamaño de las categorías indica el % de la frecuencia del viento.

### VII. Análisis exploratorio de los datos

Estudio de los niveles históricos medidos de las estaciones de monitoreo

Para conocer las concentraciones históricas de los contaminantes, se realizó un estudio exploratorio de las mediciones a lo largo de todo el período de tiempo analizado. La **Tabla 17** muestra los estadísticos descriptivos de los niveles medidos en las 21 estaciones para los contaminantes presentados en este trabajo.



**Tabla 17.** Estadísticos descriptivos para las estaciones de la Comunidad Valenciana (media, mínimo y máximo de las concentraciones horarias). Todos los valores se expresan en  $\mu\text{g}\cdot\text{m}^{-3}$ 

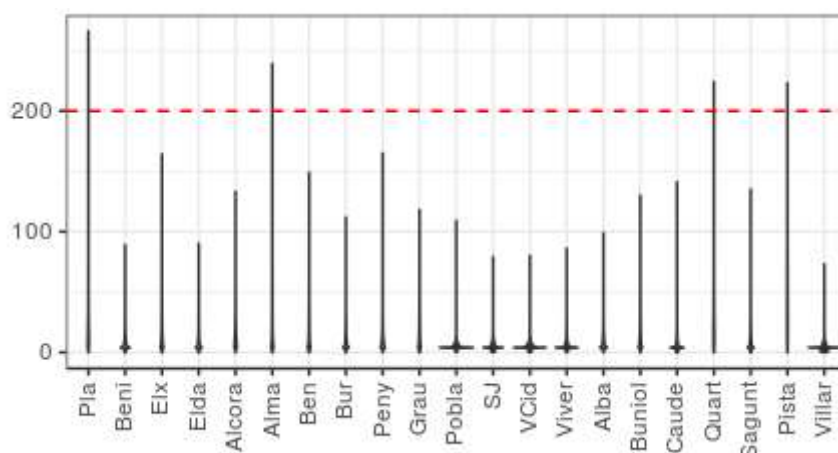
Estación	NO	NO <sub>2</sub>	O <sub>3</sub>	SO <sub>2</sub>	PM <sub>10</sub>	PM <sub>2.5</sub>
<i>Alba</i>	2.34 (0.33 - 11.67)	9.09 (0.33 - 58.17)	61.71 (3.12 - 121.88)	2.41 (0.33 - 9.04)	-	7.98 (0.78 - 44.92)
<i>Alcora</i>	6.8 (0.33 - 48.17)	13.95 (0.62 - 60.79)	56.39 (7.92 - 110.96)	4.17 (0.33 - 22.71)	-	8.9 (0.94 - 43.46)
<i>Alma</i>	15.16 (0.33 - 156.92)	20.44 (1.15 - 87.96)	-	6.33 (0.33 - 48.46)	17.18 (1.44 - 81.83)	9.37 (1.08 - 56.21)
<i>Ben</i>	7.46 (0.43 - 72.21)	15.22 (0.39 - 62)	-	3.4 (0.33 - 15.54)	10.43 (0.56 - 88.29)	7.26 (0.56 - 44.79)
<i>Beni</i>	3.11 (0.33 - 17.62)	7.46 (0.39 - 29.38)	78.91 (29.46 - 132.83)	-	-	-
<i>Buniol</i>	3.72 (0.33 - 33.38)	14 (0.82 - 60.42)	61.17 (12.08 - 113.75)	2.65 (0.33 - 14.29)	12.27 (0.56 - 97.96)	7.97 (0.56 - 59.08)
<i>Bur</i>	4.47 (0.33 - 33.21)	10.2 (0.33 - 38.33)	50.71 (3.71 - 107.12)	3.28 (0.33 - 13.62)	-	-
<i>Caude</i>	2.53 (0.33 - 14.38)	7.19 (0.33 - 36.58)	69.2 (12.25 - 125.88)	3.05 (0.33 - 16.17)	14.9 (1 - 123.21)	9.81 (0.93 - 51.38)
<i>Elda</i>	2.8 (0.33 - 35.33)	9.51 (0.33 - 52.88)	61.37 (10.79 - 106.83)	2.62 (0.33 - 11.75)	14.39 (0.81 - 118.79)	10.04 (0.75 - 52.75)
<i>Elx</i>	4.14 (0.33 - 37.92)	13.59 (0.51 - 56.08)	64.91 (7.12 - 122.79)	2.9 (0.33 - 9.08)	-	-
<i>Grau</i>	7.89 (0.39 - 91.67)	14.27 (1 - 56.88)	52.86 (4.42 - 112.08)	3.86 (0.33 - 16.75)	-	-
<i>Moli</i>	-	-	50.04 (3.38 - 105.88)	3.18 (0.33 - 12)	16.53 (1.83 - 73.08)	13.3 (0.68 - 71.46)
<i>Peny</i>	2.71 (0.33 - 28.26)	10.11 (0.33 - 46.92)	70.48 (12.65 - 134.88)	-	12.65 (0.39 - 85.58)	9.04 (0.36 - 58.08)
<i>Pista</i>	25.07 (0.85 - 154.42)	40.51 (2.22 - 120.96)	43.87 (2.96 - 101.12)	-	24.59 (2.42 - 143.92)	-
<i>Pla</i>	10.36 (0.51 - 128.04)	23.1 (0.39 - 98.58)	57.04 (5.46 - 116.62)	3.24 (0.33 - 11.96)	-	-
<i>Pobla</i>	2.28 (0.33 - 13.12)	5.03 (0.33 - 18)	84.7 (31.96 - 142.38)	4.86 (0.33 - 54.71)	-	-
<i>Poli</i>	-	-	54.77 (5.12 - 116.58)	3.15 (0.33 - 16.85)	18.35 (1.25 - 110.33)	11.74 (1 - 65.5)
<i>Quart</i>	9.39 (0.33 - 84.08)	25.94 (0.97 - 94.29)	50.65 (3.96 - 105.29)	3.73 (0.33 - 14.5)	-	-
<i>Raba</i>	-	-	62.77 (12.79 - 114.04)	2.82 (0.33 - 13.08)	14.97 (1.1 - 163.5)	7.76 (0.86 - 49.58)

<i>Sagunt</i>	3.47 (0.33 - 65.79)	10.03 (0.33 - 52.58)	53.36 (5.33 - 105.79)	2.56 (0.36 - 8.33)	14.21 (1.04 - 102.46)	8.77 (0.94 - 37.29)
<i>SJ</i>	2.26 (0.33 - 13.67)	5.42 (0.33 - 22)	70.02 (19.46 - 118.25)	-	-	-
<i>VCid</i>	2.63 (0.33 - 17)	5.28 (0.33 - 24.04)	75.21 (28.67 - 128)	3.57 (0.33 - 21.71)	-	-
<i>Villar</i>	2.05 (0.33 - 10.21)	4.61 (0.33 - 19.83)	72.82 (24.5 - 117.12)	2.78 (0.33 - 15.71)	16.01 (0.83 - 126.83)	9.28 (0.53 - 49.21)
<i>Viver</i>	2.05 (0.03 - 11.04)	6.62 (0.53 - 27)	65.88 (16.12 - 115.62)	3.02 (0.33 - 29.67)	13.7 (0.39 - 197.67)	7.97 (0.33 - 55.38)
<i>Zorita</i>	-	-	63.78 (10.12 - 135.54)	3.63 (0.33 - 19.62)	13.85 (0.9 - 205.62)	7.66 (0.69 - 50.21)

Las estaciones de monitoreo que registran los niveles medios diarios más altos de NO y NO<sub>2</sub> para todo el período son Pista, Quart, Pla y Alma. Las estaciones Pistas, Quart y Pla fueron caracterizadas como uso de suelo artificial, mientras que la última como uso de suelo agrícola. En las zonas agrícolas, la presencia de este contaminante puede provenir del uso intensivo de maquinaria agraria y del uso de fertilizantes en el suelo.

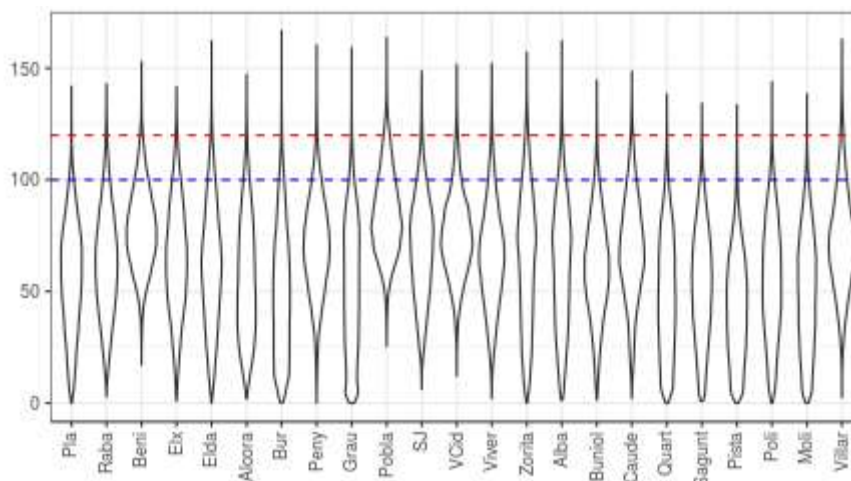
Las máximas concentraciones medias horarias para NO<sub>2</sub> se observaron en Pista (120.96 µg.m<sup>-3</sup>), Pla (98.58 µg.m<sup>-3</sup>), Quart (94.29 µg.m<sup>-3</sup>) y Alma (87.96 µg.m<sup>-3</sup>). Para NO, las máximas concentraciones medias horarias fueron de 156.92 µg.m<sup>-3</sup> para Alma, 154.42 µg.m<sup>-3</sup> para Pista, 142.83 µg.m<sup>-3</sup> para Quart y de 128.04 µg.m<sup>-3</sup> para Pla.

La **Figura 49** muestra la distribución estadística de las mediciones de NO<sub>2</sub> en cada una de las estaciones. Se puede ver que, si bien la media es baja, en todas las estaciones se presentaron eventos extremos, y en algunos casos, superando concentraciones límites establecidas por la normativa.



**Figura 49.** Gráficos violín para el conjunto de mediciones horarias para NO<sub>2</sub> del conjunto del período para las distintas estaciones. La línea punteada roja representa el límite horario que establece la UE de 200 µg.m<sup>-3</sup>

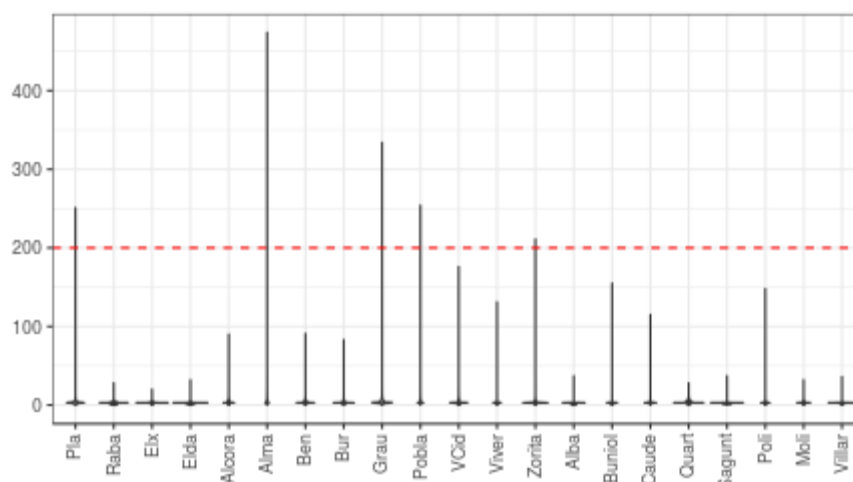
Las medias octohorarias más elevadas de  $O_3$  fueron observadas en las estaciones Pobla, Beni, VCid y Villar (**Figura 50**). Resulta llamativo que tanto Pobla como VCid tienen uso de suelo como bosque y mixto. Los valores elevados de  $O_3$  en estas áreas pueden deberse a las reacciones fotoquímicas donde intervienen compuestos orgánicos volátiles naturales, como las resinas de las coníferas [Rasmussen, 1972].



**Figura 50.** Gráfico violín para las medias móviles octohorarias de las mediciones de  $O_3$  para todas las estaciones en el período de estudio. En azul el límite recomendado por la OMS de  $100 \mu\text{g}\cdot\text{m}^{-3}$  y en rojo el límite para protección de la salud de la UE de  $120 \mu\text{g}\cdot\text{m}^{-3}$ .

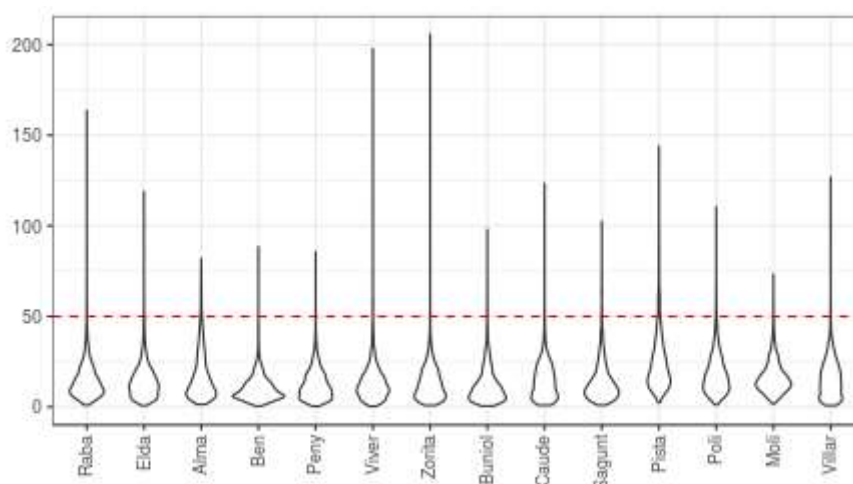
Las medias diarias más altas para el  $SO_2$  se observaron en las estaciones de Alma, Pobla y Alcora. Las máximas concentraciones medias diarias se registraron en Pobla ( $54.71 \mu\text{g}\cdot\text{m}^{-3}$ ), Alma ( $48.46 \mu\text{g}\cdot\text{m}^{-3}$ ) y Viver ( $29.67 \mu\text{g}\cdot\text{m}^{-3}$ ).

Como puede verse en la **Figura 51**, las concentraciones horarias se ubicaron dentro de un rango muy similar entre las estaciones. Sin embargo, se presentaron eventos extremos por encima de los  $200 \mu\text{g}\cdot\text{m}^{-3}$  en las estaciones Alma, Grau y Pobla, estaciones ubicadas en zonas con actividad industriales.



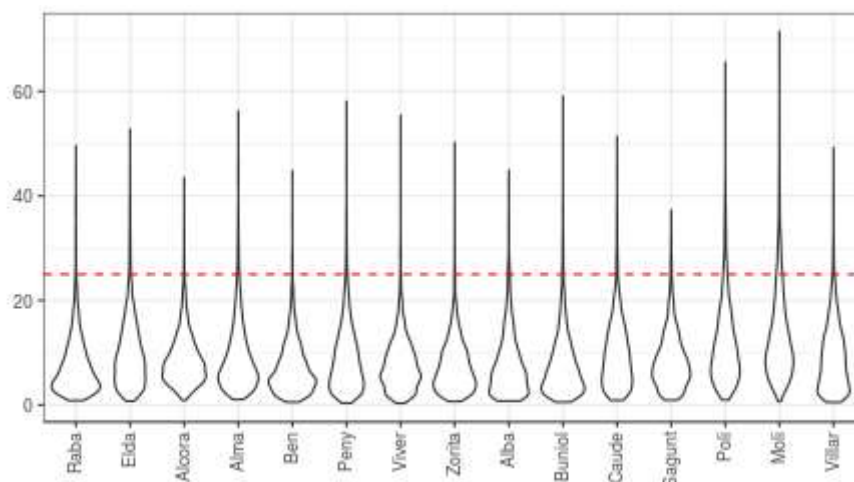
**Figura 51.** Mediciones horarias de SO<sub>2</sub> para el período 2009-2018 en las estaciones de la Generalitat Valenciana. La línea punteada indica el valor límite horarios de 200 µg.m<sup>-3</sup>, que no debe superarse en más de 24 ocasiones al año.

Las concentraciones medias diarias más elevadas de PM<sub>10</sub> se ubicaron en Pista, Poli y Alma (**Figura 52**). Los mayores valores diarios se registraron principalmente durante: el 22 y 23 de febrero del 2016 en Ben (88.29 µg.m<sup>-3</sup>), Sagunt (102.46 µg.m<sup>-3</sup>), Poli (110.33 µg.m<sup>-3</sup>), Villar (126.83 µg.m<sup>-3</sup>), Viver (197.67 µg.m<sup>-3</sup>) y Raba (163.50 µg.m<sup>-3</sup>), el 18 y 19 de octubre del 2012 en Zorita (205.62 µg.m<sup>-3</sup>) y Elda ( 118.79 µg.m<sup>-3</sup>), y 20 de marzo del 2010 en la estación Peny (85.58 µg.m<sup>-3</sup>) y Buniol (97.96 µg.m<sup>-3</sup>).



**Figura 52.** Mediciones diarias de PM<sub>10</sub> para las estaciones de la Generalitat en el período de estudio. La línea punteada roja indica el valor límite diario de 50 µg.m<sup>-3</sup>

Las concentraciones medias diarias mayores para PM<sub>2.5</sub> se dieron en Moli (71.46 µg.m<sup>-3</sup>), Poli (65.50 µg.m<sup>-3</sup>) y Buniol (59.08 µg.m<sup>-3</sup>), las cuales son también las estaciones con mayor amplitud de rango (Moli: 70.78 µg.m<sup>-3</sup>, Poli: 64.50 µg.m<sup>-3</sup>, Buniol: 58.52 µg.m<sup>-3</sup>) (**Figura 53**).



**Figura 53.** Mediciones diarias de PM<sub>2.5</sub> para las estaciones de la Generalitat en el período estudiado. La línea punteada roja indica el valor límite diario de 25 µg.m<sup>-3</sup>

Estudio del riesgo de los niveles históricos medidos de las estaciones de monitoreo

A continuación, se evalúa el cumplimiento de la normativa de la UE y las recomendaciones de la OMS.

La UE ha establecido un valor límite anual y otro horario para el NO<sub>2</sub> [Directiva 2008/50/CE]. El valor límite anual establecido por la normativa es de 40 µg.m<sup>-3</sup>, considerado el valor máximo compatible con una adecuada protección de la salud. El valor límite horario es de 200 µg.m<sup>-3</sup>, que no debe superarse más de 18 veces al año. Ambos valores límite coinciden con los recomendados por la OMS [Krzyzanowski & Cohen, 2008].

El análisis de las mediciones de NO<sub>2</sub> muestra que el límite anual se superó únicamente en la estación Pista. Los niveles de NO<sub>2</sub> en la estación Pista superaron el límite anual reiteradamente: 2008 (41,77 µg.m<sup>-3</sup>), 2009 (48,05 µg.m<sup>-3</sup>), 2010 (44,54 µg.m<sup>-3</sup>), 2013 (44,57 µg.m<sup>-3</sup>), 2014 (45,52 µg.m<sup>-3</sup>), 2015 (40,75 µg.m<sup>-3</sup>) y 2016 (42,46 µg.m<sup>-3</sup>).

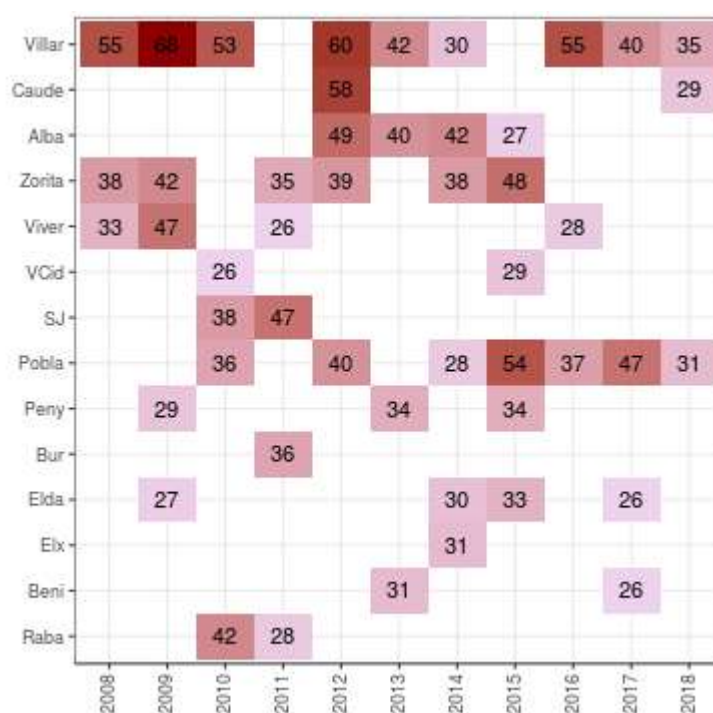
El valor límite medio horario de 200 µg.m<sup>-3</sup> se ha superado varias veces para las estaciones dentro del período, pero en ningún caso más de 18 veces por año (**Figura 51**). La estación Pla registró 7 mediciones por encima del límite en el mes de diciembre del 2014, alcanzando un máximo de 266 µg.m<sup>-3</sup>. Ninguna otra estación registró un evento en ese mismo periodo. La estación Alma midió por una única vez un evento en 2015, y luego 13 mediciones muy elevadas, con un máximo de 239 µg.m<sup>-3</sup>, durante el invierno de 2017. Hechos aislados ocurrieron también en Quart y Pista durante 2009 y 2016 para el primero y en 2009 para el segundo.

El valor objetivo de O<sub>3</sub> para la protección de la salud humana definido por la EU es de 120 µg.m<sup>-3</sup>, que no debe superarse en períodos de ocho horas (valor máximo diario de las medias móviles octohorarias) en más de 25 días al año, computados en períodos trienales. Por su parte, la OMS recomienda no sobrepasar los 100 µg.m<sup>-3</sup> en períodos de ocho horas (límite octohorario). A diferencia de la normativa de la EU no establece un máximo de ocasiones que puede sobrepasarse

este valor recomendado durante un año, ni define realizar un promedio trienal del cómputo de las superaciones.

Para comparar con la normativa, se generaron series de tiempo octohorarias para el O<sub>3</sub>, aplicando la media móvil para cada una de las estaciones. Analizando los resultados, se observó que todas las estaciones superaron en algún momento del periodo analizado el valor objetivo de la EU y el promedio diario recomendado por la OMS (**Figura 50**).

La **Figura 54** muestra la cantidad de días al año en que se superó el límite octohorario de 120 µg.m<sup>-3</sup> de O<sub>3</sub> por más de 25 veces al año. 14 estaciones de las 23 estaciones que miden O<sub>3</sub> superaron el límite legal alguna vez en los años estudiados. Las estaciones que más años superaron este límite fueron las estaciones Villar (8 años), Pobra (7 años) y Zorita (5 años). También, se puede observar que son menos las estaciones que muestran excedencias desde 2016.

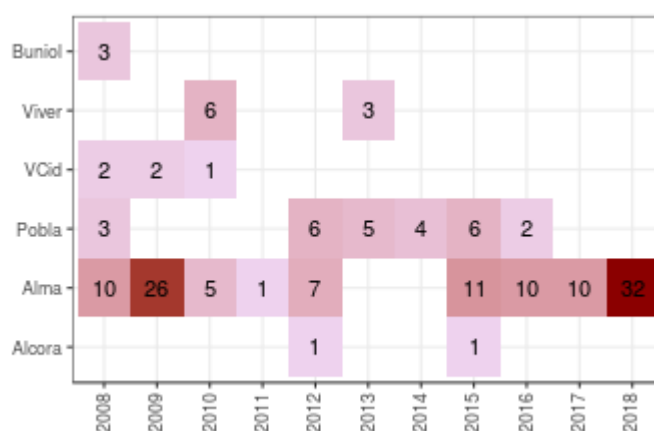


**Figura 54.** Cantidad de días al año en que las estaciones superaron el límite octohorario para el O<sub>3</sub> de 120 µg.m<sup>-3</sup>. Para simplificar la lectura del gráfico solo se muestran los años en que se superó más de 25 veces el límite establecido.

Respecto al SO<sub>2</sub>, la normativa de la EU establece un valor límite diario de 125 µg.m<sup>-3</sup>, el cual no debe superarse en más de 3 ocasiones al año. Asimismo, establece un valor límite horario de 200 µg.m<sup>-3</sup>, que no debe superarse en más de 18 ocasiones al año. Por su parte, la OMS fija como valor guía diario en 20 µg.m<sup>-3</sup> y una recomendación de no superar los 500 µg.m<sup>-3</sup> en 10 minutos. Estos valores guías no son acompañados por un número máximo de veces al año que pueden superarse estas recomendaciones, “puesto que si se respeta el nivel de 24 horas se garantizan unos niveles medios anuales bajos” [Organización Mundial de la Salud, 2006], de lo que se deduce que no debería superarse en ninguna ocasión.

Los límites diarios establecidos por la UE para el SO<sub>2</sub> no fueron superados en las estaciones estudiadas. Ningún día se alcanzó una concentración superior a 125 µg.m<sup>-3</sup> y los eventos horarios en los que se superaron los 200 µg.m<sup>-3</sup> de SO<sub>2</sub> fueron menos de 18 ocasiones al año. Los límites horarios fueron superados por la estación Alma (2 veces durante el 2009), la estación Grau (una vez durante el 2015) y por la estación Pobra en distintas ocasiones en los años 2011, 2012, 2013 y 2014.

Considerando los valores recomendados por la OMS, se superó el límite horario en Alcora, Alma, Pobra, VCid, Viver y Buniol. La estación Alma exhibe el mayor número de excesos a lo largo de los 10 años (**Figura 55**).



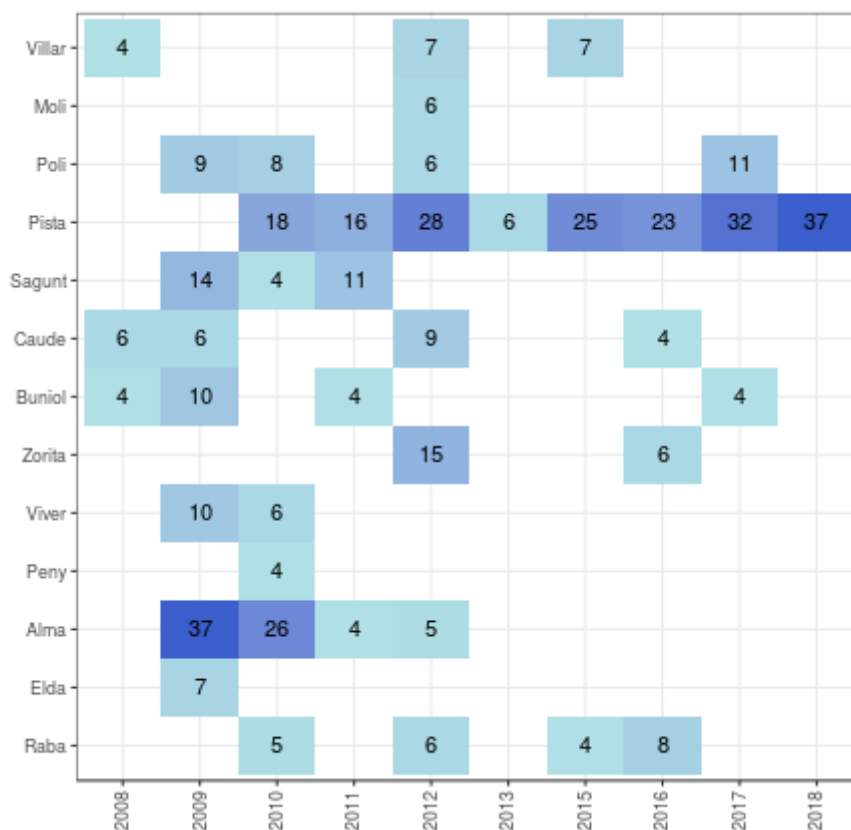
**Figura 55.** Cantidad de días al año en que las estaciones superaron el límite horario para el SO<sub>2</sub> de 20 µg.m<sup>-3</sup>

Para el PM<sub>10</sub>, al igual que con el SO<sub>2</sub>, existe una gran diferencia entre el valor reglamentado y las recomendaciones de la OMS. La anterior legislación (Directiva 1999/30/CE y Real Decreto 1073/2002) fijaba dos etapas para alcanzar de forma escalonada los valores de la OMS. La Fase I, de obligado cumplimiento desde el año 2005, establecía un valor límite anual de 40 µg.m<sup>-3</sup> y un valor límite diario de 50 µg.m<sup>-3</sup>, que no debía superarse más de 35 días en todo el año. La Fase II que debía cumplirse a partir del año 2010, establecía un valor límite anual de 20 µg.m<sup>-3</sup>, y un valor límite diario también de 50 µg.m<sup>-3</sup> pero que no debía superarse más de 7 días al año (la OMS recomienda no superar en más de 3 ocasiones).

En la proximidad del vencimiento del plazo se renunció a implementar la Fase II de la Directiva 2008/50/CE, quedando como valores límite legales los establecidos en la Fase I, considerablemente más laxos. Es importante señalar este proceso, puesto que se reconoció la importancia de los valores recomendados por la OMS para garantizar la salud de la población, si bien estos niveles resultaron inalcanzables en la práctica.

Al analizar la base de datos se observó que el límite anual de 40 µg.m<sup>-3</sup> para PM<sub>10</sub> no fue superado por ninguna estación. Sin embargo, las estaciones Raba, Elda, Alma, Viver, Buniol, Sagunt, Pista, Poli y Moli han superado los 20 µg.m<sup>-3</sup> de PM<sub>10</sub> anuales recomendados por la OMS. La estación Pista supera este límite 9 años consecutivos. Esta estación, junto con Poli, fueron las únicas 2 estaciones que mostraron superar el valor recomendado durante los últimos 3 años de estudio.

El límite diario legal para PM<sub>10</sub> sólo es superado en las estaciones Pista y Alma más de 35 días al año en 2018 y 2009, respectivamente. En la **Figura 56** se observan las estaciones que han superado ese nivel más de 3 veces al año (cantidad recomendada por la OMS). Como puede verse, la estación Pista presenta la mayor cantidad de años con excedencias.



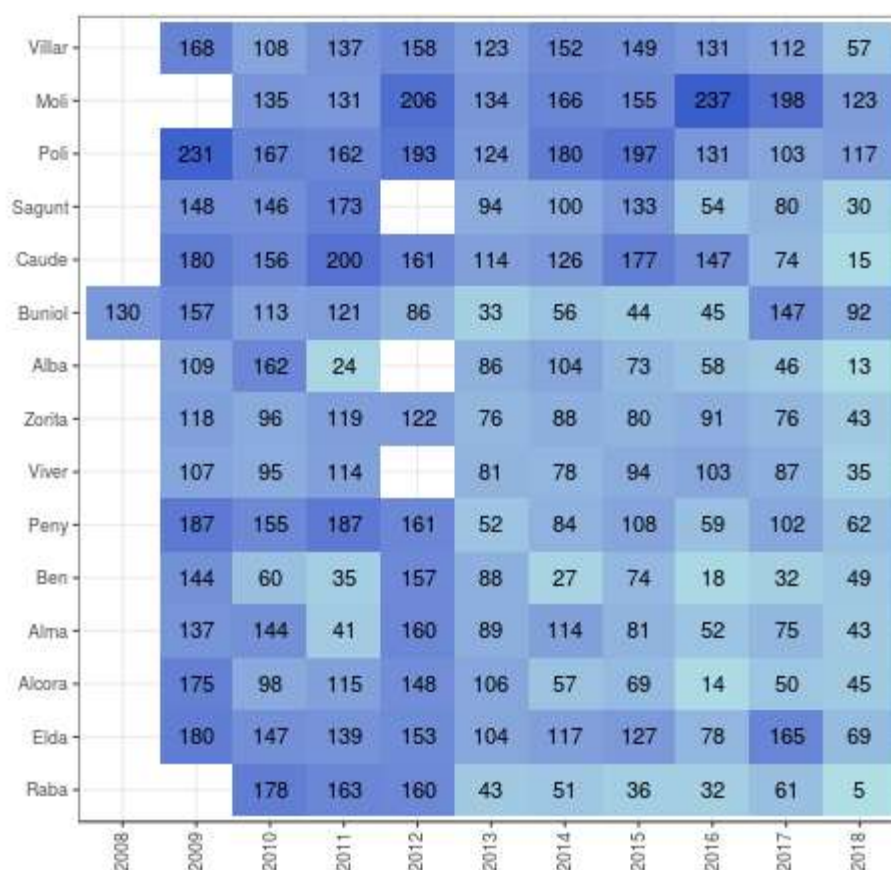
**Figura 56.** Cantidad de días al año en que las estaciones superaron el límite diario para PM<sub>10</sub> de 50 µg.m<sup>-3</sup>. Para simplificar la lectura del gráfico solo se muestran los años en que se superó más de 3 veces el límite establecido.

Por su parte, la normativa establece para el PM<sub>2.5</sub> un valor límite anual fijado en 25 µg.m<sup>-3</sup> para 2015, estando en vigor como valor objetivo desde 2010. A su vez, la Directiva 2008/50/CE introdujo objetivos adicionales para PM<sub>2.5</sub> dirigido a la exposición de la población. Estos objetivos se establecen a nivel nacional, y se determinan en función de la concentración media anual de PM<sub>2.5</sub> de 3 años promedio sobre las estaciones de monitoreo en aglomeraciones o áreas urbanas más grandes. Como concentración de exposición obligatoria se establece 20 µg.m<sup>-3</sup> desde 2015 (período 2013-2015) y se define una reducción porcentual de 0, 10, 15, 20% por año hasta alcanzar un nivel de 18 µg.m<sup>-3</sup> en 2020.

Las medias móviles anuales mayores de PM<sub>2.5</sub> se registraron durante el año 2010 y 2011 en la estación Alma (22.14 µg.m<sup>-3</sup> y 21.20 µg.m<sup>-3</sup>, respectivamente) y en la estación Moli para los años 2011 y 2012 (19.58 µg.m<sup>-3</sup> y 20.46 µg.m<sup>-3</sup>, respectivamente). Mostrando para el resto de las estaciones y años, niveles inferiores a 18 µg.m<sup>-3</sup>.



También con este contaminante los valores recomendados por la OMS se encuentran muy alejados de los establecidos por la Directiva. La OMS marca que el valor medio anual no debería sobrepasar los  $10 \mu\text{g}\cdot\text{m}^{-3}$ , además recomienda un máximo de 3 superaciones al año del valor recomendado diario de  $25 \mu\text{g}\cdot\text{m}^{-3}$ . Las recomendaciones anuales son superadas por todas las estaciones estudiadas, con excepción de la estación Zorita. En la **Figura 57** se puede observar claramente que todas las estaciones superan la recomendación diaria de la OMS en más de 3 ocasiones al año. Cabe mencionar la falta de mediciones para el año 2008 en la mayoría de las estaciones.



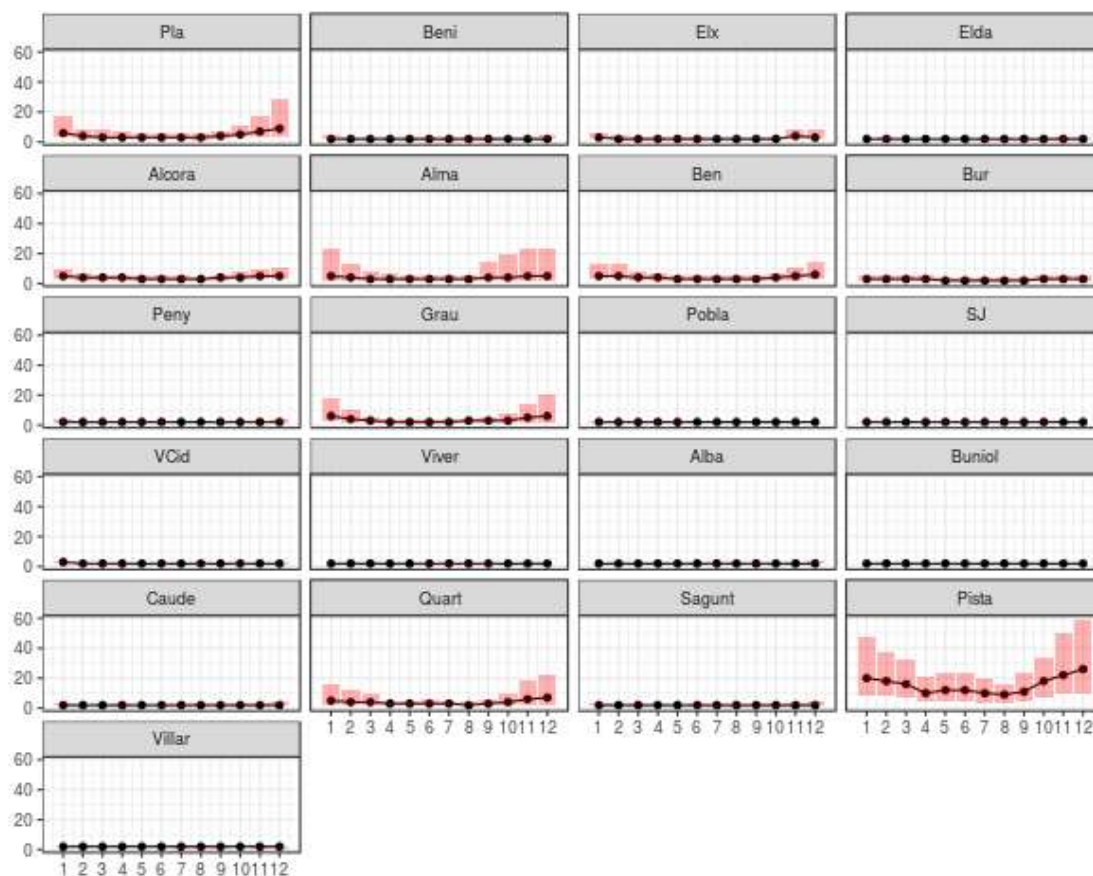
**Figura 57.** Cantidad de días al año en que las estaciones superaron el límite diario para  $\text{PM}_{2.5}$  de  $25 \mu\text{g}\cdot\text{m}^{-3}$ . Para simplificar la lectura del gráfico solo se muestran los años en que se superó más de 3 veces el límite establecido.

### Comportamiento anual

Identificar el comportamiento anual de los contaminantes permite predecir la ocurrencia de eventos extremos. Dado que la concentración de los contaminantes está sujeta a factores naturales como antrópicos, los comportamientos pueden variar de un sitio a otro. Esto hace fundamental conocer cada una de las situaciones particulares.

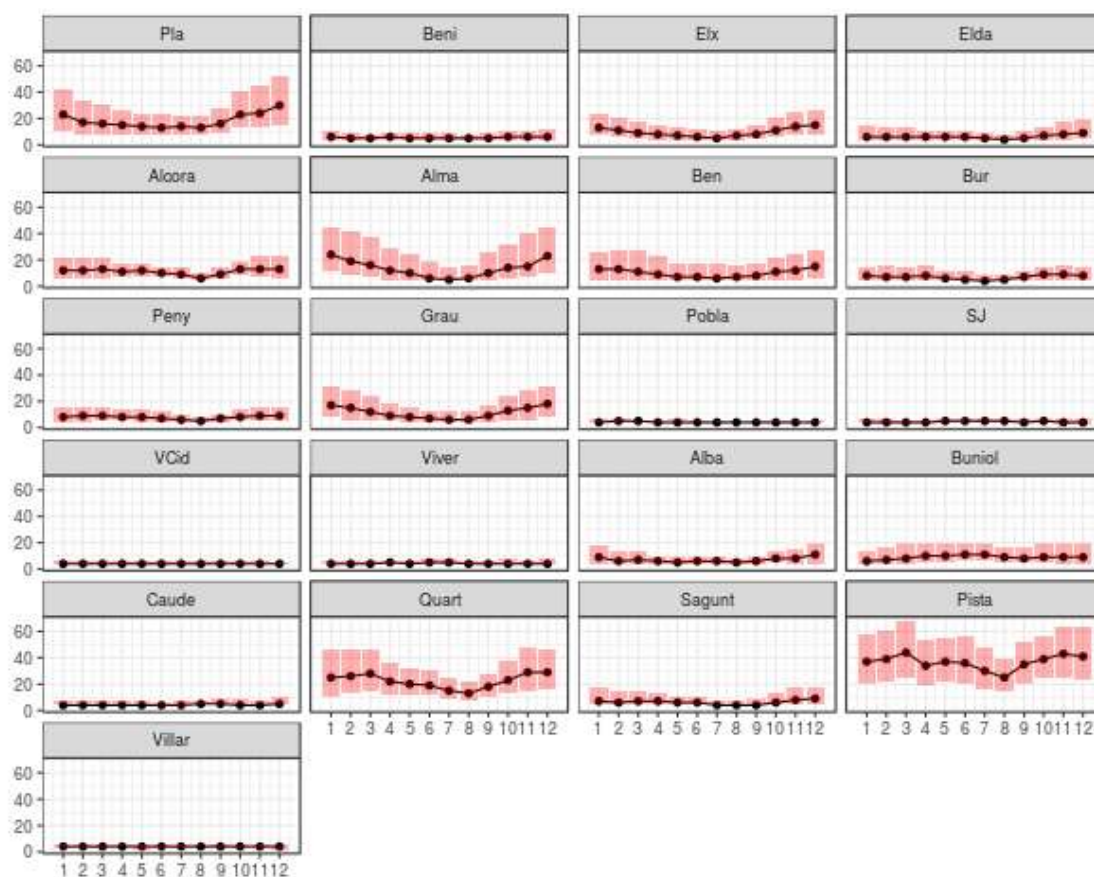
El comportamiento anual esperado para  $\text{NO}$  y  $\text{NO}_2$  es un aumento de las concentraciones en los meses fríos, dado que existe una mayor emisión por el aumento del uso de la calefacción y el

transporte. Esto puede verse de forma marcada en estaciones ubicadas en zonas urbanas o con usos múltiples que tienen un alto tránsito vehicular. Por ejemplo, las estaciones Pla, Alma, Ben, Grau, Quart y Pista presentan una marcada diferencia en sus niveles a lo largo de los meses. Por otra parte, las estaciones localizadas en zonas agrícolas o en áreas forestales no muestran grandes diferencias entre las concentraciones de NO en verano e invierno (**Figura 58**), lo cual puede atribuirse a los bajos niveles de NO y NO<sub>2</sub> que allí se presentan.



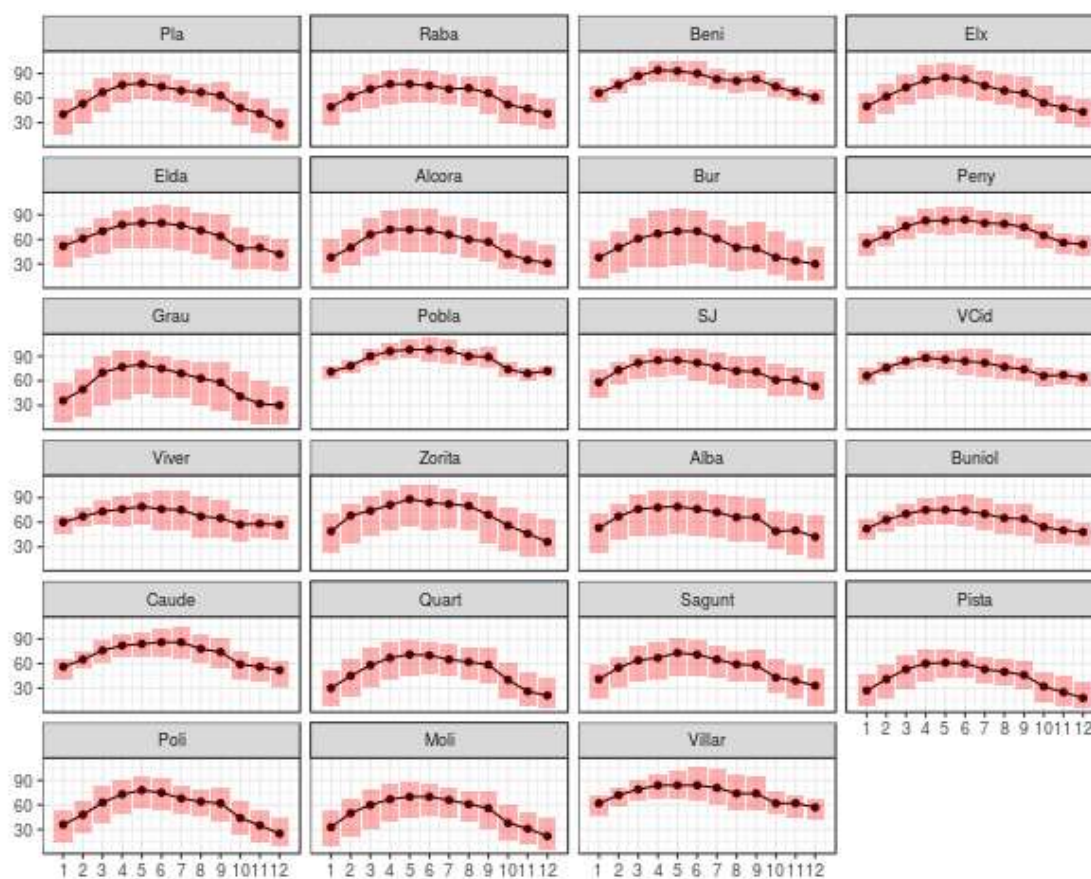
**Figura 58.** Mediciones agregadas mensualmente para el NO. El punto representa la mediana. La caja representa el rango intercuartil 0.25 - 0.75.

Las variaciones interanuales se ven acentuadas en el comportamiento del NO<sub>2</sub>. En la **Figura 59** se observa claramente que las estaciones con mayor proporción de áreas de bosques y menor proporción de área urbana muestran menores diferencias entre el primer y cuarto cuartil, como son las estaciones Pobla, VCid, Beni, SJ, Viver y Villar.



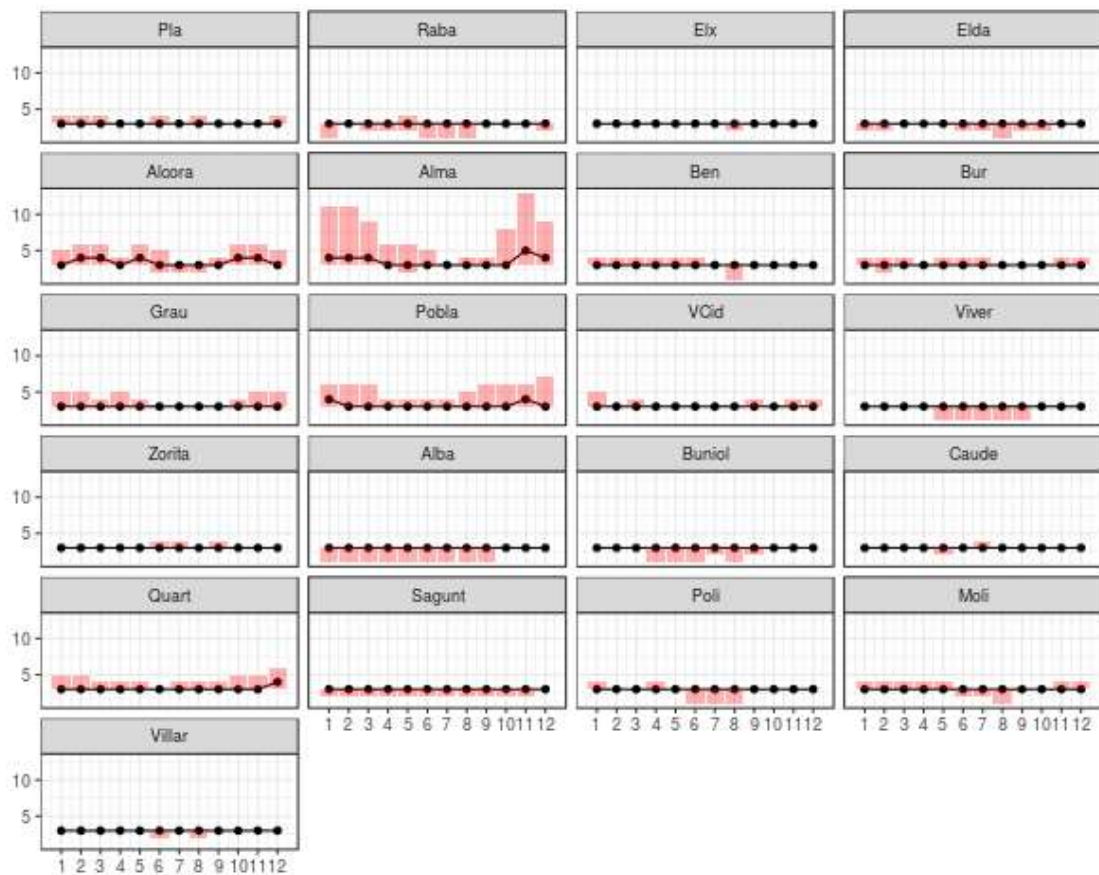
**Figura 59.** Mediciones agregadas mensualmente para el NO<sub>2</sub>. El punto representa la mediana. La caja representa el rango intercuartil 0.25 - 0.75.

Por su parte, el O<sub>3</sub> muestra una mayor independencia con las variaciones en el uso de suelo, presentando el mismo patrón interanual en todas las estaciones de monitoreo (**Figura 60**). Dado que es un contaminante secundario, cuya síntesis se ve favorecida por la radiación solar, su concentración aumenta en los meses de verano y disminuye en los meses de invierno [Bytnerowicz et al., 2007].



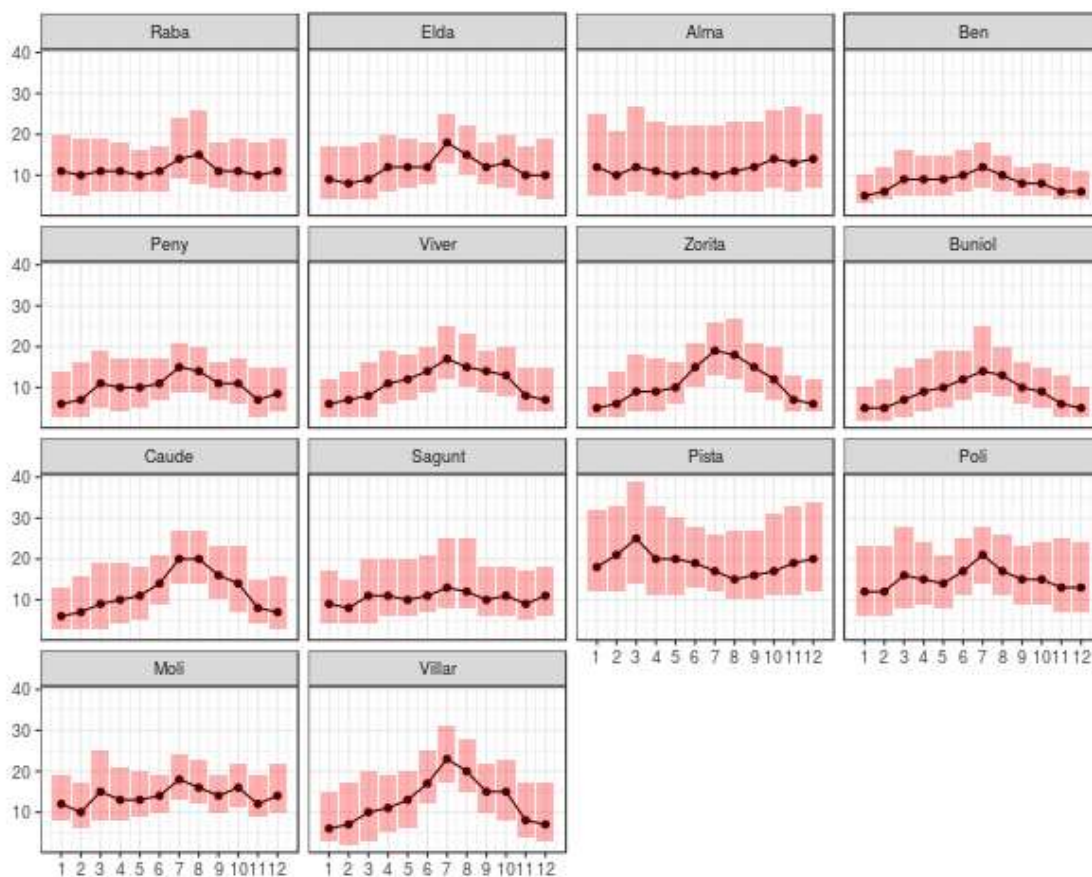
**Figura 60.** Mediciones agregadas mensualmente para el O<sub>3</sub>. El punto representa la mediana. La caja representa el rango intercuartil 0.25 - 0.75.

A diferencia de las anteriores especies químicas, el SO<sub>2</sub> no presenta un comportamiento estacional (**Figura 61**). Aunque sobresale la variación interanual que presentan las estaciones Alma, Grau y Pobla. Estas primeras dos son vecinas, mientras que la tercera, se encuentra emplazada en una zona seminatural.



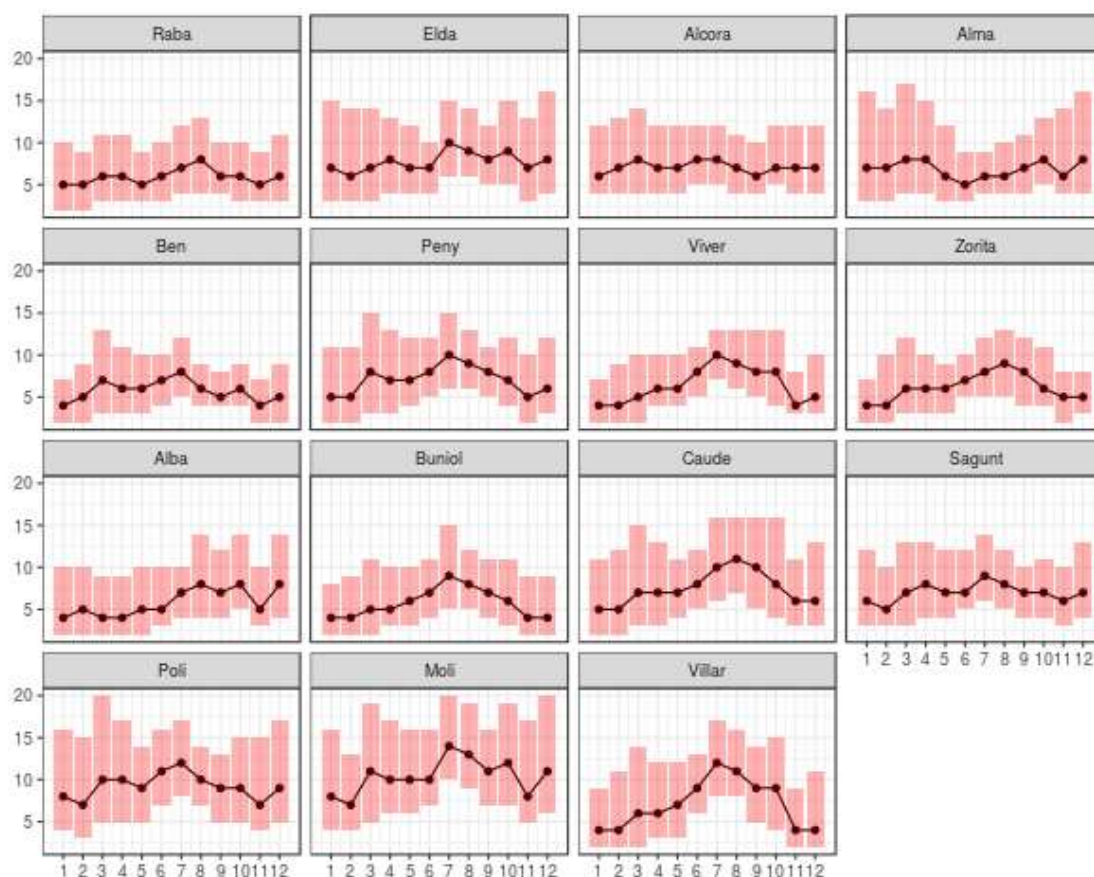
**Figura 61.** Mediciones agregadas mensualmente para el  $\text{SO}_2$ . El punto representa la mediana. La caja representa el rango intercuartil 0.25 - 0.75.

Los contaminantes  $\text{PM}_{10}$  y  $\text{PM}_{2.5}$  muestran un comportamiento estacional, con picos en los meses de verano y valles en invierno. Para el  $\text{PM}_{10}$ , las estaciones que no muestran un comportamiento claro estacional son Alma, Pista, Moli y Sagunt, estaciones emplazadas en áreas urbanas o con uso de suelo mixto. Mientras que las estaciones Zorita, Villar y Caude presentan las crestas más pronunciadas, indicando un notable aumento en los meses cálidos (**Figura 62**).



**Figura 62.** Mediciones agregadas mensualmente para el  $PM_{10}$ . El punto representa la mediana. La caja representa el rango intercuartil 0.25 - 0.75.

El  $PM_{2.5}$  no muestra grandes variaciones en las estaciones de Raba, Elda, Alcora, Alma, Alba y Sagunt (**Figura 63**). Resulta llamativo las diferencias en el rango intercuartil de los niveles de  $PM_{2.5}$  en la estación Alma, donde se observa una mayor dispersión de las mediciones en los meses fríos.



**Figura 63.** Mediciones agregadas mensualmente para el  $PM_{2.5}$ . El punto representa la mediana. La caja representa el rango intercuartil 0.25 - 0.75.

### Comportamiento semanal

La influencia de las fuentes antrópicas sobre los niveles de los distintos contaminantes se puede ver claramente en las gráficas de comportamiento semanal. Las siguientes gráficas representan la media del conjunto de las concentraciones registradas por cada una de las estaciones a lo largo de todo el período de estudio. Para los contaminantes fotoquímicos es común ver una oscilación diaria suave, correspondiente a síntesis o consumo de especies químicas en presencia del sol. Por su parte, la forma tipo silla junto a un cambio de comportamiento sobre el fin de semana se presenta cuando hay fuentes antrópicas y las concentraciones dependen de los horarios laborales. Por ejemplo, con cambios pronunciados en los horarios de entrada y salida del trabajo.

Esto se ve particularmente para  $NO$  y  $NO_2$ , dos contaminantes fuertemente asociados al tránsito vehicular. Las estaciones Pla, Pista, Quart y en menor grado la estación Alcora, muestran un máximo de  $NO$  pronunciado a comienzos del día y un segundo máximo a final de la tarde. Además, se observa un cambio en el patrón sobre el fin de semana (**Figura 64**). Aún mejor se observa en las gráficas de comportamiento del  $NO_2$  (**Figura 65**).

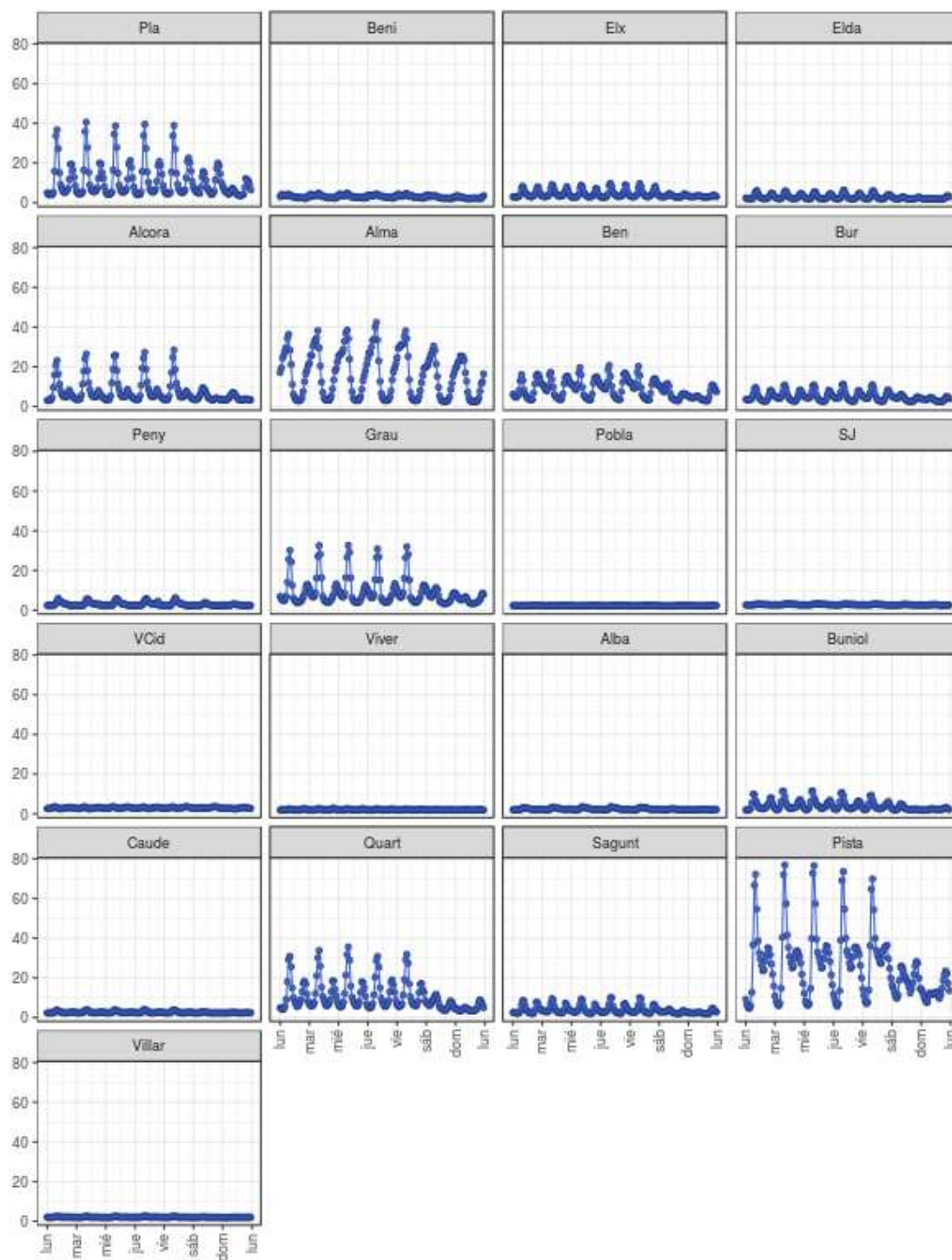


Figura 64. Mediciones agregadas horarias para el NO.



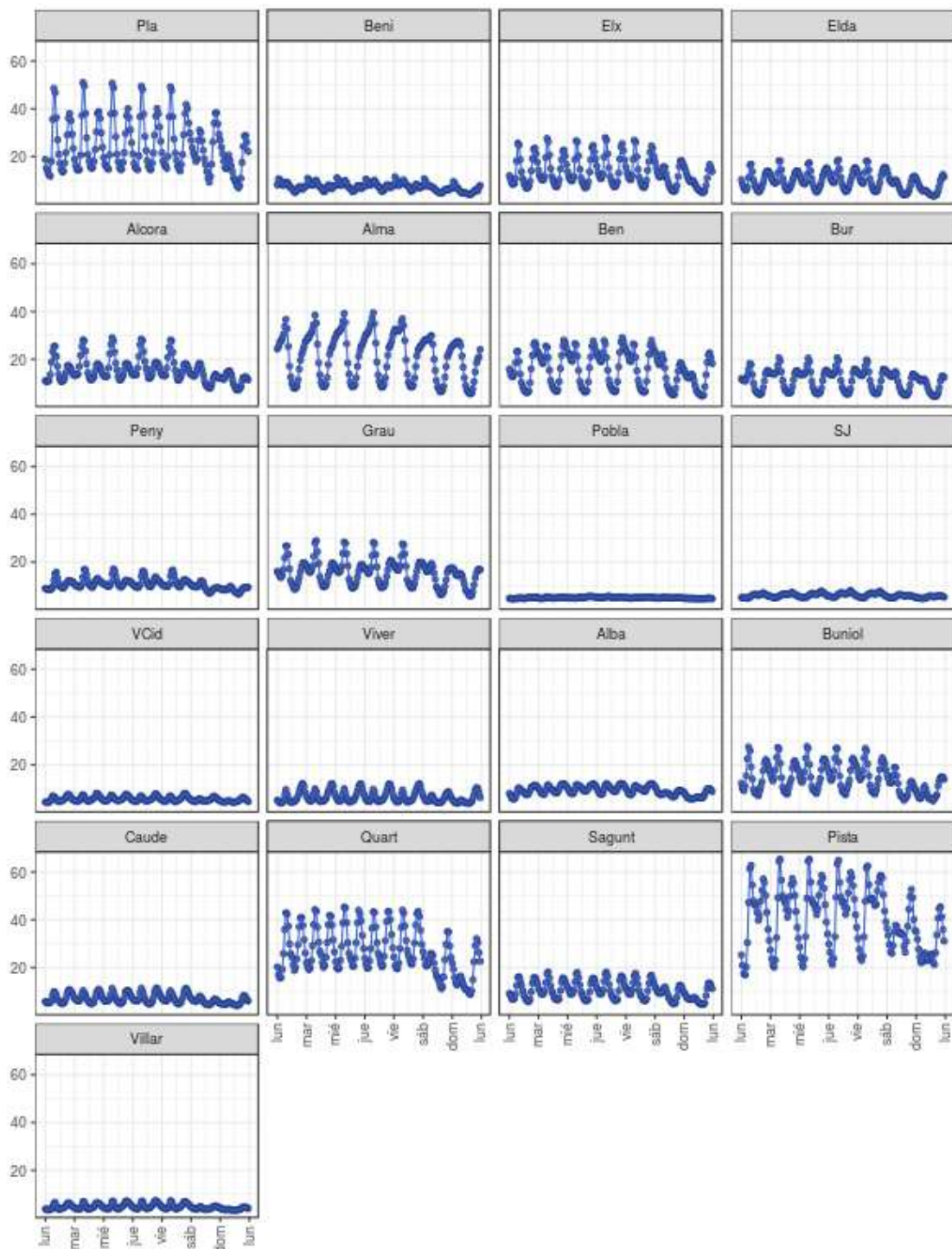
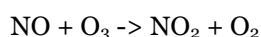
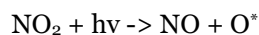


Figura 65. Mediciones agregadas horarias para el NO<sub>2</sub>.

El tiempo de vida medio del NO es más corto que el tiempo de vida del NO<sub>2</sub>, esto puede llevar a que los niveles de NO producidos lejos de una estación de monitoreo no lleguen a ser registrados por la misma, mientras que los niveles de NO<sub>2</sub> sí. Analizando la **Figura 65** se observa claramente la falta de tránsito entorno a las estaciones Beni, Pobra, SJ, VCid, Viver, Alba, Caude y Villar, estaciones con un alto porcentaje de uso destinado al bosque, zonas seminaturales y agricultura.

Como se ha indicado anteriormente, el O<sub>3</sub> junto con el NO y el NO<sub>2</sub> troposférico se encuentra en un estado fotoestacionario en presencia de la luz solar. Durante las horas diurnas, NO, NO<sub>2</sub> y O<sub>3</sub> típicamente se equilibran en la escala de tiempo de unos pocos minutos, a partir de las reacciones:



La concentración de ozono comienza a disminuir durante la mañana, después tiende a aumentar, alcanza su valor máximo durante las horas de media tarde y luego disminuye al mínimo después del atardecer.

En ausencia de una fuente de emisión, el ozono muestra un comportamiento oscilatorio acompañando el ciclo diurno. Este es el caso del comportamiento observado, por ejemplo, en las estaciones SJ, Caude o Zorita, emplazadas en un área seminatural (**Figura 66**).

En cambio, cuando existen altas emisiones de NO durante el día, se puede observar un pico desfasado del ciclo solar, dando un comportamiento en forma de silla con una mayor variación diaria de los niveles de O<sub>3</sub>. En la **Figura 66** sobresalen los valores mínimos que alcanzan las estaciones Pla, Pista, Quart y Alcora antes señaladas por presentar máximos de NO<sub>2</sub> y NO.

Analizando estas gráficas resulta llamativo que no se monitoree NO y NO<sub>2</sub> en las estaciones Poli y Moli, 2 estaciones ubicadas dentro del entramado urbano, que presentan un comportamiento de los niveles de O<sub>3</sub> indicativo de un marcado efecto del tránsito intenso.

Al analizar el comportamiento del SO<sub>2</sub> se observa un comportamiento típicamente antrópico para las estaciones Alcora, Grau, Quart y Alma (**Figura 67**). El principal aporte de SO<sub>2</sub> a la atmósfera es antrópico por la quema de hidrocarburos las cuales pueden provenir del tráfico vehicular o la actividad industrial. Esto es razonable con el hecho que Grau, Alcora y Alma se encuentran en una zona de producción de cerámicas y Quart en una zona de alto tránsito.

Por otra parte, no se observan grandes excedencias en los niveles en las estaciones de Zorita, Poblá y VCid que se encuentran en la zona de influencia de la Central Termoeléctrica de Andorra.

PM<sub>10</sub> tiene un comportamiento marcado antrópico en las estaciones Alma y Viver con picos máximos por la tarde y en Pista se observan picos máximos por la mañana (**Figura 68**). Esto podría estar relacionado con el flujo de vehículos en horarios claves, como son la salida de camiones o la entrada al trabajo.

Resulta llamativa la tendencia creciente que se observa para PM<sub>2.5</sub> (**Figura 69**) en las estaciones Poli, Moli, Villar y Viver ubicadas en áreas fuertemente influenciadas por las emisiones del tráfico e industrias. Esto puede ser interpretado como una acumulación de PM<sub>2.5</sub> en la atmósfera a lo largo de la semana que desciende el fin de semana cuando las emisiones decaen.

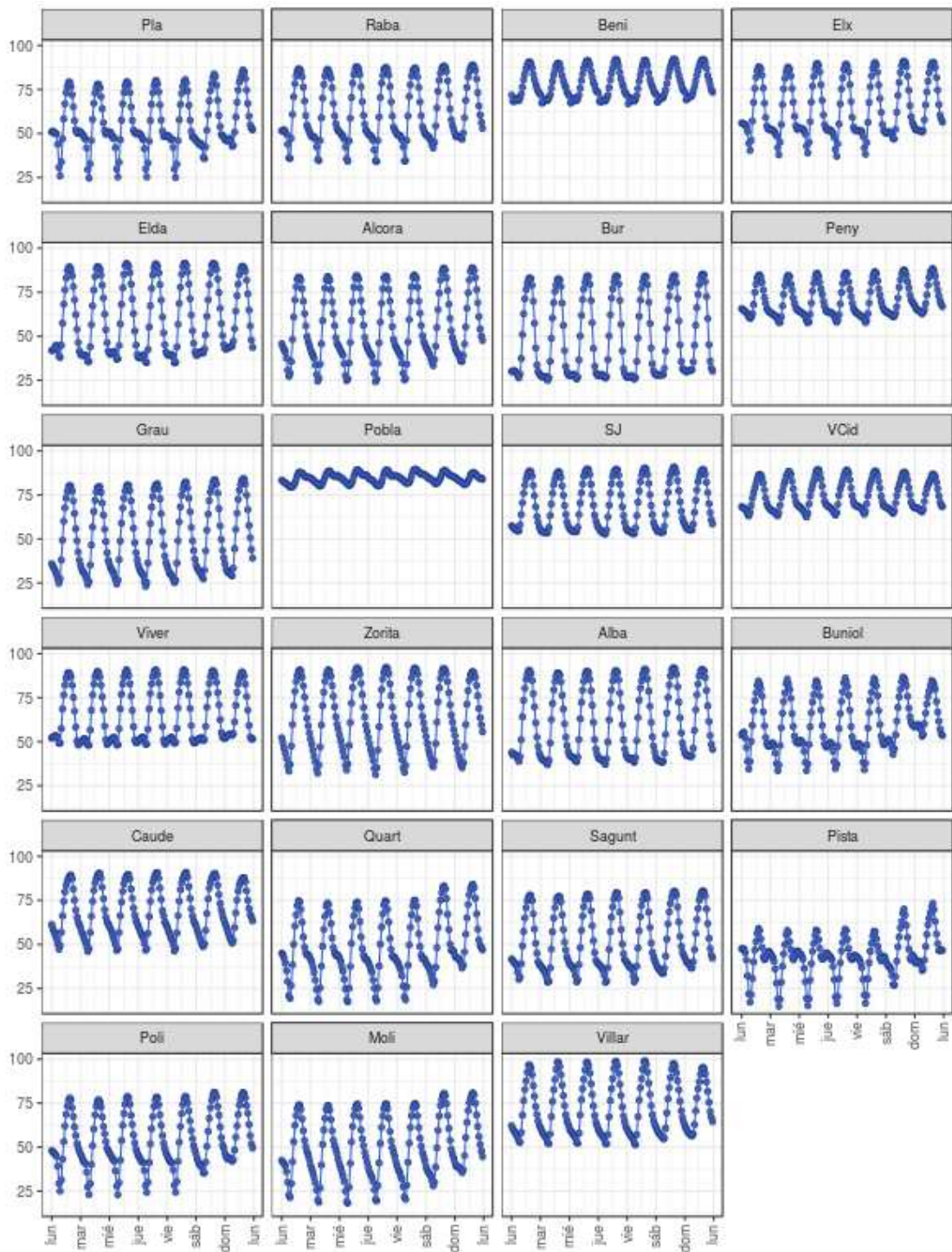


Figura 66. Mediciones agregadas horarias para el O<sub>3</sub>.

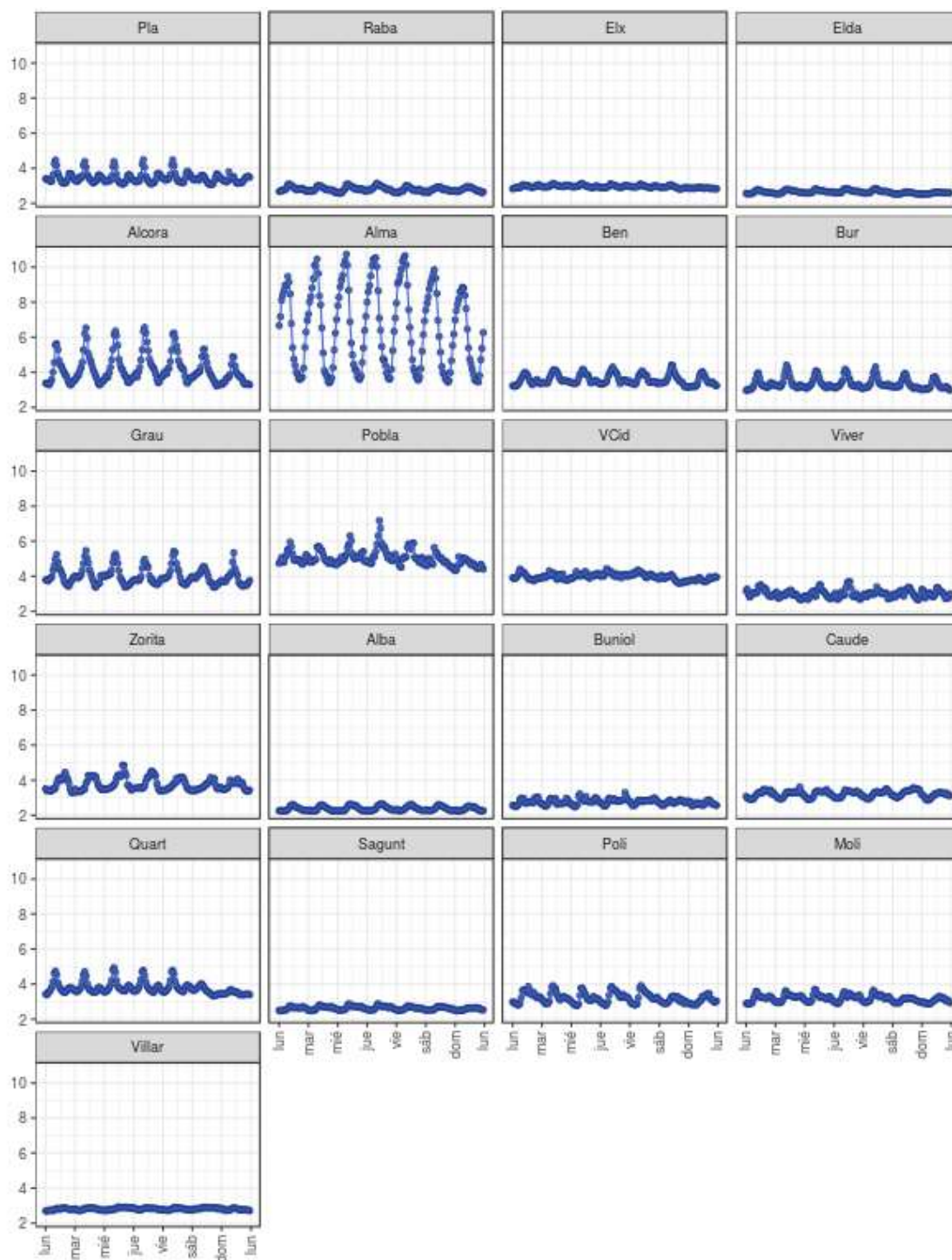


Figura 67. Mediciones agregadas horarias para el SO<sub>2</sub>.

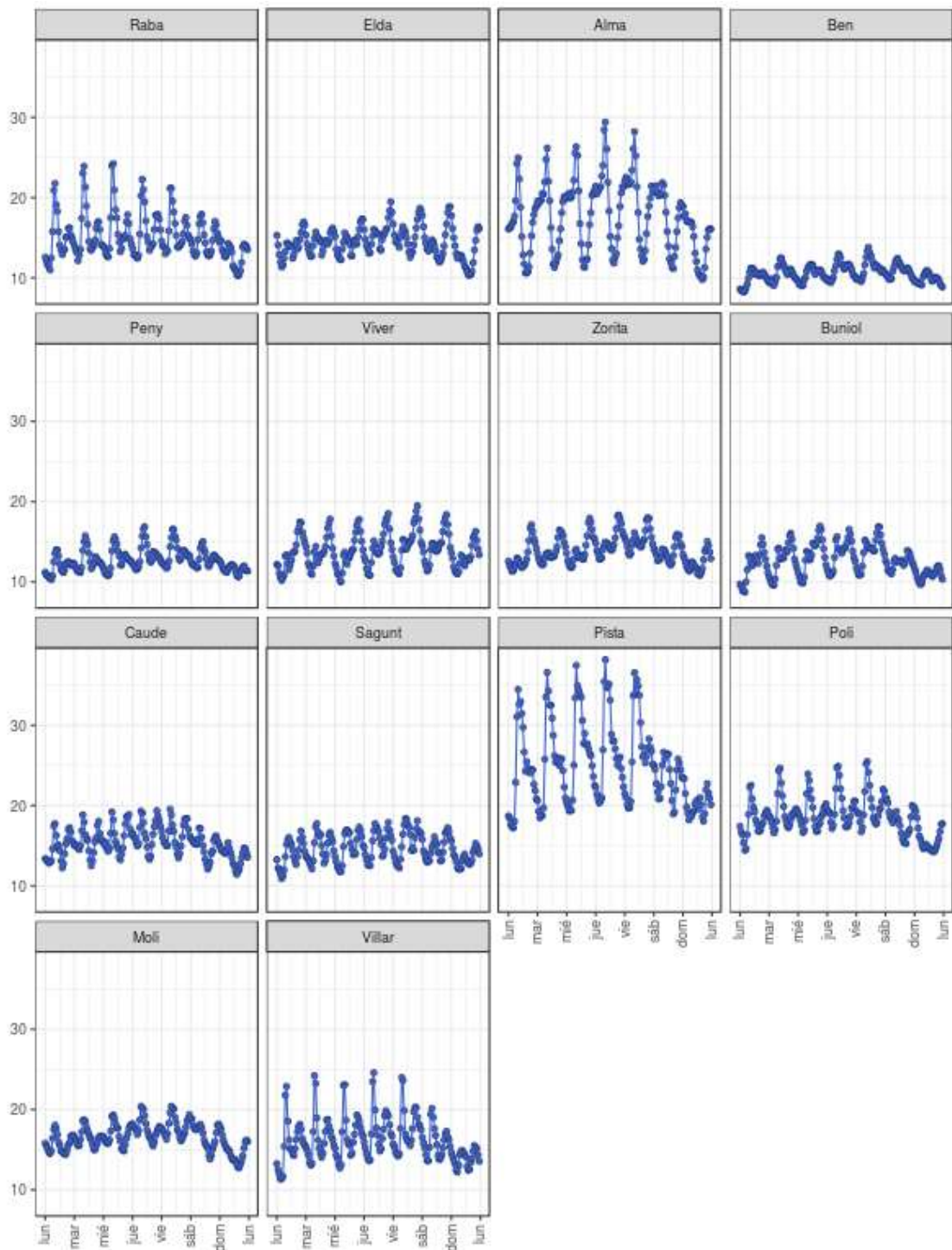


Figura 68. Mediciones agregadas horarias para el PM<sub>10</sub>.

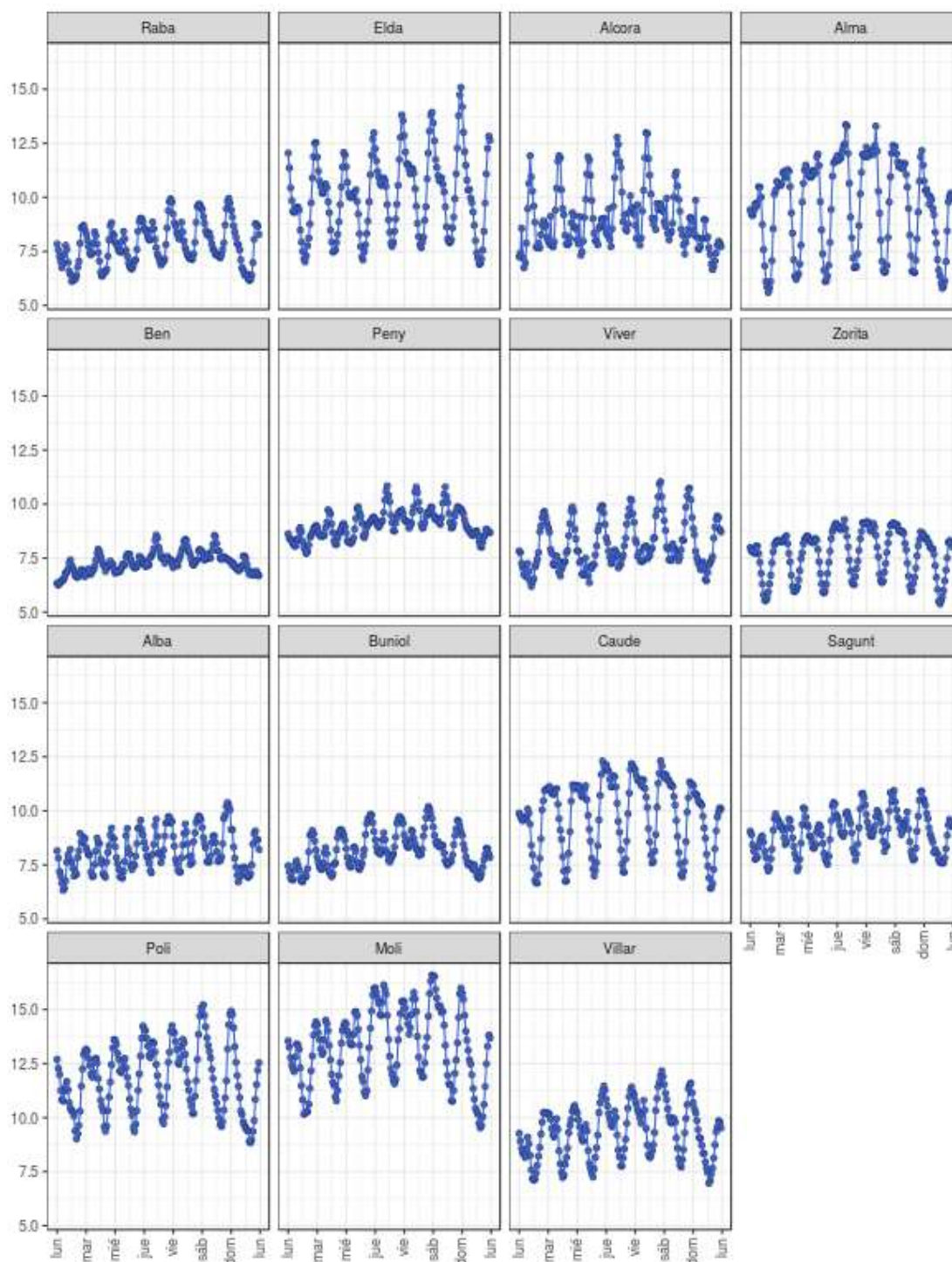


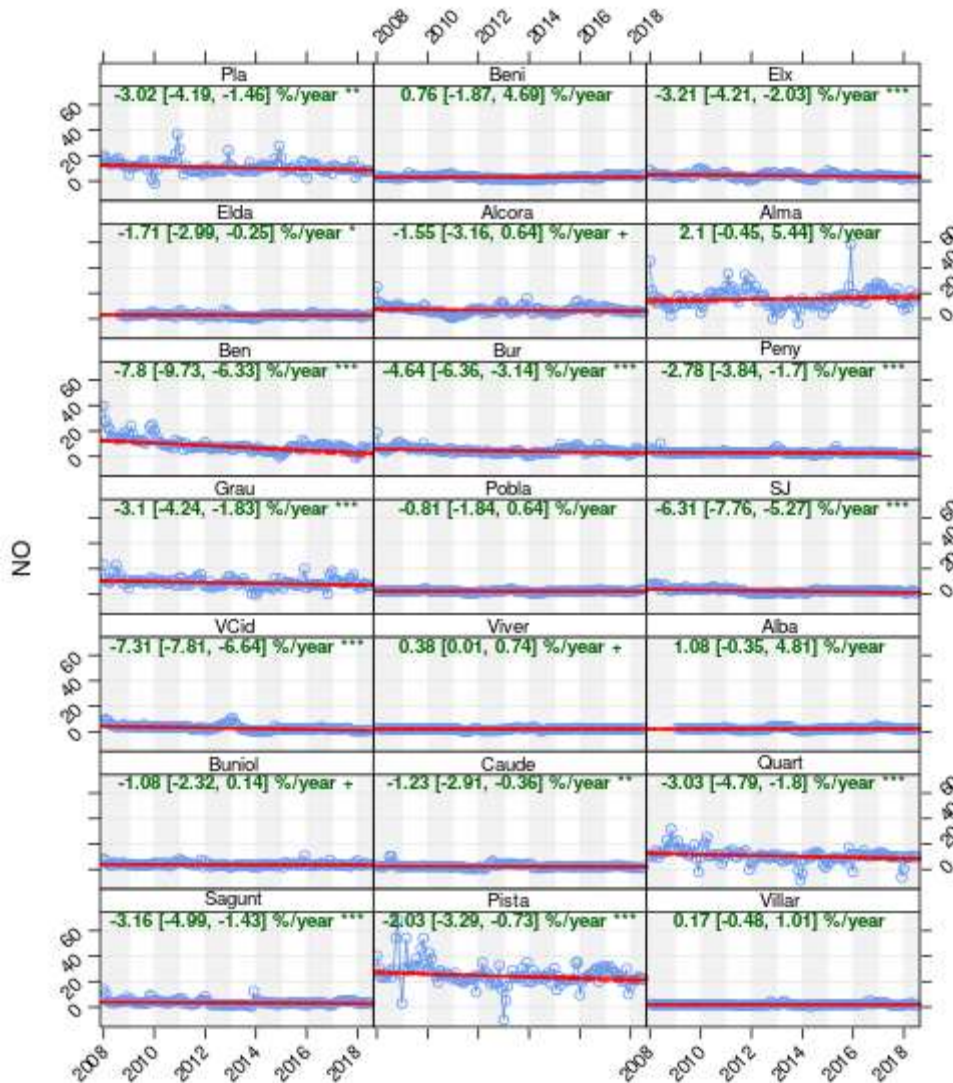
Figura 69. Mediciones agregadas horarias para el PM<sub>2.5</sub>.

### VIII. Evaluación de la tendencia temporal a largo plazo

A continuación, se detalla la tendencia de los contaminantes en cada una de las estaciones.

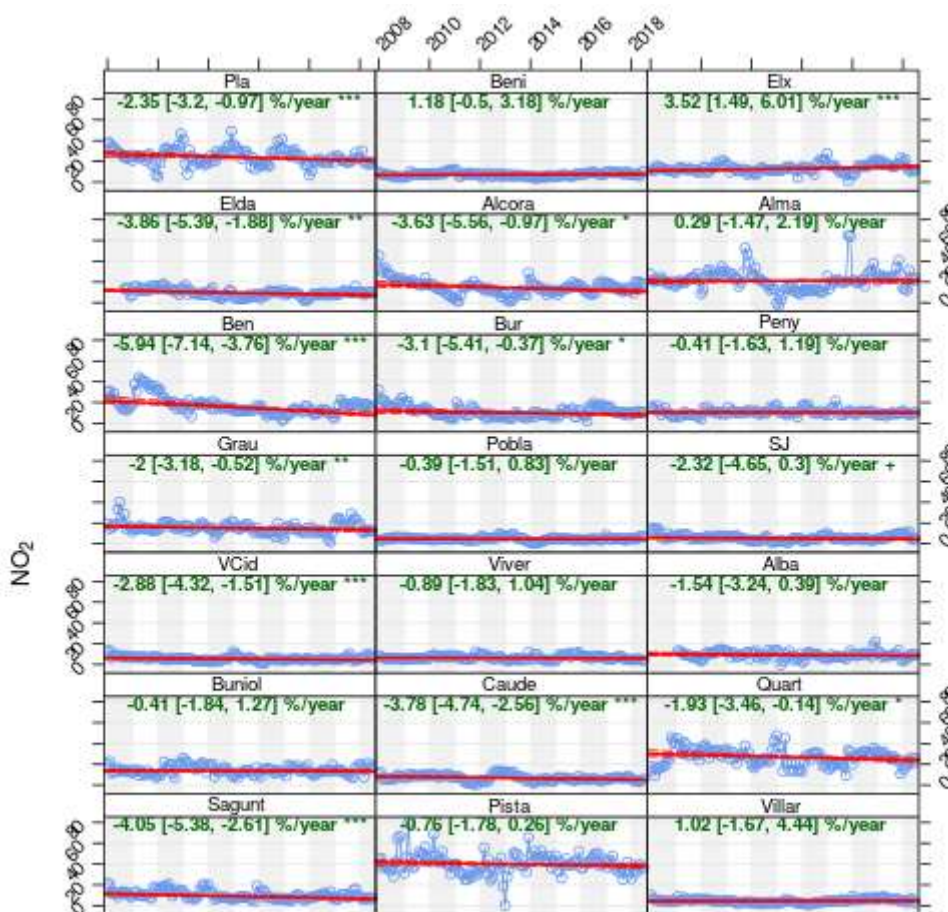
El NO presentó tendencia negativa para gran parte de las estaciones (**Figura 70**): Elda (-0.05  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Caude (-0.03  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Pla (-0.38  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Ben (-0.99  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Bur (-

0.26  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Elx (-0.16 [-0.239, -0.096]  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Grau (-0.32  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Peny (-0.09  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Pista (-0.55  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Quart (-0.38  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Sagunt (-0.13  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), SJ (-0.23  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), VCid (-0.31  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ).



**Figura 70.** Gráficas de tendencia porcentual para el NO por estación de monitoreo a lo largo del período 2008-2018. El grado de significancia: + 0.1, \* 0.05, \*\* 0.01, \*\*\* 0.001.

10 de las 21 estaciones mostraron una tendencia negativa significativa para el  $\text{NO}_2$ . La **Figura 71** muestra las tendencias porcentuales para todas las estaciones. Las tendencias significativas se presentaron en Ben (-1.3  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Caude (-0.32  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Pla (-0.64  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Sagunt (-0.45  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), VCid (-0.18  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Elda (-0.44  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Grau (-0.34  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Alcora (-0.64  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Bur (-0.38  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Quart (-0.58  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ). Sin embargo, la estación Elx mostró tendencia positiva (0.38  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ).



**Figura 71.** Gráficas de tendencia porcentual para el NO<sub>2</sub> por estación de monitoreo a lo largo del período 2008-2018. El grado de significancia: + 0.1, \* 0.05, \*\* 0.01, \*\*\* 0.001.

El O<sub>3</sub> ha demostrado ser un gran problema dentro de la Comunidad Valenciana, por la elevada cantidad de excedencias respecto a los límites establecidos por la normativa (**Figura 72**). La tendencia de este contaminante fue negativa sólo en Raba (-0.80 µg.m<sup>-3</sup>.año), mientras que 9 estaciones presentaron una tendencia positiva: Elx (0.62 µg.m<sup>-3</sup>.año), Pla (0.92 µg.m<sup>-3</sup>.año), Pobla (0.66 µg.m<sup>-3</sup>.año), Quart (0.98 µg.m<sup>-3</sup>.año), Sagunt (1.05 µg.m<sup>-3</sup>.año), Buniol (0.61 µg.m<sup>-3</sup>.año), Beni (0.43 µg.m<sup>-3</sup>.año), Bur (0.43 µg.m<sup>-3</sup>.año) y Caude (0.27 µg.m<sup>-3</sup>.año). El resto de las estaciones no mostraron una tendencia significativa (p > 0.05) dentro del período en estudio.



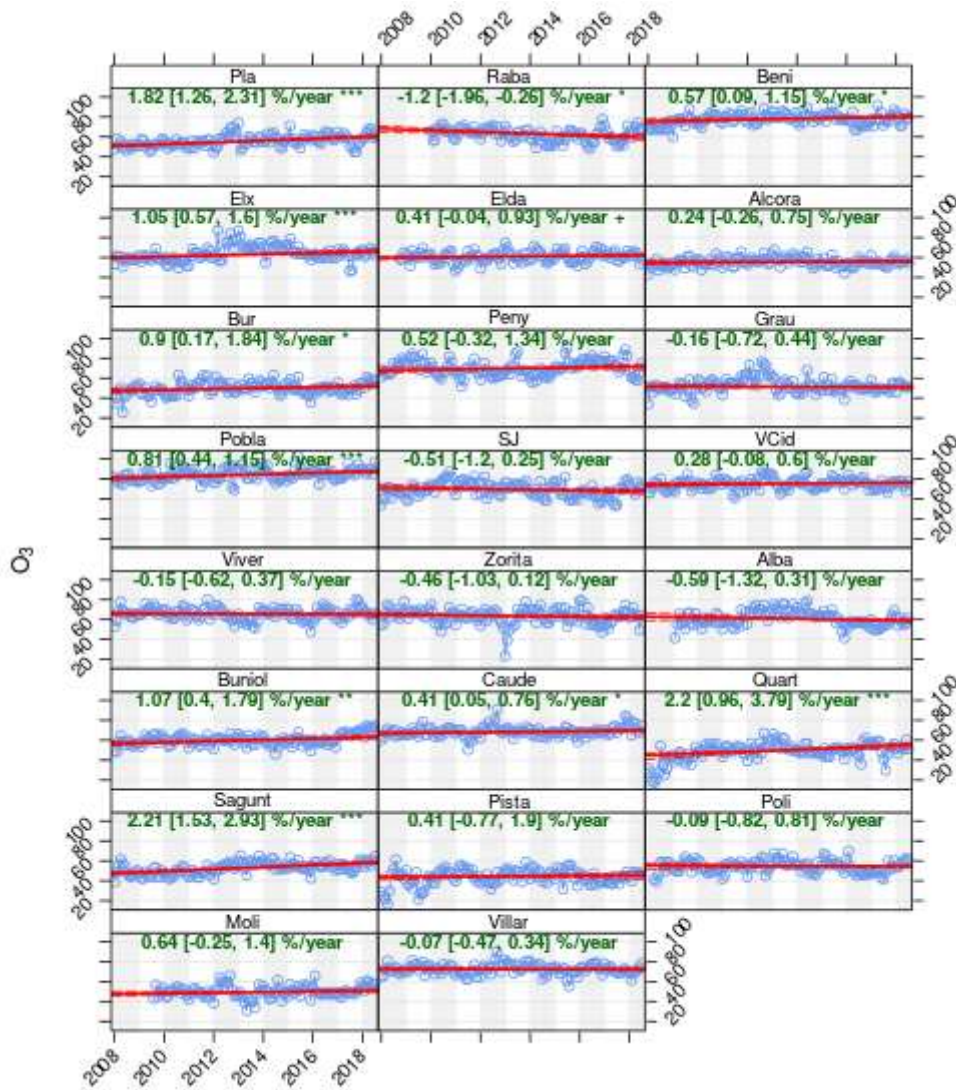
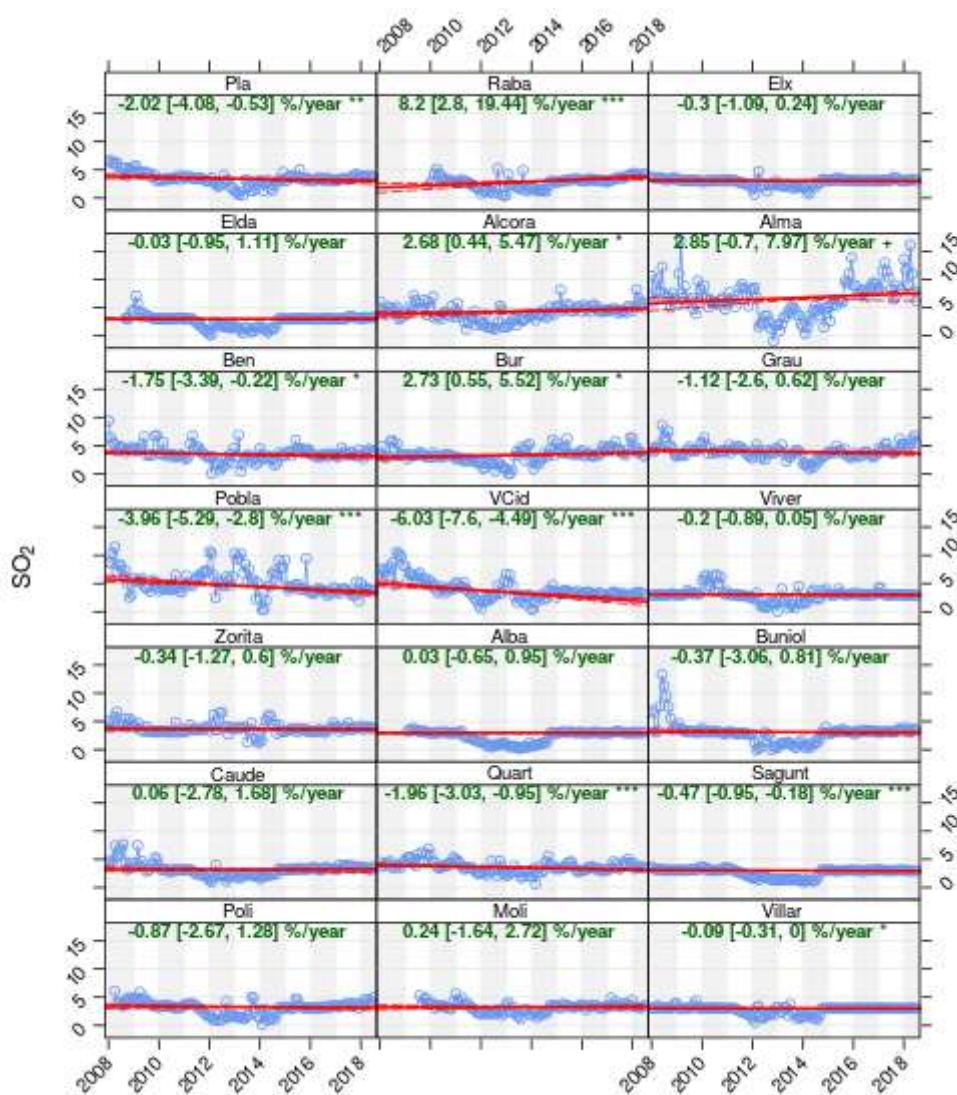


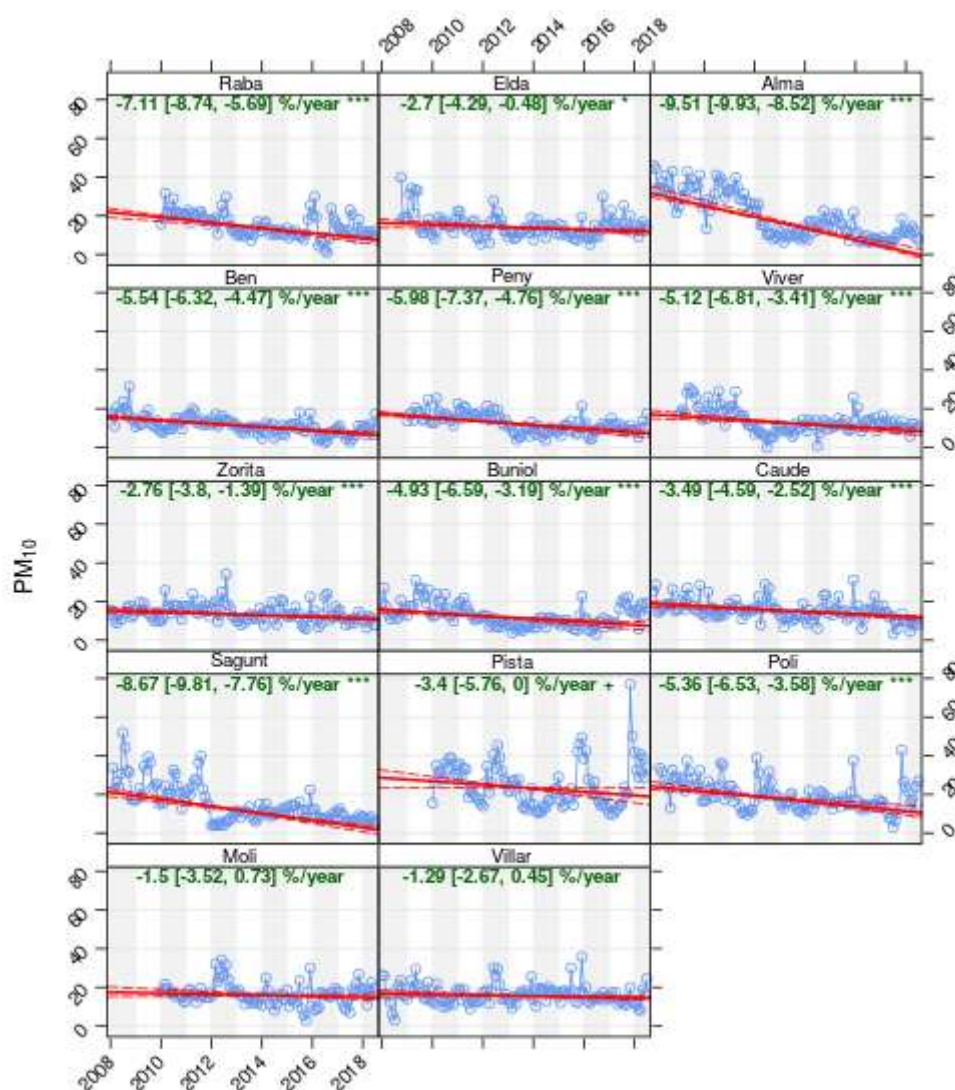
Figura 72. Gráficas de tendencia porcentual para el O<sub>3</sub> por estación de monitoreo a lo largo del período 2008-2018. El grado de significancia: + 0.1, \* 0.05, \*\* 0.01, \*\*\* 0.001.

El SO<sub>2</sub> mostró una tendencia negativa en las estaciones Pobla (-0.23 µg.m<sup>-3</sup>.año), Quart (-0.08 µg.m<sup>-3</sup>.año), Sagunt (-0.02 µg.m<sup>-3</sup>.año), VCid (-0.30 µg.m<sup>-3</sup>.año), Pla (-0.08 µg.m<sup>-3</sup>.año), Ben (-0.07 µg.m<sup>-3</sup>.año), Villar (-0.003 µg.m<sup>-3</sup>.año); y positiva para Raba (0.18 µg.m<sup>-3</sup>.año), Alcora (0.10 µg.m<sup>-3</sup>.año) y Bur (0.08 µg.m<sup>-3</sup>.año) (Figura 73).



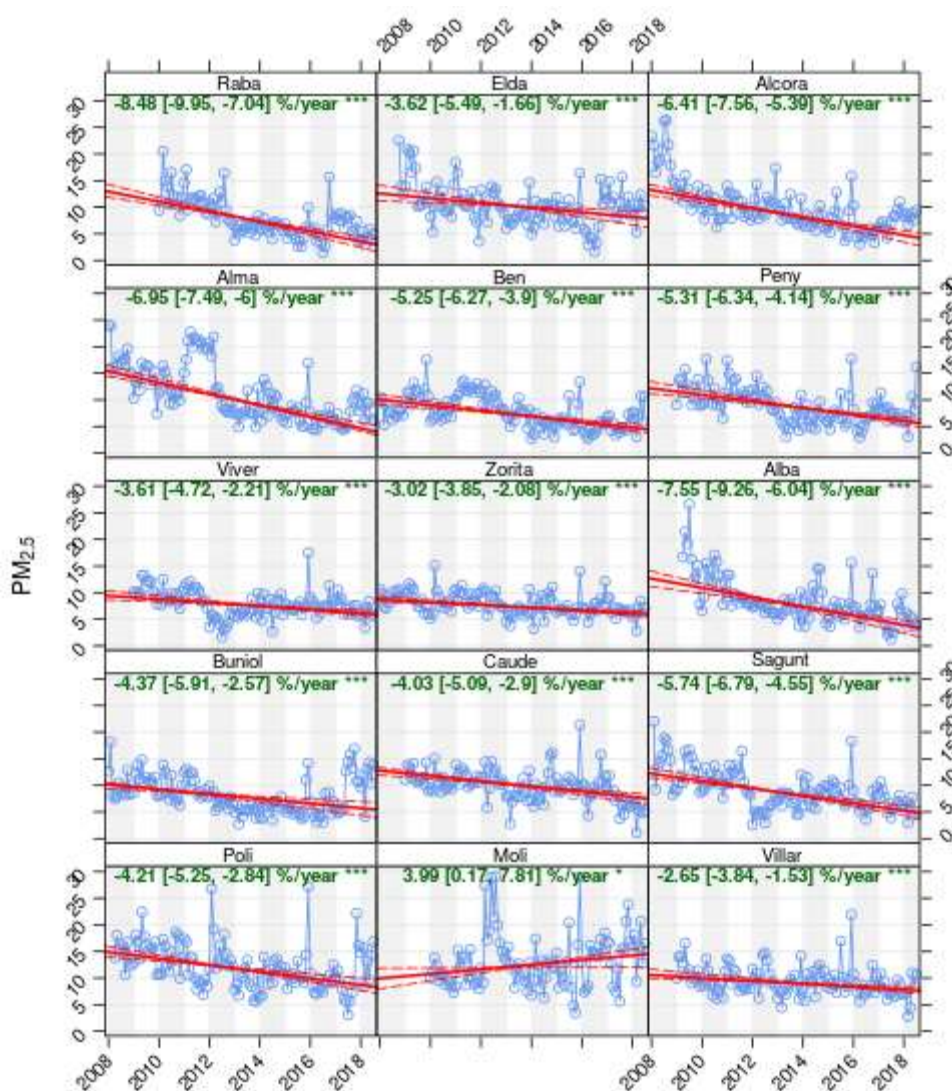
**Figura 73.** Gráficas de tendencia porcentual para el SO<sub>2</sub> por estación de monitoreo a lo largo del período 2008-2018. El grado de significancia: + 0.1, \* 0.05, \*\* 0.01, \*\*\* 0.001.

En la **Figura 74** se pueden ver los resultados para PM<sub>10</sub>, donde las siguientes estaciones mostraron tendencia negativa significativa ( $p < 0.05$ ): Alma (-2.97  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Ben (-0.87  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Buniol (-0.77  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Caude (-0.65  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Elda (-0.43  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Peny (-0.98  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Poli (-1.27  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Raba (-1.36  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Sagunt (-1.82  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Viver (-0.83  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ) y Zorita (-0.42  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ). Las estaciones Moli, Villar y Pista no mostraron una tendencia significativa ( $p > 0.05$ ).



**Figura 74.** Gráficas de tendencia porcentual para el  $PM_{10}$  por estación de monitoreo a lo largo del período 2008-2018. El grado de significancia: + 0.1, \* 0.05, \*\* 0.01, \*\*\* 0.001.

La tendencia de  $PM_{2.5}$  en el período de tiempo fue mayoritariamente negativa (**Figura 75**): Alba (-0.87  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Alcora (-0.85  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Alma (-1.07  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Ben (-0.53  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Buniol (-0.44  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Caude (-0.49  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Elda (-0.44  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Peny (-0.61  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Poli (-0.62  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Raba (-0.94  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Sagunt (-0.70  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Villar (-0.27  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ), Viver (-0.33  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ) y Zorita (-0.26  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ ). Excepto para la estación Moli cuya tendencia fue de 0.43  $\mu\text{g}\cdot\text{m}^{-3}\cdot\text{año}$ .



**Figura 75.** Gráficas de tendencia porcentual para el  $PM_{2.5}$  por estación de monitoreo a lo largo del período 2008-2018. El grado de significancia: + 0.1, \* 0.05, \*\* 0.01, \*\*\* 0.001.

### VIII. Análisis del comportamiento espacial del material particulado

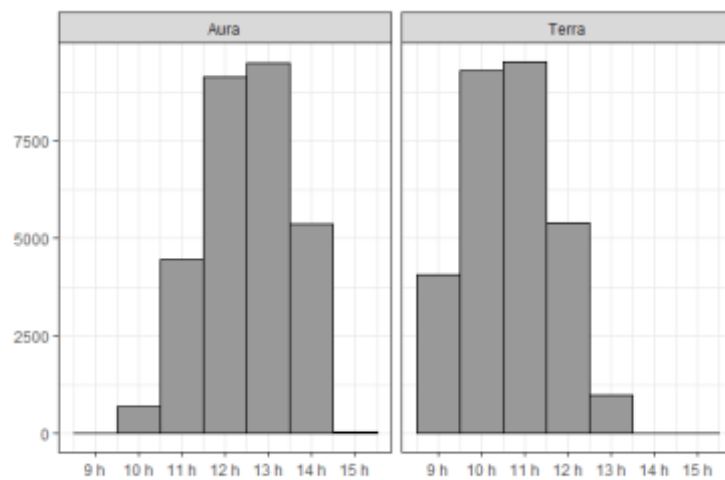
El material particulado atmosférico es un contaminante importante en la atmósfera por su efecto en el clima y en la salud de la población. Como se describió anteriormente, es posible cuantificar la concentración de aerosoles presentes en la atmósfera mediante el parámetro AOD. Sin embargo, la relación entre AOD y el material particulado es compleja y se ve afectada por la distribución vertical de los aerosoles, la eficiencia de la dispersión, la concentración del material particulado y de las propiedades químicas de los aerosoles [Li et al., 2018].

#### Obtención y validación de las imágenes satelitales

El producto MCD19A2 provisto por la NASA es una elaboración diaria compuesta de distintas capturas realizadas por los satélites Terra y Aqua con una resolución espacial de 1 km [Lyasputin

& Wang, 2018]. Para el período comprendido entre el 01 de enero del 2008 y el 31 de julio de 2018 se descargaron un total de 15705 imágenes del producto satelital MCD19A2 provisto por la NASA.

En la **Figura 76** se observan las horas de paso de los satélites Aura y Terra sobre la Comunidad Valenciana. Para evaluar la fuerza de la relación entre AERONET y MODIS se tomaron los valores dentro de los 30 min en un buffer de 3km, se calcularon los factores de correlación de Pearson, Kendall y Spearman, los cuales dieron: 0.78 (p-value < 0.001), 0.61 (p-value < 0.001) y 0.81 (p-value < 0.001), respectivamente. Esto mismo se realizó para un buffer de 7 km de diámetro, encontrándose que los factores de correlación de Pearson (0.80, p-value < 0.001), Kendall (0.63, p-value < 0.001) y Spearman (0.83, p-value < 0.001) fueron mejores.



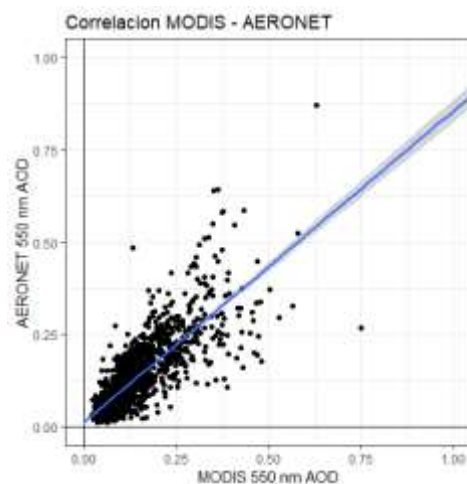
**Figura 76.** Horas de sobrevuelo de los satélites Aura y Terra sobre Valencia que componen el producto satelital MCD19A2 (UTC)

Finalmente, se consideró la correspondencia con mediciones realizadas en un intervalo de 15 min de paso del satélite y con un buffer de 7 km. Los resultados no mostraron mejoras en los factores de correlación de Pearson (0.80, p-value < 0.001), de Kendall (0.63, p-value < 0.001), de Spearman (0.83, p-value < 0.001), ofreciendo resultados similares.

El modelo lineal generado fue:

$$AOD_{AERONET} = a + b \cdot AOD_{MODIS}$$

donde  $a = 0.012 \pm 0.003$  (p-value < 0.001) y  $b = 0.85 \pm 0.02$  (p-value < 0.001), el cual se observa en la **Figura 77**. La regresión lineal presentó un  $R^2$  ajustado = 0.64. Esta ecuación fue la utilizada para la validación de las imágenes.

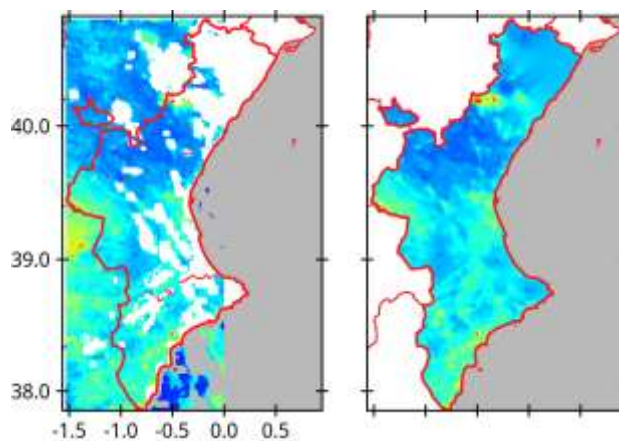


**Figura 77.** Regresión lineal entre los datos MODIS y AERONET a 550 nm.

#### Generación de un modelo

Para la confección de mapas de calidad de aire a partir de imágenes satelitales fue necesario estudiar la correlación entre las variables satelitales y terrestres, para luego aplicar un modelo multivariable. Como primer paso se analizó la correlación lineal entre AOD y  $PM_{10}$  y AOD y  $PM_{2.5}$  utilizando las mediciones diarias de AOD provenientes del producto satelital MCD19A2 y las mediciones en las estaciones de monitoreo terrestres de  $PM_{10}$  y  $PM_{2.5}$  dentro de la hora de cada paso del satélite. Sin embargo, la pérdida de registros por falta de datos de PM o AOD fue mayor al 90% (93% para  $PM_{2.5}$  y 94% para  $PM_{10}$ ) de los días.

Para saldar esta falta de datos en las imágenes satelitales, debido a sombras, nubes u otros tipos de interferencia que reduzcan la calidad de la estimación de AOD, se rellenaron los pixeles faltantes mediante una interpolación por el método Ponderado de Distancia Inversa (IDW) con una distancia de alisamiento de 5 km [Goldberg et al., 2019] (**Figura 78**). En la validación cruzada del método de interpolación IDW, se obtuvo un RMSE promedio de  $0.009 \pm 0.006$ .



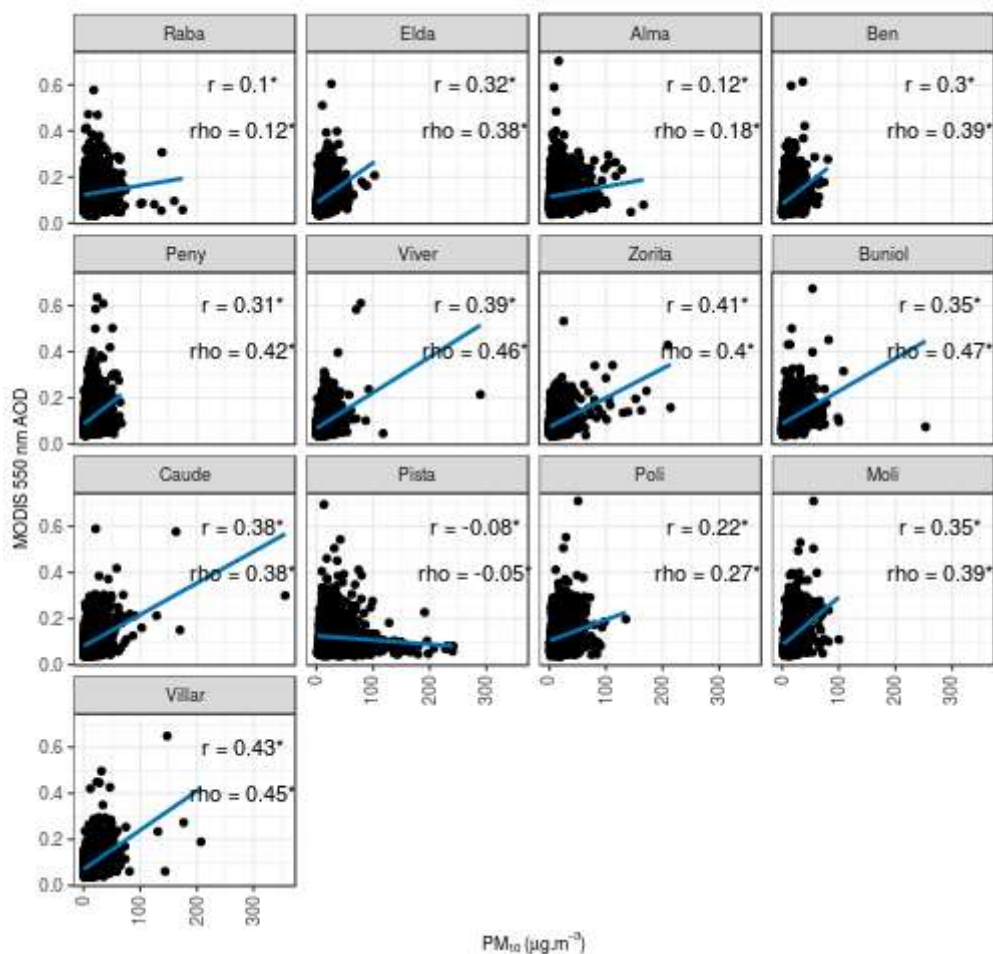
**Figura 78.** a) Imagen de AOD MAIAC e b) interpolación de AOD con el método IDW para el 1 de enero del 2008

Para relacionar las mediciones satelitales con las medidas terrestres de material particulado se evaluó la co-linealidad con los valores del pixel de las imágenes de MODIS que coincidía con el sitio de monitoreo terrestre y con un promedio de los valores de los pixel de las imágenes MODIS que coincidían con un buffer circular de diámetro igual a la resolución de la imagen satelital (1 km) en las estaciones de monitoreo terrestre.

Sin aplicar el buffer, los factores de correlación de Spearman y Pearson dieron para la relación entre AOD y PM<sub>2.5</sub>: 0.41 ( $p < 0.001$ ) y 0.34 ( $p < 0.001$ ), respectivamente. Mientras que para PM<sub>10</sub>, fueron inferiores los factores de correlación de Spearman (0.34,  $p < 0.001$ ) y Pearson (0.24,  $p < 0.001$ ) con un n total de 13948 mediciones.

Utilizando los datos de AOD promedio dentro de un buffer de 1 km entorno a la estación de monitoreo, los factores de correlación de Spearman y Pearson dieron para AOD-PM<sub>2.5</sub>, respectivamente, 0.42 ( $p < 0.001$ ), 0.34 ( $p < 0.001$ ). En este análisis también fueron similares los factores de correlación de Spearman (0.34,  $p < 0.001$ ) y de Pearson (0.24,  $p < 0.001$ ) para PM<sub>10</sub>. Como los factores de correlación fueron similares sin aplicar el buffer sobre las estaciones, se continuó trabajando con este conjunto de datos en la confección del modelo de PM ya que requiere menos procesamiento.

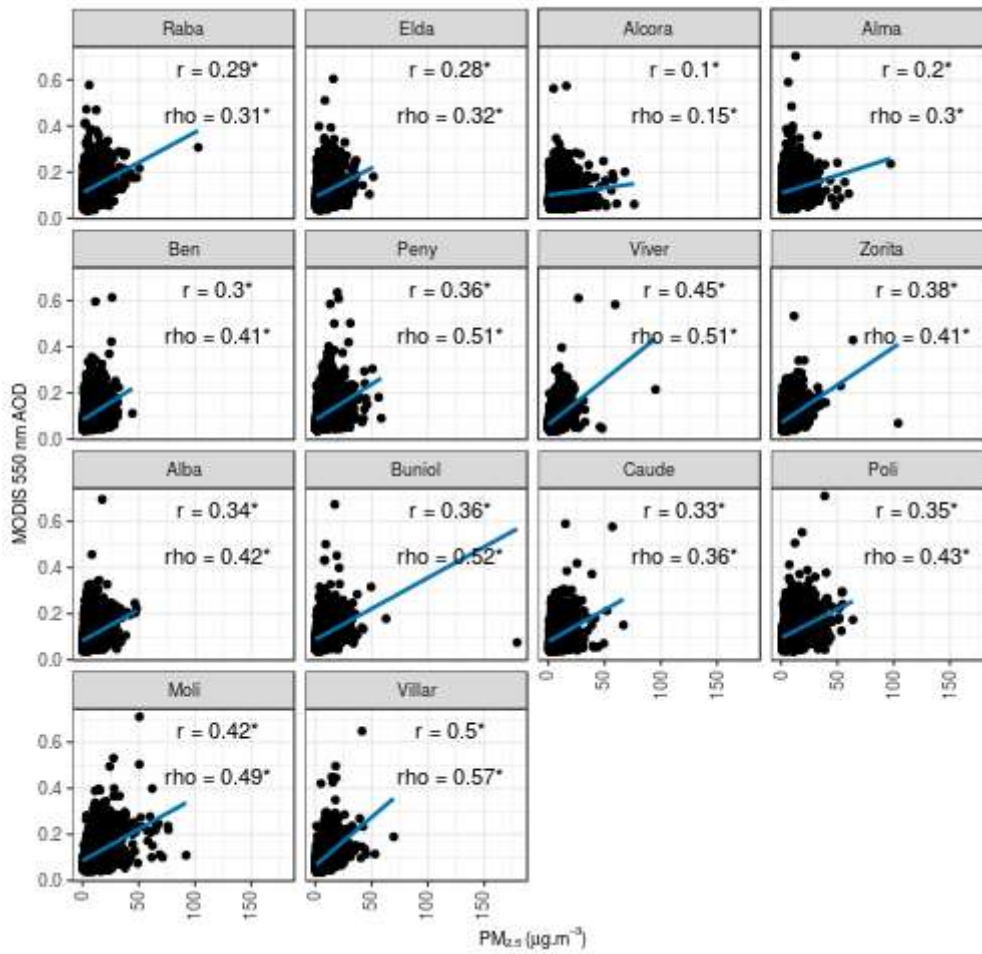
Por lo general, la correlación es mayor cuando se analiza de forma individual cada una de las estaciones. En la **Figura 79** se observa el factor de correlación de Pearson y Spearman en cada sitio de monitoreo de PM<sub>10</sub>. Llamativamente, las estaciones Raba, Alma y Pista presentaron factores de correlación extremadamente bajos. Esto puede deberse a una fuerte influencia de las fuentes locales sobre las mediciones de la estación.



**Figura 79.** Correlación entre las mediciones de AOD de MODIS y las mediciones promedio de  $PM_{10}$  dentro de las horas de paso de órbita del satélite para cada una de las estaciones en todo el período de estudio ( $p < 0.05$ )

En la **Figura 80** se puede observar el gráfico de correlación entre AOD y  $PM_{2.5}$  por estaciones de monitoreo. En general, se observa que la correlación es significativa ( $p < 0.05$ ) y baja para las mediciones diarias. Donde la mayor correlación se presenta en la estación Zorita ( $r = 0.49$ ,  $\rho = 0.58$ ) y la menor en la estación Alcora ( $r = 0.18$ ,  $\rho = 0.27$ ), ambas estaciones con un número equiparable de mediciones.



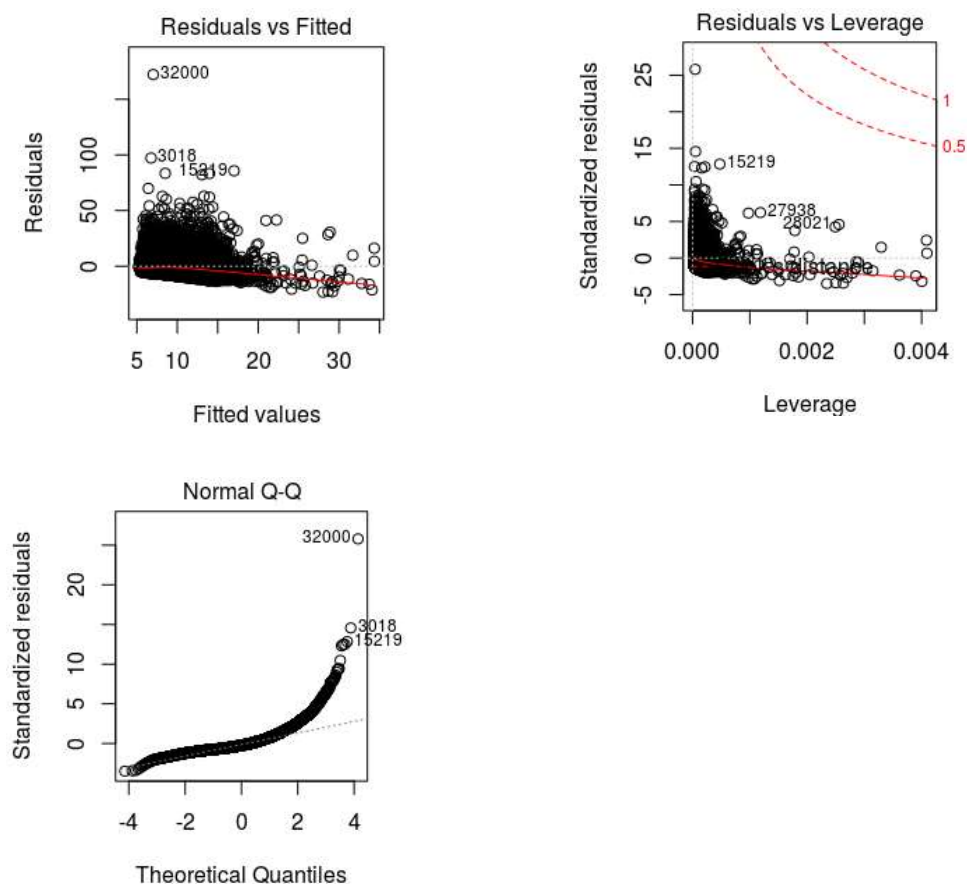


**Figura 80.** Correlación entre las mediciones de AOD de MODIS y las mediciones promedio de  $PM_{2.5}$  dentro de las horas de paso de órbita del satélite para cada una de las estaciones en todo el período de estudio ( $p < 0.05$ )

Por presentar una mejor correlación, se decidió elaborar un mapa de concentración de  $PM_{2.5}$ . Como primer modelo se consideró un modelo de regresión lineal del tipo:

$$Y = \beta_0 + \beta_1 X$$

donde X representa a el valor de AOD diario en los puntos de monitoreo e Y la concentración de  $PM_{2.5}$ . El modelo obtenido arrojó los parámetros  $\beta_0 = 3.83 (\pm 0.08, p\text{-value} < 0.001)$  y  $\beta_1 = 42.98 (\pm 0.7, p\text{-value} < 0.001)$ , con un factor de determinación muy bajo ( $R^2 = 0.11$ ).

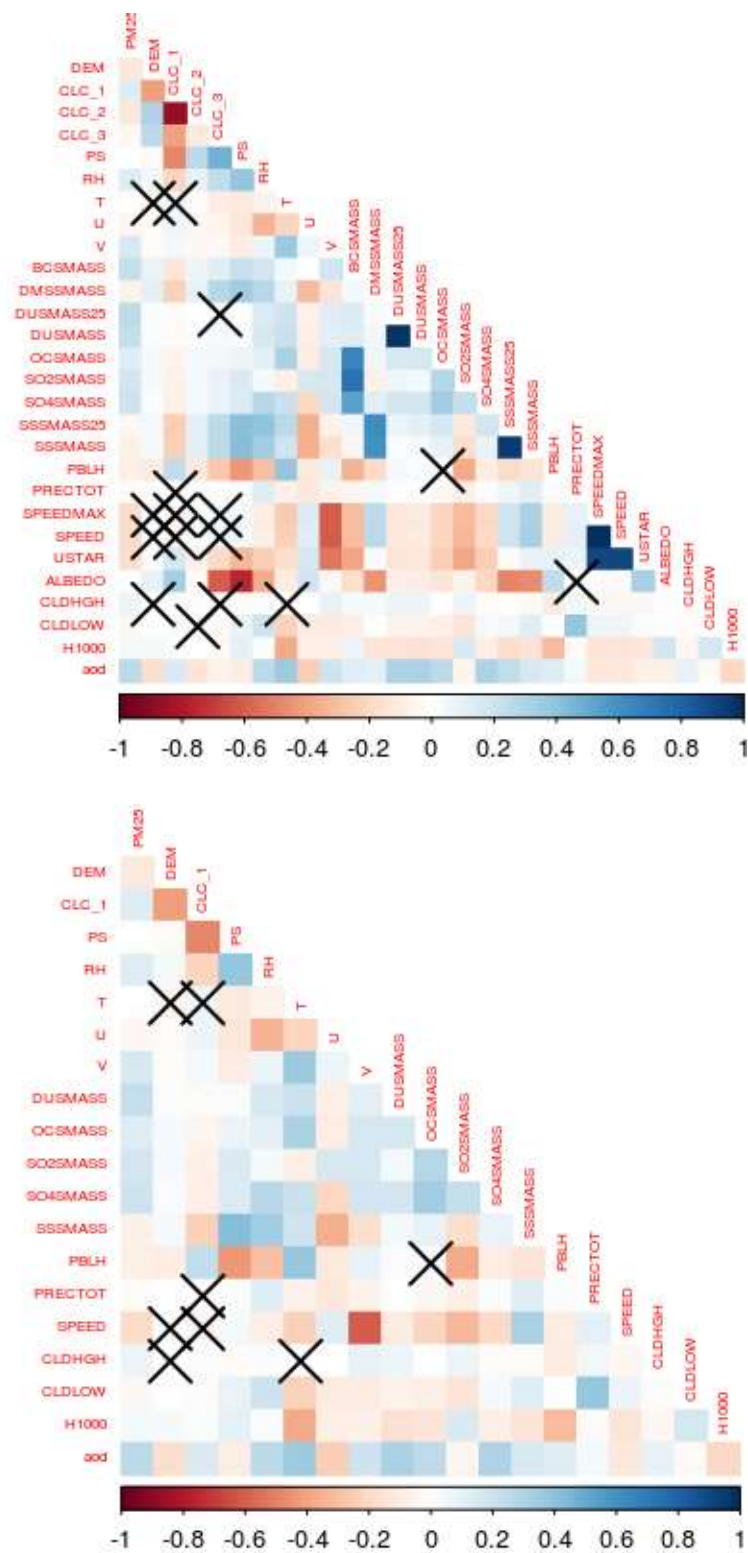


**Figura 81.** Gráficos evaluatorios del modelo. a) Residuos del modelo, b) Distancia de Cook, c) QQ-Plot de los residuos estandarizados.

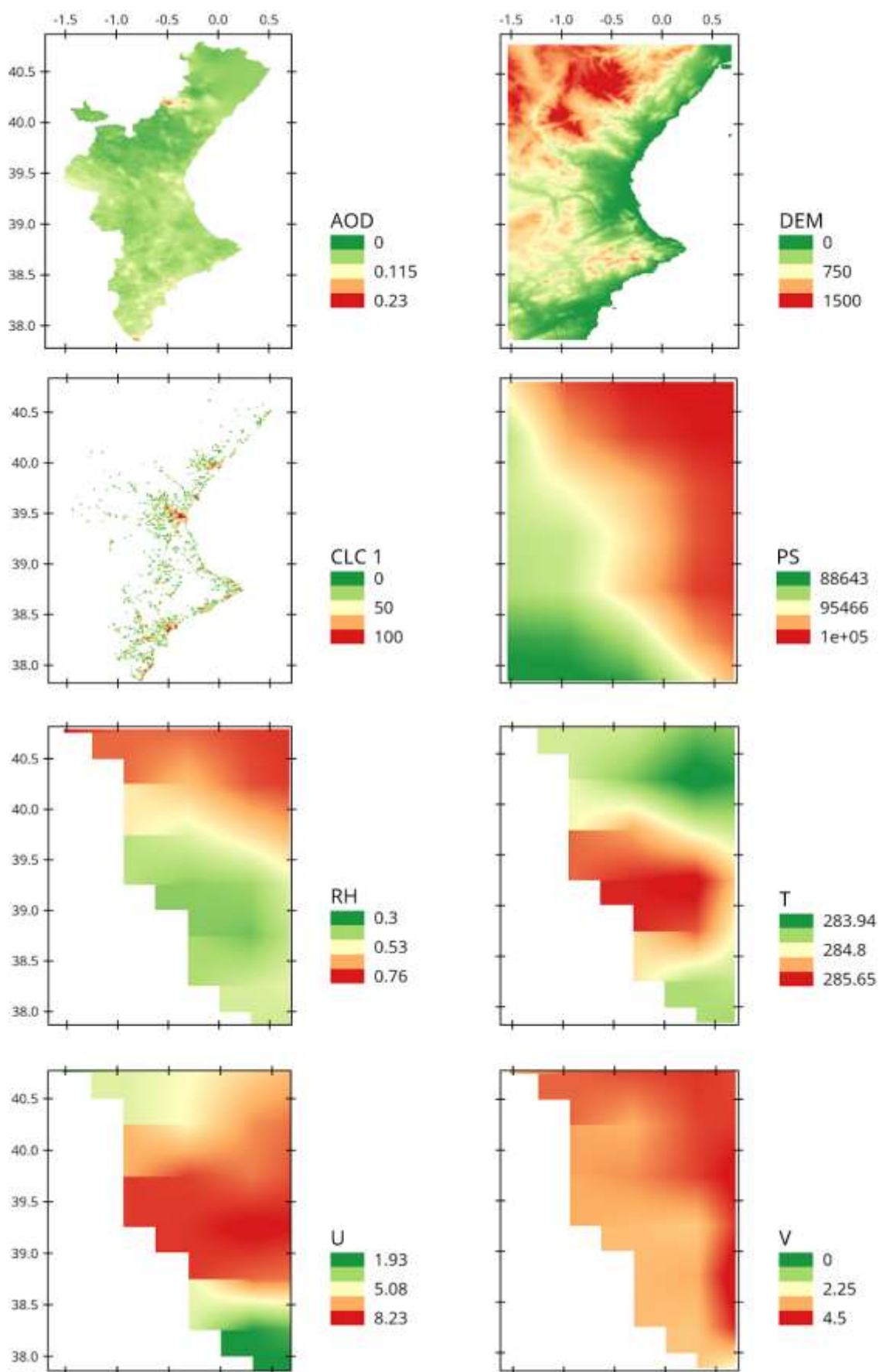
La **Figura 81** muestra la falta de normalidad de los residuos estandarizados. Además de la inspección visual de los residuos, se evaluó su normalidad con una prueba de Shapiro, mediante la cual se rechazó la hipótesis nula de normalidad de los residuos ( $W = 0.84$ ,  $p\text{-value} < 0.001$ ). En base a esto se procedió a trabajar con un modelo multivariable que pueda explicar mejor la relación entre estas variables.

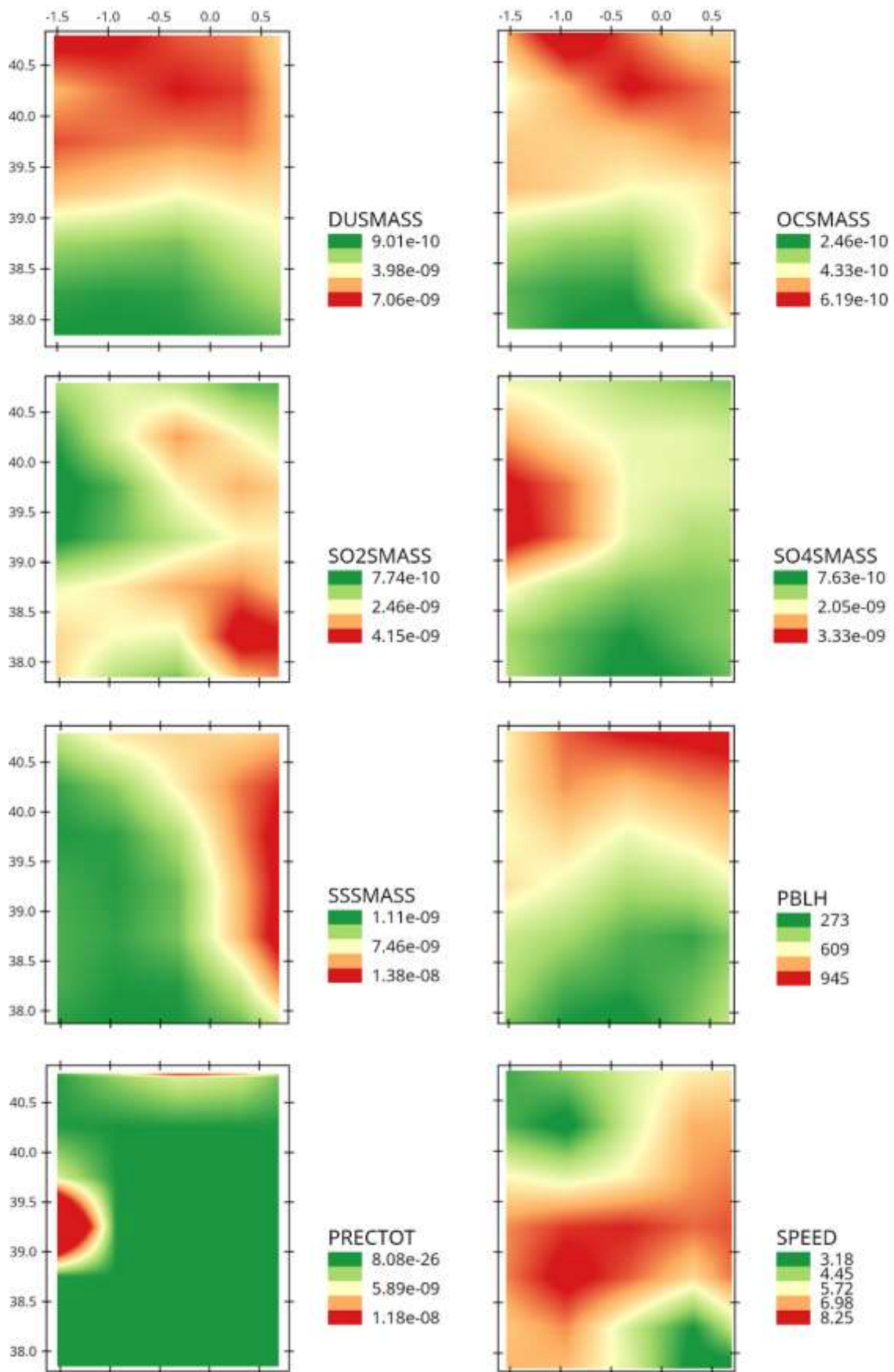
En la elaboración del modelo multivariable se trabajó con las variables accesorias detalladas en el **Módulo II, Sección II- XI**. Previo al desarrollo del modelo, se analizó la correlación entre las variables accesorias (**Figura 82**). Se seleccionaron 19 variables a modelar  $PM_{2.5}$ : DEM, CLC\_1, PS, RH, T, U, V, DUSMASS, OCSMASS, SO<sub>2</sub>SMASS, SO<sub>4</sub>SMASS, SSSMASS, PBLH, PRECTOT, SPEED, CLDHGH, CLDLOW, H1000, AOD.

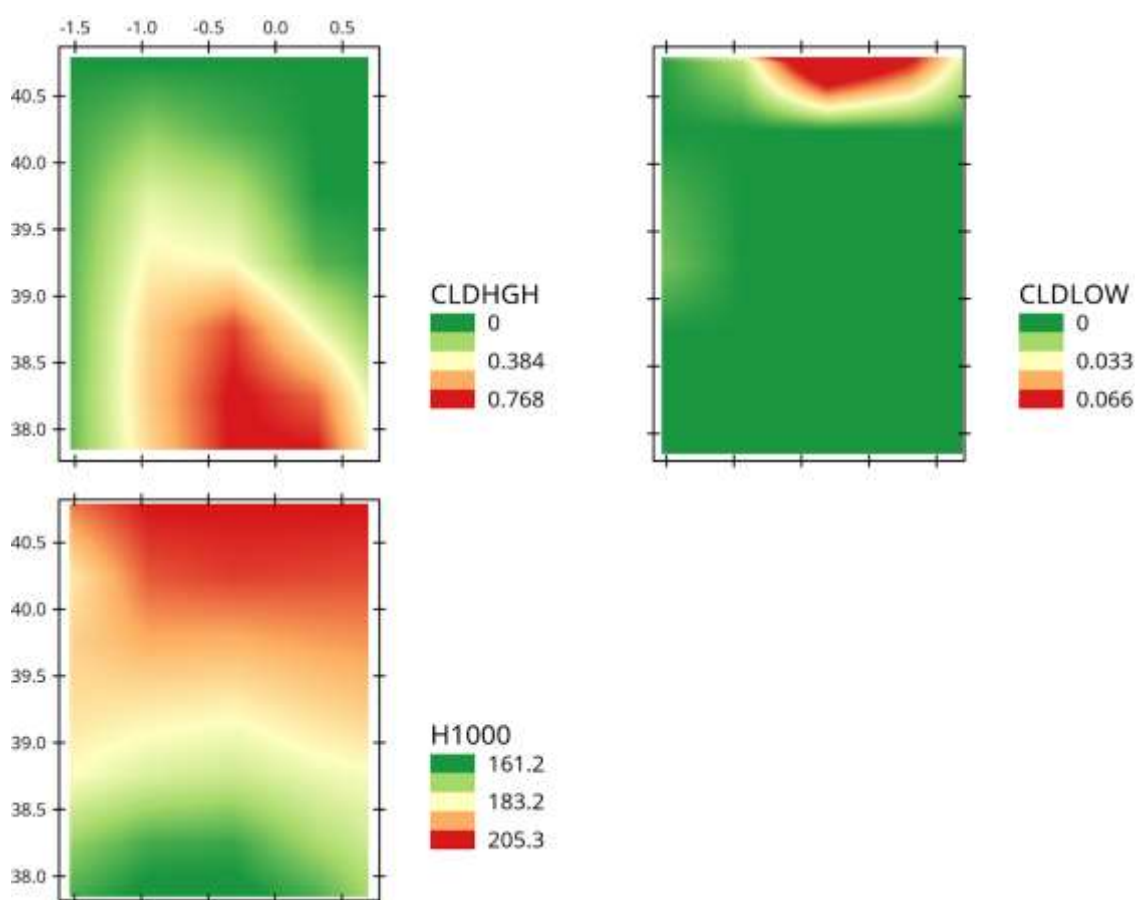
En la **Figura 83** se muestra una visualización de las variables para la fecha 01 de enero del 2008, en la cual se puede observar que los parámetros modelados por MERRA-2 pueden presentar discontinuidades (ver RH, T, U, V).



**Figura 82.** Análisis de correlación entre las variables. a) Entre todas las variables, b) Entre las variables seleccionadas para el modelo. De rojo a azul se indica en factor de correlación de Pearson. Las cruces indican correlación no significativa.







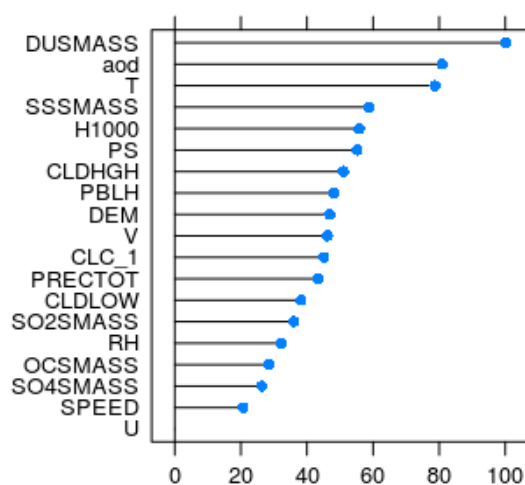
**Figura 83.** Visualización de las variables incluidas en el modelo multivariable para una fecha determinada. DEM = Modelo de elevación (m); CLC\_1 = proporción de superficie artificial (%); PS = presión en superficie (hPa); RH = humedad relativa a 1000 hPa (%); T = temperatura del aire a 1000 hPa (K); U = componente del viento hacia el este ( $\text{m}\cdot\text{s}^{-1}$ ); V = componente del viento del norte ( $\text{m}\cdot\text{s}^{-1}$ ); DUSMASS = concentración másica superficial modelada de polvo ( $\text{kg}\cdot\text{m}^{-3}$ ); OCSMASS = concentración másica superficial modelada de carbono orgánico ( $\text{kg}\cdot\text{m}^{-3}$ ); SO<sub>2</sub>SMASS = concentración másica superficial modelada de SO<sub>2</sub> ( $\text{kg}\cdot\text{m}^{-3}$ ); SO<sub>4</sub>SMASS = concentración másica superficial modelada de SO<sub>4</sub> ( $\text{kg}\cdot\text{m}^{-3}$ ); SSSMASS = concentración másica superficial modelada de sal marina ( $\text{kg}\cdot\text{m}^{-3}$ ); PBLH = altura de la capa límite planetaria (m); PRECTOT = precipitación superficial total ( $\text{kg}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ ); SPEED = velocidad superficial del viento ( $\text{m}\cdot\text{s}^{-1}$ ); CLDHGH = Fracción de área cubierta por nubes altas (%); CLDLOW = Fracción de área cubierta por nubes bajas (%); H1000 = altura a 1000 hPa (m); aod = densidad óptica de aerosoles.

Como primera prueba de un modelo LUR se realizó una regresión lineal múltiple. Se tomó un set de entrenamiento con el 80% de los datos (98087 registros) y se utilizó el 20% restante (24520 registros) para calcular el error del modelo. Como resultado se obtuvo un RMSE =  $6.98 \mu\text{g}\cdot\text{m}^{-3}$ ,  $R^2 = 0.19$  y MAE =  $4.60 \mu\text{g}\cdot\text{m}^{-3}$  para el set de entrenamiento, y un RMSE =  $6.98 \mu\text{g}\cdot\text{m}^{-3}$  y MAE =  $4.62 \mu\text{g}\cdot\text{m}^{-3}$  para el set de prueba. Si bien el RMSE y MAE son aceptables, aún existe una gran varianza de los datos sin explicar.

Por tanto, se evaluó la técnica regresión de *machine learning* no paramétrica, Random Forest. Como parámetros, se utilizaron un total de 1000 árboles de entrenamiento sobre los 98087 registros (80% del set de datos). La verificación del modelo se llevó a cabo utilizando la técnica de

10-fold cross validation y una evaluación de un conjunto de pruebas con el 20% de la información recopilada.

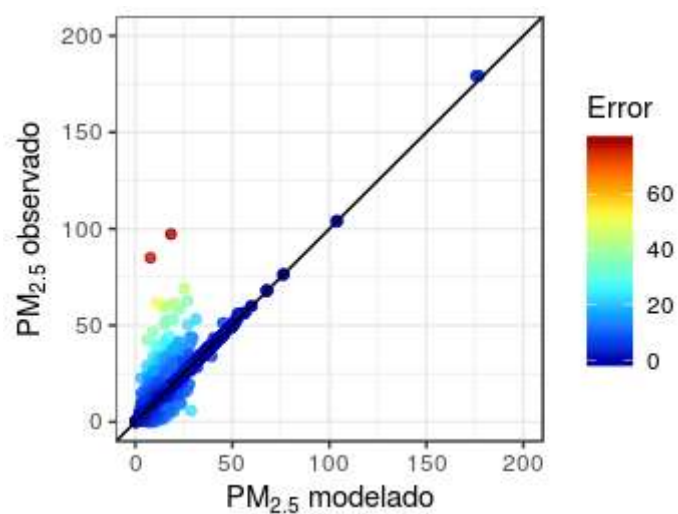
Al analizar los resultados se vio que las variables DUSMASS, AOD, T fueron las variables más importantes del modelo (**Figura 84**). Como se ha señalado con anterioridad, las mediciones de AOD satelital son un parámetro óptimo para la elaboración de modelos de aerosoles sobre la superficie terrestre. Por su parte, la variable DUSMASS es la predicción del modelo MERRA de las concentraciones másicas en la superficie de polvo, y T la temperatura, la cual influye en la dispersión de los contaminantes. A estas le siguen por importancia las variables SSSMAS, H1000 y PS, la primera corresponde al modelado de las sales marinas, y las otras dos, a variables relacionadas con la estabilidad atmosférica y la dispersión de los contaminantes en la atmósfera.



**Figura 84.** Peso de las variables en la construcción del modelo de  $PM_{2.5}$  superficial para la Comunidad de Valencia

El modelo presentó un muy buen ajuste con un  $R^2 = 0.90$ , un MAE de  $0.8 \mu\text{g}\cdot\text{m}^{-3}$  y un RMSE de  $2.4 \mu\text{g}\cdot\text{m}^{-3}$  en el ajuste del modelo, y un MAE de  $0.7 \mu\text{g}\cdot\text{m}^{-3}$  y RMSE de  $2.4 \mu\text{g}\cdot\text{m}^{-3}$  para el conjunto de datos de prueba. Asimismo, las estaciones no incluidas en el modelo debido a su alta tasa de datos faltantes obtuvieron un ajuste igualmente satisfactorio (RMSE de  $5.8 \mu\text{g}\cdot\text{m}^{-3}$  y MAE de  $4.5 \mu\text{g}\cdot\text{m}^{-3}$ ). Como un parámetro extra del ajuste del modelo se evaluó el MAPE para los datos de prueba, obteniendo un ajuste excelente (MAPE = 15%). A su vez, el BIAS del modelo fue bajo (5.54%) resultando un modelo confiable para realizar estimaciones en el territorio de la Comunidad Valenciana.

En la **Figura 85** se observa un scatterplot del ajuste entre las concentraciones predichas por el modelo y las mediciones de  $PM_{2.5}$ . Estos resultados muestran que la metodología propuesta puede utilizarse para generar mapas con estimaciones de concentraciones diarias de  $PM_{2.5}$  con una resolución de 1 km con muy buena precisión ( $< 6 \mu\text{g}\cdot\text{m}^{-3}$  para todo el dominio de simulación).



**Figura 85.** Scatterplot entre las concentraciones predichas por el modelo y las mediciones de PM<sub>2.5</sub>

A partir de este modelo se confeccionaron la serie de mapas diarios para todo el período de estudio. En la **Figura 86** se muestran mapas con las medias anuales de PM<sub>2.5</sub> para la Comunidad Valenciana. Puede observarse que la falta de información en ciertas áreas afecta la cobertura de toda el área de la Comunidad Valenciana. Se cuenta con períodos de tiempo donde las irregularidades son menores como 2016 y 2017, y otros años donde fueron mayores (2008, 2009 y 2010). Por esta falta de información, sólo a partir de 2013 se pudo simular las concentraciones sobre la localidad de Elche. También se observa un marcado escalonamiento, el cual puede ser producto de la menor resolución que presentaban algunas de las variables incluidas.



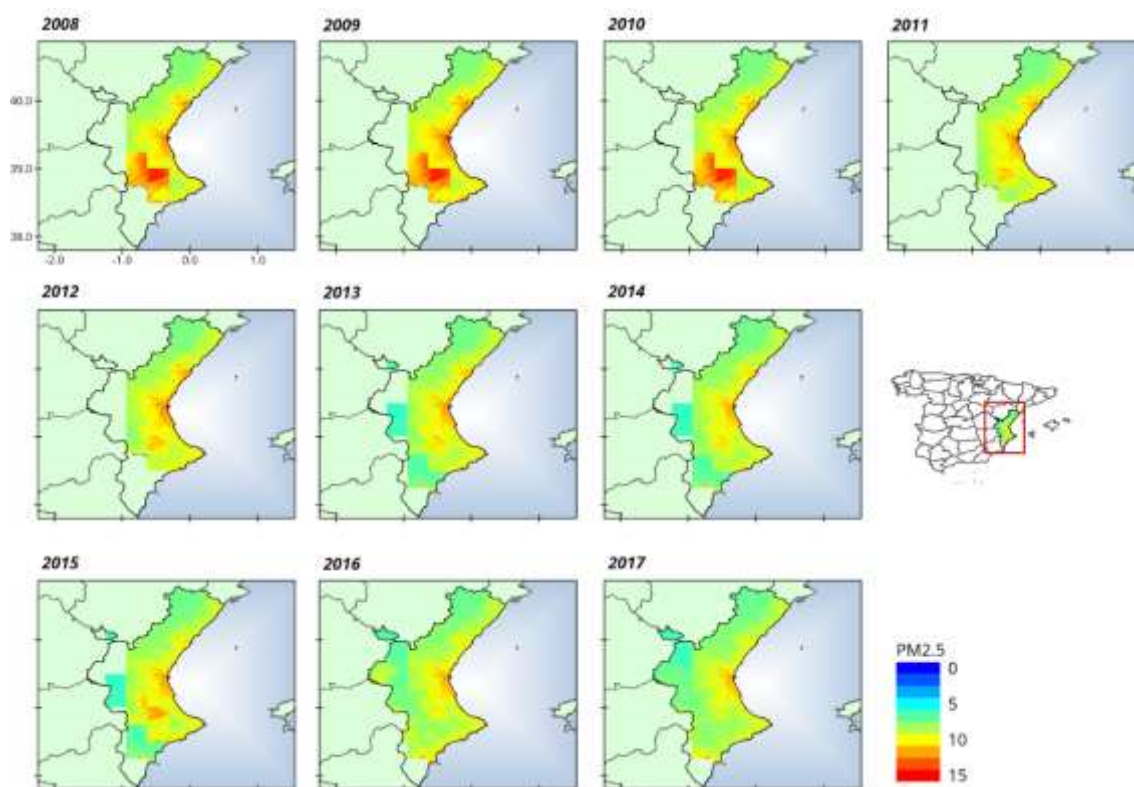
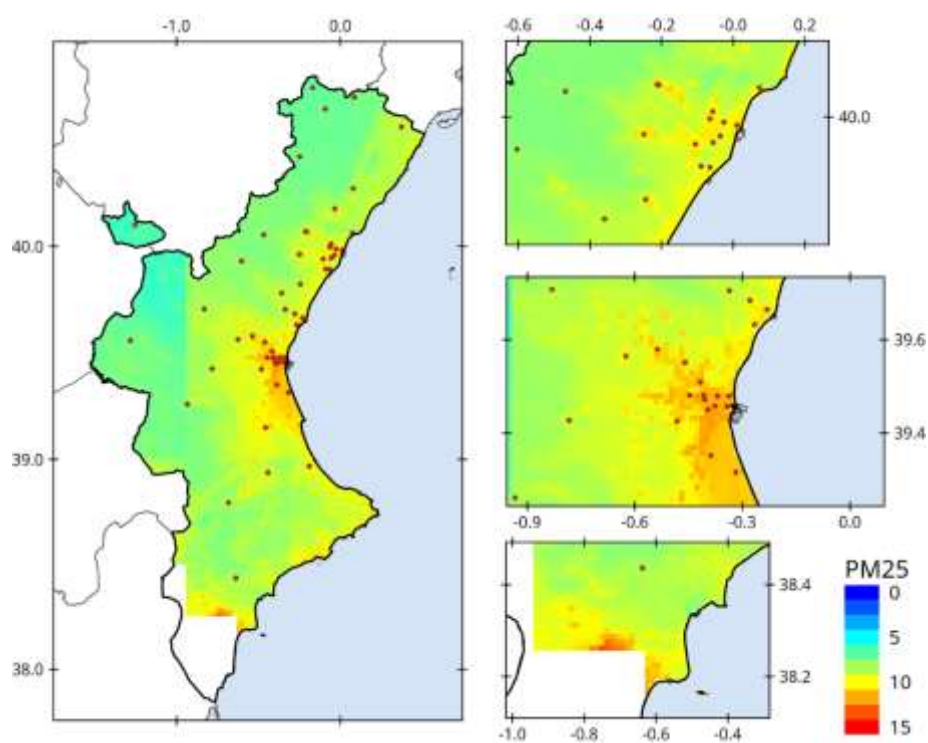


Figura 86. Mapa de concentraciones medias anuales de PM<sub>2.5</sub> (µg.m<sup>-3</sup>)

En la **Figura 87** se puede ver con mayor detalle las concentraciones medias anuales para el 2017. Las concentraciones medias anuales mayores se encontraron en Elx 13.68 µg.m<sup>-3</sup>, Valencia 12.03 µg.m<sup>-3</sup>, Alcira 11.31 µg.m<sup>-3</sup> y en Benidorm 11.37 µg.m<sup>-3</sup>.



**Figura 87.** Mapa de las concentraciones medias anuales de PM<sub>2.5</sub> ( $\mu\text{g}\cdot\text{m}^{-3}$ ) para el año 2017. En rojo las estaciones de monitoreo de calidad del aire.

### *IX. Mapas de riesgo*

A partir de los mapas de concentración de PM<sub>2.5</sub> se aplicó el índice de calidad de aire para cada píxel, generando mapas de calidad de aire. En la **Figura 88** puede verse la calidad del aire para la Comunidad de Valencia calculada a partir de las concentraciones medias anuales.

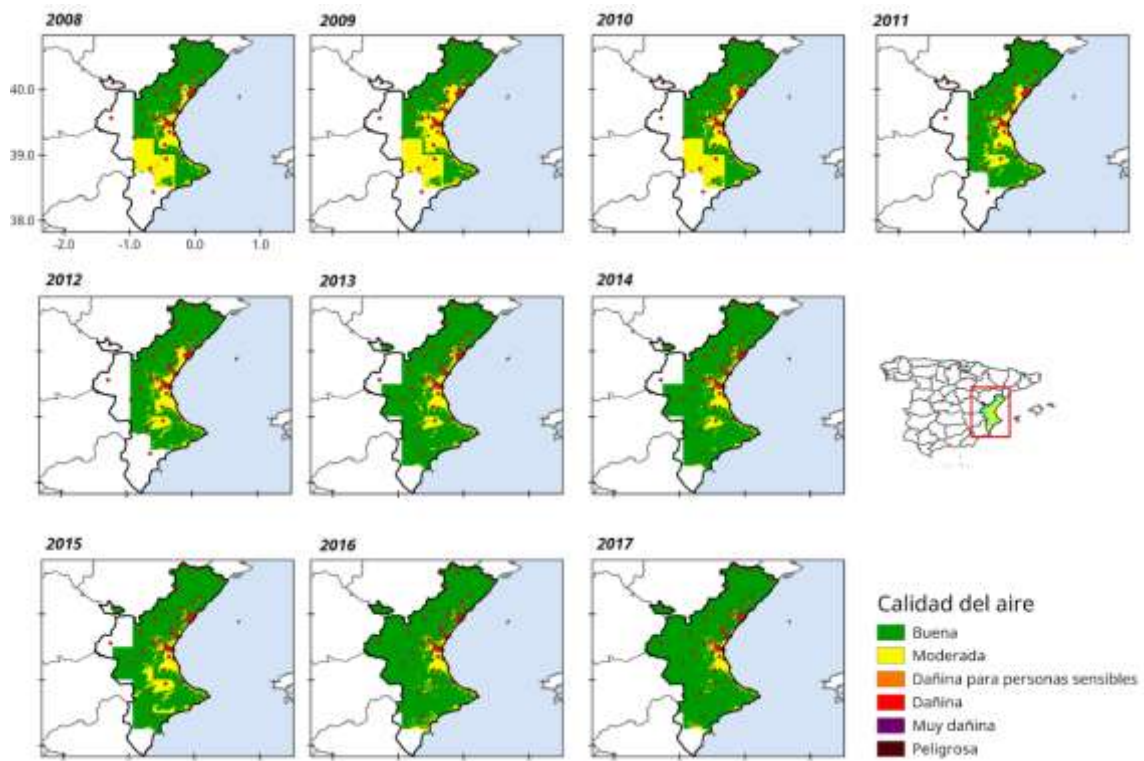


Figura 88. Mapa de calidad de aire para la Comunidad de Valencia

La **Figura 89** muestra con mayor detalle de la calidad del aire durante el año 2017 para la Comunidad Valenciana. Se observa que la calidad del aire en los principales núcleos urbanos es moderada, sin resultar dañina para la salud.

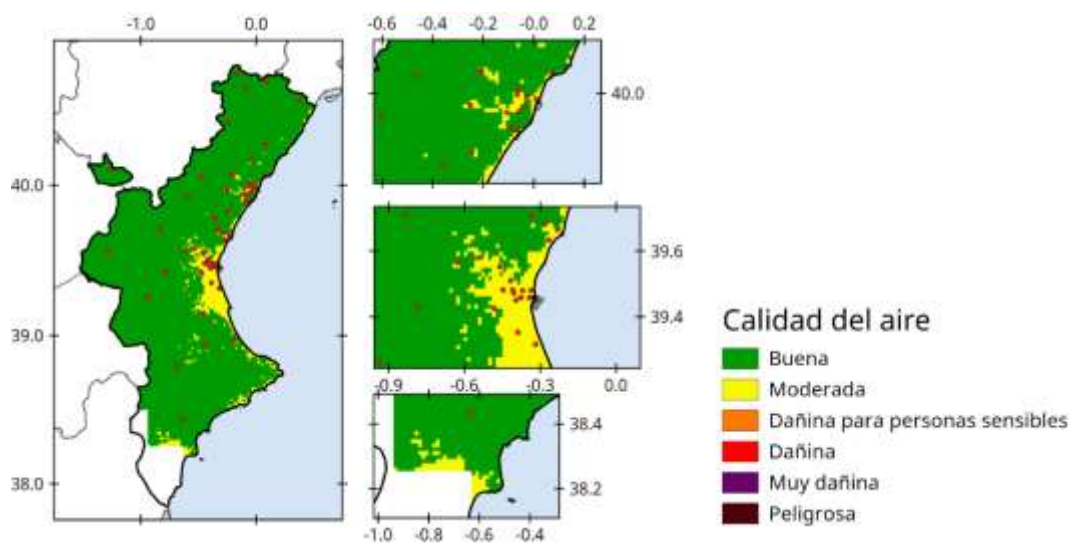
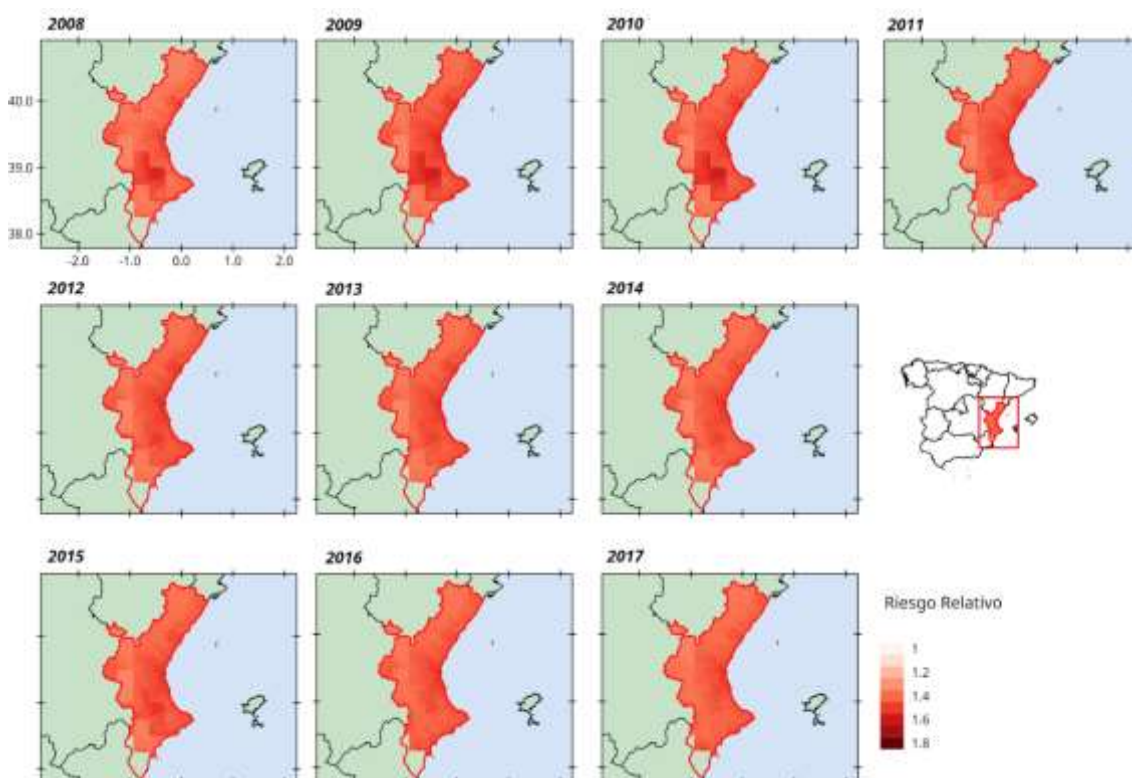


Figura 89. Mapa de calidad de aire para la Comunidad de Valencia del año 2017

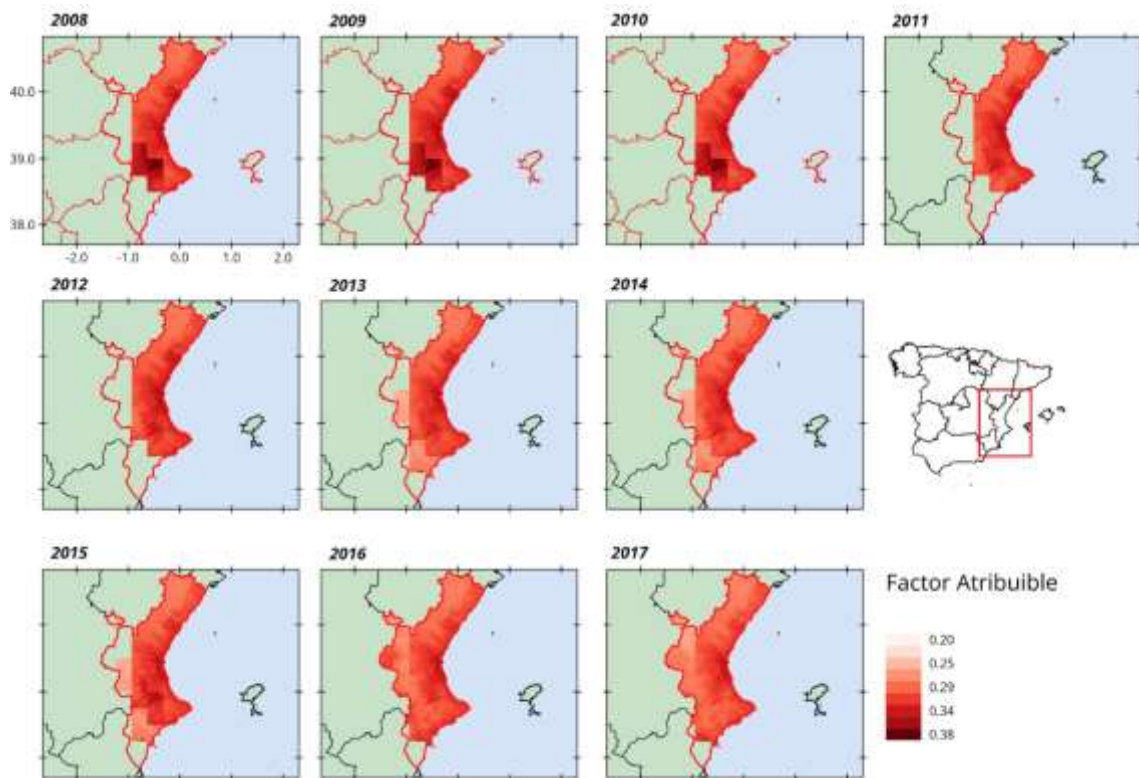
Conocer el riesgo al cual está expuesta la población es un excelente indicador para diagramar las políticas públicas de una región. Para evaluar el riesgo a la exposición de  $PM_{2.5}$  se tomó como métrica el riesgo a padecer cáncer de pulmón por exposición a largo plazo para adultos mayores de 30 años. En la **Figura 90** se muestra la variación del riesgo relativo (RR) a lo largo de los distintos años incluidos en nuestro estudio. El mayor RR se registró en Xativa (1.63), Onteniente (1.62), Albaida (1.54), Montaberner (1.61), Valencia (1.56) y Castellón de la Plana (1.58) para el 2008 disminuyendo levemente a lo largo de los años.



**Figura 90.** Mapa de riesgo relativo para personas mayores de 30 años de padecer cáncer de pulmón por exposición a largo plazo a  $PM_{2.5}$

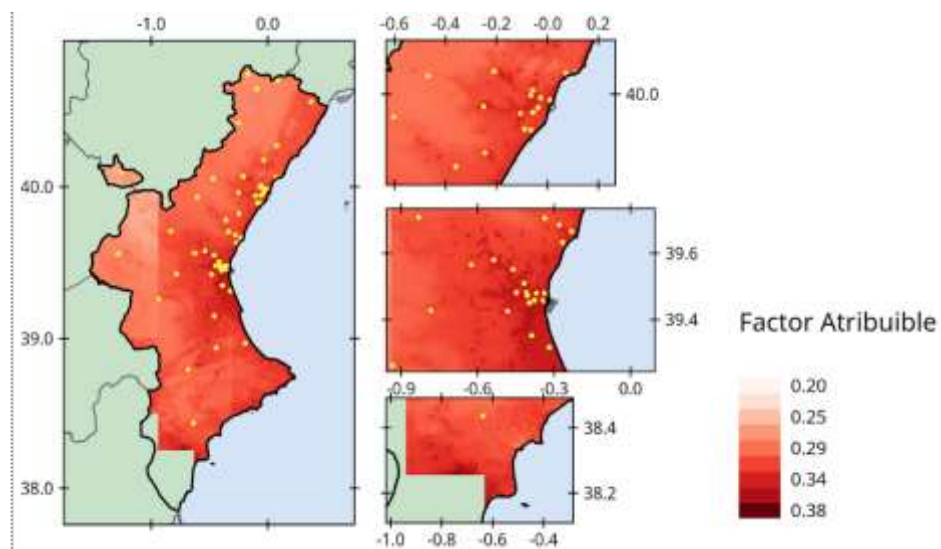
Analizando lo sucedido en el 2017, el RR a padecer cáncer de pulmón por motivos de la exposición a  $PM_{2.5}$  en la ciudad de Valencia se estimó de 1.53, esto es un riesgo del 53% mayor de una población sin exponer, mientras que en la ciudad de Elche se estimó en 1.59.

Tras esto, se calculó la fracción atribuible de muerte (FAM). Esto es, la fracción de muertes atribuidas a la exposición de  $PM_{2.5}$  sobre el total de las muertes ocurridas en el área analizada. En la **Figura 91** es visible que la FAM tiene valores máximos en las localidades antes mencionadas (Xativa, Onteniente, Albaida, Montaberner, Valencia y Castellón de la Plana) durante los primeros años en estudio.



**Figura 91.** Mapa de la fracción atribuible de muertes prematuras por cáncer de pulmón por exposición a largo plazo a  $PM_{2.5}$

En la **Figura 92** se puede ver la distribución de la fracción atribuible de muerte durante el 2017. La FAM por exposición a largo plazo de  $PM_{2.5}$  en Elche es del 37% y en Valencia de 34% sobre el total de muertes. Es evidente como al alejarse de los grandes conglomerados urbanos la FAM decae.



**Figura 92.** Mapa de la fracción atribuible de muertes prematuras por cáncer de pulmón por exposición a largo plazo a  $PM_{2.5}$  obtenido para el 2017

## III.II. Informe para la Comunidad Valenciana

### *Aspectos del monitoreo terrestre*

- **¿Qué diseño sigue la red de monitoreo?**

Una distribución no homogénea, estratificada en las áreas con mayor población de la Comunidad Valenciana. La mayor densidad de estaciones se presenta en la zona este-noreste de la Comunidad Valenciana, en las ciudades de Valencia y Castellón. Una segunda concentración de estaciones se da al sur, en las cercanías de la ciudad de Alicante.

- **¿Cuáles son las técnicas utilizadas para el monitoreo?**

Los métodos de referencia que se siguen son:

- El método de referencia para la medición de dióxido de azufre es el que se describe en la Norma UNE-EN 14212:2013 “Aire ambiente. Método normalizado de medida de la concentración de dióxido de azufre por fluorescencia de ultravioleta”.

- El método de referencia para la medición de dióxido de nitrógeno y óxidos de nitrógeno es el que se describe en la Norma UNE-EN 14211:2013 “Aire ambiente. Método normalizado de medida de la concentración de dióxido de nitrógeno y monóxido de nitrógeno por quimioluminiscencia”.

- El método de referencia para la toma de muestras y la medición de  $PM_{10}$  y  $PM_{2,5}$  es el que se describe en la Norma UNE-EN 12341:2015 “Aire ambiente. Método de medición gravimétrico normalizado para la determinación de la concentración másica  $PM_{10}$  o  $PM_{2,5}$  de la materia particulada en suspensión”.

- El método de referencia para la medición de ozono es el que se describe en la Norma UNE-EN 14625:2013 “Aire ambiente. Método normalizado de medida de la concentración de ozono por fotometría ultravioleta”.

- **¿Cuál es la calidad de la información recolectada por la red de monitoreo?**

Respecto a la calidad de los datos, de las 60 estaciones solo 25 cumplen con los siguientes criterios: presentar menos del 30% de mediciones diarias con al menos 12 mediciones horarias, a lo largo de más de 6 años y registrando como mínimo 3 contaminantes.

- **¿Hay redundancia de información entre las estaciones?**

Sí. Se observó una redundancia significativa entre las estaciones:

- Pla, Elx, Bur, Grau y Quart
- Alma y Ben
- Poli y Moli

- **¿Cuáles son las características del área de emplazamiento de las estaciones de monitoreo?**

Se observa que hay 3 grupos de estaciones bien diferenciadas:

- Superficie artificial (A): Pla, Raba, Grau, Quart, Pista, Poli y Moli
- Zonas agrícolas (B): Elx, Alcora, Bur, Viver, Alba, Caude y Villar
- Bosques y áreas seminaturales (C): Pobla, VCid y Zorita

Y un grupo de estaciones que responde a un uso de suelo mixto (Beni, Elda, Alma, Ben, Peny, SJ, Buniol y Sagunt).

- **¿Cómo influyen los vientos locales sobre los niveles registrados en las estaciones?**

Las situaciones de calma dificultan la dispersión de los contaminantes, favoreciendo los escenarios de mayores concentraciones de los contaminantes. En general, para todas las estaciones los vientos son más calmos durante el verano y más intensos en el invierno. Las estaciones con vientos medios más calmos fueron Pista ( $0.83 \text{ m.s}^{-1}$ ), Alma ( $0.93 \text{ m.s}^{-1}$ ) y Raba ( $0.95 \text{ m.s}^{-1}$ ).

La velocidad máxima de viento máximo para todo el período se registró en la estación Villar y alcanzó los  $24.8 \text{ m.s}^{-1}$ , mientras que la velocidad media máxima se registró en la estación Pobra ( $4.14 \text{ m.s}^{-1}$ ). En la tabla de valoración de vientos de Beaufort, estos son vientos tranquilos, considerados bajo la denominación de ventolinas.

### *Calidad del aire en los sitios de medición*

- **¿Cuáles son los niveles más frecuentes?**

Las estaciones que registran los niveles medios diarios más altos de  $\text{NO}_2$  para todo el período son Pista ( $40.51 \mu\text{g.m}^{-3}$ ), Quart ( $25.94 \mu\text{g.m}^{-3}$ ), Pla ( $23.10 \mu\text{g.m}^{-3}$ ) y Alma ( $20.44 \mu\text{g.m}^{-3}$ ), las mismas que para  $\text{NO}$  (Pista:  $25.07 \mu\text{g.m}^{-3}$ , Alma:  $15.16 \mu\text{g.m}^{-3}$ , Pla:  $10.36 \mu\text{g.m}^{-3}$  y Quart:  $9.39 \mu\text{g.m}^{-3}$ ). En todas las estaciones se presentaron eventos extremos de  $\text{NO}_2$ , y en algunos casos, alcanzando concentraciones límites establecidas por la normativa.

Las medias más elevadas de  $\text{O}_3$  fueron observadas en las estaciones Pobra ( $84.70 \mu\text{g.m}^{-3}$ ), Beni ( $78.91 \mu\text{g.m}^{-3}$ ), VCid ( $75.21 \mu\text{g.m}^{-3}$ ) y Villar ( $72.82 \mu\text{g.m}^{-3}$ ). En general, las medidas de  $\text{O}_3$  fueron elevadas para todas las estaciones.

La media más baja de  $\text{SO}_2$  ocurrió en la estación Alba ( $2.41 \mu\text{g.m}^{-3}$ ) mientras que la media máxima fue en Alma ( $6.33 \mu\text{g.m}^{-3}$ ), presentando un rango muy pequeño de variación entre las estaciones. Este contaminante presentó eventos extremos por encima de los  $200 \mu\text{g.m}^{-3}$  en las estaciones Alma, Grau y Pobra, estaciones ubicadas en zonas con actividad industriales.

Las medias diarias más elevadas de  $\text{PM}_{10}$  se ubicaron en Pista ( $24.59 \mu\text{g.m}^{-3}$ ), Poli ( $18.35 \mu\text{g.m}^{-3}$ ) y Alma ( $17.18 \mu\text{g.m}^{-3}$ ). Mientras que, las estaciones Viver ( $0.39 \mu\text{g.m}^{-3}$ ), Peny ( $0.39 \mu\text{g.m}^{-3}$ ) y Buniol ( $0.56 \mu\text{g.m}^{-3}$ ) presentaron las medias diarias menores, situándose muchas mediciones por debajo del límite de detección del equipo.

Las medias diarias más elevadas para  $\text{PM}_{2.5}$  corresponden a las estaciones Moli ( $13.30 \mu\text{g.m}^{-3}$ ), Poli ( $11.74 \mu\text{g.m}^{-3}$ ) y Elda ( $10.04 \mu\text{g.m}^{-3}$ ), mientras que las concentraciones medias diarias menores se dieron en Ben ( $7.26 \mu\text{g.m}^{-3}$ ), Zorita ( $7.66 \mu\text{g.m}^{-3}$ ) y Raba ( $7.76 \mu\text{g.m}^{-3}$ ).

- **¿La concentración de los contaminantes excede los límites legales?**

- $\text{NO}_2$ : la estación Pista fue la única estación que superó el límite anual reiteradamente (2009, 2010, 2013, 2014, 2015, 2016). Para este contaminante, el valor

límite horario se ha superado varias veces, pero en ningún caso más de las 18 veces por año recomendadas.

- $O_3$ : todas las estaciones superaron en algún momento del período analizado el promedio diario recomendado por la OMS, pero solo 14 de las 23 estaciones superaron el límite legal alguna vez en el período estudiado.

- $SO_2$ : los límites diarios no fueron superados y los límites horarios no fueron superados en ninguna estación por más de 18 veces en el año.

- $PM_{10}$ : el límite anual establecido por la UE no fue superado en ninguna ocasión. El límite horario se superó solo en las estaciones Pista y Alma para los años 2018 y 2009 respectivamente.

- $PM_{2.5}$ : Se observan pocos casos de superación del límite legal. En la estación Alma durante 2010 y 2011, y en la estación Moli durante 2011 y 2012

- ***¿La concentración pone en riesgo a la población?***

- $NO_2$ : solo en la estación Pista.

- $O_3$ : en todas las estaciones se supera lo recomendado por la OMS.

- $SO_2$ : El límite horario recomendado por la OMS se superó en Alcora, Alma, Pobra y Viver. En la estación Alma los excesos se repiten a lo largo de todo el período.

- $PM_{10}$ : Distintas estaciones han superado el valor recomendado por la OMS para la protección de la salud tanto anual como diario. En particular, la estación Pista superó estas recomendaciones varios años consecutivos.

- $PM_{2.5}$ : Todas las estaciones superan los niveles anuales recomendados por la OMS, con excepción de la estación Zorita. Las recomendaciones diarias son superadas por todas las estaciones.

### *Evolución temporal de los contaminantes*

- ***¿A cuál hora del día es mayor la concentración de los contaminantes?***

Para el  $NO$  y  $NO_2$  se observan picos en Pla, Alcora y Pista sobre las primeras horas del día, mientras que para Alma y Grau se observaron picos en las primeras horas de la tarde. El  $O_3$  alcanzó en todas las estaciones valores máximos al medio día. El mismo comportamiento se observó con las mediciones de  $SO_2$ , en particular para las estaciones Alcora, Alma, Grau, Zorita, Quart y Poli. Respecto a  $PM_{10}$ , en Pista, Villar y Raba se observaron máximos en la mañana, mientras que para Viver, Buniol, Poli y Moli los máximos fueron por la tarde. En el caso de  $PM_{2.5}$ , se observaron niveles máximos cercanos a la media noche.

- ***¿Las concentraciones máximas varían a lo largo de la semana?***

Los contaminantes asociados con el uso de combustible ( $NO$ ,  $NO_2$ ,  $SO_2$  y  $PM_{10}$ ) mostraron máximas concentraciones durante los días laborales, descendiendo sus niveles para los fines de semana. El  $O_3$  mostró valores máximos los fines de semana, lo cual puede explicarse por el equilibrio dinámico que mantiene con los  $NO_x$ . Finalmente, se observó una acumulación de  $PM_{2.5}$  a lo largo de la semana, siendo el viernes el día con mayores concentraciones.

- ***¿En cuál época del año ocurren los episodios de máximas concentraciones?***



- Para el NO y NO<sub>2</sub> se observa un aumento de las concentraciones en los meses fríos (noviembre, diciembre, enero). Esto se ve mayoritariamente en la estación Pista.
  - Por el contrario, el O<sub>3</sub> tiene su máximo en los meses cálidos con es mayor la radiación solar.
  - Las concentraciones de SO<sub>2</sub> son mayores en los meses fríos, probablemente debido a una mayor emisión de este contaminante.
  - La estacionalidad no es muy visible para PM<sub>10</sub> y PM<sub>2.5</sub> pero existe una tendencia a presentarse mayores concentraciones en los meses cálidos.
- **¿Cómo han cambiado los niveles de los contaminantes con los años?**  
Mientras las concentraciones de NO decrecieron para gran parte de las estaciones, solo 10 de las 21 estaciones mostraron una tendencia negativa para el NO<sub>2</sub>. Para el O<sub>3</sub> la tendencia fue negativa solo en Raba, mientras que en 9 estaciones fue positiva. La tendencia del SO<sub>2</sub> fue negativa en 7 estaciones y positiva en 3 (Raba, Alcora y Bur). Para PM<sub>10</sub> y PM<sub>2.5</sub> la tendencia fue mayoritariamente decreciente. Reviste especial interés la estación Moli que presentó una tendencia en aumento de las concentraciones de PM<sub>2.5</sub> y donde no se monitorea NO<sub>x</sub>.

### *Distribución espacial de PM<sub>2.5</sub>*

- **¿Dónde ocurren las máximas concentraciones de PM<sub>2.5</sub>?**  
En función de los mapas de concentración de PM<sub>2.5</sub> elaborados, las concentraciones medias anuales mayores durante el 2017 se encuentran en Elche 13.68 µg.m<sup>-3</sup>, Valencia 12.03 µg.m<sup>-3</sup>, Alcira 11.31 µg.m<sup>-3</sup> y en Benidorm 11.37 µg.m<sup>-3</sup>.
- **¿Se observa un patrón espacial de las máximas concentraciones?**  
Las máximas concentraciones se observaron en los núcleos urbanos de Castellón de la Plana, Valencia, Alcira, Elche y Alicante.
- **¿Cómo es la calidad del aire donde no se mide?**  
En particular, resultan llamativos los niveles modelados para Elche los cuales son muy elevados (en 2017 una media anual de 13.68 µg.m<sup>-3</sup>).
- **¿La calidad del aire representa un riesgo para la población?**  
La calidad del aire para la Comunidad Valenciana es buena en zonas poco pobladas, y moderada sin resultar dañina para la población en las zonas urbanas. La calidad del aire mejora al alejarnos de los núcleos urbanos principales.

*Caso de aplicación II*

**Ciudad de Buenos Aires,  
Argentina**

---



## Introducción

Como segundo caso para la aplicación de la propuesta se eligió la Ciudad Autónoma de Buenos Aires, capital nacional de la República Argentina. Esta elección se debe a la importancia política, social, cultural, histórica y económica que tiene esta ciudad para Argentina.

Argentina presenta pocos estudios sobre la calidad del aire [Lijteroff et al., 1999; Mazzeo et al., 2005; Colman Lerner et al., 2014b]. En el mayor de los casos, las mediciones con las que se cuenta pertenecen a campañas científicas las cuales fueron pensadas para estudiar eventos particulares [García Ferreyra, 2013; Puliafito et al., 2010].

Estos trabajos de investigación han abordado cuestiones diversas a lo largo del país, como la evaluación del impacto de fuentes específicas a escala local [Olcese & Toselli, 1997; Tombesi et al., 2014; Puliafito et al., 2010; Wichmann et al., 2009; Massolo et al., 2010; Colman Lerner et al., 2014a]; caracterización de la composición de aerosoles [Sosa et al., 2017; Achad et al., 2014; López et al., 2011; Vasconcellos et al., 2011; Giuliani et al., 2018]; estimación de inventarios de emisiones [Puliafito et al., 2015; Mazzeo et al., 2005; D'Angiola et al., 2010; Allende et al., 2012]; implementación de modelos de transporte de contaminantes [Abril et al., 2016; Venegas y Mazzeo, 2006; Allende et al., 2012; García Ferreyra, 2016]; desarrollo de estudios de biovigilancia [Carreras & Pignata, 2002; Rodríguez et al., 2011; Estrabou et al., 2011] e identificación de fuentes mediante modelos receptores [Puliafito et al., 2011]. Sin embargo, no proporcionan una imagen de conjunto que permita evaluar la calidad del aire en el país.

Como caso excepcional en el país, Buenos Aires es la única ciudad con un monitoreo de calidad del aire sostenido a lo largo de los años. La Agencia de Protección Ambiental (APrA) del Gobierno de la Ciudad de Buenos Aires evalúa la calidad del aire desde 2009, mediante una red de monitoreo compuesta por tres estaciones y que informan las concentraciones de NO<sub>2</sub>, CO y PM<sub>10</sub> con una frecuencia horaria.

Buenos Aires es la ciudad más importante del país. Alberga casi 3 millones de habitantes de forma permanente [INDEC, 2010], a ello deben adicionarse 1 millón de trabajadores del Gran Buenos Aires que ingresan diariamente a CABA [Ciudad de Buenos Aires, 2019]. La economía de la ciudad de Buenos Aires está basada principalmente en el sector servicios (65% de las empresas), siendo los principales rubros el comercio, los servicios inmobiliarios y los servicios financieros [CAC, 2011]. Sin embargo, también cuenta con industrias farmacéuticas y de fabricación de coque, productos de la refinación de petróleo, combustible nuclear y la fabricación de productos elaborados de metal [Ciudad de Buenos Aires, 2014].

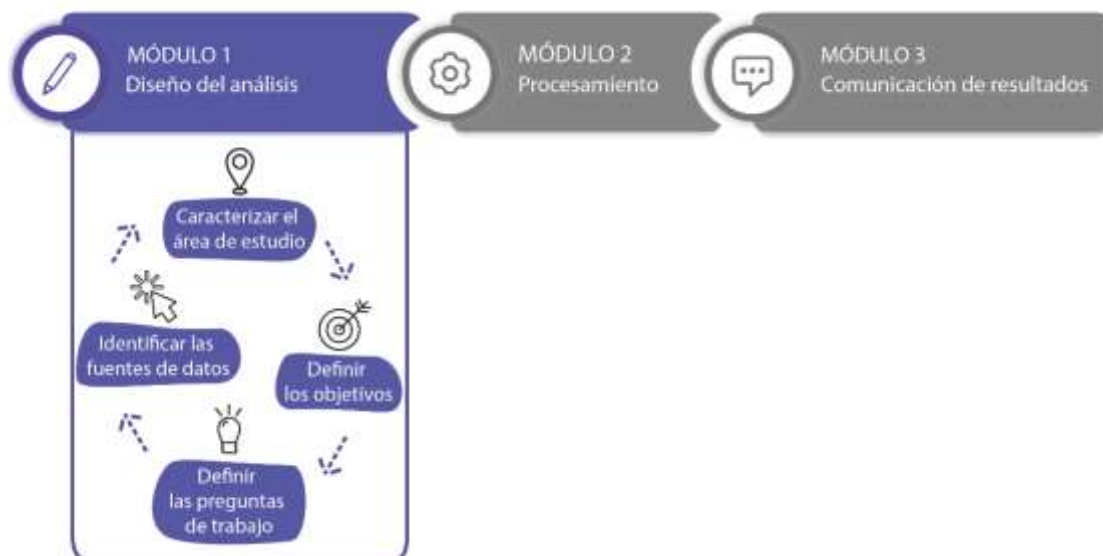
Alrededor de 1.6 millones de vehículos circulan diariamente en la ciudad, creciendo de forma sostenida en el área metropolitana [AFAC, 2018]. Según el inventario presentado por [D'Angiola et al., 2010], las emisiones del tráfico son responsables del 72% de las emisiones totales de partículas inhalables, seguidas de las de las centrales eléctricas, que representan el 18% [Resquin et al., 2018].

Finalmente, podemos señalar algunos trabajos que describen efectos en la salud debido a la exposición de contaminantes atmosféricos [Orona et al., 2016; Abrutzky et al., 2012], poniendo en evidencia el peligro potencial que la calidad del aire presenta para los habitantes de la ciudad.

En base a lo antes mencionado, nos interesa poner a prueba la metodología desarrollada con el fin de realizar un análisis exhaustivo de la información recolectada en la ciudad, y generar nueva información que sea de utilidad para el público.

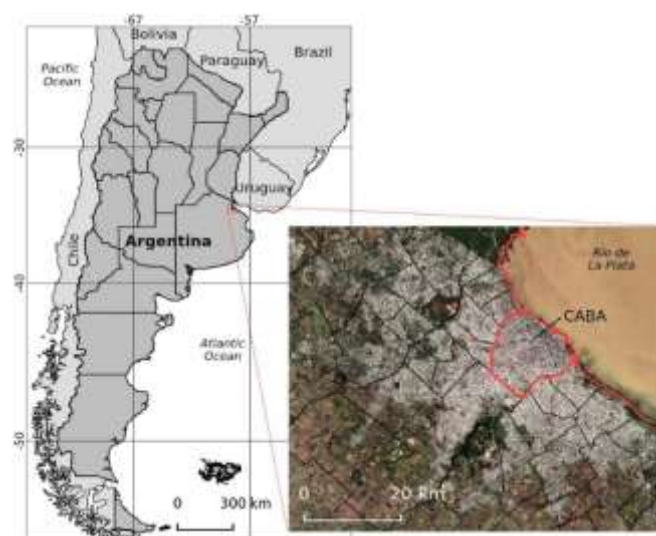
# Aplicación de la metodología

## Módulo I - Etapa de diseño del análisis



### I. Definición y caracterización del área de estudio

La ciudad de Buenos Aires, capital de Argentina, está ubicada a  $34.61^{\circ}\text{S}$ ,  $58.37^{\circ}\text{O}$ , en la ribera oeste del Río de la Plata (**Figura 93**). Se trata de un estuario de llanura costera poco profundo y de gran escala que cubre una superficie aproximada de  $35000 \text{ km}^2$  [Licursi et al., 2006]. Todo el terreno es llano, con pequeñas diferencias de altura de menos de 30 m.



**Figura 93.** Mapa de la Ciudad de Buenos Aires (CABA) y el Área Metropolitana de Buenos Aires

La región presenta un clima subtropical, donde la Cordillera de los Andes ejerce una influencia importante en el clima regional. Ubicada a 900 km de la capital, con una altura promedio de más de 4000 m, impide el paso de la humedad del Océano Pacífico hacia Argentina desde el sector oeste [Barros et al., 2015].

Los vientos derivan predominantemente del N y del NE [Barros et al., 2006; Resquin et al., 2018], como resultado de la influencia de los sistemas semipermanentes, incluyendo el sistema de alta presión del Atlántico Sur (SAS), que trae aire húmedo y a veces caliente del E y del N y eventos sinópticos y de mesoescala tales como frentes y tormentas convectivas [Garreaud y Falvey, 2008], provocando una marcada variación climatológica estacional [Barros et al., 2015].

La Ciudad de Buenos Aires está inmersa en un área urbana continua, conocida como el Área Metropolitana de Buenos Aires. Con una densidad de población de 3600 habitantes por km<sup>2</sup>, es la zona más poblada del país [INDEC, 2010], la décima megalópolis del mundo y la tercera de América Latina. Arkouli et al. [2010] han señalado que las principales fuentes de contaminación atmosférica en la ciudad de Buenos Aires son el tráfico y el centro industrial que incluye centrales termoeléctricas al sureste de la ciudad.

Una buena ventilación y dilución de los contaminantes han sido frecuentemente asociados con la planicie del terreno de la ciudad, su ubicación costera y los vientos moderados medios superficiales provenientes del Río de la Plata [Arkouli et al., 2010].

### *Contexto jurídico-normativo*

En primer lugar, la República Argentina cuenta con la Constitución Nacional, máximo instrumento legal de la Nación. En su Art. 41 se establece el “derecho a un ambiente sano, equilibrado, apto para el desarrollo humano y para que las actividades productivas satisfagan las necesidades presentes sin comprometer las de las generaciones futuras; y tienen el deber de preservarlo”. En el tercer párrafo se introduce un concepto novedoso al marco de la distribución de competencias Nación-Provincias. Se lee: “Corresponde a la Nación dictar las normas que contengan los presupuestos mínimos de protección, y a las provincias, las necesarias para complementarlas, sin que aquéllas alteren las jurisdicciones locales”.

En base a esto, la Ley Nacional N° 25.675 conocida como la “Ley General del Ambiente” establece el ambiente como un bien jurídicamente protegido. En su primer artículo, establece su finalidad, el cual es establecer “los presupuestos mínimos para el logro de una gestión sustentable y adecuada del ambiente, la preservación y protección de la diversidad biológica y la implementación del desarrollo sustentable”.

Cabe resaltar que la competencia ambiental es preeminentemente local, dado el Art. 124 de la Constitución Nacional que explicita el dominio de las provincias sobre sus recursos naturales. Sin embargo, los presupuestos mínimos de protección ambiental implican una delegación de ciertas facultades de las provincias a la Nación, como son controlar el impacto ambiental de todas las actividades que perjudiquen al ecosistema, promover acciones que eviten la contaminación del aire, garantizar el derecho a solicitar y recibir la adecuada información y participar en la defensa del ambiente, de los recursos naturales y culturales, entre otras.

Por su parte, la Ciudad de Buenos Aires se constituye como ciudad autónoma, tal como fue incorporada en la Constitución Nacional en su Art. 129. Por esto, CABA presenta su propia ley

regulatoria en materia de preservación del recurso aire y la prevención y control de la contaminación atmosférica, la Ley 1356/04.

Entre otras cuestiones, la Ley 1356/04 establece los estándares de calidad del aire (**Tabla 18**) y crea un programa de monitoreo permanente de contaminantes atmosféricos. En base a esto, se implementa para la Ciudad de Buenos Aires una red de monitoreo de la calidad del aire.

**Tabla 18.** Estándares de calidad de aire ambiente para la Ciudad de Buenos Aires

Contaminante	Concentración		Período
	mg.m <sup>-3</sup>	ppm	
SO <sub>2</sub>	0.08	0.03	Media aritmética anual
	0.365	0.14	Promedio 24 horas
PM <sub>10</sub>	0.050	-	Media aritmética anual
	0.15	-	Promedio 24 horas
PM <sub>2.5</sub>	0.015	-	Media aritmética anual
	0.065	-	Promedio 24 horas
CO	10	-	Promedio 8 horas
	40	-	Promedio 1 hora
O <sub>3</sub>	0.157	0.08	Promedio 8 horas
	0.235	0.12	Promedio 1 hora
NO <sub>2</sub>	0.100	0.053	Media aritmética anual

## I.II. Determinación de los objetivos

Para la Ciudad de Buenos Aires existe una acotada provisión de información que limita los tipos de análisis que se pueden realizar.

Para este trabajo se toman como objetivos:

- *Caracterizar aspectos del monitoreo terrestre*

Descripción general de las estaciones. Calidad de los datos. Características del sitio de emplazamiento. Estudiar las direcciones de viento que contribuyen con mayores concentraciones de los contaminantes

- *Determinar la calidad del aire en los sitios de medición*

Evaluar si las concentraciones respetan las recomendaciones de la OMS y si se encuentran dentro de los parámetros fijados por ley.

- *Conocer la evolución temporal de los contaminantes*



Calcular la tendencia de las concentraciones a largo plazo. Estudiar la tendencia en función de las direcciones del viento. Evaluar la presencia de eventos extremos en el período estudiado. Analizar el comportamiento diario, semanal e interanual de los contaminantes.

- *Evaluar la distribución espacial de los contaminantes*

Comparar las concentraciones medidas entre las estaciones. Confeccionar mapas de  $PM_{10}$  mediante información satelital. Analizar el comportamiento espacial de  $PM_{10}$ .

### I.III. Delimitación de las preguntas de estudio

A continuación, se detallan las preguntas que se pueden responder a partir de la información existente a la fecha:

#### **Caracterizar aspectos del monitoreo terrestre**

- ¿Cuál es la calidad de la información recolectada por la red de monitoreo?
- ¿Hay redundancia de información entre las estaciones de monitoreo?
- ¿Cuáles son las características del área de emplazamiento de las estaciones de monitoreo?
- ¿Cómo influyen los vientos locales sobre los niveles registrados en las estaciones?
- ¿De qué sector del viento llegan las concentraciones mayores?

#### **Determinar la calidad del aire en los sitios de medición**

- ¿Cuáles son los niveles más frecuentes?
- ¿La concentración de los contaminantes excede los límites legales?
- ¿La concentración registrada de los contaminantes pone en riesgo a la población?

#### **Conocer la evolución temporal de los contaminantes**

- ¿A cuál hora del día es mayor la concentración de los contaminantes?
- ¿Las concentraciones máximas varían a lo largo de la semana?
- ¿Durante cuál época del año ocurren los episodios de máximas concentraciones?
- ¿Los contaminantes presentan un comportamiento asociado?
- ¿Cómo han cambiado los niveles de los contaminantes con los años?

#### **Evaluar la distribución espacial de los contaminantes**

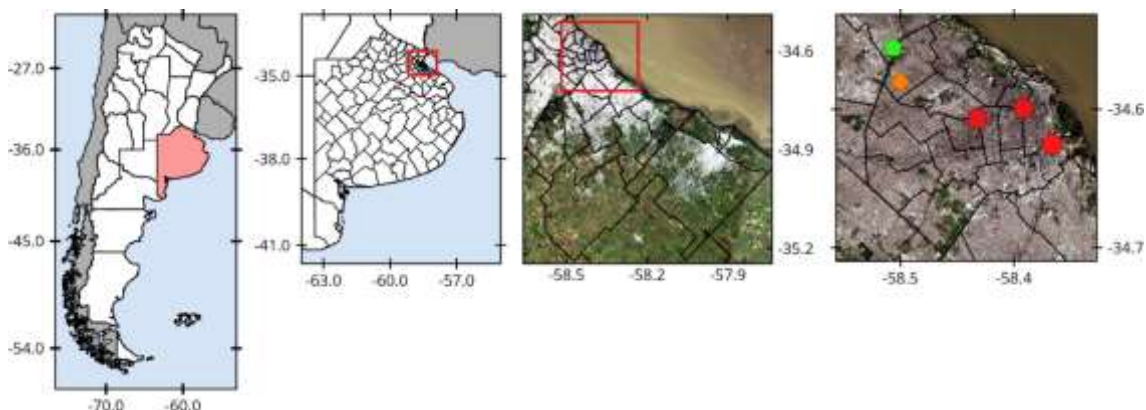
- ¿Dónde ocurren las máximas concentraciones?
- ¿Las máximas concentraciones ocurren en diferentes sitios a distintos tiempos?
- ¿Cómo es la calidad del aire donde no se mide?

### I.IV. Adquisición de los datos

#### *Estaciones de monitoreo terrestre*

Los datos de calidad del aire fueron descargados del sitio oficial del gobierno de la Ciudad de Buenos Aires (CABA) a través del sitio web (<https://data.buenosaires.gob.ar/>, visitado por última vez en marzo de 2018). Los datos horarios comunicados por Agencia de Protección Ambiental (APrA) para la concentración de CO corresponden a la media móvil de las 8 horas anteriores; para  $NO_2$ , los 60 minutos anteriores, y para  $PM_{10}$  las 24 horas anteriores. El período analizado se definió desde el 1 de octubre del 2009 (comienzo del monitoreo continuo) hasta el 1 de octubre del 2017 (fecha de descarga y procesamiento de los datos).

La red de monitoreo de la calidad del aire está compuesta por tres estaciones tipo EPA (**Figura 94**), que proporcionan datos horarios de  $\text{NO}_2$ ,  $\text{CO}$  y  $\text{PM}_{10}$ . Las técnicas de análisis utilizadas son: quimioluminiscencia en fase gaseosa para  $\text{NO}_2$ , fotometría infrarroja no dispersiva para  $\text{CO}$  y monitorización de atenuación beta para  $\text{PM}_{10}$ .



**Figura 94.** Mapa de Ciudad Autónoma de Buenos Aires con los sitios de medición de calidad de aire (puntos rojos), la estación meteorológica (rombo naranja) y estación de AERONET CEILAP (pentágono verde). Mapa base: Imagen Sentinel-2, RGB (4, 3, 2).

## AERONET

La estación AERONET más cercana se encuentra en CEILAP, a unos 15 km al norte del centro de la ciudad de Buenos Aires ( $34.56^\circ$  S,  $58.51^\circ$  W). Los datos se descargaron del sitio oficial de AERONET (<https://aeronet.gsfc.nasa.gov>) con Nivel 2.0 (datos sin nubes, calibrados e inspeccionados manualmente).

### *Datos meteorológicos*

La información meteorológica que registran las estaciones de monitoreo no se encontraba disponible debido a problemas de calibración de los sensores [comunicación personal APrA]. Por tanto, se recurrió a los datos meteorológicos de superficie por hora para la velocidad y dirección del viento provistos por el Servicio Meteorológico Nacional [SMN, 2017]. El conjunto de datos SMN proviene de la estación del Observatorio de Buenos Aires ( $34.58^\circ$  S,  $58.50^\circ$  W) que es la estación meteorológica más cercana, ubicada a unos 7, 10 y 13 km de las estaciones de monitoreo de la calidad del aire (**Figura 94**).

### *Datos modelados de atmósfera*

Los datos de las condiciones de la atmósfera y contaminantes modelados se descargaron del sitio web de MERRA-2 (<https://disc.gsfc.nasa.gov/daac-bin/FTPSubset2.pl>). El MERRA-2 es un reanálisis atmosférico utilizando el GEOS-5 con su ADAS, versión 5.12.4, a una resolución espacial de  $0.5^\circ \times 0.625^\circ$  [Gelaro et al., 2017]. Los datos brindan un análisis climático histórico para una amplia gama de escalas de tiempo [Dee et al., 2011]. Como estos datos se utilizaron para

los modelos multivariantes, se descargaron los sets de datos de las horas más cercanas al momento del paso del satélite.

Los productos de MERRA-2 presenta simulaciones de cinco especies de  $PM_{2.5}$  (carbono negro, carbono orgánico, sal marina, polvo, sulfato) basándose en el modelo de aerosol GOCART (del inglés, *Goddard Chemistry Aerosol Radiation and Transport*) [Chin, 2005]. El producto del modelo se llama `tavg1_2d_aer_Nx`. Las variables seleccionadas fueron las concentraciones máxicas de sulfuro de dimetilo, polvo, carbón orgánico, sal marina, sulfatos y dióxido de azufre, para el intervalo de tiempo de 13.30 - 16.30 UTC.

También se trabajó con los modelos de MERRA-2 de la altura de la capa límite, flujo total de precipitación superficial, velocidad de viento y velocidad máxima de viento sobre la superficie (producto `tavg1_2d_flg_Nx` para el intervalo 10.30 - 13.30 UTC). Se incorporaron las variables radiométricas como porcentaje del albedo y fracción de área de nube para nubes altas y bajas del producto `tavg1_2d_rad_Nx` para el intervalo 10.30 - 13.30 UTC); y datos de superficie como presión de superficie, humedad relativa, componentes del viento (u,v), temperatura del aire (`inst3_3d_asm_Np` para el intervalo horario 12.00 - 15.00 UTC).

### *Imágenes satelitales utilizadas*

El producto de datos MCD19A2 Versión 6 es un producto combinado de MODIS Terra y Aqua con la implementación del algoritmo *Multi-angle Implementation of Atmospheric Correction* (MAIAC) *Land Aerosol Optical Depth* (AOD) en cuadrícula de Nivel 2 producido diariamente con una resolución de 1 km de píxel. El producto de datos MCD19A2 se descargó del sitio oficial de LAADS <https://ladsweb.modaps.eosdis.nasa.gov/> en formato TIFF como un mosaico para todo el Área Metropolitana de Buenos Aires, tomando un polígono de coordenadas (59° S, 34.3° W, 57.8° S, 35° W).

## Módulo II - Etapa de procesamiento



Para el análisis de los datos se utilizaron algunas de las herramientas descritas en el **Capítulo 3** “Propuesta metodológica para el estudio de la calidad del aire”. La inclusión de las mismas se realizó en función de los datos disponibles para la Ciudad de Buenos Aires.

### I. Preparación de los datos

La preparación de los datos requirió la adecuación del formato de fecha, la verificación de las series de tiempo y la generación de series de tiempo agregadas en medias diarias, mensuales y anuales. Se mantuvieron las unidades de concentración reportadas por APrA, las cuales fueron  $\mu\text{g.m}^{-3}$  para  $\text{PM}_{10}$ , ppm para CO, y ppb para  $\text{NO}_2$ .

### II. Selección de algoritmos para el análisis

Como primer paso, se realizó un análisis de la calidad de los datos evaluando la continuidad temporal de las mediciones. Como indicador se trabajó con el porcentaje de datos faltantes a lo largo del período de estudio. Se utilizaron visualizaciones gráficas para facilitar la evaluación de este parámetro.

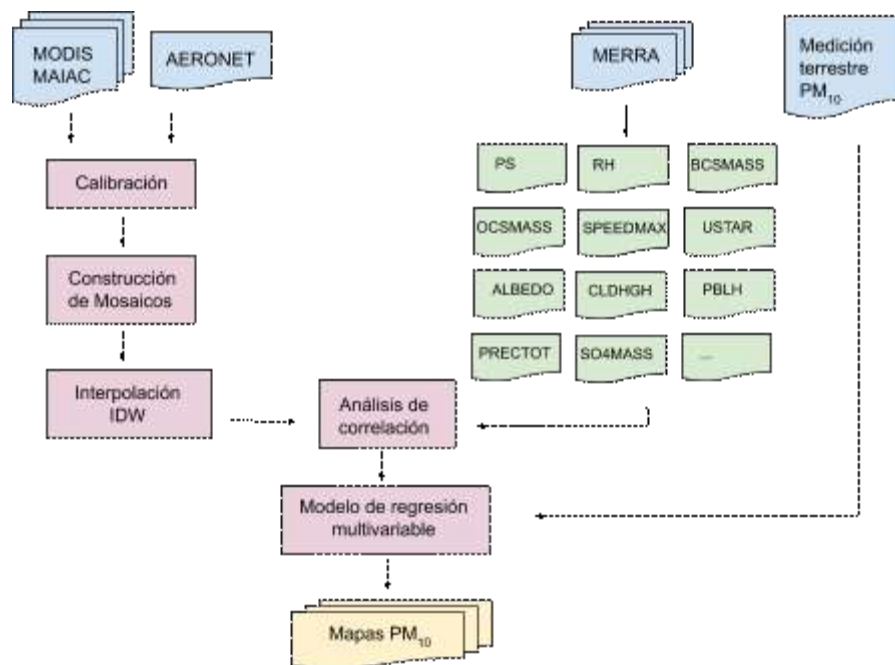
Para conocer el grado de redundancia entre las mediciones se utilizó el factor de correlación de Pearson y se aplicó un análisis de conglomerados jerárquicos entre los coeficientes de correlación de las estaciones. Los resultados fueron expuestos mediante un dendograma.

El análisis del área de influencia se realizó en base a la información obtenida por un análisis visual del área aledaña a las estaciones de monitoreo. Para comprender la dinámica de los vientos, se analizaron las frecuencias de los vientos para el área de estudio a nivel diario y estacional mediante la elaboración de rosas de vientos.

El análisis estadístico descriptivo de las mediciones de la calidad del aire se realizó con los datos horarios. La visualización de las series de tiempo permitió la detección de eventos particulares en los que las concentraciones de contaminantes alcanzaron valores extremadamente altos. Estos eventos atípicos fueron contextualizados, en la medida de lo posible, con la bibliografía relevada.

Para evaluar la variación interanual, se calcularon las tendencias de las concentraciones mediante un método no paramétrico basado en la pendiente de Mann-Kendall con el método de Theil-Sen [Masiol et al., 2017]. El intervalo de confianza para la estimación de la pendiente se obtuvo con un 95% de significancia.

Para la elaboración de mapas de contaminación atmosférica se realizó un modelo de regresión multivariable para el período comprendido entre 2009-10-01 y 2017-10-01. En la **Figura 95** se puede ver un resumen de la metodología llevada adelante.



**Figura 95.** Resumen de la metodología desarrollada para confeccionar mapas de PM<sub>10</sub>

El producto satelital utilizado fue MCD19A2 de MODIS que tiene una resolución espacial de 1 km. Se descargaron un total de 12055 imágenes para un mosaico que comprende la Ciudad de Buenos Aires y el área metropolitana hasta la ciudad de La Plata. Los productos se descargaron en formato TIFF con un sistema de referencia WGS84.

Para garantizar la máxima calidad en las medidas de AOD se aplicó como máscara la capa SDS de AOD AQ. Esta capa se encuentra codificada como un número binario de 16 bits. Su utilización requirió la recodificación de la capa como una capa booleana.

Para la calibración de AOD determinado a 550 nm se utilizaron los datos provistos por la red AERONET recolectados en la estación CEILAP. La calibración se realizó mediante una

interpolación cuadrática de las mediciones de AOD utilizando las longitudes de onda 440, 500 y 675 nm.

Una vez obtenido el conjunto de todas las imágenes calibradas y recortadas para el área de interés, se procedió a completar los valores faltantes. Solo las imágenes con menos de un 30% de datos faltantes fueron completadas. El método empleado para la interpolación espacial fue el IDW.

Las variables incorporadas del modelo multivariante pueden verse en la **Tabla 19**, de las cuales una gran parte de estas son parámetros aportados por el modelo de revisión MERRA-2. En todos los casos donde fue necesario, la información de entrada se recortó al área de análisis y se adaptó a la grilla de MAIAC de dimensiones 120 filas por 168 columnas con una resolución de 0.008°, mediante un remuestreo utilizando el método de interpolación del vecino más cercano.

**Tabla 19.** Lista completa de variables accesorias analizadas para modelar PM<sub>10</sub>.

Código	Variable	Unidades	Resolución espacial	Resolución temporal
<i>AOD</i>	AOD	-	1km x 1km	Diaria
<i>BCSMAS</i>	Concentración másica superficial de carbono negro	kg.m <sup>-3</sup>	0.5° x 0.625°	3 horas
<i>DMSSMAS</i>	Concentración másica superficial de sulfuro de dimetilo	kg.m <sup>-3</sup>	0.5° x 0.625°	3 horas
<i>DUSMAS</i>	Concentración másica superficial de polvo	kg.m <sup>-3</sup>	0.5° x 0.625°	3 horas
<i>DUSMAS25</i>	Concentración másica superficial de polvo de tamaño PM <sub>2.5</sub>	kg.m <sup>-3</sup>	0.5° x 0.625°	3 horas
<i>OCSMAS</i>	Concentración másica superficial de carbono orgánico	kg.m <sup>-3</sup>	0.5° x 0.625°	3 horas
<i>SO2SMAS</i>	Concentración másica superficial de SO <sub>2</sub>	kg.m <sup>-3</sup>	0.5° x 0.625°	3 horas
<i>SO4SMAS</i>	Concentración másica superficial de SO <sub>4</sub>	kg.m <sup>-3</sup>	0.5° x 0.625°	3 horas
<i>SSSMAS</i>	Concentración másica superficial de sal marina	kg.m <sup>-3</sup>	0.5° x 0.625°	3 horas
<i>SSSMAS25</i>	Concentración másica superficial de sal marina tamaño PM <sub>2.5</sub>	kg.m <sup>-3</sup>	0.5° x 0.625°	3 horas
<i>PBLH</i>	Altura de la capa límite planetaria	m	0.125° x 0.125°	3 horas
<i>PRECTOT</i>	Precipitación superficial total	kg.m <sup>-2</sup> .s <sup>-1</sup>	0.125° x 0.125°	3 horas
<i>SPEED</i>	Velocidad superficial del viento	m.s <sup>-1</sup>	0.125° x 0.125°	3 horas
<i>SPEEDMAX</i>	Velocidad superficial máxima del viento	m.s <sup>-1</sup>	0.125° x 0.125°	3 horas
<i>USTAR</i>	Escala de velocidad de superficie	m.s <sup>-1</sup>	0.125° x 0.125°	3 horas
<i>H1000</i>	Altura a 1000 hPa	m	0.125° x 0.125°	1 hora

<i>ALBEDO</i>	Surface albedo	%	0.125° x 0.125°	6 horas
<i>CLDHGH</i>	Fracción de área cubierta por nubes bajas	%	0.125° x 0.125°	3 horas
<i>CLDLOW</i>	Fracción de área cubierta por nubes altas	%	0.125° x 0.125°	3 horas
<i>PS</i>	Presión en superficie	hPa	0.125° x 0.125°	3 horas
<i>T</i>	Temperatura del aire a 1000 hPa	K	0.125° x 0.125°	3 horas
<i>RH</i>	Humedad relativa a 1000 hPa	%	0.125° x 0.125°	3 horas
<i>U</i>	Componente de viento hacia el este	m.s <sup>-1</sup>	0.125° x 0.125°	3 horas
<i>V</i>	Componente de viento hacia el norte	m.s <sup>-1</sup>	0.125° x 0.125°	3 horas

Previo a ser incluidas en el modelo, se hizo un análisis de correlación para evitar la generación de resultados espurios por información redundante. Esto permitió descartar un amplio porcentaje de las variables consideradas inicialmente.

Se evaluó la correlación entre AOD y PM<sub>10</sub> y se confeccionó un modelo lineal. En base a los resultados, se evaluó el ajuste de un modelo lineal multivariable y, finalmente, se generó un modelo Random Forest. El grado de ajuste se comparó utilizando R<sup>2</sup>, RMSE y MAE con un set de datos de prueba (20% de los datos totales).

Luego de obtenidos los mapas de PM<sub>10</sub>, se realizó un análisis del comportamiento espacial. Para esto, se confeccionaron mapas de medias anuales y mapas de medias anuales promedio para cada comuna de la ciudad y se compararon los niveles modelados para cada área.

Como último paso, se generaron mapas de riesgo relativo y de factor atribuible de muerte (FAM) por exposición a PM<sub>10</sub>. El riesgo relativo se calculó como:

$$RR = e^{\beta (X - X_0)}$$

donde  $\beta = 0.0008$  para todas las causas de muerte para todas las edades y  $x_0 = 10 \mu\text{g}\cdot\text{m}^{-3}$  de PM<sub>10</sub> y la FAM como:

$$FAM = \frac{(RR - 1)}{RR}$$

Todas las elaboraciones estadísticas de este estudio se realizaron utilizando el lenguaje de programación R [R Core Team, 2017], trabajando principalmente con los paquetes *open-air* [Carslaw y Ropkins, 2012], *dplyr* [Wickham, 2019], *ggplot2* [Wickham, 2009], *raster* [Hijmans, 2019], *lubridate* [Grolemund & Wickham 2011], *dplyr* [Wickham et al., 2019], *foreign* [R Core Team, 2018], *reshape2* [Wickham, 2007], *corrplot* [Wei & Simko, 2017], *colorRamps* [Keitt, 2012], *rgdal* [Bivand et al., 2019]. Se utilizó el paquete *RandomForest* [Liaw y Wiener, 2002] para el entrenamiento y estimación de RF, y el paquete *caret* [Kuhn, 2019] para la elaboración de los sets de entrenamiento y de prueba. Para los procesos intensivos de cálculo se utilizó el paquete

*parallel* que permite realizar cálculos en paralelo aprovechando los recursos informáticos multinúcleo [R Core Team, 2018b].



## Módulo III - Etapa de comunicación de los resultados

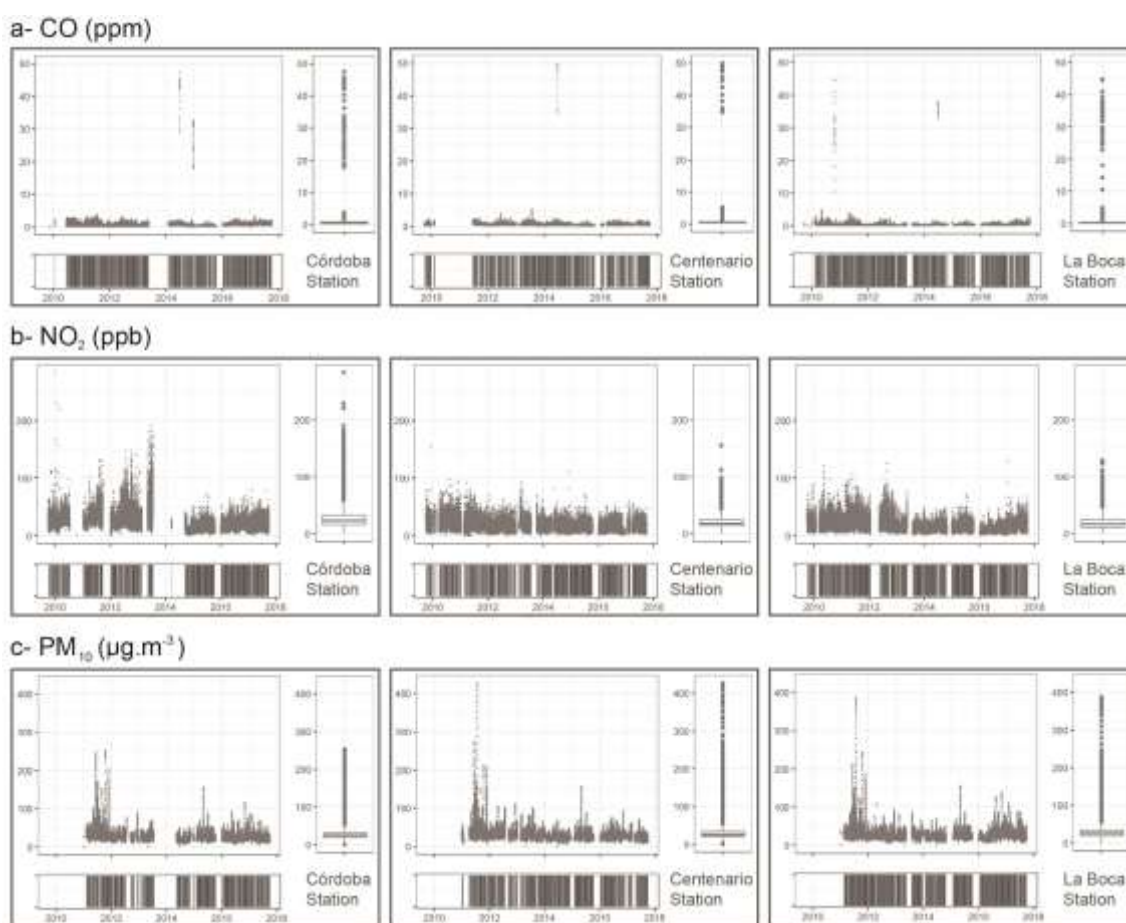


### III.I. Resultados

#### *I. Continuidad temporal de las mediciones*

En este estudio se analizaron las concentraciones de tres contaminantes (CO, NO<sub>2</sub> y PM<sub>10</sub>) registrados en 3 sitios de fondo urbano en la ciudad de Buenos Aires. El período analizado incluye las mediciones realizadas entre el 1 de octubre de 2009 y el 1 de octubre de 2017, con un total de 70115 registros de por contaminante para cada estación de monitoreo.

La **Figura 96** muestra las series cronológicas de los contaminantes medidos y la disponibilidad de los datos de las tres estaciones. Analizando los registros, se pueden observar largos períodos de ausencia de mediciones. La estación de Córdoba tiene el menor número de registros (sólo el 65% en todo el período), mientras que la estación de La Boca es la más completa (76%). La mayor falta de mediciones se dio en la estación de Córdoba durante el año 2013, faltando más del 50% de los datos horarios.



**Figura 96.** Serie de tiempo de cada uno de los contaminantes en las estaciones de la Ciudad de Buenos Aires

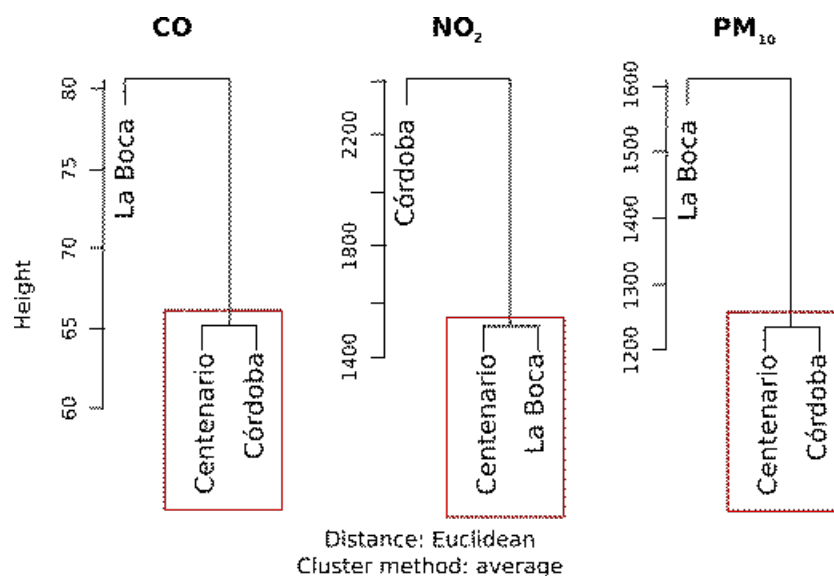
También se observó una alta proporción de datos faltantes para los últimos meses de 2015 en todas las estaciones. En particular, las mediciones de PM<sub>10</sub> sólo se encuentran disponibles a partir de 2011 (**Figura 96, c**). También, resalta la presencia de casos extremos los cuales se verán con mayor detalle en la próxima sección.

## II. Análisis de redundancia

Dado la proximidad entre las estaciones, uno de los principales intereses ha sido evaluar la calidad de la información recolectado por la red de monitoreo en función de la redundancia registrada.

En un primer paso, se evaluó la fuerza de la relación entre las estaciones para cada uno de los contaminantes calculando el coeficiente de correlación de Pearson. En todos los casos, las correlaciones fueron significativas y altas. Para el CO, las tres estaciones mostraron una  $r > 0.99$  ( $p < 0.05$ ). Lo mismo ocurrió para el PM<sub>10</sub>, donde  $r > 0.85$  en todos los casos ( $p < 0.05$ ). Una correlación más baja presentó NO<sub>2</sub>, donde la relación más alta fue entre Centenario y La Boca (0.69,  $p < 0.05$ ) y la más baja entre Centenario y Córdoba (0.56,  $p < 0.05$ ).

Tras esto se aplicó un análisis de conglomerados jerárquicos para agrupar los sitios de monitoreo en función de la similitud de los valores de concentración de los contaminantes (**Figura 97**). CO y PM<sub>10</sub> mostraron resultados similares, donde las estaciones de Centenario y Córdoba constituyeron un solo grupo. Sin embargo, fueron las estaciones de Centenario y La Boca las que formaron un cluster para NO<sub>2</sub>.



**Figura 97.** Dendrogramas para cada contaminante. En el eje vertical se observa la distancia euclídea. En el recuadro las asociaciones significativas entre las estaciones.

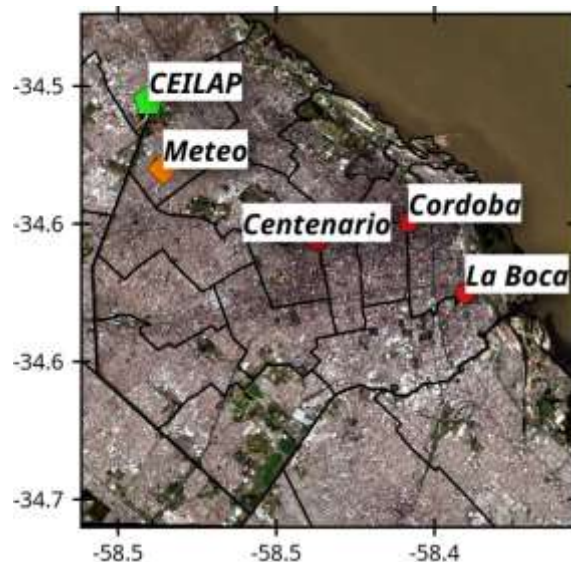
Los resultados indican que las estaciones realizaron mediciones similares para los diferentes contaminantes, con patrones únicos para los tres contaminantes a la vez. Es decir, la información que se recolecta presenta redundancia por contaminantes. Este es el caso de Centenario y Córdoba para CO y PM<sub>10</sub>, y para Centenario y La Boca para NO<sub>2</sub> donde pese a esto no se podría descartar ninguna, puesto que presentan un distinto patrón considerando los 3 contaminantes en conjunto.

Dada la falta de recursos, y buscando un máximo aprovechamiento de la tecnología disponible, resultaría interesante estudiar otros sitios potenciales para ubicar la estación de monitoreo, y de esta forma generar mediciones que brinden información independiente.

### III. Análisis del área de influencia de las estaciones de monitoreo

Las estaciones están ubicadas en diversas zonas de la ciudad de Buenos Aires (**Figura 98**) y monitorean el casco urbano, un tejido homogéneo, con ligeras diferencias en el uso del suelo. La estación de Córdoba (34.60° S, 58.39° W) está situada en una avenida de alto tránsito, en una zona comercial de edificios de gran altura. La estación de La Boca (34.63° S, 58.37° W) está ubicada en el área de la cuenca Matanza-Riachuelo, cerca de una autopista y un centro industrial (complejo petroquímico Dock Sud), por lo que es una zona mixta de alto flujo vehicular y fuentes estacionarias como refinerías de petróleo, dos centrales termoeléctricas, un área de

almacenamiento de hidrocarburos y de productos líquidos no inflamables, industrias químicas, entre otras.



**Figura 98.** Imagen de la Ciudad Autónoma de Buenos Aires con la ubicación de los tres sitios de monitoreo (en rojo), la estación AERONET (en verde) y la estación meteorológica (en naranja).

Mientras que las estaciones de La Boca y Córdoba han estado en operación desde 2009, la estación Centenario ( $34.61^{\circ}$  S,  $58.43^{\circ}$  W) comenzó a operar en 2005. Esta última estación se encuentra en las afueras del parque Centenario, un área arbolada, en una zona comercial con un flujo vehicular medio y baja influencia de fuentes estacionarias.

En función de lo expuesto anteriormente, resultaría conveniente evaluar la instalación de una estación de monitoreo en la zona norte de la ciudad, en las proximidades de la estación CEILAP o de la estación meteorológica.

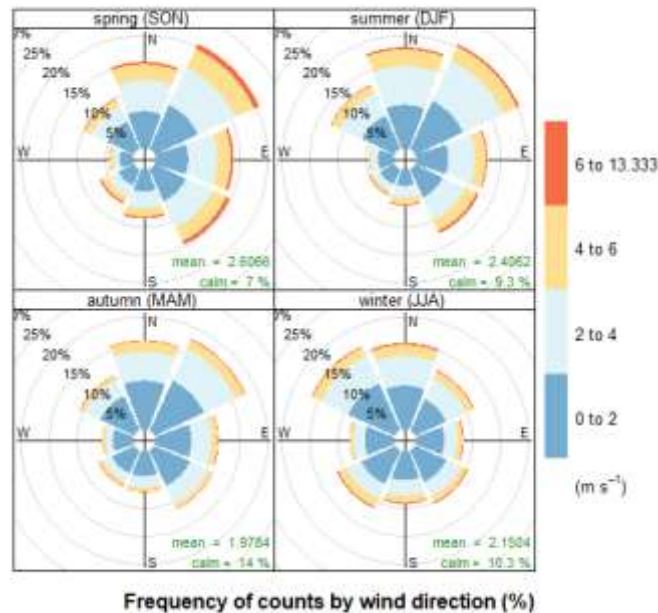
#### *IV. Análisis de influencia de la meteorología*

La **Tabla 20** muestra el resumen estadístico estacional de los parámetros meteorológicos, incluyendo la velocidad del viento, la temperatura ambiente y la humedad relativa. Como se mencionó, la región tiene un clima subtropical húmedo [Peel et al., 2007]. Los inviernos y veranos son húmedos con una humedad media anual del 71%.

**Tabla 20.** Parámetros meteorológicos estacionales provistos por el SMN: velocidad del viento ( $m.s^{-1}$ ), temperatura ambiente ( $^{\circ}C$ ) y humedad relativa (%) para el período analizado.

Estación	Velocidad del viento ( $m.s^{-1}$ )			Temperatura ( $^{\circ}C$ )			Humedad relativa (%)		
	Min	Media	Max	Min	Media	Max	Min	Media	Max
Otoño	0.0	2.0	10.3	3.2	18.2	34.8	24	75	100
Primavera	0.0	2.6	11.4	3.4	18.0	35.2	20	69	100
Verano	0.0	2.1	13.3	9.7	24.6	38.6	18	66	100
Invierno	0.0	2.6	12.8	-0.6	12.2	30.5	21	76	100

Los vientos tuvieron una tasa de calma anual del 3.8% y una velocidad media anual del viento de  $2.3 m.s^{-1}$ . A lo largo del año los vientos predominantes provienen del NE, excepto en invierno cuando el viento del sector NW se hace más frecuente (**Figura 99**). El mayor valor medio estacional de la velocidad del viento se observó durante el verano ( $13.3 m.s^{-1}$ ). En trabajos anteriores, durante esta temporada, se han encontrado los mayores valores estacionales medios diarios del factor de recirculación [Venegas & Mazzeo, 1999; Ulke & Mazzeo, 1998].



**Figura 99.** Rosas de viento estacionales para el período 2009 -2017. Las velocidades del viento se dividen en los intervalos que se muestran en la escala. Los círculos grises muestran la frecuencia de conteos por dirección del viento.

Es importante señalar que la baja variabilidad estacional puede afectar el proceso de deposición húmeda, en particular para  $PM_{10}$  [Gramsch et al., 2006]. Las variaciones en la velocidad del viento también influyen en los niveles de contaminantes en el aire. Los períodos de calma prolongados proporcionan una estabilidad atmosférica que evita la dispersión de contaminantes. En este contexto, se espera que se produzcan concentraciones persistentes de contaminación.

#### V. Análisis exploratorio de los datos

En la **Tabla 21** se presentan los resultados de las estadísticas descriptivas de las mediciones horarias obtenidas para los diferentes contaminantes atmosféricos durante todo el período de medición.

Analizando el CO de todo el período se puede observar que mientras que los valores máximos (50.0, 47.3 y 44.7 ppm para las estaciones de Centenario, Córdoba y La Boca, respectivamente) difieren entre las tres estaciones en más del doble de sus desviaciones estándar (0.9; 1.1 y 1.0 ppm), las medianas y medias (0.5 y 0.6 ppm) son iguales para las estaciones de Centenario y Córdoba, mientras que son inferiores para la estación de La Boca (0.2 y 0.3 ppm).

**Tabla 21.** Estadísticos descriptivos de los valores de concentración obtenidos para los diferentes contaminantes atmosféricos en cada una de las estaciones, para el período de tiempo analizado (LD = límite de detección).

Contaminante	Estación	Min	Mediana	Media	Max	Sd
CO (ppm)	Centenario	< LD	0.5	0.6	50.0	0.9
	Córdoba	< LD	0.5	0.6	47.3	1.1
	La Boca	< LD	0.2	0.3	44.7	1.0
NO <sub>2</sub> (ppb)	Centenario	< LD	18.0	19.6	155.0	9.7
	Córdoba	< LD	23.0	26.7	283.0	16.9
	La Boca	< LD	17.0	19.6	129.0	11.8
PM <sub>10</sub> (μg.m <sup>-3</sup> )	Centenario	< LD	26.0	30.0	427.0	17.9
	Córdoba	< LD	25.0	28.5	253.0	16.2
	La Boca	< LD	26.0	29.3	387.0	17.2

En cuanto al NO<sub>2</sub>, las concentraciones medias están dentro de las esperadas para las zonas urbanas [European Environment Agency, 2017; Li et al., 2016], en el rango de 19.6-26.7 ppb

(Tabla 21). Es notable la amplia dispersión de los datos de NO<sub>2</sub> de la estación Centenario antes de la interrupción de las mediciones durante 2013 (Figura 96).

Si bien las concentraciones medidas están dentro de los límites legales para CABA, hay excedencias respecto a las recomendaciones para el cuidado de la salud que propone la OMS. Las concentraciones medias horarias de NO<sub>2</sub> fueron superiores a 106 ppb (200 µg.m<sup>-3</sup>) en distintas ocasiones en cada una de las estaciones (Figura 100).

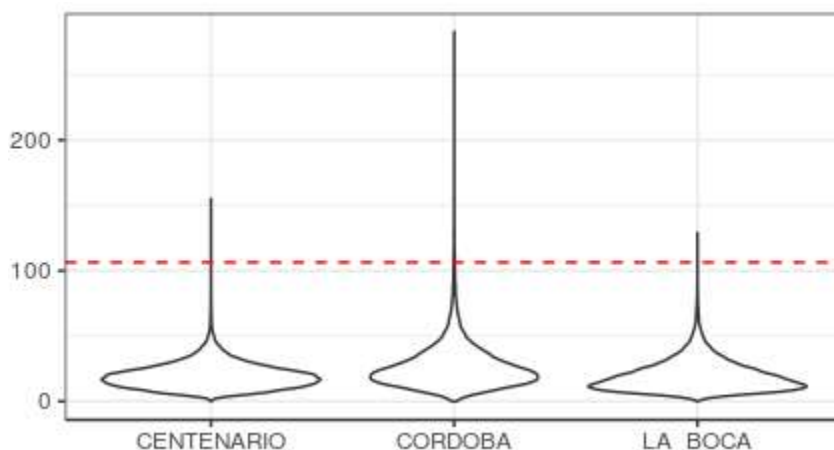


Figura 100. Gráfico violín para el conjunto de las mediciones de NO<sub>2</sub> realizadas en el período 2010-2017 (ppb). La línea roja punteada indica la concentración recomendada horaria por la OMS de 106 ppb.

La estación con mayores excedencias fue la estación Córdoba que sobrepasó el límite recomendado en el 2009, 2010, 2011, 2012 y 2013. El año con mayor número de excedencias fue el 2013, con un total de 32 para todo el año. El resto de las estaciones, si bien sobrepasaron el límite recomendado por la OMS, no superaron las 3 veces al año (Figura 101).

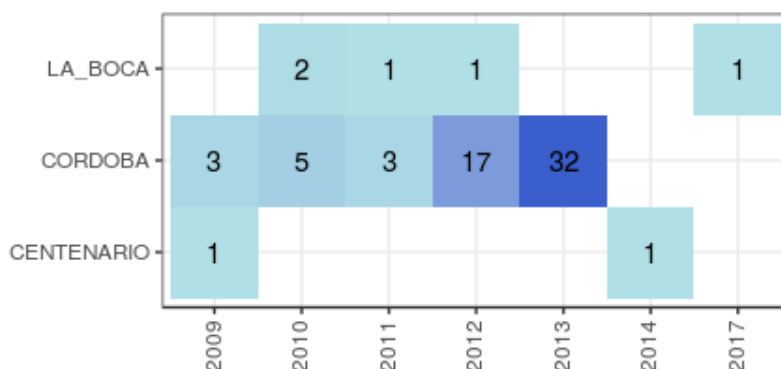
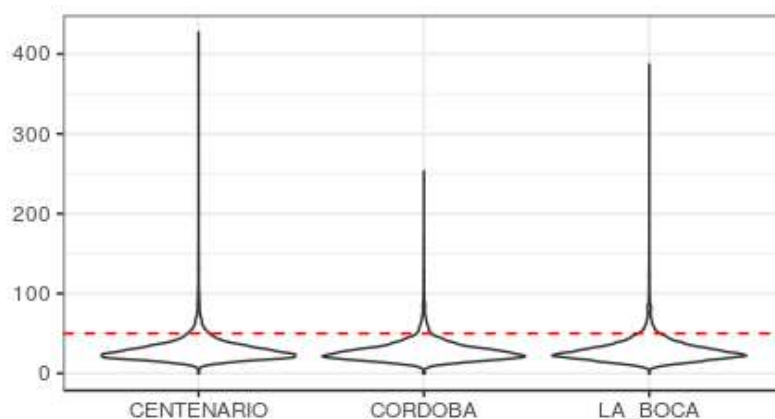


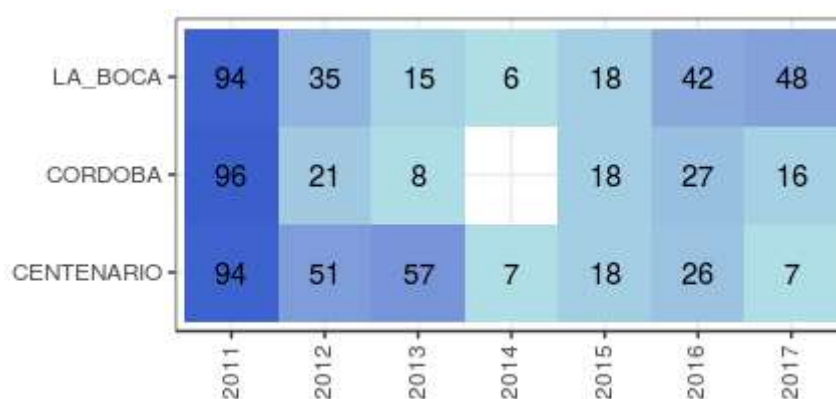
Figura 101. Cantidad de días al año en que las estaciones superan el límite horario para el NO<sub>2</sub> recomendado por la OMS.

Con respecto a la media de las concentraciones de  $PM_{10}$ , las tres estaciones de monitoreo presentaron resultados cercanos entre los 28.5 y 30  $\mu g.m^{-3}$  (**Figura 102**). Para todas las especies se registraron valores mínimos inferiores al límite de detección del instrumento.



**Figura 102.** Gráfico violín para el conjunto de las mediciones de  $PM_{10}$  realizadas en el período 2010-2017 ( $\mu g.m^{-3}$ ). La línea roja punteada indica el límite recomendado por la OMS de 20  $\mu g.m^{-3}$ .

Respecto a las excedencias, las concentraciones medias anuales de  $PM_{10}$  superaron el valor recomendado por la OMS de 50  $\mu g.m^{-3}$  para cada año en las tres estaciones de control (**Figura 103**). El año con mayores excedencias fue el 2011.



**Figura 103.** Cantidad de días al año en que las estaciones superan el límite anual para el  $PM_{10}$  recomendado por la OMS.

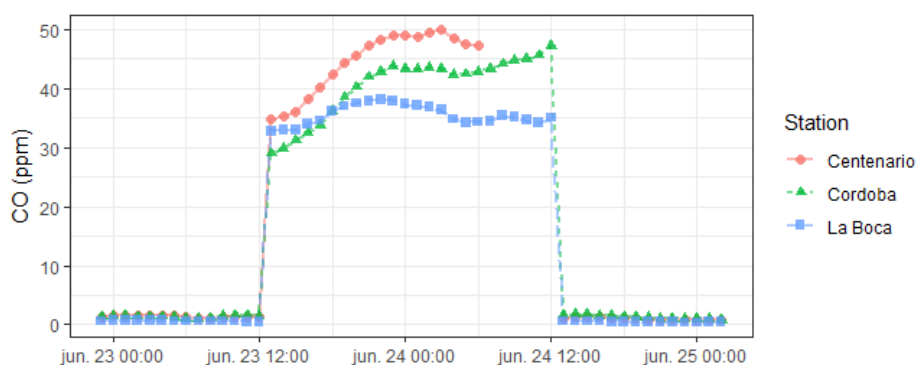
A su vez, los diferentes contaminantes presentaron valores extremadamente altos durante el período evaluado. Los fenómenos de contaminación atmosférica extrema son especialmente



preocupantes, ya que representan un riesgo para la salud humana y pueden producir cambios en los ecosistemas [Zhou et al., 2012]. Es por esto que los valores extremos del período analizado fueron estudiados en su contexto, lo cual permitió la identificación y caracterización de los eventos extremos a lo largo de los 8 años de mediciones.

En cuanto al CO, hay tres ocasiones diferentes en las que las concentraciones superan el valor límite (25 ppm) indicado por la OMS, poniendo en peligro la salud de las personas expuestas [OMS, 1999]. Estos eventos tuvieron lugar los días 25 y 26 de octubre de 2010, 23 y 24 de junio y 23 y 24 de diciembre de 2014.

De estos, el evento más importante tuvo lugar del 23 al 24 de junio de 2014, al mediodía, y fue detectado por las tres estaciones de monitoreo. La **Figura 104** ilustra los niveles de CO medidos en los diferentes sitios. Los niveles más altos de CO se registraron en la estación Centenario a las 3 de la mañana del 24 de junio (50 ppm) y al mediodía del mismo día en la estación de Córdoba (47 ppm). En la estación de La Boca, se mantuvo una concentración promedio de 32 ppm durante todo el evento, en contraste con las mediciones regulares de 0.3 ppm.



**Figura 104.** Registro horario del evento registrado por las 3 estaciones entre el 23 y 25 de junio de 2014

Las condiciones climáticas asociadas a este evento fueron una velocidad media del viento de 0.9 m.s<sup>-1</sup>, una temperatura media de 14 °C y una humedad relativa del 94%. Después del mediodía del 24 de junio, los niveles de CO volvieron a sus niveles anteriores.

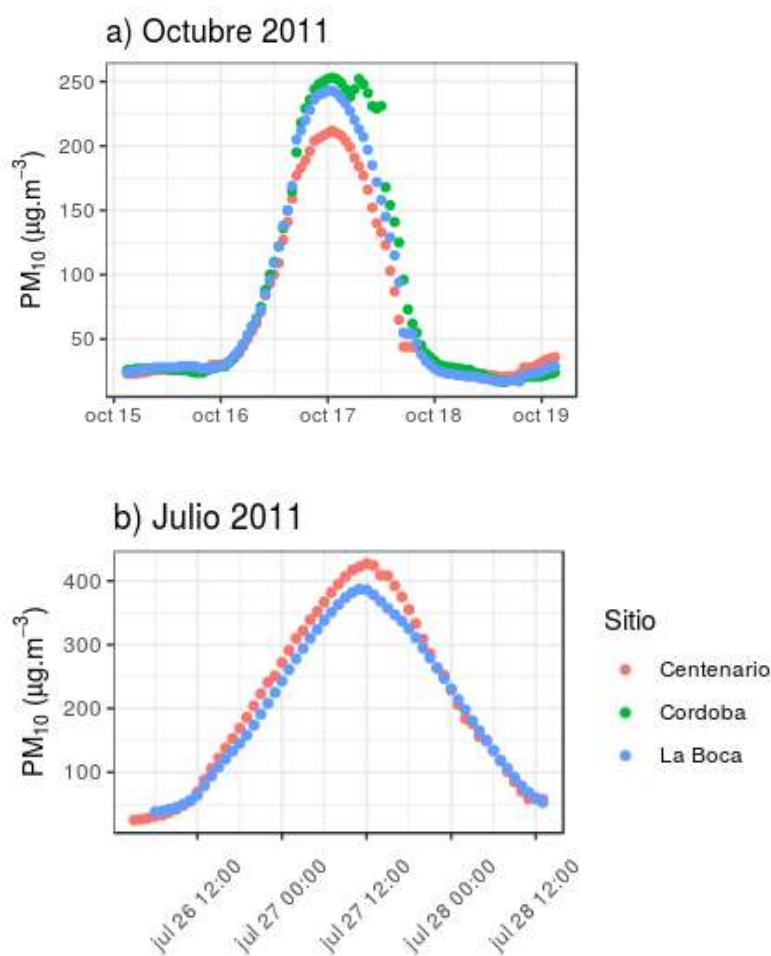
Además, una sola estación midió altas concentraciones de CO en otras dos ocasiones durante el período analizado. Los días 25 y 26 de octubre de 2010, la estación de La Boca registró un evento que alcanzó una concentración de 45 ppm. Por su parte, la estación de Córdoba midió valores altos entre el 23 y 24 de diciembre de 2014, registrando 33 ppm de CO como el valor más alto.

Estos eventos no se mencionaron en trabajos científicos anteriores ni en informes públicos/institucionales. La inexistencia de registros públicos de emisiones accidentales dificulta la explicación de estas mediciones inusuales. Al mismo tiempo, dada la insuficiencia de los registros sistemáticos del sistema de salud, no es posible analizar las implicancias de estos hechos para la población expuesta.

Los valores extremos de NO<sub>2</sub> son frecuentes en el período analizado, alcanzando concentraciones extremadamente altas que superan los niveles recomendados por la OMS [OMS, 2006]. Estos

niveles extremos ( $> 104$  ppm para el corto plazo) ocurrieron 232 veces en la estación de Córdoba, 7 veces en la estación de La Boca y 2 veces en la estación de Centenario. El mayor número de rebasamientos de la estación de Córdoba se explica por el hecho de que está situada en una avenida de alto tráfico circunscrita por edificios altos. Los óxidos de nitrógeno son emitidos por el tráfico rodado, la industria y la calefacción. Dentro de la atmósfera, y en presencia de compuestos orgánicos y radiación solar, el NO reacciona con el  $O_3$  para formar  $NO_2$  secundario. Por lo tanto, las concentraciones de  $NO_2$  pueden fluctuar considerablemente con el tiempo debido a la fuerte influencia de factores antropogénicos y naturales.

Además, los períodos de aumento de las concentraciones de  $PM_{10}$  fueron consistentes con los eventos de actividad volcánica. Los valores máximos de  $PM_{10}$  de las series temporales fueron registrados en julio de 2011, en las estaciones de Centenario y La Boca, y en octubre del mismo año en las tres estaciones (**Figura 105**). Estas fechas son consistentes con los meses de la erupción del complejo volcánico chileno Puyehue-Cordón Caulle en la Cordillera de los Andes, ubicado a 1400 km de Buenos Aires. La erupción comenzó el 4 de junio de 2011 y duró varios meses emitiendo cenizas. Las cenizas tuvieron un transporte circumpolar, viajando hacia el este y pasando por Sudáfrica, el sur de Australia y Nueva Zelanda para llegar de nuevo a Chile después de 10 días [Collini et al., 2013; Ulke et al., 2016].



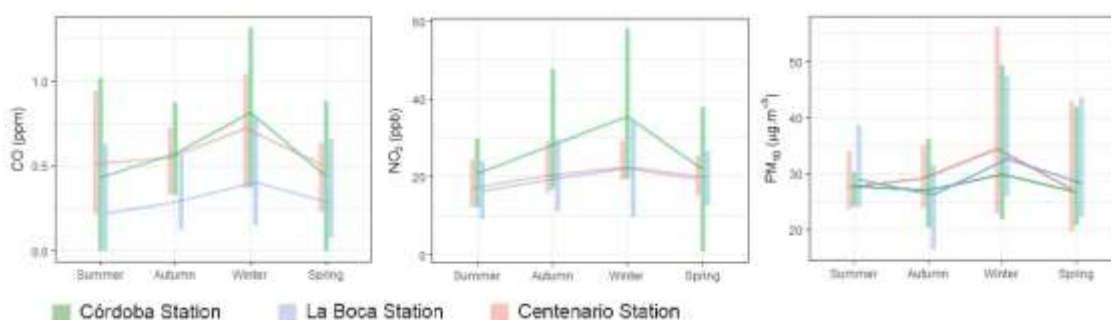
**Figura 105.** Serie de tiempo de las concentraciones medidas de  $PM_{10}$  en los eventos particulares de julio y octubre de 2011.

Otro período de altas concentraciones de  $PM_{10}$  coincidiendo con la actividad volcánica ocurrió en abril de 2015, cuando un enjambre de terremotos volcánicos-tectónicos, incluyendo el volcán Calbuco, entró en erupción después de décadas de inactividad. Varios días después, los vientos se llevaron las cenizas de la erupción del volcán y el sábado 25 de abril, la nube de cenizas llegó a la ciudad de Buenos Aires y al sur de Uruguay [Reckziegel et al., 2016].

Es importante señalar que los patrones y mecanismos de arrastre, transporte y deposición de material particulado siguen siendo temas poco explorados para esta región [Ristori et al., 2010; Otero et al., 2011; Otero et al., 2012]. Investigaciones anteriores han demostrado que el polvo patagónico puede viajar miles de kilómetros [McConnell et al., 2007; Gassó et al., 2010; Li & Ramaswamy, 2010], y en ciertas condiciones climáticas, es transportado en dirección este-noreste [Johnson et al. 2011; Gasso & Stein, 2007]. En este sentido, sería interesante promover estudios en la región que puedan relacionar las mediciones de la calidad del aire con el transporte de aerosoles.

#### Variaciones estacionales

Analizando los promedios estacionales de los contaminantes en todas las estaciones (**Figura 106**), se observó un patrón de disminución para las concentraciones de CO y  $NO_2$ , siguiendo el orden invierno > otoño > primavera > verano, con diferencias significativas entre ellas (valor  $p < 0.001$ ). La comparación de los datos de la media estacional de CO mostró el pico máximo en invierno (0.8 ppm) en la estación de Córdoba, y un valor mínimo en verano (0.2 ppm) en la estación de La Boca. Por su parte, el  $NO_2$  presentó un máximo de promedios estacionales durante el invierno (33.7 ppb) en la estación de Córdoba, mientras que el valor estacional mínimo (15.7 ppb) se registró durante el verano en la estación de La Boca. Este tipo de comportamiento es consistente con los reportados por otros autores en diferentes ciudades [Henschel et al., 2015; Henschel et al., 2015; Mavroidis & Iliá, 2012]; en invierno, las condiciones climáticas y el uso de la calefacción central conducen a un aumento de las emisiones y acumulación de contaminantes.



**Figura 106.** Concentraciones medias mensuales para cada contaminante en las 3 estaciones de monitoreo.

No obstante, las concentraciones de  $PM_{10}$  tuvieron un patrón diferente para los promedios estacionales, con los valores más altos observados en invierno y las concentraciones más bajas en otoño. Las  $PM_{10}$  alcanzaron el valor medio estacional máximo ( $33.7 \mu g \cdot m^{-3}$ ) durante el invierno en la estación Centenario y el mínimo ( $26.2 \mu g \cdot m^{-3}$ ) en otoño en la estación La Boca. Esto no se

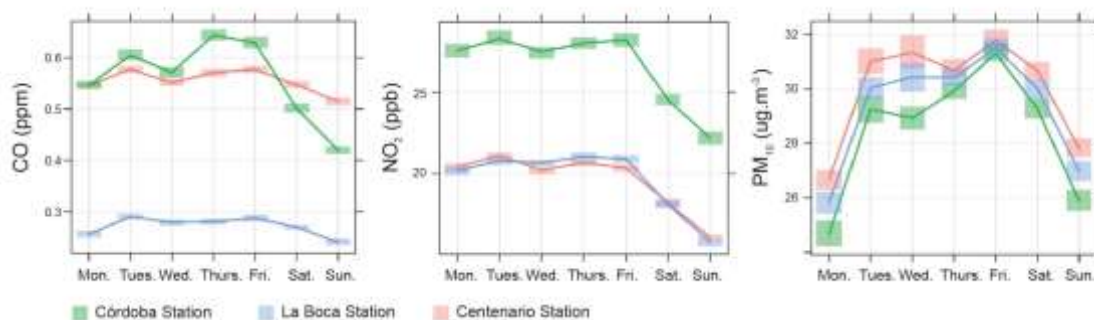
condice con lo informado por [Bogo et al., 2003], donde se señalaba que los niveles de PM en Buenos Aires disminuyen de verano a otoño-invierno.

Mientras que el CO y el NO<sub>2</sub> mostraron un comportamiento característico asociado con el aumento de las emisiones invernales, la reducción de la radiación solar y la disminución de la altura de la capa de mezcla, las concentraciones de PM<sub>10</sub> no mostraron un descenso estacional marcado (**Figura 106**). Esto puede deberse a la humedad relativa constante y a la baja velocidad del viento superficial a lo largo del año que mantiene las partículas de aerosol dentro de la región de origen.

#### Análisis semanal, diario y horario

El comportamiento semanal y diario de los contaminantes es el resultado de la interacción de diferentes factores, incluyendo la interacción entre la intensidad de las emisiones y los procesos fotoquímicos [Seinfeld y Pandis, 2016]. En este sentido, se espera que los niveles de contaminantes analizados disminuyan durante el fin de semana y durante las horas no laborables cuando las actividades productivas y el tráfico vehicular disminuyan.

La **Figura 107** muestra el comportamiento semanal de los contaminantes por estación de monitoreo. En el caso de las mediciones de CO, Centenario y Córdoba mostraron concentraciones similares a lo largo de la semana, mientras que en La Boca se presentaron menores mediciones. La estación de Córdoba tuvo un fuerte descenso en los niveles de CO hacia el fin de semana, cuando para el resto de las estaciones los niveles cambiaron ligeramente. Esto indica una fuerte relación entre los niveles de CO y las actividades productivas que tienen lugar alrededor de esta estación.



**Figura 107.** Concentraciones medias diarias para cada contaminante en las 3 estaciones de monitoreo.

Se observa un patrón diferente para el NO<sub>2</sub>, donde todas las estaciones reportaron una fuerte reducción en los niveles al final de la semana. La estación de Córdoba mostró mayores concentraciones de NO<sub>2</sub> que las estaciones de La Boca y Centenario. Este resultado cumple la predicción ya que la estación de Córdoba se encuentra en un punto de alto tránsito de la ciudad de Buenos Aires.

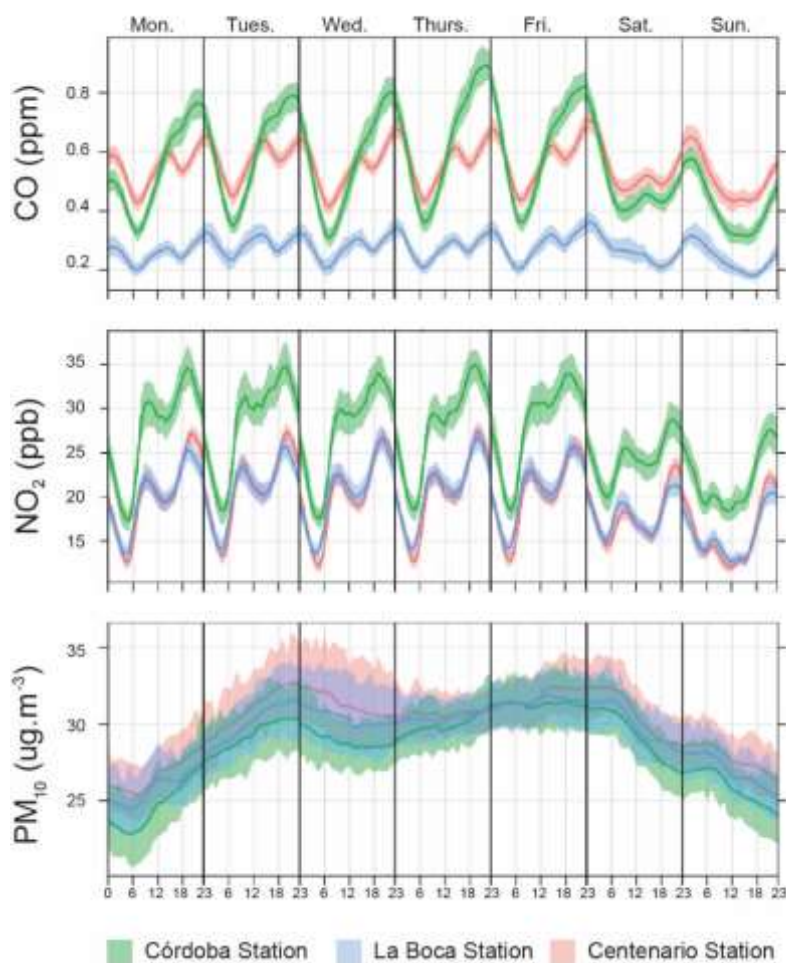
A diferencia de otros contaminantes, las concentraciones de PM<sub>10</sub> fueron similares durante la semana en las tres estaciones de control, con niveles medios más bajos los lunes y niveles máximos los viernes.

Los comportamientos horarios de CO y NO<sub>2</sub> presentaron patrones típicos para áreas urbanas como picos relacionados con el alto flujo vehicular durante el mediodía y la tarde y pronunciados descensos durante la noche. Este comportamiento fue anteriormente señalado por [Bogo et al., 2003], donde determinaron que los niveles de CO y NO<sub>x</sub> son muy bajos por la noche durante todo el año.

Es posible ver en la **Figura 108** un aumento en las concentraciones desde las 6 a.m. hasta el mediodía, cuando las concentraciones disminuyen ligeramente. Posteriormente, se observa un nuevo aumento de las concentraciones con una concentración máxima diaria cercana a las 6 p.m., y luego los niveles vuelven a bajar.

La disminución entre los dos picos podría estar relacionada con diferentes factores, entre los que podemos destacar: menores emisiones, mayor potencial de dispersión atmosférica y procesos fotoquímicos, especialmente para el NO<sub>2</sub>. Este tipo de comportamiento ha sido descrito en otras publicaciones sobre el tema [Henschel et al., 2015; Mavroidis & Ilia, 2012].

Los días del fin de semana presentaron un comportamiento particular. Los valores máximos se desplazaron hacia las primeras horas del día. Particularmente los domingos, los niveles de CO alcanzaron valores mínimos entre 16-18 h mientras que para el NO<sub>2</sub> ocurre a las 12-14 h.



**Figura 108.** Concentraciones medias horarias a lo largo de la semana para cada contaminante en las 3 estaciones de monitoreo.

En el caso de  $PM_{10}$ , se observó con mayor intensidad una capacidad de acumulación de este contaminante. Las concentraciones aumentaron al principio de la semana, luego se mantuvieron en un promedio de  $30 \mu\text{g}\cdot\text{m}^{-3}$  los días laborables y disminuyeron en las primeras horas del sábado. El comportamiento inusual observado para  $PM_{10}$  puede explicarse a partir del pretratamiento de los datos, donde se realiza el promedio móvil de 24 horas de las concentraciones, haciendo más sutiles las variaciones que pueden estar presentes dentro de este rango de tiempo. Este es un aspecto importante para destacar ya que implica una exposición continua a un nivel de partículas superior al nivel medio anual recomendado por la OMS [2006].

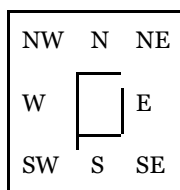
## VI. Evaluación de las tendencias a largo plazo

La cuantificación de las tendencias en el tiempo se evaluó computando el estimador no paramétrico de pendiente Theil-Sen en las medias mensuales. Para examinar los cambios en la pendiente, se analizaron las tendencias de los contaminantes totales y para cada dirección del viento.

La tendencia total en la estación Córdoba para  $PM_{10}$  fue de  $-3.48 [-5.88, -1.3] \text{ \%}\cdot\text{año}^{-1}$  ( $p < 0.05$ ,  $n = 70126$ ), para CO de  $9.04 [0.65, 21.93] \text{ \%}\cdot\text{año}^{-1}$  ( $p < 0.05$ ,  $n = 70126$ ) y para  $NO_2$  de  $-4.27 [-6.67, -2.49] \text{ \%}\cdot\text{año}^{-1}$  ( $p < 0.05$ ,  $n = 70126$ ). Igualmente, en la estación Centenario la tendencia total para  $PM_{10}$  fue de  $-5.42 [-7.27, -3.67] \text{ \%}\cdot\text{año}^{-1}$  ( $p < 0.05$ ,  $n = 70126$ ), para  $NO_2$  de  $-5.23 [-6.63, -3.89] \text{ \%}\cdot\text{año}^{-1}$  ( $p < 0.05$ ,  $n = 70126$ ) y para CO no significativa. Por su parte, en la estación La Boca, la tendencia total fue también no significativa para  $PM_{10}$ , mientras que para  $NO_2$  fue de  $-6.97 [-8.31, -5.25] \text{ \%}\cdot\text{año}^{-1}$  ( $p < 0.05$ ,  $n = 70126$ ) y para CO fue positiva de  $17.13 [13.21, 131.18] \text{ \%}\cdot\text{año}^{-1}$  ( $p < 0.05$ ,  $n = 70126$ ).

En la **Tabla 22** se observan las pendientes por dirección de viento. Al desagregar los datos por dirección de viento, se obtuvo que la tendencia del CO fue positiva en las estaciones de Córdoba y La Boca. En el caso de la estación de Córdoba, las contribuciones fueron significativas desde el NW y NE. En cuanto a la estación de La Boca, se observó una tendencia creciente con aportes significativos de los NE, E, SE, S, SW y W. Aunque las tendencias fueron positivas, la proyección de la concentración de CO no se aproxima en el corto plazo al valor límite recomendado por la OMS para la protección de la salud de la población [OMS, 1999]. Por otra parte, se observaron disminuciones consistentes en la media mensual desde los sectores NW, N, NE y E en la estación Centenario.

Las concentraciones de  $NO_2$  tuvieron una clara tendencia a la baja con una alta significación estadística durante el período estudiado en todas las estaciones y en todas las direcciones del viento. Las concentraciones medias mensuales de  $PM_{10}$  también presentaron una tendencia global a la baja en las estaciones de Córdoba (aunque no en las direcciones NE y E) y Centenario (no en la dirección E). Sin embargo, la tendencia para la estación de La Boca no fue significativa en ninguna dirección del viento. Las disminuciones de  $NO_2$  y  $PM_{10}$  pueden atribuirse a mejoras en la tecnología de las flotas de automóviles, como el uso de convertidores catalíticos en los tubos de escape y el uso menos común de vehículos diesel que emiten una mayor fracción de  $NO_2$  en comparación con los vehículos de gasolina [Mavroidis & Ilia, 2012].

**Tabla 22.** Tendencias significativas a largo plazo para diferentes direcciones de viento con un intervalo de confianza del 95% entre corchetes. ns, no significativo; \* p<0.001; \*\* p<0.05

	Centenario			Córdoba			La Boca		
CO (ppm.año <sup>-1</sup> )	-0.03 [-0.06, -0.01]**	-0.02 [- 0.05, 0]**	-0.02 [- 0.04, 0]**	0.04 [0.01, 0.06]*	ns	0.05 [0.02, 0.07]*	ns	ns	0.03 [0.02, 0.04]*
	ns		-0.03 [- 0.05, 0.01]**	ns		ns	0.03 [0.01, 0.05]*		0.03 [0.02, 0.04]*
	ns	ns	ns	ns	ns	ns	0.04 [0.02, 0.05]*	0.03 [0.01, 0.04]*	0.03 [0.01, 0.04]*
NO <sub>2</sub> (ppb.año <sup>-1</sup> )	-1.78 [- 2.25, - 1.27]*	-1.44 [- 1.85, - 0.97]*	-1.32 [- 1.73, - 0.97]*	-1.69 [- 2.35, - 1.06]*	-1.48 [- 2.34, - 0.74]*	-1.68 [- 2.61, - 1.04]*	-2.51 [- 3.21, - 1.84]*	-2.15 [- 2.63, - 1.60]*	-1.87 [- 2.35, - 1.29]*
	- 1.23 [- 1.66, - 0.77]*		-1.06 [- 1.48, - 0.61]*	-1.18 [- 1.91, - 0.62]*		-1.12 [- 2.31, - 0.32]**	-1.67 [- 2.32, - 0.91]*		-1.47 [- 1.88, - 0.87]*
	-1.10 [- 1.48, - 0.70]*	-1.36 [- 1.82, - 0.87]*	-1.01 [- 1.45, - 0.57]*	-1.26 [- 1.97, - 0.66]*	-1.59 [- 2.32, - 0.83]*	-1.25 [- 2.23, - 0.25]**	-1.17 [- 1.71, - 0.56]*	-1.25 [- 1.70, - 0.69]*	-1.29 [- 1.82, - 0.66]*
PM <sub>10</sub> (μg.m <sup>-3</sup> .año <sup>-1</sup> )	-2.68 [- 3.74, - 1.82]*	-2.00 [- 2.77, - 1.31]*	-1.63 [- 2.22, - 1.09]*	-1.79 [- 2.55, - 0.97]*	-0.91 [- 1.80, - 0.19]**	ns	ns	ns	ns
	-1.72 [- 2.83, - 0.82]*		ns	-1.74 [- 2.93, - 0.70]*		ns	ns		ns
	-2.04 [- 3.16, - 1.11]*	-2.15 [- 3.24, - 1.32]*	-1.92 [- 2.79, - 1.09]*	-1.79 [- 2.98, - 0.84]*	-1.78 [- 2.76, - 0.71]*	-1.13 [- 1.93, - 0.13]**	ns	ns	ns

## VII. Análisis de los aportes

Las rosas de la contaminación se trazaron utilizando la concentración de los diferentes contaminantes en cada estación (**Figura 109**). De esta manera es posible observar las direcciones de los vientos que más contribuyeron a las concentraciones medias totales, permitiendo la identificación de los sectores de procedencia de las fuentes más relevantes.

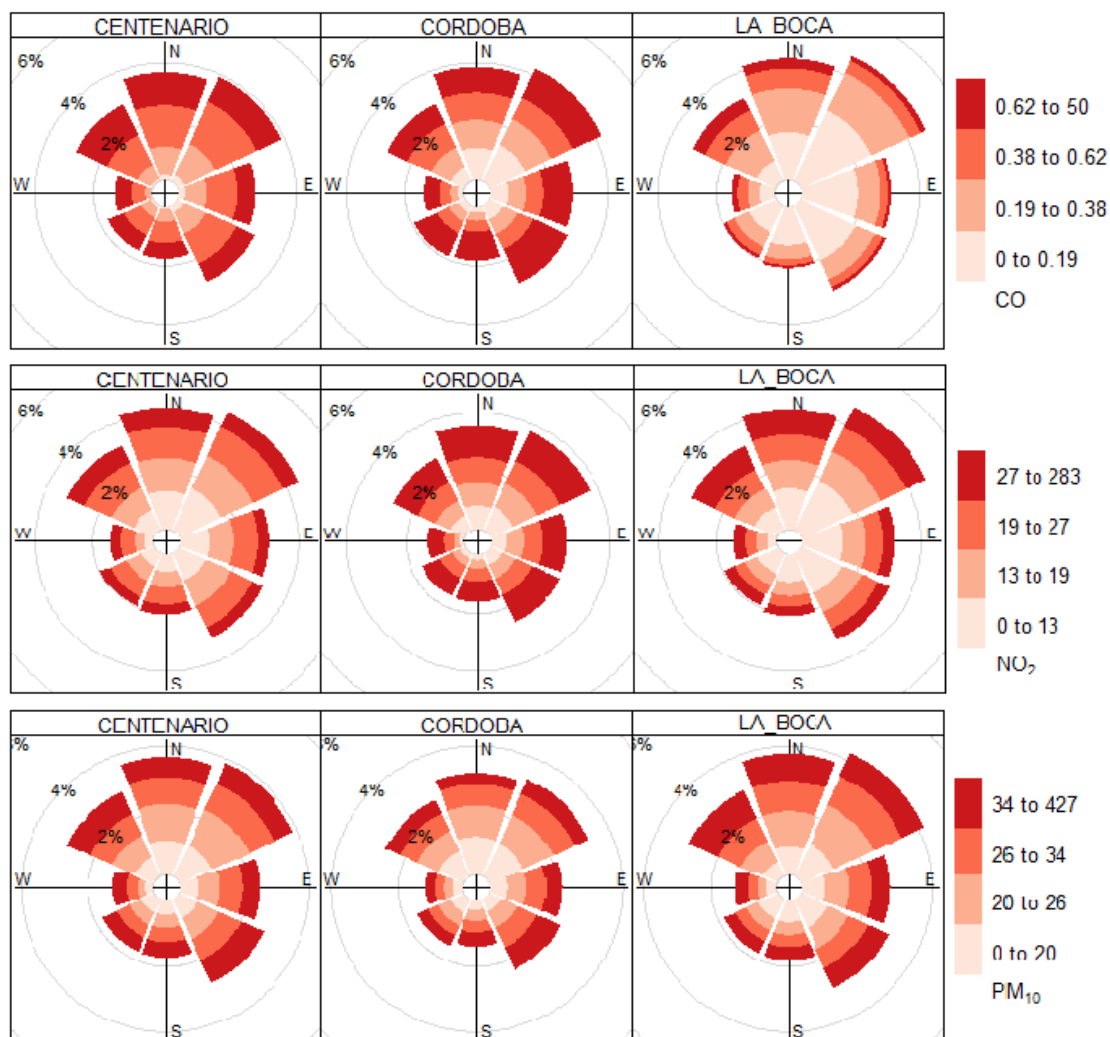


Figura 109. Rosas de contaminación

En la estación Centenario, las mayores concentraciones de CO y NO<sub>2</sub> llegaron con mayor frecuencia desde el sector NW, aunque los sectores N y NE hicieron una contribución considerable, mientras que las principales contribuciones de PM<sub>10</sub> provinieron de los sectores NW y SE. En cuanto a la estación de Córdoba, la mayor concentración de aportes de CO proviene de SE, mientras que las mayores concentraciones de NO<sub>2</sub> y PM<sub>10</sub> provienen de NE, E, SE y S. Finalmente, en la estación de La Boca las mayores concentraciones de CO y NO<sub>2</sub> provienen de NW > N > NE, mientras que las contribuciones de PM<sub>10</sub> también son mayores de NW, sin embargo, las contribuciones de N, NE y SE son comparables.

No hubo sectores que contribuyesen sustancialmente a las concentraciones medias de las tres estaciones. Como se describió anteriormente, estas estaciones de monitoreo fueron ubicadas en el centro urbano, donde el flujo de aire es bajo-moderado y se localizan diferentes fuentes con intensidad relativamente similar. En vista de estas características, la metodología propuesta tiene algunas limitaciones [Ruiz et al., 2014] que fueron superadas por el análisis de rosas de viento para diferentes rangos de concentración (Figura 110).



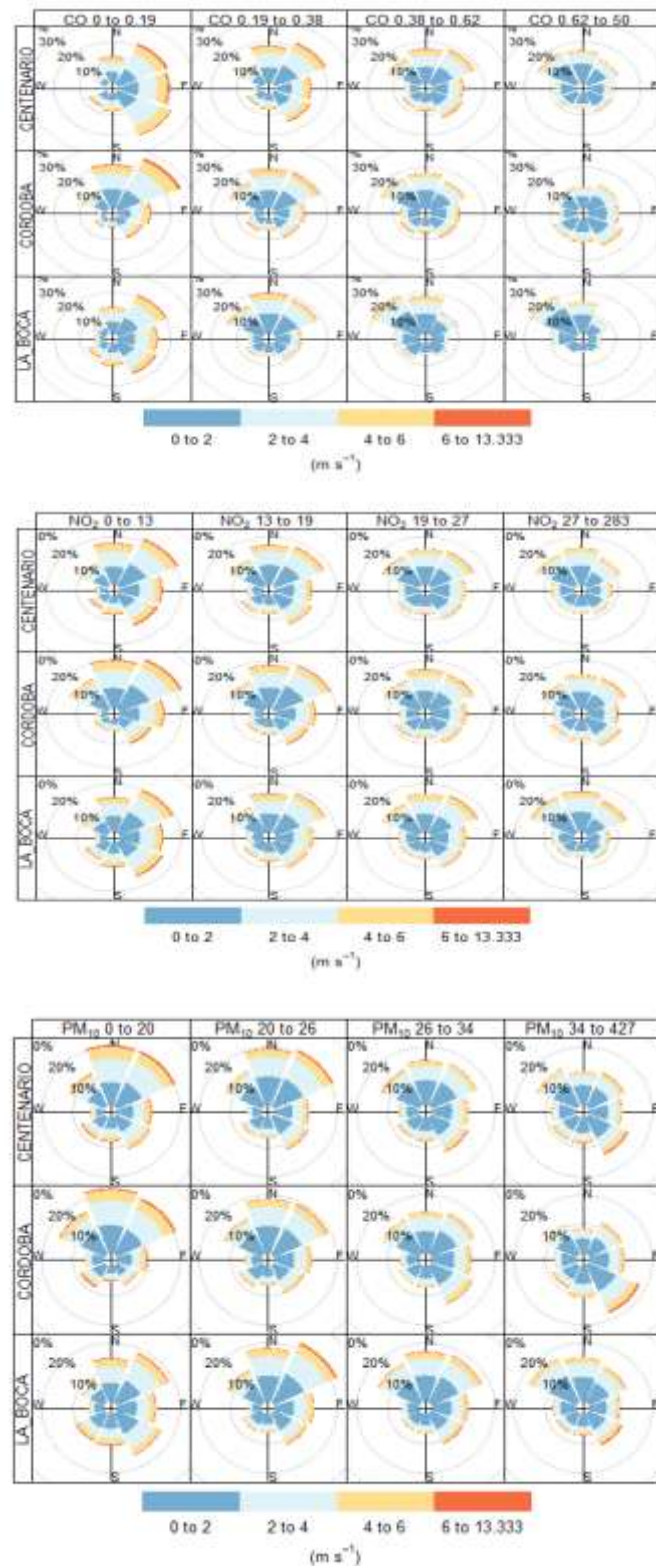


Figura 110. Rosas de viento para diferentes rangos de concentración de a) CO b) NO<sub>2</sub> y c) PM<sub>10</sub>

## VII. Análisis del comportamiento espacial

Para evaluar el comportamiento espacial de los contaminantes en el aire se procedió a elaborar mapas de concentraciones de  $PM_{10}$ , entendiendo que el  $PM_{10}$  es un contaminante criterio que funciona como un buen indicador de la contaminación atmosférica

### Obtención y validación de imágenes espaciales

Para la confección de mapas de material particulado se trabajó con el producto satelital MCD19A2 de 1 km para el período comprendido entre 2009-10-01 y 2017-10-01 provisto por la NASA. Se descargaron un total de 12055 imágenes diarias. El producto AOD MCD19A2 es un producto diario que resume la información de distintas órbitas de los satélites Terra y Aqua, las horas de sobrepaso de los satélites resultó entre las 12 y las 15 hs UTC sobre la ciudad de Buenos Aires.

De las imágenes se extrajo la información en un área circular (3 y 7 km de diámetro) a las horas de sobrepaso de los satélites (12-15 hs UTC) y se promediaron los valores de AOD de los pixeles que cayeron dentro del área. A su vez, se generó un promedio con las mediciones AERONET en el lapso temporal de sobrevuelo del satélite para relacionarlo con los valores medidos por MODIS. Como se señaló anteriormente, este paso requirió la interpolación cuadrática de las mediciones de AOD registradas por AERONET a la longitud de onda utilizada por MODIS (550 nm).

Como primera cuestión, cabe resaltar que se observó una gran falta de información. El período comprende más de 9 mil días, y se observó una pérdida del 77% por falta de registros de AERONET o por presencia de nubes u otra aberración que aumentaba el error de la recuperación de AOD de forma satelital.

Como se mencionó anteriormente, la correlación se realizó aplicando un área circular de 3 y 7 km de diámetro entorno a la estación de AERONET y promediando los valores de AOD obtenidos a partir de las mediciones de MODIS en esa área. Para el área de 3 km de diámetro, se obtuvo un factor de correlación de Pearson negativo ( $r = -0.08$ ,  $p < 0.001$ ) y un coeficiente de Spearman no significativo. A su vez, considerando los valores obtenidos en el área circular de 7 km de diámetro, el factor de correlación de Pearson y de Spearman también fueron negativos ( $r = -0.08$ ,  $p < 0.001$ ;  $\rho = -0.06$ ,  $p < 0.05$ ). Estos resultados muestran una muy mala relación entre las mediciones satelitales y las realizadas por AERONET, independiente del radio que se considere.

Analizando con mayor detalle el set de datos, se observó una gran cantidad de valores altos con signo negativo de AOD medidos por MODIS. Estos valores pueden advertir de una falla en el algoritmo de recuperación de los aerosoles de la columna atmosférica en función de la reflectancia medida en el tope de atmósfera. Si bien los valores negativos aparecen en atmósferas muy limpias porque MODIS no tiene sensibilidad sobre el terreno para recuperar los valores de AOD con una precisión mejor que  $\pm 0.05$ , los valores altos negativos no se explicarían por este motivo puesto que superaban ampliamente el margen de precisión. Estos resultados anómalos abren nuevas preguntas para futuros trabajos en la región.

Por otra parte, cabe mencionar que no se realizó un tratamiento adicional para evaluar la calidad de los datos de CEILAP AERONET. En [Ristori et al., 2003], se menciona que cuando se realizó un prefiltro de los datos de esta misma estación mediante el cálculo del error cuadrático medio sobre el ajuste de una regresión parabólica de la Ley de Ångström, se descartaron el 88% de los

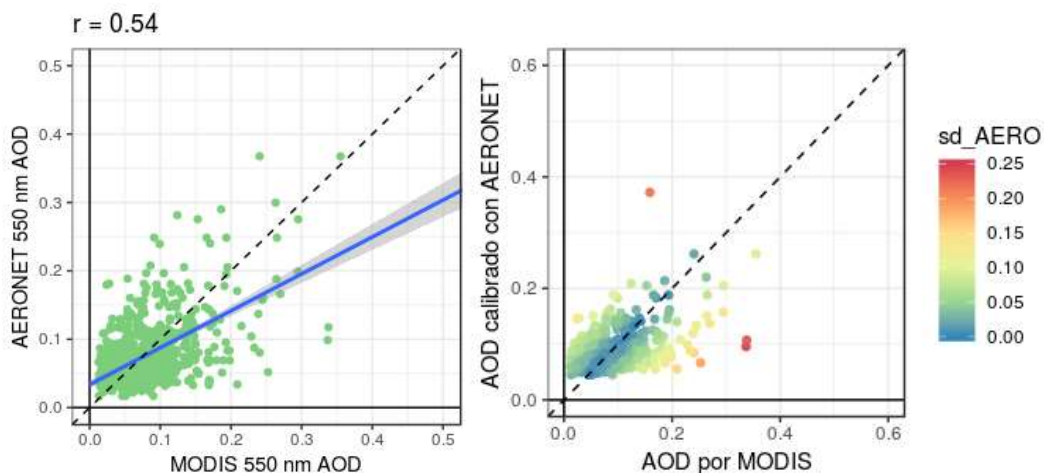
datos provistos para un período temporal distinto al utilizado en esta tesis. Resulta una etapa interesante a incorporar como mejora a este trabajo.

Continuando con el análisis, se decidió no considerar las mediciones negativas para la calibración de MODIS. De esta forma, se evaluó la fuerza de la relación entre AERONET y MODIS y se calcularon los factores de correlación de Pearson y Spearman para el buffer de 3 km, los cuales dieron: 0.54 (p-value < 0.001) y 0.34 (p-value < 0.001), respectivamente. Esto mismo se realizó para un buffer de 7 km de diámetro, encontrándose que los factores de correlación de Pearson (0.47, p-value < 0.001) y Spearman (0.30, p-value < 0.001) fueron levemente inferiores.

Para la calibración, el modelo lineal realizado responde a la fórmula:

$$\text{AOD}_{\text{AERONET}} = a + b \cdot \text{AOD}_{\text{MODIS}}$$

donde  $a = 0.034 \pm 0.003$  (p-value < 0.001) y  $b = 0.62 \pm 0.03$  (p-value < 0.001), el cual se observa en la **Figura 111**. La regresión lineal presentó un  $R^2$  ajustado = 0.34. Estos resultados son llamativamente bajos, pero no se cuenta con otro sistema de validación para las imágenes en la zona. La recta de validación presentó un RMSE = 0.05 y un MAE = 0.03 unidades de AOD para un set de datos de prueba (n = 1840 observaciones).



**Figura 111.** a) Correlación entre los datos MODIS y AERONET a 550 nm, b) Scatterplot entre las medidas de AOD y las medidas de AOD calibradas con AERONET .

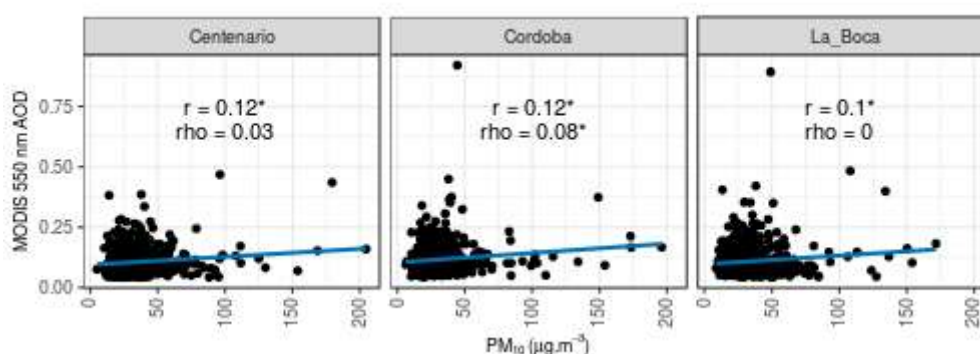
Las imágenes MODIS fueron recortadas, calibradas y se eliminaron los valores negativos. Tras esto se generaron mosaicos diarios. Cuando las imágenes presentaron una superposición en algunas de sus partes, se reemplazó el valor de píxel por la media.

Para completar la falta de datos en las imágenes satelitales, debido a la presencia de sombras, nubes o por la eliminación de los valores anómalos negativos, se rellenaron las imágenes mediante el método de interpolación Ponderado de Distancia Inversa (IDW) con una distancia de alisamiento de 3 km [Goldberg et al., 2019]. Solo se rellenaron imágenes que tuvieran más del 30% de los valores para el área. Las imágenes diarias que no cumplieron con este porcentaje se

descartaron para no introducir un error al modelo. En la validación cruzada del método de interpolación IDW, se obtuvo un RMSE promedio de  $0.007 \pm 0.003$  unidades de AOD, un orden de magnitud menor que la media de los valores de AOD y su desviación estándar.

De las imágenes diarias completadas por el método de interpolación se extrajeron los valores de AOD sobre las estaciones de monitoreo. Utilizando los datos diarios de AOD promedio dentro de un buffer de 1 km entorno a la estación de monitoreo de PM<sub>10</sub>, los factores de correlación de Pearson y Spearman entre AOD-PM<sub>10</sub> dieron 0.11 ( $p$ -value  $< 0.001$ ) y no significativo. Sin aplicar el buffer, se obtuvieron los mismos valores en los factores de correlación de Pearson y Spearman entre AOD-PM<sub>10</sub>. Este nivel de correlación es bajo respecto a la Comunidad Valenciana. Se continuó trabajando sin aplicar buffer, ya que requiere menos procesamiento.

No se observaron diferencias en la correlación AOD y PM<sub>10</sub> en las estaciones de monitoreo (**Figura 112**).

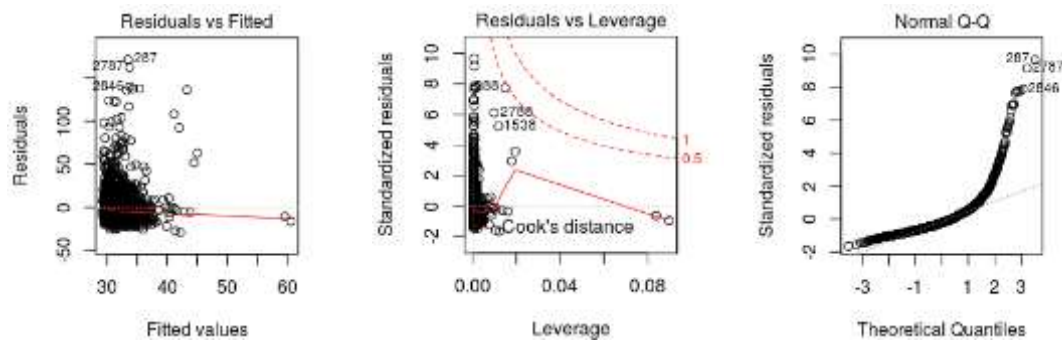


**Figura 112.** Correlación entre las mediciones de AOD de MODIS y las mediciones promedio de PM<sub>10</sub> dentro de las horas de paso de órbita del satélite para cada una de las estaciones en todo el período de estudio ( $p < 0.05$ )

Como primer modelo se consideró un modelo de regresión lineal del tipo:

$$Y = \beta_0 + \beta_1 X$$

donde X representa a las mediciones de AOD diarias en los puntos de monitoreo e Y el logaritmo de las mediciones de PM<sub>10</sub>. Los parámetros obtenidos fueron un factor de determinación  $R^2 = 0.01$



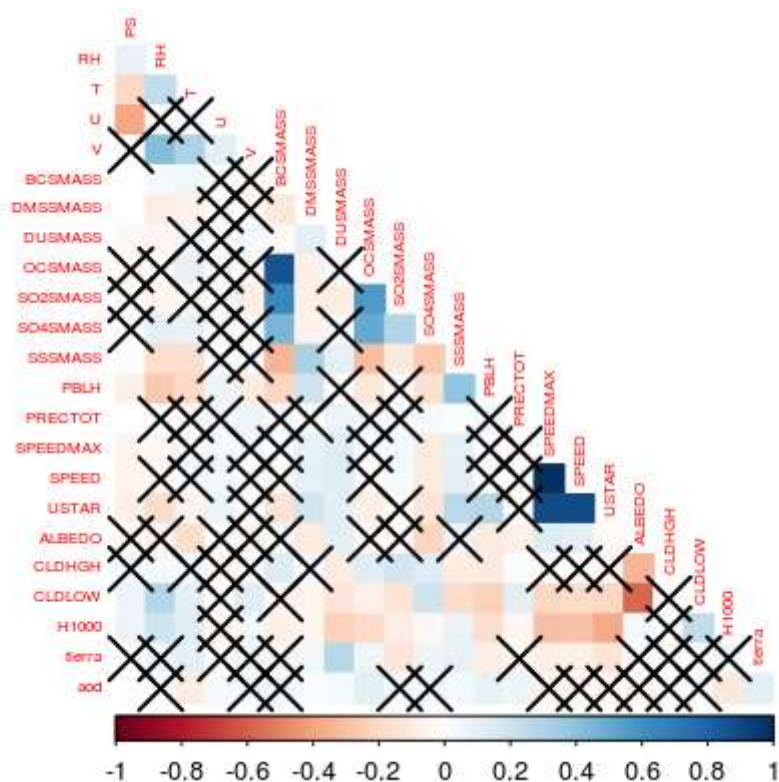
**Figura 113.** Gráficos evaluatorios del modelo. Izquierda) Residuos del modelo, centro) QQ-Plot de los residuos estandarizados, derecha) Distancia de Cook o añade letras en la parte gráfica.

Analizando el modelo lineal, se observó una falta de normalidad de los residuos estandarizados (**Figura 113, izquierda**). Esto se comprobó con un Test de normalidad de Shapiro ( $W = 0.70$ ,  $p\text{-value} < 0.001$ ), rechazando la hipótesis nula de normalidad. En base a esto se procedió a elegir un modelo no paramétrico para la confección del modelo. Se seleccionó el método no paramétrico de *machine learning* Random Forest para desarrollar el modelo LUR.

#### Generación de un modelo espacial para $PM_{10}$

Para la confección del modelo espacial de  $PM_{10}$  se trabajó con las variables PS, RH, T, U, V, BCSMASS, DMSSMASS, DUSMASS, OCSMASS,  $SO_2$ SMASS,  $SO_4$ SMASS, SSSMASS, PBLH, PRECTOT, SPEEDMAX, SPEED, USTAR, ALBEDO, CLDHGH, CLDLOW, H1000 obtenidas mediante modelación con MERRA-2, explicadas anteriormente.

Como paso previo a la confección del modelo, se evaluó la correlación entre las variables para descartar aquellas que mantienen una fuerte correlación. En la **Figura 114** se ve el coeficiente de correlación entre todas las variables consideradas.



**Figura 114.** Correlación entre las variables iniciales. De rojo a azul se indica el factor de correlación de Pearson. Las cruces indican correlación no significativa.

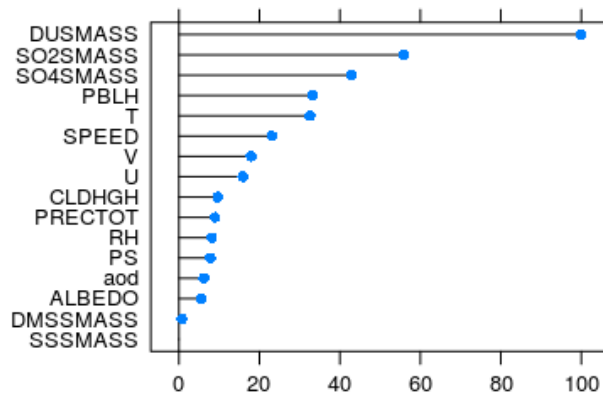
En función de la alta correlación, descartamos las variables BCSMASS, OCSMASS, SPEEDMAX, USTAR, H1000 y CLDLGW, quedando el modelo final con las variables:

$$PM_{10} \sim AOD + PS + RH + T + U + V + DMSSMASS + DUSMASS + SO_2SMASS + SO_4SMASS + SSSMASS + PBLH + PRECTOT + SPEED + ALBEDO + CLDHGH$$

Como primer paso, se evaluó el ajuste con un modelo lineal multivariable. El modelo dio un  $R^2 = 0.13$ ,  $RMSE = 16.79 \mu\text{g}\cdot\text{m}^{-3}$  y  $MAE = 10.44 \mu\text{g}\cdot\text{m}^{-3}$ . En base a estos resultados, se decidió probar un modelo no-paramétrico Random Forest.

El modelo Random Forest se entrenó usando un 80% de los datos ( $n = 1840$ ), reservándose un 20% para su evaluación. En base a distintas pruebas, se definió trabajar con 1000 árboles de entrenamiento utilizando como método de validación la técnica de validación cruzada por 10 veces.

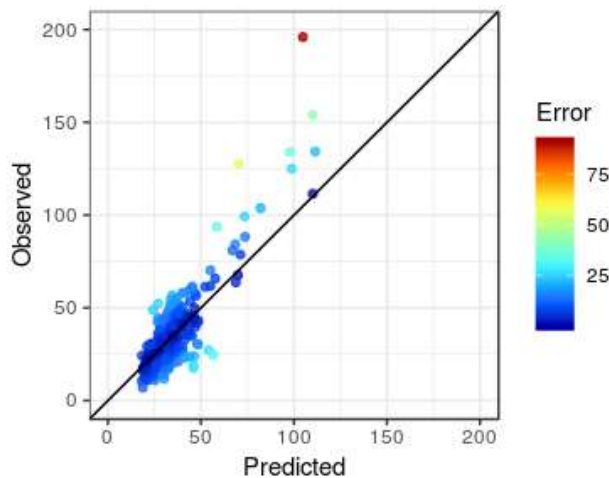
En la **Figura 115** se pueden ver las variables con mayor relevancia en el modelo. La variable AOD no presentó una gran importancia para la predicción de  $PM_{10}$ . Por el contrario, fueron las variables provistas por el modelo MERRA-2 referidas a masa de polvo, masa de  $SO_4$  y masa de  $SO_2$  las más importantes.



**Figura 115.** Peso de las variables en la construcción del modelo de PM<sub>10</sub> superficial para Buenos Aires

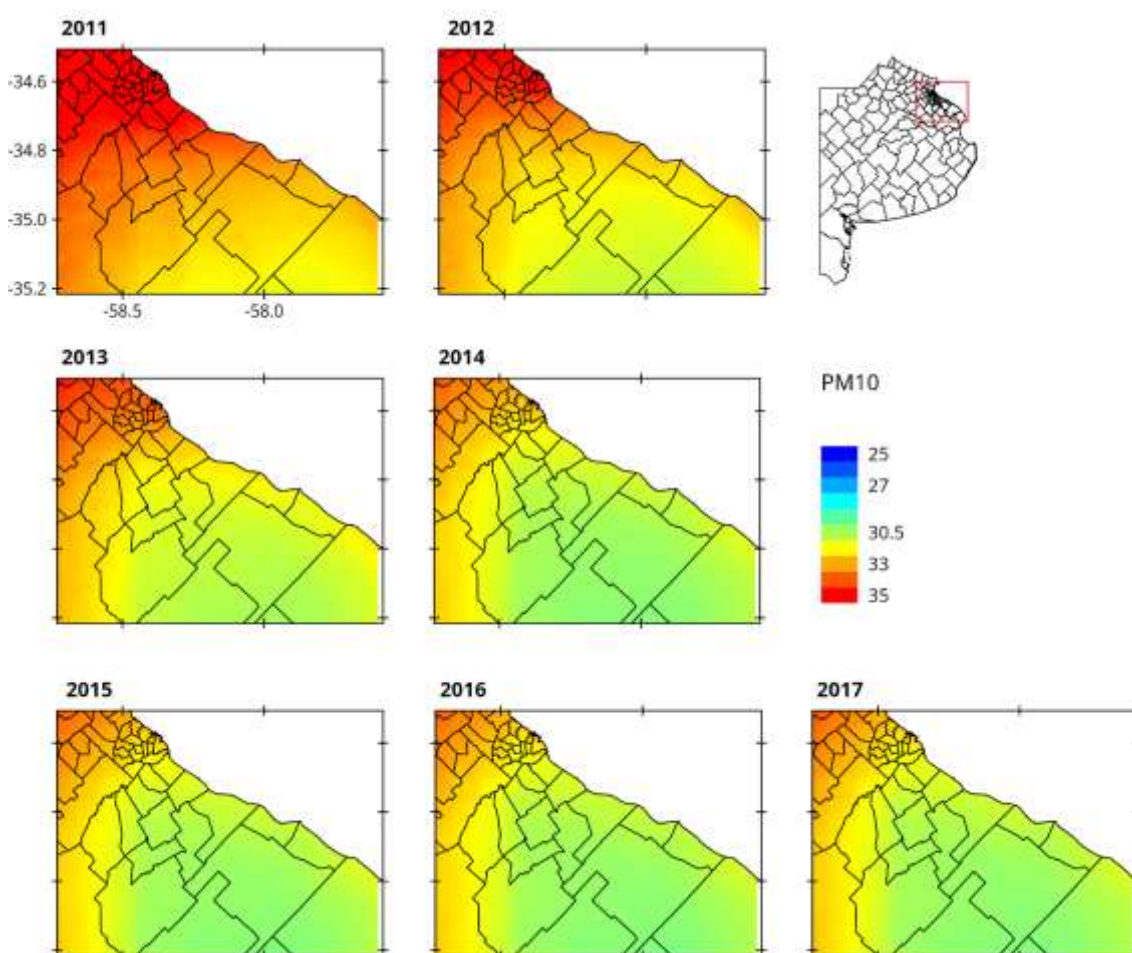
El ajuste del modelo para PM<sub>10</sub> usando Random Forest mostró un  $R^2 = 0.96$ , con un RMSE = 4.93  $\mu\text{g.m}^{-3}$  y MAE = 2.92  $\mu\text{g.m}^{-3}$  para los datos de entrenamiento. En la evaluación con el set de prueba se obtuvo un RMSE = 10.16  $\mu\text{g.m}^{-3}$  y un MAE = 6.83  $\mu\text{g.m}^{-3}$ , mostrando un mejor grado de predictibilidad que el modelo lineal multivariable, al igual que lo observado en el caso de la Comunidad Valenciana. El MAPE para los datos de prueba fue de un 24.4% mostrando que el error del modelo es aceptable para esta región. Por otra parte, se obtuvo un BIAS de -38% indicando una subestimación de los datos modelados.

Al analizar el scatterplot entre los datos medidos y modelados presentado en la **Figura 116**, se observa que el mayor error de predicción se da con los valores extremos medidos a campo. Esto es consonante con los fenómenos locales que pueden estar infiriendo en las estaciones de monitoreo y no ser registradas por el sensor satelital dada la dilución del contaminante.



**Figura 116.** Scatterplot entre las mediciones predichas por el modelo y las mediciones de PM<sub>10</sub>

En la **Figura 117** se presenta la concentración media anual modelada para  $PM_{10}$  en cada uno de los años en estudio. Los valores más elevados de  $PM_{10}$  se encuentran predominantemente al N y NW de la ciudad de Buenos Aires. Las concentraciones más elevadas afectan los municipios de General San Martín, Tres de Febrero, Morón, Ituzaingó, Hurlingham y San Miguel. Durante este período se observó un máximo de  $PM_{10}$  hacia el año 2011 y luego un descenso paulatino. Además, es evidente los menores niveles de  $PM_{10}$  en áreas menos densamente pobladas, como en los municipios de Coronel Brandsen o Magdalena al sureste. En toda el área de estudio y a lo largo de los años analizados se excedió la concentración media anual recomendada por la OMS ( $20 \mu\text{g}\cdot\text{m}^{-3}$ ), esto representa un riesgo para la salud de la población.

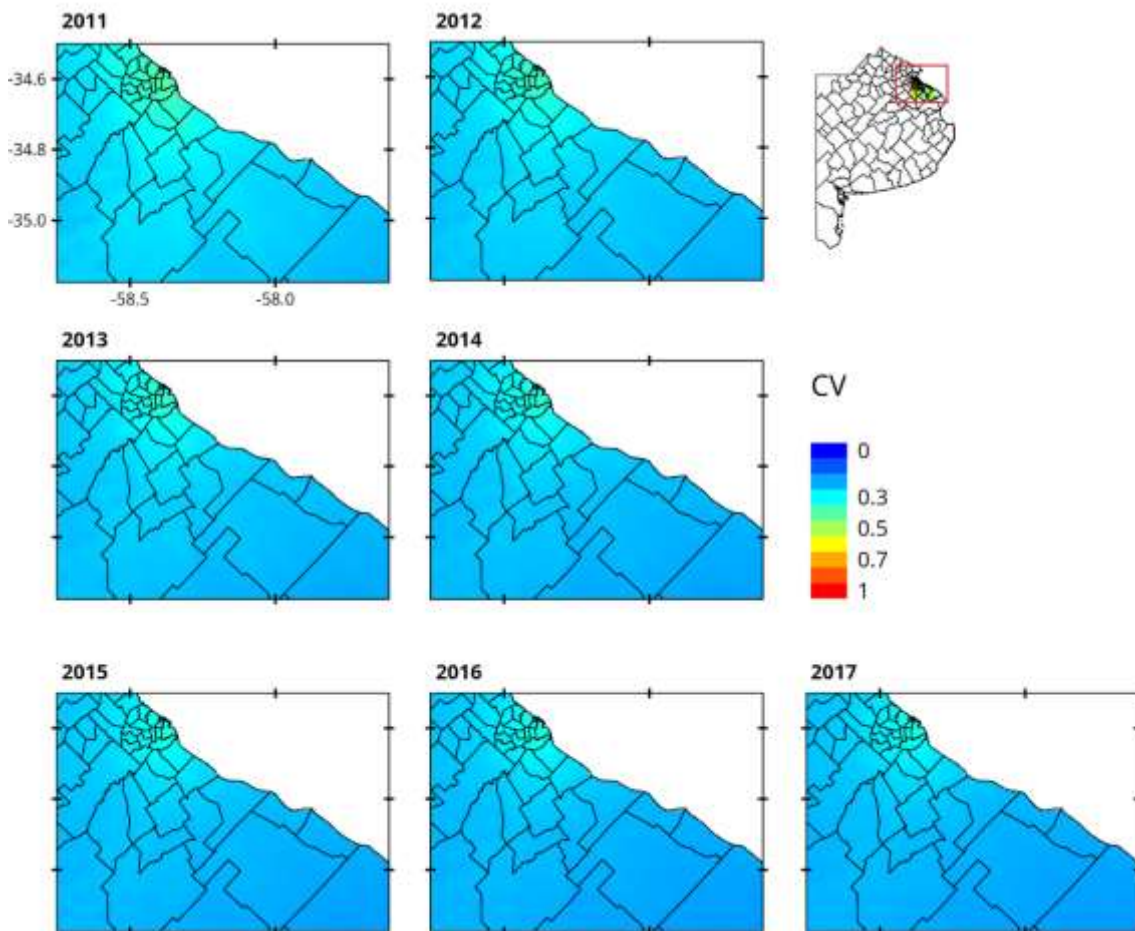


**Figura 117.** Mapas de concentración media anual de  $PM_{10}$  modeladas para la Ciudad de Buenos Aires y el Área Metropolitana de la Provincia de Buenos Aires en el período 2011-2017 (Unidades:  $\mu\text{g}\cdot\text{m}^{-3}$ ).

Para analizar la dispersión de la concentración media anual, se calculó el coeficiente de variación ( $CV = \text{desviación estándar} / \text{media}$ ) que expresa la dispersión relativa de los datos modelados respecto a la media aritmética. A diferencia de la desviación estándar, este parámetro aporta una mejor interpretación de la variabilidad de los datos ya que presenta una menor influencia de los valores extremos (**Figura 118**).

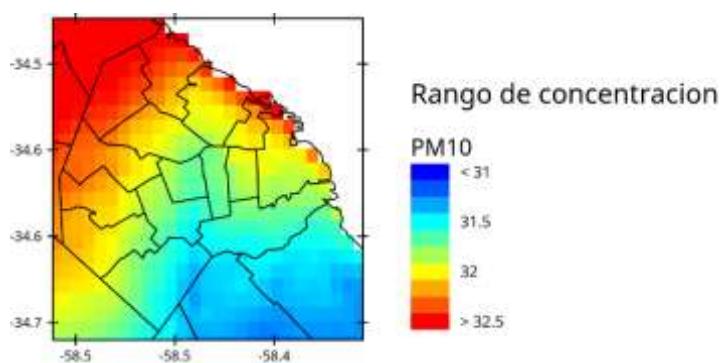


Los resultados obtenidos muestran desvíos entre un 10-40% para el período estudiado. La mayor variabilidad se observó en la Ciudad Autónoma de Buenos Aires, correspondiente al área más densamente urbanizada. Los resultados obtenidos a partir del coeficiente de variación evidencian la variabilidad intra-anual de los valores modelos para  $PM_{10}$ .



**Figura 118.** Mapas de coeficiente de variación para la concentración media anual de  $PM_{10}$  modelada para Ciudad de Buenos Aires y el Área Metropolitana de la Provincia de Buenos Aires en el período 2011-2017.

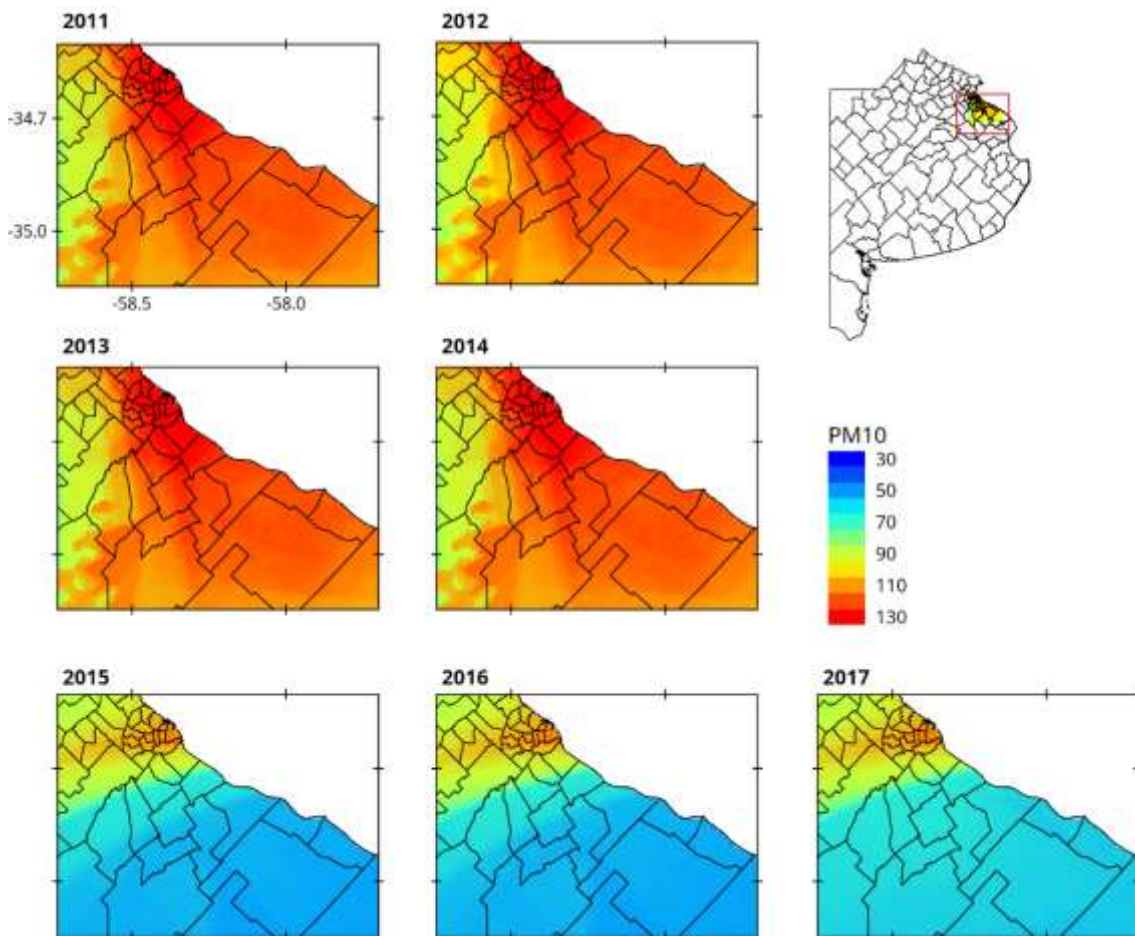
Con una escala de color diferente, se observa en la **Figura 119** las concentraciones medias anuales de  $PM_{10}$  para la Ciudad de Buenos Aires. Se observa que las zonas más afectadas son los barrios de Puerto Madero, Retiro, la zona norte de Palermo, Belgrano y Nuñez, Saavedra, Villa Urquiza, Villa Pueyrredón y Villa Devoto.



**Figura 119.** Mapa de concentraciones medias de  $PM_{10}$  en  $\mu g.m^{-3}$  durante el 2017 para la Ciudad de Buenos Aires

También, resulta interesante analizar la **Figura 120** donde se muestran mapas de las máximas concentraciones modeladas de  $PM_{10}$  para el área de estudio. Se puede ver que si bien la media anual de  $PM_{10}$  se encuentra entre  $25-35 \mu g.m^{-3}$  a lo largo de los distintos años, las concentraciones máximas diarias modeladas alcanzan los  $150 \mu g.m^{-3}$  y se observaron en la zona céntrica de Buenos Aires.

Se puede ver que el patrón espacial cambia a lo largo de los años. Entre 2011 y 2014 las mayores concentraciones se encontraron al E y SE de la Ciudad de Buenos Aires, mientras en 2015, 2016 y 2017 se ubicaron al NW.

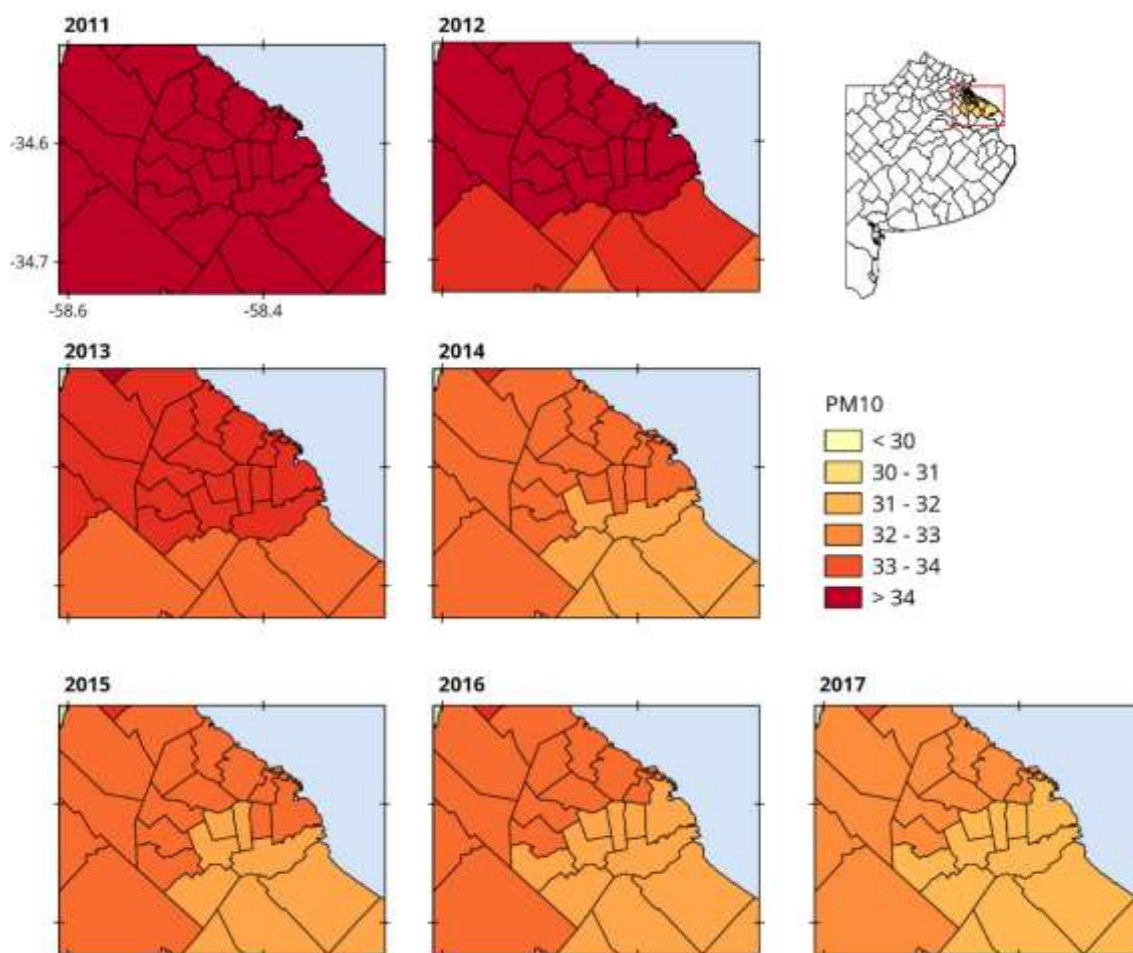


**Figura 120.** Mapas de concentraciones máximas modeladas anualmente de  $PM_{10}$  en  $\mu g.m^{-3}$

*IX. Mapas de calidad de aire, de riesgo y de FAM por exposición a  $PM_{10}$*

Como se mencionó en el Capítulo 2, la OMS establece como valor recomendado para la protección de la salud una concentración media anual de  $20 \mu g.m^{-3}$  de  $PM_{10}$ . En todo el período se ha superado en toda el área analizada el valor recomendado poniendo en riesgo la población expuesta.

En la **Figura 121** se observan los mapas de medias anuales de  $PM_{10}$  promediados en el área de cada comuna de la Ciudad de Buenos Aires. Se puede observar que las comunas ubicadas al N y NW alcanzan las mayores concentraciones medias anuales con respecto a las comunas ubicadas al S y SE. En particular, el 2017 fue el año menos afectado y el 2011 el mayor.



**Figura 121.** Mapas de concentraciones media anual de  $PM_{10}$  promediadas por comuna en Ciudad de Buenos Aires.

Se evaluó si existía una correlación entre las medias anuales y datos de salud entre los años 2011-2014 por comuna. Las variables fueron defunciones totales, defunciones de menores de un año y defunciones fetales [Dirección General de Estadísticas y Censos, 2016]. Los resultados no fueron concluyentes, mostrando una correlación no significativa. Sería interesante profundizar esta línea de trabajo desde la mirada de la epidemiología ambiental.

Para una mejor visualización se generaron mapas de calidad del aire a partir de los mapas de concentración de  $PM_{10}$ . En la **Figura 122** puede verse el mapa de calidad de aire para las concentraciones medias anuales durante el 2017 en Buenos Aires. Los resultados muestran que la situación media en Buenos Aires es potencialmente dañina para personas sensibles, ya sea por afecciones cardiovasculares o del aparato respiratorio.



Figura 122. Mapa de calidad del aire en Buenos Aires durante el 2017

El riesgo relativo a muerte para todas las edades por exposición a corto plazo de  $PM_{10}$  fue de 1.02 a lo largo del período de estudio, sin grandes variaciones espaciales o interanuales. En la **Figura 123** se observa la fracción atribuible de muerte por exposición a  $PM_{10}$ , la cual se ubica en el rango de 0.015 a 0.02 a lo largo del período de estudio. Esto implica entre un 1.5 - 2% de muertes atribuibles del total de muertes producidas.

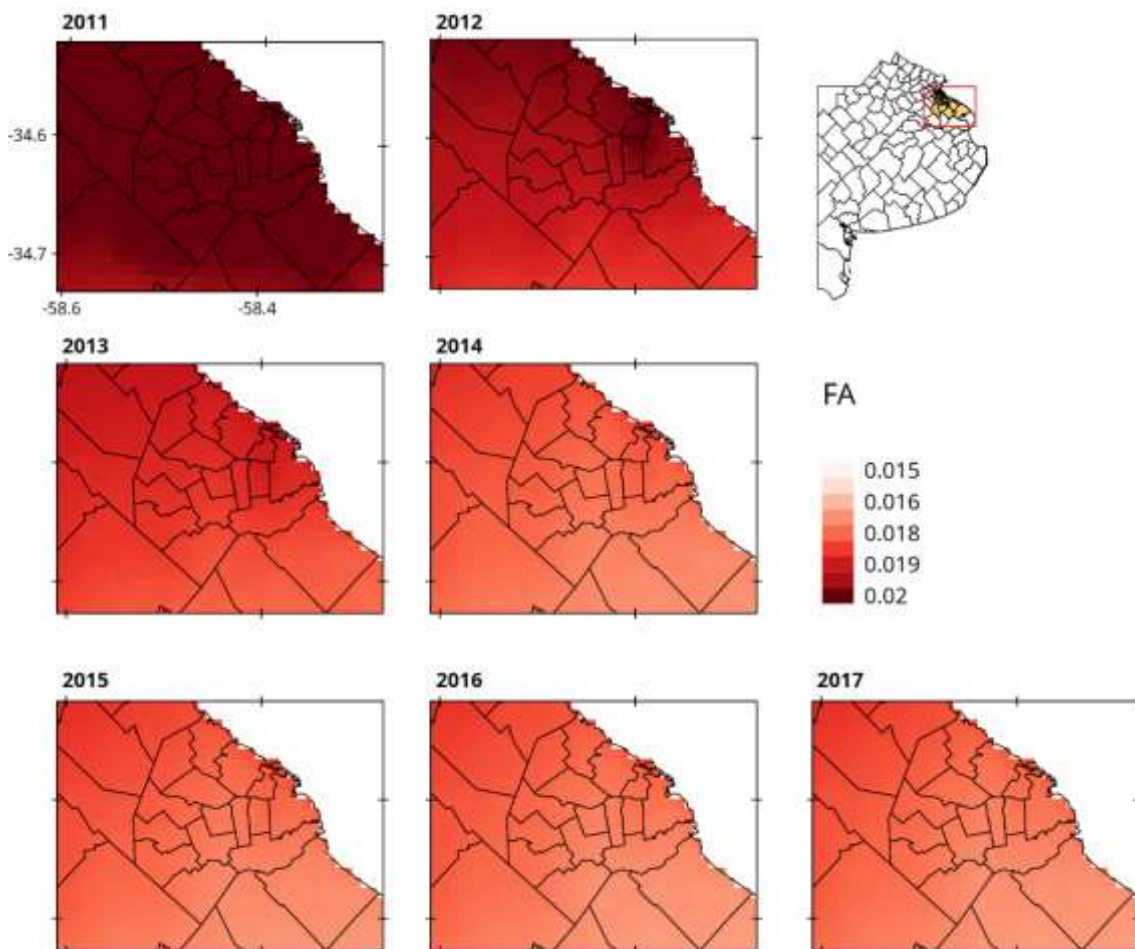


Figura 123. Mapa de la fracción atribuible de muertes prematuras por exposición a largo plazo a  $PM_{10}$

## III.II. Informe para la Ciudad de Buenos Aires

### *Aspectos del monitoreo terrestre*

- **¿Cuál es la calidad de la información recolectada por la red de monitoreo?**  
Se observaron largos períodos de ausencia de mediciones en la base de datos. La estación de Córdoba tiene el menor número de registros con sólo el 65% de los datos. La mayor falta de mediciones se dio durante el año 2013, faltando más del 50% de los datos horarios.

A su vez, se encontró una alta proporción de datos faltantes para los últimos meses de 2015 en todas las estaciones y todos los contaminantes. Esta anomalía resulta llamativa y alarmante. Como caso particular, las mediciones de PM<sub>10</sub> sólo se encuentran disponibles a partir de 2011.

- **¿Hay redundancia de información entre las estaciones?**  
Existe una redundancia entre la información recolectada por las estaciones. Sin embargo, no podemos descartar ninguna estación de monitoreo de la red porque no se repite el mismo patrón para los distintos contaminantes en la misma estación.
- **¿Cuáles son las características del área de emplazamiento de las estaciones de monitoreo?**  
Las 3 estaciones se encuentran dentro del entramado urbano, siendo fuertemente influenciadas por su entorno.
- **¿Cómo influyen los vientos locales sobre los niveles registrados en las estaciones?**  
Buenos Aires presenta una baja variabilidad estacional con una alta humedad relativa a lo largo de todo el año, lo cual puede afectar el proceso de deposición húmeda, en particular para PM<sub>10</sub>. Los períodos de calma prolongados proporcionan una estabilidad atmosférica que evita la dispersión de contaminantes. En este contexto, se espera que se produzca una persistencia de los contaminantes a lo largo del día.
- **¿De qué sector del viento llegan las concentraciones mayores?**  
En general, no hubo sectores que contribuyesen sustancialmente a las concentraciones medias de las tres estaciones.

En la estación Centenario, las mayores concentraciones de CO y NO<sub>2</sub> llegaron con mayor frecuencia desde el sector NW, con aporte de los sectores N y NE. Mientras que las principales contribuciones de PM<sub>10</sub> para esta estación provinieron de los sectores NW y SE.

En cuanto a la estación de Córdoba, los mayores aportes de CO provinieron del SE, mientras que las mayores concentraciones de NO<sub>2</sub> y PM<sub>10</sub> provienen de NE, E, SE y S.

Finalmente, en la estación de La Boca las mayores concentraciones de CO y NO<sub>2</sub> provienen de NW > N > NE. Las concentraciones de PM<sub>10</sub> también provienen mayoritariamente del NW, con contribuciones comparables de los sectores N, NE y SE.

### *Calidad del aire en los sitios de medición*

- **¿Cuáles son los niveles más frecuentes?**

Las concentraciones medias de CO fueron iguales para las estaciones de Centenario y Córdoba las medianas y medias (0.6 ppm) y levemente inferiores para la estación de La Boca (0.3 ppm). Las concentraciones medias de NO<sub>2</sub> están en el rango de 19.6-26.7 ppb, las esperadas para las zonas urbanas. Con respecto a la mediana de las concentraciones de PM<sub>10</sub>, las tres estaciones de monitoreo presentaron resultados cercanos (28.5-30 µg.m<sup>-3</sup>).

- **¿La concentración de los contaminantes excede los límites legales?**

Las concentraciones medidas están dentro de los límites legales para CABA.

- **¿Los contaminantes medidos ponen en riesgo a la población?**

Las concentraciones medias anuales de PM<sub>10</sub> superaron el valor recomendado por la OMS de 20 µg.m<sup>-3</sup> para cada año en las tres estaciones de control. Por otro lado, las concentraciones medias anuales de NO<sub>2</sub> también fueron superiores al valor recomendado (21 ppb) para al menos una de las estaciones en los años 2009, 2010, 2011, 2012, 2013 y 2017.

En cuanto al CO, se vieron tres ocasiones diferentes en las que las concentraciones superan el valor límite (25 ppm) indicado por la OMS, poniendo en peligro la salud de las personas expuestas [OMS, 1999]. Estos eventos tuvieron lugar los días 25 y 26 de octubre de 2010, 23 y 24 de junio y 23 y 24 de diciembre de 2014.

### *Evolución temporal de los contaminantes*

- **¿A cuál hora del día es mayor la concentración de los contaminantes?**

Los comportamientos horarios de CO y NO<sub>2</sub> presentaron patrones típicos para áreas urbanas como picos relacionados con el alto flujo vehicular durante el mediodía y la tarde y pronunciados descensos durante la noche. En el caso de PM<sub>10</sub>, se observó con mayor intensidad una capacidad de acumulación de este contaminante. Las concentraciones aumentaron al principio de la semana, luego se mantuvieron en un promedio de 30 µg.m<sup>-3</sup> los días laborables y disminuyeron en las primeras horas del sábado.

- **¿Las concentraciones máximas varían a lo largo de la semana?**

Las concentraciones de NO<sub>2</sub> y PM<sub>10</sub> disminuyeron para el fin de semana lo cual demuestra la fuerte vinculación de estos contaminantes al flujo vehicular. Para el CO, la estación de Córdoba tuvo un fuerte descenso también hacia el fin de semana, a diferencia de las estaciones La Boca y Centenario que no se vieron influenciadas por los días de la semana.

- **¿Durante cuál época del año ocurren los episodios de máximas concentraciones?**

Se observó un patrón de disminución para las concentraciones de CO y NO<sub>2</sub>, siguiendo el orden invierno > otoño > primavera > verano, con diferencias significativas entre ellas (valor p < 0.001).

Las concentraciones de  $PM_{10}$  tuvieron un patrón diferente para los promedios estacionales, con los valores más altos observados en invierno y las concentraciones más bajas en otoño.

- **¿Cómo han cambiado los niveles de los contaminantes con los años?**

La tendencia total en la estación Córdoba fue negativa para  $PM_{10}$  y  $NO_2$  y positiva para CO. Las contribuciones de CO fueron significativas desde el NW y el NE. En la estación Centenario la tendencia total fue negativa para  $PM_{10}$  y  $NO_2$  negativa, y para CO, la tendencia fue no significativa. Por su parte, en la estación La Boca, la tendencia total fue no significativa para  $PM_{10}$ , mientras que para  $NO_2$  fue negativa y para CO fue positiva con aportes significativos desde el NE, E, SE, S, SW y W.

Las concentraciones medias mensuales de  $PM_{10}$  presentaron una tendencia global a la baja en las estaciones de Córdoba y Centenario, se observó una tendencia significativa en aumento en las direcciones NE y E y desde el E para cada una de ellas.

### *Distribución espacial de los contaminantes*

- **¿Se observa un patrón espacial de las concentraciones? ¿Dónde ocurren las máximas concentraciones?**

El patrón espacial cambia a lo largo de los años. Entre 2011 y 2014 las mayores concentraciones se encontraron al E y SE de la Ciudad de Buenos Aires, mientras en 2015, 2016 y 2017 se ubicaron al NW.

- **¿Cómo es la calidad del aire donde no se mide?**

Las simulaciones realizadas muestran que las concentraciones de  $PM_{10}$  aumentan hacia el N, O y SO del área de estudio.

- **¿La calidad del aire en la región representa un riesgo para la población?**

La calidad del aire ha mostrado ser dañina para personas sensibles. La fracción atribuible de muerte por exposición a  $PM_{10}$  se ubica en el rango de 0.015 a 0.02 a lo largo del período de estudio. Esto implica entre un 1.5 - 2% de muertes atribuibles del total de muertes producidas.





*Caso de aplicación III*

## **La Plata, Argentina**

---



*“Los hechos son inciertos, los valores son discutibles,  
la relevancia es alta, las decisiones son urgentes.”*

Funtowicz & Ravetz (1991)

## Introducción

Como vimos en el caso de aplicación anterior, Argentina cuenta con un somero monitoreo de la calidad del aire centralizado en la ciudad de Buenos Aires. En el año 2018 se incorporó una nueva estación de monitoreo automático de calidad de aire en la Ciudad de Córdoba, lugar que ha sido monitoreado de forma interrumpida en la última década. A esta escueta red se le suma la información que publican 2 de los polos petroquímicos más grandes del país (Dock Sud y Bahía Blanca), ubicados en la Provincia de Buenos Aires.

La Plata, a diferencia de los casos anteriores, no presenta un monitoreo diario de la calidad del aire. Únicamente se cuenta con las mediciones que el equipo del Centro de Investigaciones del Medioambiente (CIM, UNLP-CONICET) ha podido recolectar mediante un monitoreo manual, que se ha intentado sostener a lo largo de los años de forma periódica.

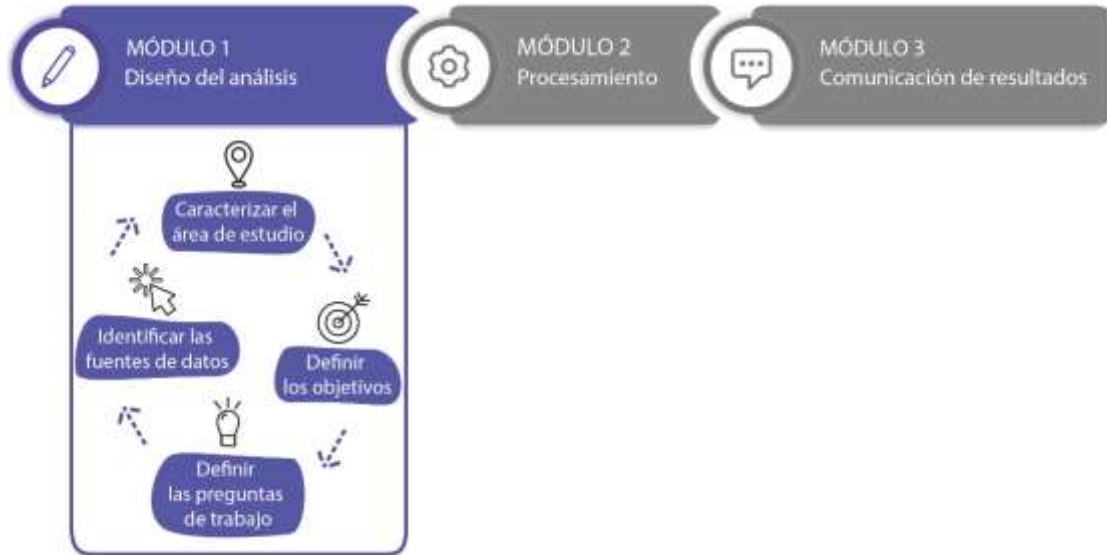
El caso de La Plata se presenta como un ejemplo representativo de la situación de la mayoría de las ciudades del mundo, como afirma la OMS [World Health Organization, 2019], donde no se cuenta con información de la calidad del aire. Obtener mapas de los contaminantes del aire y del riesgo asociado en esta ciudad fue el objetivo motor de esta tesis, ya que resulta inminente contar con información respecto a la calidad del aire que se respira. En ese sentido, este capítulo avanza sobre las expectativas sembradas hace 4 años atrás, cuando se buscó desde la ciencia y la tecnología saldar una deuda pública.

La falta de inversión durante años hizo que desde la ciencia se busque completar aquello que desde la administración estatal no se pudo proveer: información confiable para evaluar el riesgo de la población a la exposición de contaminantes atmosféricos. Más allá del esfuerzo que acompañó este proceso, parte del aprendizaje ha sido que para poder completar un análisis de calidad de aire se precisan datos completos y confiables.

A continuación, se aplica la metodología propuesta para el estudio de la calidad del aire en el Partido de La Plata. Se trabajó con las mediciones de  $PM_{10}$  mediante monitoreo manual y con los modelos generados para la Ciudad de Buenos Aires. A diferencia de los casos anteriores no se cuenta con datos de otros contaminantes.

# Aplicación de la metodología

## Módulo I - Etapa de diseño del análisis



### I.I. Definición y caracterización del área de estudio

El área del Gran La Plata se encuentra integrada por los partidos de Berisso, Ensenada y La Plata, en el noreste de la provincia de Buenos Aires, a 50 km al sureste de la Ciudad Autónoma de Buenos Aires (**Figura 124**). Esta región se encuentra en la margen sur del estuario del Río de la Plata, extenso cuerpo de agua con forma de embudo que tiene 300 km de largo en dirección NW-SE y anchos de 200 (límite exterior) y 40 km (límite interior). La Plata, ciudad capital de la Provincia, junto a las localidades adyacentes, alberga una población superior a los 800 mil habitantes [INDEC, 2010].

Diferentes trabajos científicos han demostrado que se exceden en la región los valores recomendados por la OMS para el material particulado (20 y 10  $\mu\text{g}\cdot\text{m}^{-3}$  para  $\text{PM}_{10}$  y  $\text{PM}_{2.5}$ , respectivamente), y que se presentan distintos compuestos orgánicos volátiles en diferentes grados de concentración [Colman et al., 2014; Orte et al., 2015]. También, se cuenta con estudios de efectos a la salud [Whichmann et al., 2009; Colman Lerner et al., 2014]. Pese a esto, no existe un monitoreo continuo de la calidad del aire por fuera del monitoreo que realiza el Polo Petroquímico como control interno y por expreso pedido del Organismo Provincial de Desarrollo Sostenible (OPDS), datos que no son de acceso público.



**Figura 124.** Los partidos de La Plata, Berisso y Ensenada componen el área del Gran La Plata

### *Relevamiento de fuentes*

En el área central del conglomerado urbano, entre los Partidos de La Plata, Berisso y Ensenada, se encuentra un complejo petroquímico que posee la principal refinería de petróleo del país, con los mayores volúmenes de producción a nivel nacional (38 mil m<sup>3</sup> de petróleo refinado por día) [Represa et al., 2016]. Este complejo incluye también seis plantas petroquímicas, con una producción superior a las 650 mil Tn por año de diversos compuestos tales como compuestos aromáticos (benceno, tolueno, xilenos), disolventes alifáticos (n-pentano, n-hexano, n-heptano), polipropileno, anhídrido maleico, ciclohexano, metanol y coque de petróleo [Giuliani et al., 2017].

En zonas aledañas al complejo petroquímico se halla la central termoeléctrica “Central Térmica Ensenada de Barragán”. Esta central tiene una capacidad de generación de 560 MW y es una de las grandes generadoras de energía eléctrica del país [Ratto, 2016]. La inauguración de la primera turbina se realizó en enero del 2012 y fue habilitada para operar comercialmente en abril del mismo año.

Sobre la vera del Río, se encuentran una empresa siderúrgica de gran envergadura, un astillero y un puerto con gran movimiento naviero, junto con un conjunto de industrias de distintos rubros. El Puerto de La Plata es un puerto fluvial dedicado preponderantemente al transporte de cargas con potencial importancia para el Mercosur. Parte de este puerto está compuesto por el Astillero Río Santiago, uno de los más grandes de América Latina [ARS, 2015], dedicado a la construcción de barcos mercantes y de guerra, así como de material ferroviario.

La Plata tiene una alta densidad de población con la mayor tasa automóvil por persona de Argentina (1 automóvil cada 2 personas). El centro de la ciudad se caracteriza por el tráfico intenso, con un aumento del número de rodados cerca del 75% en la última década [Massolo et al., 2010].

### *Características meteorológicas*

El clima del Gran La Plata es de tipo templado-húmedo, con precipitaciones medias anuales de 1079 mm. año<sup>-1</sup>, humedad relativa del 77% y temperatura media de 15°C [Kruse et al. 2014]. Según la clasificación de Köppen-Geiger, el área presenta un clima templado (C), subtropical, sin estación seca (pampeano) y con veranos cálidos (fa). De acuerdo con la clasificación en el sistema de Thornthwaite el clima es húmedo (B2) y microtermal (C'2), con valores de evapotranspiración mayores a 570 mm [Fucks et al., 2017].

Las temperaturas medias anuales se encuentran entre los 18°C en el norte de la región y 16.7°C en el sur, con precipitaciones total anuales de alrededor de 1.000 mm [Barros et al., 2005]. Las precipitaciones tienen su máximo durante el verano y su mínimo durante el invierno.

Por su parte, las circulaciones eólicas dependen principalmente del anticiclón subtropical del Océano Atlántico Sur. La ubicación de este sistema de alta presión cambia durante el año al mismo tiempo que interactúa con diferentes centros ciclónicos a lo largo del territorio argentino, esto produce que en invierno la dirección media del viento sea NW mientras que en verano E-NE. El ciclo anual implica una rotación en el sentido de las agujas del reloj a través de las cuatro estaciones, desde el invierno hasta el verano [Barros et al., 2005].

### *Geografía del área*

La ciudad de La Plata se encuentra ubicada en la Región Pampeana, en el extremo oriental de la Pampa Deprimida [Frenguelli, 1950]. Es una planicie de escasa pendiente con paisaje de llanura caracterizado por su bajo relieve.

Los partidos de La Plata, Berisso y Ensenada se encuentran ubicados en dos unidades geomorfológicas bien diferenciadas: la planicie continental, desarrollada al SW, también denominada zona interior [Fidalgo y Martínez 1983] y la planicie o llanura costeras marginal, al noreste [Fidalgo y Martínez, 1983; Cavallotto, 1995].

La planicie continental es la zona con mayor desarrollo urbanístico y productivo de la región, donde la altura promedio sobre el nivel del mar es mayor (un máximo de 50 msnm al oeste del Partido de La Plata) y se alcanza una pendiente media de ~0.27% [Fucks et al., 2017].

### *Marco normativo*

Como se mencionó en el capítulo anterior, en Argentina, el dominio sobre el ambiente y los recursos naturales recae en las Provincias. Son estas quienes tienen como obligación controlar el impacto ambiental de todas las actividades que perjudiquen al ecosistema; promover acciones que eviten la contaminación del aire; garantizar el derecho a solicitar y recibir la adecuada información y participar en la defensa del ambiente, de los recursos naturales y culturales, entre otras.

Recientemente el Organismo Provincial de Desarrollo Sostenible de la Provincia de Buenos Aires (OPDS), actualizó los niveles guías para contaminantes específicos mediante el Decreto N°

1.074/188 que reglamenta la Ley N°5.965/58. El mismo es objeto de controversias, dado que si bien incorpora el  $PM_{2.5}$  al grupo de los contaminantes regulados, plantea un esquema menos restrictivo que el decreto anterior para las diferentes emisiones.

Respecto a los valores norma para los estándares de calidad de aire, se fijaron etapas de reducción progresiva a los 2, 3 y 4 años de publicado el Decreto, pudiendo la Autoridad de Aplicación prorrogar la entrada en vigor de alguna de las etapas detalladas. Para  $PM_{10}$  se fijó un límite a las 24hs de  $150 \mu\text{g.m}^{-3}$  para no ser superado en más de una vez al año, y una concentración media anual de  $50 \mu\text{g.m}^{-3}$  para todas las etapas.

## I. II. Determinación de los objetivos

La falta de información limita los objetivos que se pueden abordar en el caso del Gran La Plata. En función de esta falta de información, se pueden abordar solamente los siguientes objetivos:

- *Caracterizar aspectos del monitoreo terrestre*

Describir los sitios en donde se ha monitoreado y las técnicas utilizadas. Analizar la influencia del área de monitoreo sobre las mediciones realizadas.

- *Conocer la evolución temporal de los contaminantes*

Evaluar la variación de las concentraciones a lo largo de los años. Estudiar las variaciones interanuales.

- *Evaluar la distribución espacial de los contaminantes*

Identificar las zonas donde se producen las máximas concentraciones y estudiar si las mismas cambian en el tiempo. Estudiar mediante mapas de concentración si las zonas donde no se monitorean presentan concentraciones elevadas. Determinar la calidad del aire y si la población se ve expuesta a un riesgo.

## I.III. Delimitación de las preguntas de estudio

Las preguntas planteadas para este caso de estudio son:

### **Caracterizar aspectos del monitoreo terrestre**

¿Cuáles son las técnicas utilizadas para el monitoreo?

¿Cuáles son las características del área de emplazamiento de las estaciones de monitoreo?

### **Conocer la evolución temporal de los contaminantes**

¿En cuál época del año ocurren los episodios de máximas concentraciones?

¿Cómo han cambiado los niveles de los contaminantes con los años?

### **Evaluar la distribución espacial de los contaminantes**

¿Dónde ocurren las máximas concentraciones?

¿Se observa un patrón espacial de las concentraciones?



¿Cómo es la calidad del aire donde no se mide?

¿La calidad del aire en la región representa un riesgo para la población?

## I. IV. Adquisición de los datos

La información disponible para la ciudad de La Plata y sus zonas aledañas es realmente limitada. La municipalidad de La Plata no realiza mediciones de la calidad del aire, y el OPDS no brinda información respecto a los contaminantes que monitorea o supervisa, su lugar de medición, técnicas utilizadas, ni sitio de emplazamiento. Solamente, se cuenta con las mediciones periódicas que realiza el Centro de Investigación del Medio Ambiente (CIM, CONICET-UNLP) en distintos puntos de la ciudad, las cuales son objeto de estudio para distintas líneas de investigación. A partir de esas campañas se desprenden varios trabajos que dan cuenta del alto impacto sobre la salud de la población que genera la contaminación de la región [Massolo et al., 2010].

En vistas a generar mapas de calidad del aire y mapas de riesgo para esta región, se procedió a solicitar los resultados de las distintas campañas realizadas. A su vez, por la cercanía a la Ciudad de Buenos Aires se utilizaron las imágenes satelitales calibradas en el CEILAP y el modelo espacial para PM<sub>10</sub> generado en el **Capítulo 6**.

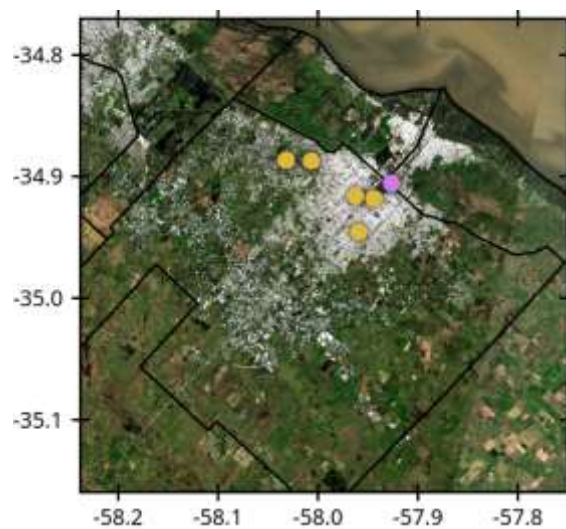
### *Teledetección y datos de monitoreo satelital*

Dado que la ciudad de La Plata no cuenta con un fotómetro en sus cercanías para la validación de las imágenes satelitales, se utilizaron las imágenes MCD19A2 Versión 6 calibradas en el CEILAP de Buenos Aires (**Capítulo 6**). Como se mencionó anteriormente este producto satelital implementa el algoritmo MAIAC en una cuadrícula con una resolución de 1 km por píxel. Se descargaron en formato TIF un total de 12055 imágenes para el período comprendido entre 2009-10-01 y 2017-10-01. El sistema de referencia utilizado fue WGS84.

### *Datos de monitoreo terrestre*

Para evaluar el ajuste del modelo confeccionado para la Ciudad de Buenos Aires en la localidad de La Plata se utilizaron las mediciones realizadas en el marco de la tesis doctoral de la Lic. Daniela Giuliani. Estas mediciones se realizaron en distintos períodos comprendidos entre el 11/2011 al 10/2017, contándose con un total de 30 mediciones de PM<sub>10</sub>.

Los sitios de monitoreo fueron variando a lo largo de los años, centrándose en el área urbana y en el área industrial en las proximidades del Polo Petroquímico (**Figura 125**). Se utilizó un equipo muestreador de MP de bajo volumen MiniVol TAS (Airmetrics), a un flujo constante de 5 l.min<sup>-1</sup>, con una duración de 3 días y se utilizaron filtros de politetrafluoroetileno (PTFE) como medio colector. La razón de la extensión del muestreo es obtener una masa crítica medible por gravimetría [Colman-Lerner et al., 2011].



**Figura 125.** Sitios de monitoreo en La Plata. Se observan en amarillo los sitios de monitoreo urbanos y en rosa el sitio de monitoreo industrial.

### *Datos meteorológicos*

Los datos meteorológicos abarcan el período 2010-2018 y fueron proporcionados por el Servicio Meteorológico Nacional (SMN) bajo solicitud. Los datos consisten en promedios horarios de las direcciones, velocidades y calmas del viento. El sitio de medición corresponde al Aeropuerto de La Plata (zona rural) que se encuentra a unos 8 km del centro de la ciudad, a 36 msnm.

## Módulo II - Etapa de procesamiento



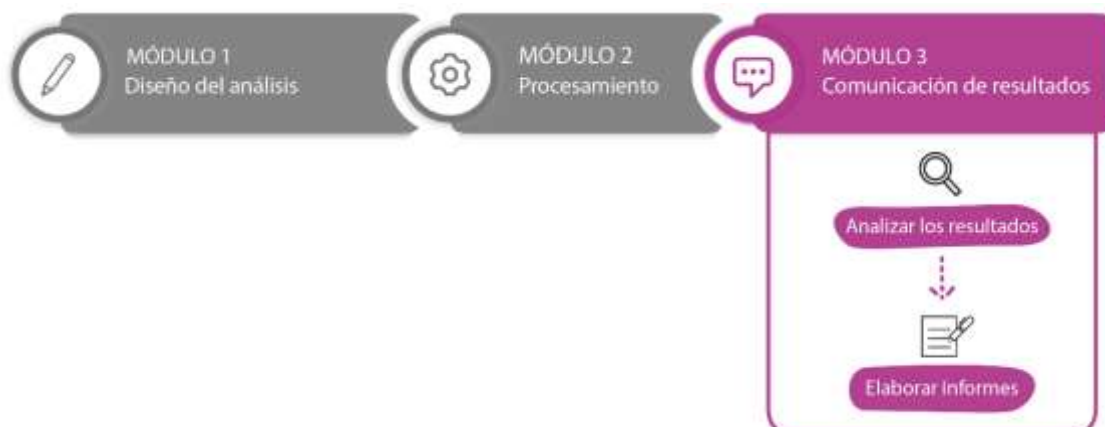
### II. I. Preparación de los datos y selección de algoritmos

Para esta zona se trabajó con los mapas de concentración de  $PM_{10}$  elaborados en el Capítulo 6 para la Ciudad de Buenos Aires. Dadas que son las únicas mediciones que se cuentan para la región de estudio, se utilizaron las mediciones realizadas de forma manual por el equipo del CIM para evaluar el ajuste del modelo a esta región. Las métricas utilizadas fueron el RMSE y el MAE.

Gracias a la información suministrada por el SMN, se analizaron los vientos predominantes en la región utilizando rosas de vientos. A su vez, se realizó un análisis descriptivo de la evolución de la concentración de  $PM_{10}$  en los distintos sitios monitoreados.

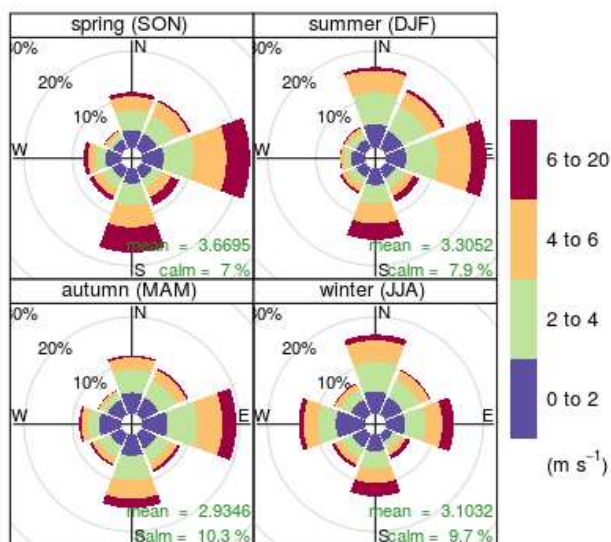
El cálculo del riesgo relativo (RR) por todas las causas de muerte por exposición a corto plazo de  $PM_{10}$  y de la fracción atribuible de muertes (FAM) para todas las edades se realizó mediante la metodología descrita en la propuesta metodológica.

## Módulo III - Etapa de comunicación de los resultados



### III.I. Análisis de los resultados

Tal como puede observarse en la **Figura 126**, los vientos en la región son leves -moderados, con un 8.7% de calma en todo el año y vientos medios de 3.25 m.s<sup>-1</sup>.

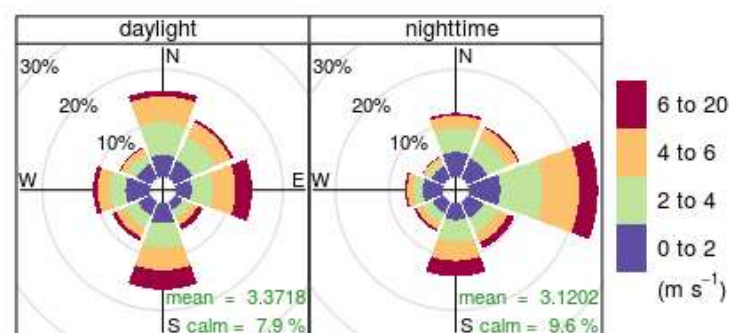


**Figura 126.** Rosa de Vientos estacional. Frecuencia de la dirección del viento en %

La frecuencia de los vientos NE, S y N son muy elevadas respecto al resto de las direcciones a lo largo de las estaciones del año, diferente a las direcciones señaladas para otros sectores de la ciudad en [Ratto et al., 2012]. Según [Barros et al., 2005], estas tres direcciones del viento son originadas por el flanco oeste del anticiclón subtropical del Océano Atlántico Sur.

El área de estudio es una amplia llanura con leves variaciones en el terreno, donde la única fuente significativa de variabilidad atmosférica local es el contraste de temperatura entre el suelo y la superficie del Río de la Plata [Ratto et al., 2010]. Si bien se espera que la brisa marina se produzca entre NNW y ESE (en el sentido de las agujas del reloj) dentro de un ciclo diurno y que su influencia sea pronunciada durante el verano debido al mayor contraste de temperatura entre las superficies [Ratto et al., 2012], esto no se observa en los datos registrados por la estación meteorológica ubicada a 12 km de la costa.

Durante la noche el viento E es más frecuente (**Figura 127**), esto puede ser atribuible a la brisa terrestre porque esta dirección del viento es algo perpendicular a la costa. Durante las primeras horas de la mañana la frecuencia el viento E disminuye, aumentando la influencia de los vientos meridionales alcanzando un máximo pasado el mediodía.



**Figura 127.** Rosa de vientos diario. Frecuencia de la dirección del viento en %

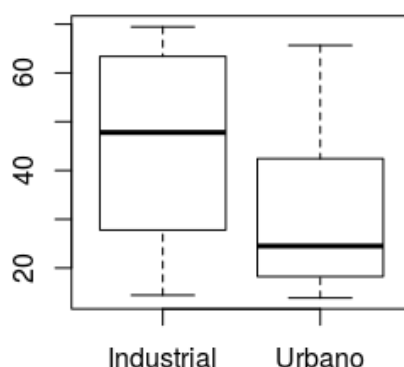
Para este trabajo se contó con las mediciones realizadas a lo largo del período comprendido entre el 11/2011 al 10/2017 mediante un método de monitoreo discreto, el cual permite conocer la concentración diaria promedio de material particulado tras 3 días de medición. En la **Tabla 22** pueden observarse los estadísticos descriptivos de las concentraciones de PM<sub>10</sub> registradas.

**Tabla 22.** Análisis descriptivo de las concentraciones de PM<sub>10</sub> medido en La Plata

Sitio	n	min	media	mediana	máximo
<i>Industrial</i>	12	14.4	44.7	46.4	69.4
<i>Urbano</i>	18	13.9	33.8	26.3	65.6

La distribución de las mediciones no es heterogénea en todo el período. Se cuenta con pocas mediciones para 2011 (n = 1), 2012 (n = 3) y 2013 (n = 2). A su vez, se encuentran algunos meses infrarrepresentados (enero 1, febrero 2, diciembre 1, agosto = 1, septiembre = 1). Esto imposibilita sacar conclusiones respecto a tendencias inter o intra-anales.

Como las mediciones se realizaron en 2 sitios bien diferenciados como industrial y urbano, se evaluó el efecto del área circundante sobre las concentraciones analizando su varianza. El análisis ANOVA no mostró que haya efectos significativos del sitio de estudio sobre los niveles medidos en ese período de estudio, aunque puede verse una tendencia a presentar mayores niveles en el área industrial (**Figura 128**).



**Figura 128.** Boxplot de las mediciones realizadas de  $PM_{10}$  para La Plata.

### Evaluación de los mapas de concentración de $PM_{10}$

Para corroborar el ajuste del modelo realizado con las mediciones históricas de la Ciudad de Buenos Aires en la ciudad de La Plata, se utilizaron las mediciones del monitoreo discreto. Como se mencionó anteriormente, con el monitoreo discreto se obtiene una concentración promedio de los días que llevó la medición. Por tanto, fue necesario condensar los valores modelados para cada día que duró el monitoreo en un único estadístico descriptivo.

De las 30 mediciones con las que se contaba, solo 26 mediciones coincidieron con al menos una imagen completa del modelo de  $PM_{10}$  realizado para el área metropolitana de Buenos Aires. En estos casos, se extrajo el valor máximo y el valor promedio de la serie temporal para el sitio donde se realizó el monitoreo discreto. De esta forma se obtuvo un set de datos con los máximos y los promedios de las mediciones modeladas de  $PM_{10}$  en cada sitio donde se midió  $PM_{10}$  en la ciudad de La Plata.

Como primer paso se analizaron las correlaciones entre los valores máximos y los valores promedios simulados con las mediciones. La correlación no fue significativa con las medias modeladas, pero sí con las concentraciones máximas de  $PM_{10}$  ( $r = 0.45$ ,  $p < 0.05$ ).

En la **Figura 129** se puede ver el gráfico de dispersión entre las concentraciones predichas y las realizadas. El ajuste para el conjunto de datos resultó en un  $RMSE = 18.7 \mu g.m^{-3}$  y  $MAE = 14.9 \mu g.m^{-3}$ , valores considerablemente mayores que la validación obtenida en Buenos Aires. A su vez, el MAPE fue del 41.5% sobre los datos observados.

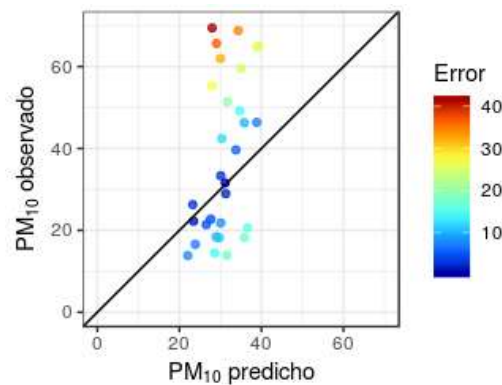


Figura 129. Ajuste del modelo respecto a mediciones

Al analizar por categorías se observan ciertas diferencias entre los sitios de monitoreo. El modelo presenta un mayor error (RMSE = 21.3  $\mu\text{g}\cdot\text{m}^{-3}$ , MAE = 17.2  $\mu\text{g}\cdot\text{m}^{-3}$ ) en la zona industrial, mientras que en la zona urbana estos errores se reducen (RMSE = 16.5  $\mu\text{g}\cdot\text{m}^{-3}$ , MAE = 13.3  $\mu\text{g}\cdot\text{m}^{-3}$ ). Esto nos indica que la representatividad es mejor en áreas similares a las empleadas para la confección del modelo. Sin embargo, se realizó una prueba ANOVA para evaluar si el sitio ejerce efectos sobre el error, y se observó que los mismos no son significativos ( $F = 0.712$ ,  $df = 1$ ,  $p\text{-valor} > 0.4$ ).

Por otra parte, cabe recordar que el método empleado en el monitoreo de la ciudad de La Plata es un método discreto, donde la medida obtenida resulta de la integración de lo observado a lo largo de varios días (entre 2 y 4 días). Este método es distinto al empleado para las mediciones de PM<sub>10</sub> en Buenos Aires pudiendo haber diferencias entre las técnicas empleadas.

A su vez, los datos ingresados en el modelo obedecen a las condiciones atmosféricas de una ventana temporal próxima al paso de los satélites, las cuales no son constantes a lo largo del día. Por el contrario, la variación diaria puede ser amplia. Por ejemplo, en la **Figura 130** se observan los niveles modelados para 3 fechas consecutivas en las que se registró 68.77  $\mu\text{g}\cdot\text{m}^{-3}$  y 59.54  $\mu\text{g}\cdot\text{m}^{-3}$  en el monitoreo discreto para el área industrial y urbana (estas fechas presentan el mayor error entre el modelo y las mediciones). Si bien el modelo da cuenta de las altas concentraciones al segundo día, es factible imaginar que los máximos niveles ocurrieron entre los momentos modelados.

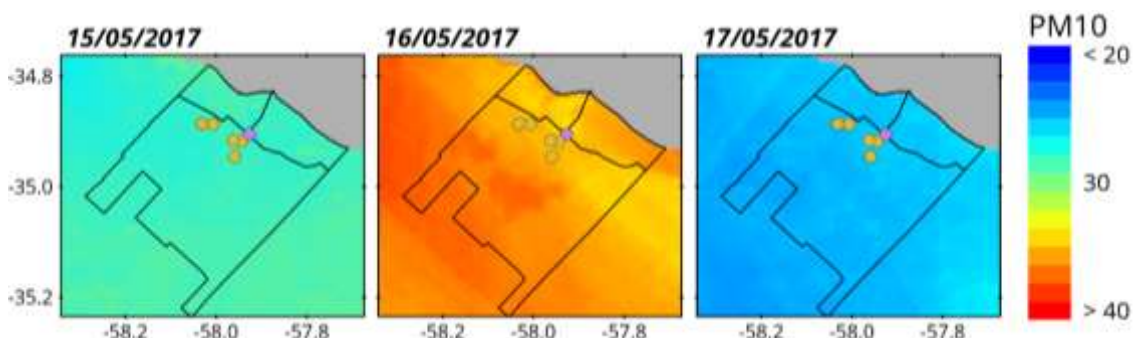
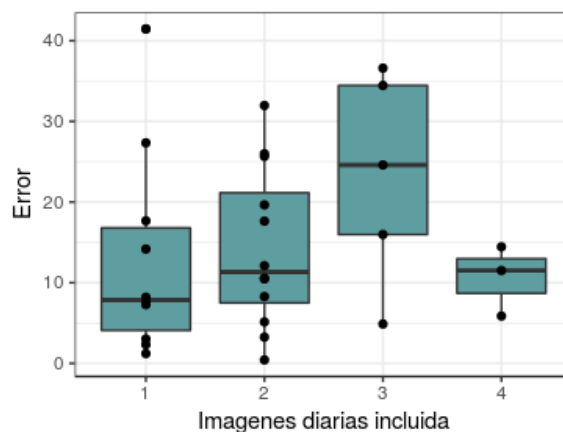


Figura 130. Variación al cabo de tres días de medición con el equipo discreto.

Por otra parte, faltaron salidas diarias del modelo para todos los períodos en que se produjeron las mediciones de  $PM_{10}$ . En algunos casos se contó solo con una única imagen para los días que se realizaron mediciones. En la **Figura 131** se puede observar la variación del error en función del número de salidas del modelo de  $PM_{10}$  incluidas en la confección de la media. Si bien se observa que existen diferencias en la variación del error de las estimaciones cuando se utilizan distinto número de imágenes, estas diferencias no son significativas ( $F = 1.39$ ,  $df = 3$ ,  $p\text{-valor} > 0.27$ ).



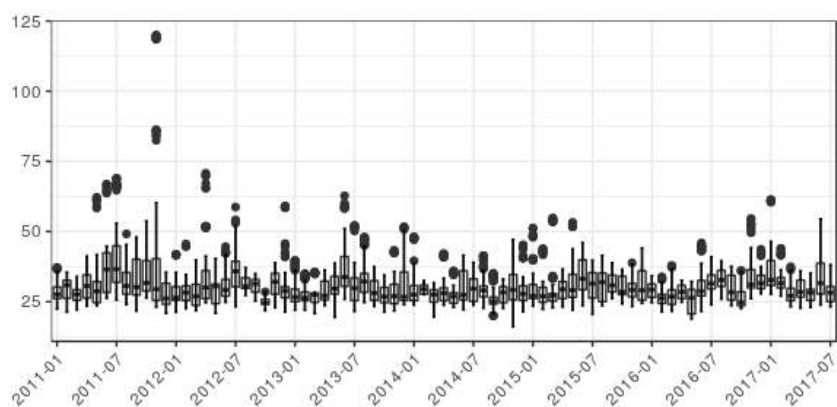
**Figura 131.** Variación del error respecto a la cantidad de imágenes utilizadas para confeccionar la media

Como antecedente en esta área de estudio, cabe mencionar el sistema de modelado WRF/CALMET/CALPUFF aplicado por [Allende et al., 2016]. Este modelo fue realizado en función de un inventario de emisiones generado para la ciudad presentando muy buenos resultados: se obtuvo un  $RMSE = 9 \mu g.m^{-3}$ , mostrando ligeras sobreestimaciones de las concentraciones de  $PM_{10}$ . Sin embargo, los autores señalaron que posiblemente existan fuentes no contempladas en el inventario, ya que en uno de los sitios de control se subestimó la concentración de  $PM_{10}$ . Resultaría interesante incorporar la salida simulada de este modelo como fuente de información a una nueva simulación sobre el área.

### Variación temporal de $PM_{10}$

De las concentraciones modeladas, no se observó un efecto significativo de los sitios, pero sí se observaron efectos por el factor mes ( $df = 11$ ,  $F = 44.5$ ,  $p\text{-valor} < 0.05$ ), año ( $df = 8$ ,  $F = 39.0$ ,  $p\text{-valor} < 0.05$ ) y una interacción entre mes-año ( $df = 74$ ,  $F = 12.8$ ,  $p\text{-valor} < 0.05$ ). Se detectó una tendencia estacional con un aumento de las concentraciones durante los meses fríos (junio, julio) y con una disminución de los valores en los meses más cálidos (noviembre, diciembre) (**Figura 132**). Esto había sido señalado con anterioridad por [Bilos et al., 2001].





**Figura 132.** Boxplot mensuales de PM<sub>10</sub> simulado para todo el período en las estaciones urbanas.

Finalmente, se analizó la tendencia de las mediciones modeladas mediante el método de Theil-Sen. Los resultados indicaron que la tendencia es no significativa para el sitio en el caso urbano y para el área industrial.

### Variación espacial de PM<sub>10</sub>

En la **Figura 133** se pueden ver los mapas de las concentraciones medias anuales de PM<sub>10</sub> para el área del Gran La Plata. Como se veía en el caso anterior, las concentraciones son mayores los años 2011 y 2012. A su vez, se observan mayores niveles de PM<sub>10</sub> en las áreas costeras de Berisso y Ensenada.

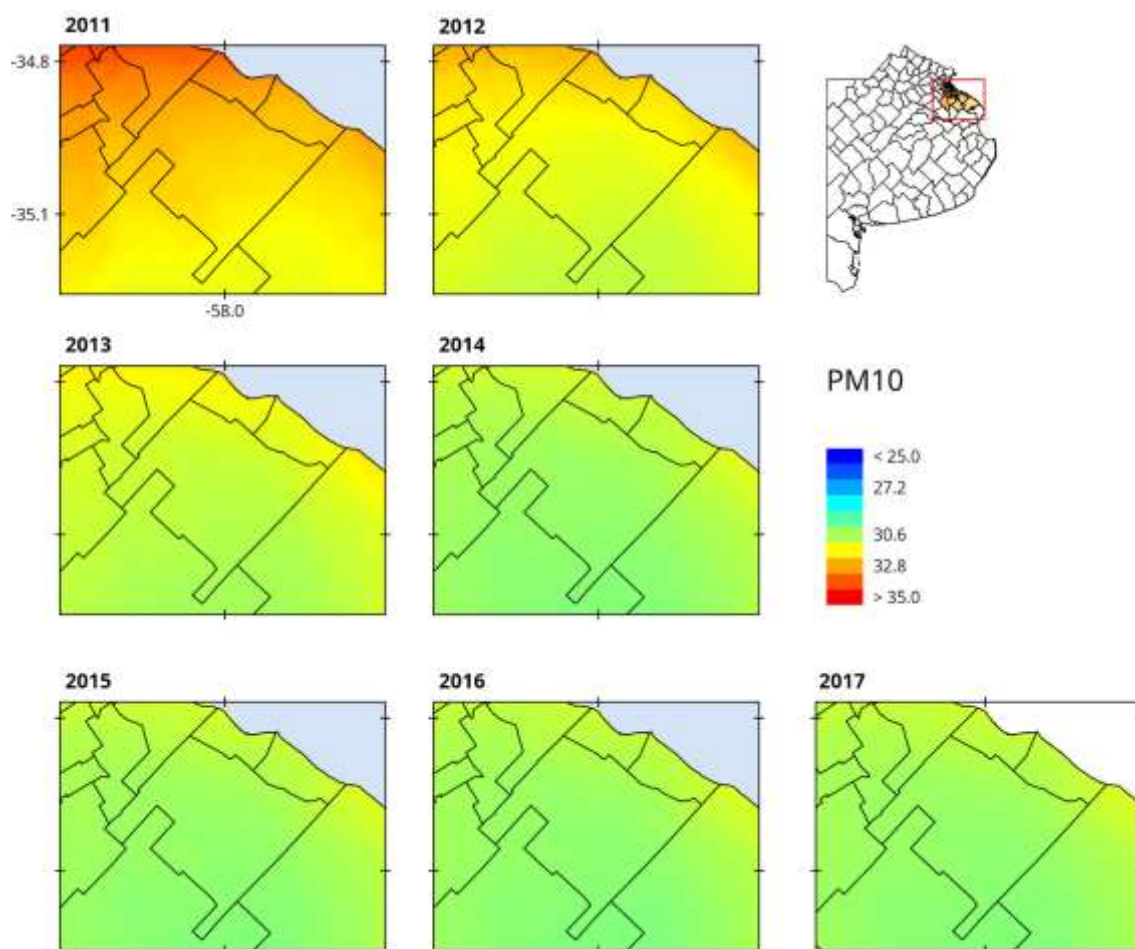


Figura 133. Mapas de concentraciones medias anuales de PM<sub>10</sub>

En la **Figura 134** se observan las concentraciones medias modeladas para el área del Gran La Plata en el año 2017. El MAPE fue menor este año para el área industrial (29.6%) por sobre el área urbana (47.2%), disminuyendo sustancialmente el error del modelo respecto a las mediciones realizadas. En esta figura se puede ver claramente la mayor afectación del área costera, en particular las localidades de Berisso y Ensenada, sobre la calidad del aire en el sector N de La Plata.

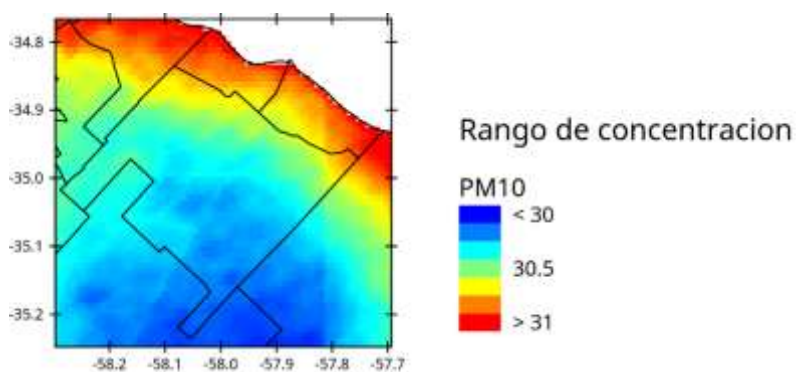
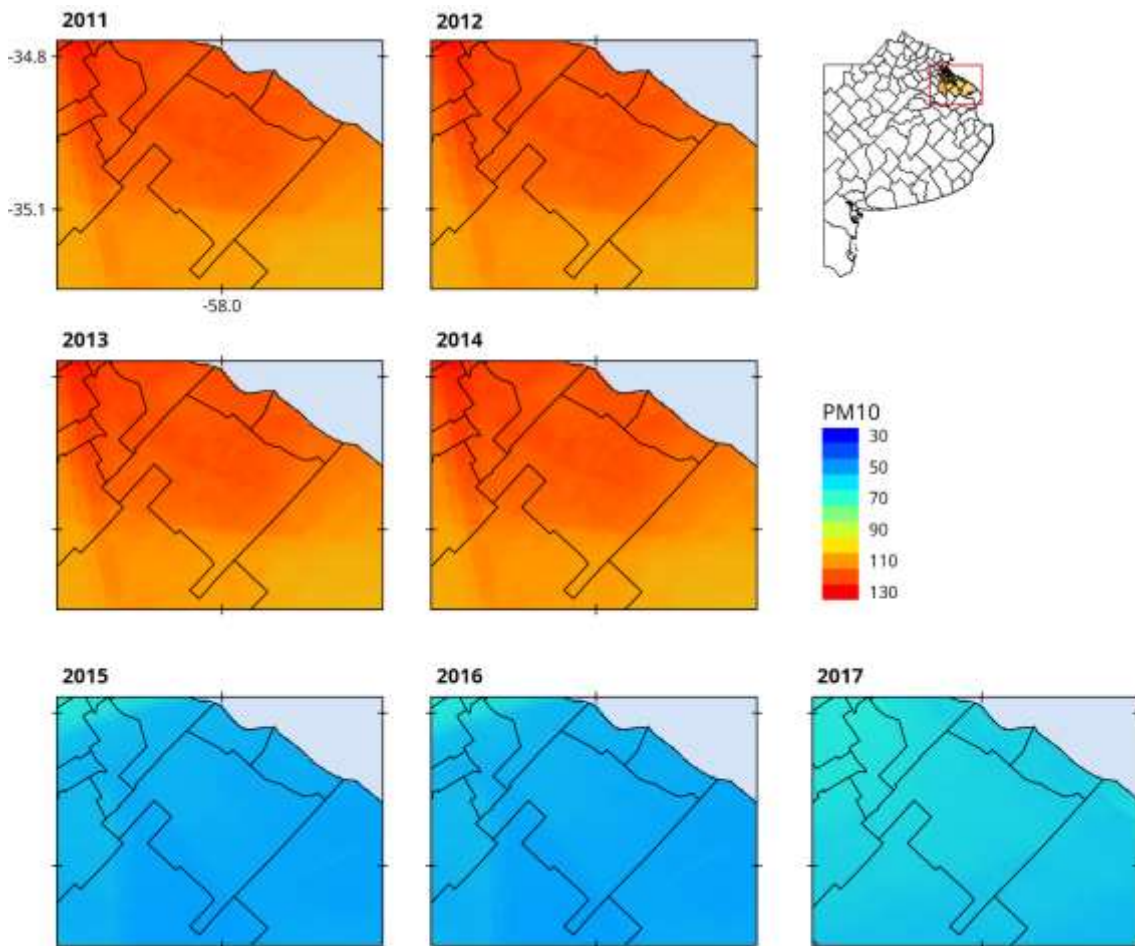


Figura 134. Mapa de la concentración media anual de PM<sub>10</sub> para el año 2017

Al analizar las concentraciones máximas modeladas para cada año se observaron áreas que alcanzan los  $120 \mu\text{g}\cdot\text{m}^{-3}$  entre 2011 y 2014 (**Figura 135**). Resulta interesante analizar el origen de estos altos niveles, para identificar las fuentes de emisión y prever su impacto.



**Figura 135.** Mapas de concentraciones máximas anuales de PM<sub>10</sub>

### Mapas de calidad del aire, de riesgo y de FAM por exposición a PM<sub>10</sub>

Para la elaboración de mapas de riesgo se calculó la concentración media de PM<sub>10</sub> en cada partido. En la **Figura 136** se observa que los rangos de concentraciones medias anuales modeladas se encuentran por encima de los  $10 \mu\text{g}\cdot\text{m}^{-3}$  recomendados. Las ciudades más afectadas son las ubicadas al N y NW de la La Plata.

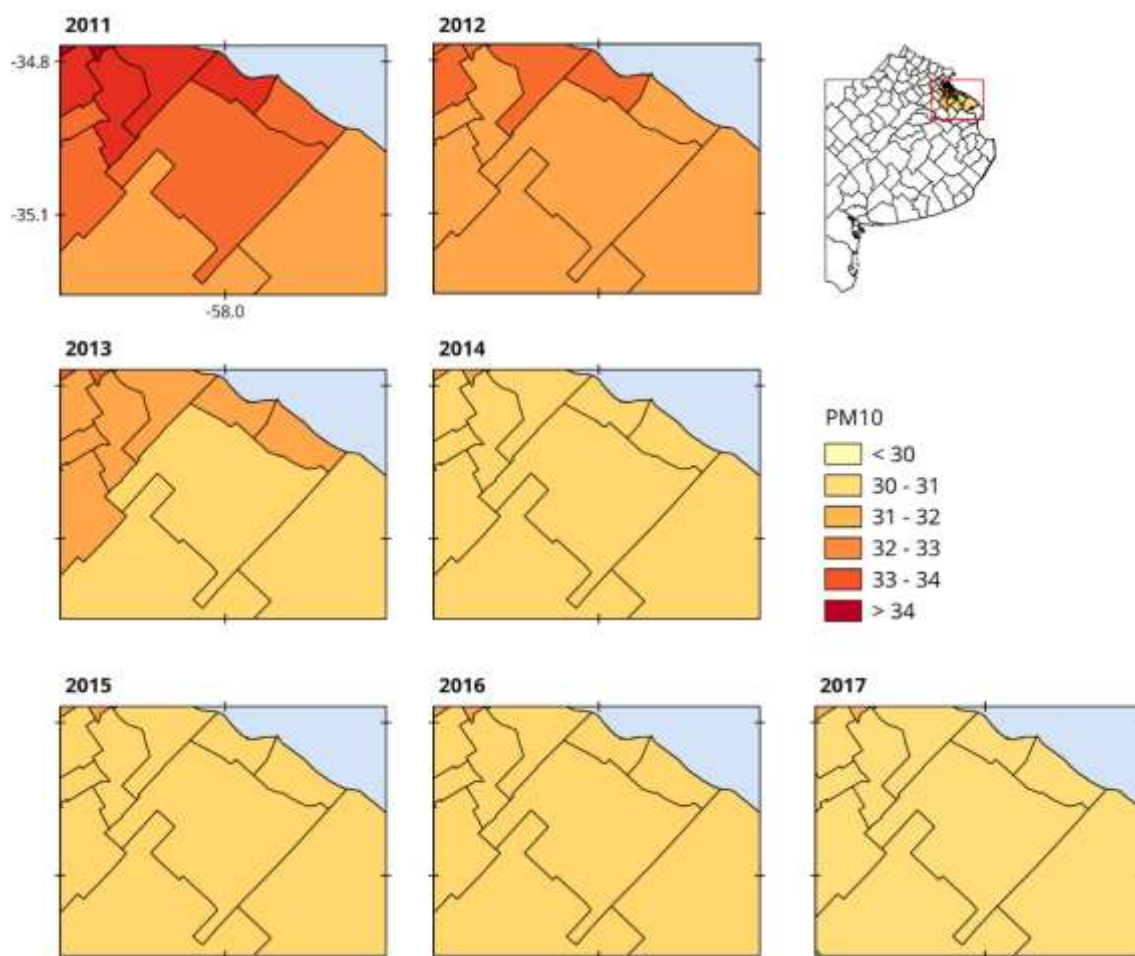


Figura 136. Mapas de concentraciones media anual de  $PM_{10}$  promediada por ciudad

Como siguiente paso se confeccionaron los mapas de calidad del aire. Tal como puede ver en la **Figura 137**, la calidad del aire para el Gran La Plata durante el 2017 representa un potencial daño para las personas sensibles.



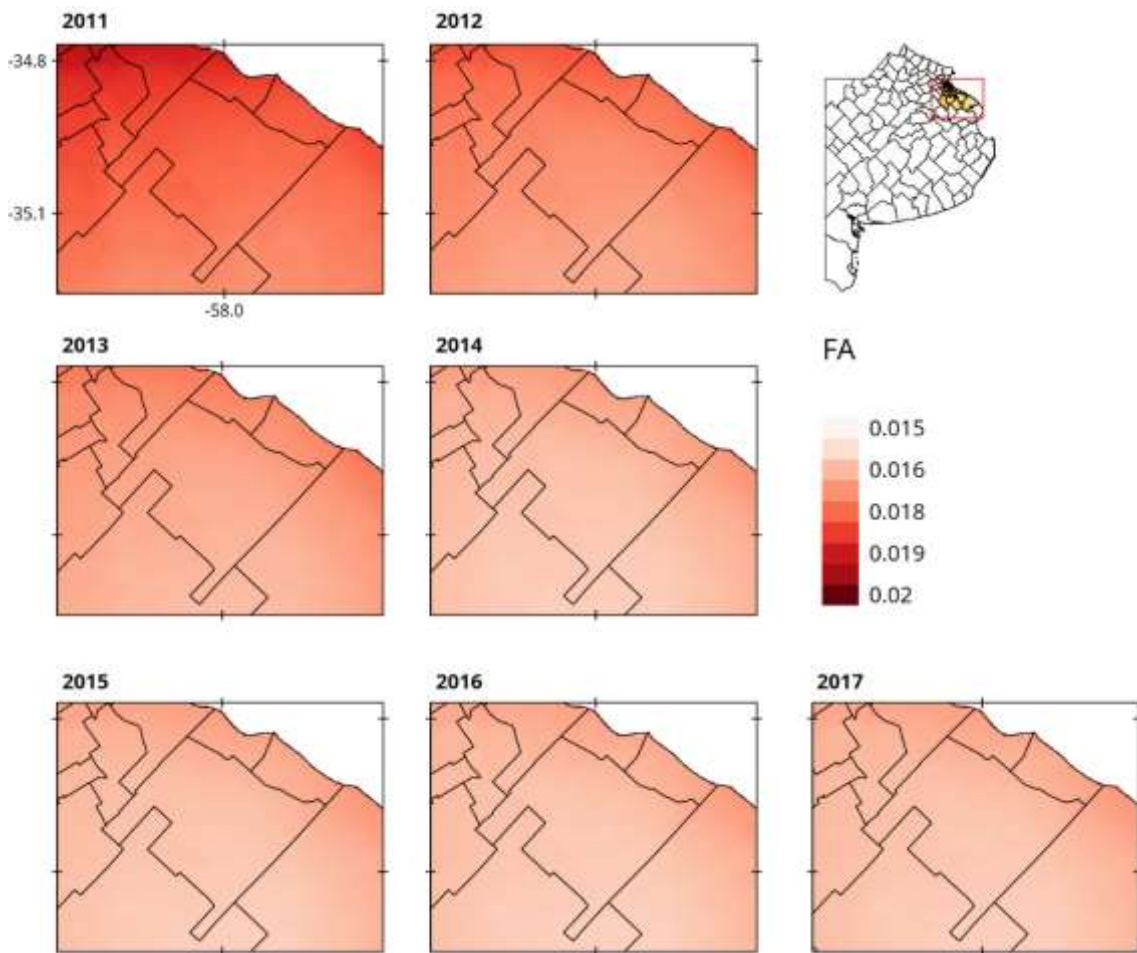
Figura 137. Mapas de calidad del aire en el Gran La Plata durante el 2017

En base a estos resultados se estimó el riesgo relativo de muerte por exposición a  $PM_{10}$ . El riesgo relativo se calculó como:

$$RR = e^{\beta (X - X_0)}$$

donde  $\beta = 0.0008$  para todas las causas de muerte para todas las edades y  $X_0 = 10 \mu g.m^{-3}$  de  $PM_{10}$ . Para todos los años se observó un riesgo relativo constante de 1.02 para el Gran La Plata.

Finalmente, la **Figura 138** muestra un mapa de la fracción atribuible de muertes prematuras por exposición a largo plazo a  $PM_{10}$ . Se observa que la FAM varía entre 0.015 y 0.02 para toda la región, esto es entre un 1.5% y un 2% de las muertes son atribuibles a la contaminación del aire, viéndose más afectada la localidad de Ensenada y el área costera sobre el resto.



**Figura 138.** Mapa de la fracción atribuible de muertes prematuras por exposición a largo plazo a  $PM_{10}$

### III.II. Informe para La Plata

#### *Caracterizar aspectos del monitoreo terrestre*

- **¿Cuáles son las técnicas utilizadas para el monitoreo?**

Se realiza un monitoreo discreto el cual cuenta con un equipo muestreador de PM de bajo volumen MiniVol TAS (Airmetrics), que mide a un flujo constante de  $5 \text{ l}\cdot\text{min}^{-1}$ , a lo largo de 3 días. La técnica es gravimétrica, por diferencia de pesadas. El PM se recolecta en filtros de politetrafluoroetileno (PTFE) que actúa como medio colector.

- **¿Cuáles son las características del área de emplazamiento de las estaciones de monitoreo?**

El monitoreo de la calidad del aire se realizó en sitios diferenciados como urbanos industriales. Sin embargo, no se observaron diferencias significativas entre los sitios de monitoreo.

#### *Conocer la evolución temporal de los contaminantes*

- **¿Durante cuál época del año ocurren los episodios de máximas concentraciones?**

En base a las modelaciones de  $\text{PM}_{10}$  se pudo observar que las concentraciones mayores se registraron en los meses fríos.

- **¿Cómo han cambiado los niveles de los contaminantes con los años?**

No existen tendencias significativas en las concentraciones modeladas de  $\text{PM}_{10}$ . Sin embargo, se registraron evidencias del fenómeno de erupción del volcán ocurrido en 2011.

#### *Evaluar la distribución espacial de los contaminantes*

- **¿Dónde ocurren las máximas concentraciones?**

Las máximas concentraciones se registran en los partidos de Ensenada y Berisso.

- **¿Se observa un patrón espacial de las concentraciones?**

La concentración modelada de  $\text{PM}_{10}$  disminuye hacia el sur del Partido de La Plata.

- **¿Cómo es la calidad del aire donde no se mide?**

La calidad del aire modelada alcanza los niveles de dañina para personas sensibles.

- **¿La calidad del aire en la región representa un riesgo para la salud de la población?**

Sí, las concentraciones modeladas indicaron un nivel de riesgo relativo constante de 1.02 para el Gran La Plata, con un FAM entre 0.015 y 0.02 para toda la región.



*"Don't be afraid to start,  
and don't be afraid to finish"*

Schabenberger & Pierce, 2002.

## Conclusiones generales

---

En la gestión, la administración pública se enfrenta a diferentes retos a la hora de llevar a cabo estudios sobre la calidad del aire: un conocimiento limitado de la literatura científica actualizada, deficiencias en la integración de información de distintas fuentes, dificultad para abordar la gran cantidad de datos e informar de forma concluyente. La constitución de una metodología de análisis de los datos de calidad del aire aporta una nueva mirada a un problema complejo.

En esta tesis se propuso una metodología para el estudio y la evaluación de la calidad del aire. Tomando el enfoque de la ciencia de datos se presentó un proceso para la identificación de fuentes de datos, selección de objetivos y preguntas de estudios, una serie de algoritmos y métodos para su abordaje y análisis, junto a una forma sintética de comunicación de resultados.

La metodología desarrollada se encuentra dividida en 3 módulos: diseño del análisis, procesamiento y comunicación. Cada módulo se compone por una serie de tareas y pasos que completan una unidad independiente del proceso. Cada unidad puede ser pensada de forma independiente, ya sea en períodos de tiempo espaciados o realizada por diferentes usuarios.

Se trabajó con un enfoque lineal para el proceso completo, incluyendo un enfoque circular en el primer módulo (diseño del análisis). Como ventaja, este enfoque permitió estructurar y organizar el análisis, ya que los pasos del proceso se organizan en un orden fijo. Una posible desventaja es que la naturaleza paso a paso puede limitar la profundización de los análisis, centrándose demasiado en cumplir con la etapa y perdiendo el objetivo central del proceso. Para evitar esto, se reforzó la idea de trabajar con preguntas orientativas que se propongan en el primer módulo (módulo recursivo) y que conduzcan los siguientes.

El enfoque cíclico del primer módulo proporciona orden y flexibilidad a la naturaleza de la toma de decisiones y también incluye más y diferentes tipos de decisiones que pueden tomarse. Si bien, existe un enfoque secuencial, existe la posibilidad de "aprender haciendo" y mejorar el proceso antes de pasar al siguiente módulo [MSHS, 2016].

Con respecto a la etapa del diseño del análisis, se observó que la utilización de preguntas de investigación simplificó notablemente la selección de algoritmos para el análisis y el abordaje de la interpretación de los resultados. A su vez, permitió diferenciar las preguntas que fueron originarias del estudio de aquellas que surgieron tras el análisis de los datos. Esto concuerda con



[Wentz et al., 2014], quienes señalaron que el uso de bases de datos de gran tamaño plantea como principal reto la extracción de información útil, precisa y concisa del mar de datos del Big Data.

La gran limitación para el análisis de la calidad del aire resultó la falta de mediciones sistemáticas de los contaminantes atmosféricos. Los datos que se incluyeron en el análisis fueron seleccionados por su disponibilidad como por la calidad que presentaron. La mayor dificultad fue trabajar en ausencia de datos (La Plata) o con pocos datos disponibles de calidad baja (Buenos Aires). A su vez, la recolección de datos resultó un proceso laborioso, ya que los datos se encontraron de muchos tipos diferentes y de una amplia variedad de fuentes (ver Capítulo III, Sección I. IV).

Los datos masivos e interconectados fomentan la investigación interdisciplinaria, ya que están menos limitados por el marco teórico inicial. A lo largo de este trabajo, hemos visto que desde la ciencia de datos es posible explorar, extraer valor y darles sentido a los datos, permitiendo generar modelos y teorías más holísticas y extensas de sistemas complejos como la atmósfera y la calidad del aire. Sin embargo, este análisis requiere una gran cantidad de datos, y los costos asociados y la capacidad requerida para adquirirlos son altos.

La forma de acceso a las mediciones de calidad de aire ha sido distinta en cada caso de estudio. Para Buenos Aires, la serie completa de datos se encontraba para su libre descarga, mientras que para Valencia se precisó realizar una extracción (*web scraping*) de la web oficial para obtener la base de datos histórica. En el caso de La Plata, donde el municipio no asume la tarea de monitorear la calidad del aire, ha sido la universidad e instituciones de investigación quienes desarrollaron las mediciones que se utilizaron en este estudio.

Con esta metodología se pudo procesar la información recolectada por las 60 estaciones de la Comunidad Valenciana. Se analizaron distintos aspectos de la red de monitoreo y la calidad de los datos recolectados. A su vez, estas mediciones se integraron con los datos del re-análisis atmosférico MERRA-2 y datos de uso de suelo para generar un muy buen modelo de  $PM_{2.5}$  (MAPE 15%). En base a este, se obtuvieron mapas de calidad del aire, de riesgo y de la fracción atribuible de muerte por exposición a este contaminante.

Entre los resultados más relevantes cabe mencionar la redundancia de mediciones entre las estaciones, la falta de mediciones de contaminantes relevantes en sitios fuertemente afectados por la contaminación antropogénica (estación Moli) y una ausencia de control en localidades que se ven afectadas con una inferior calidad del aire, como es el caso de Elche o Benidorm.

En la Ciudad de Buenos Aires, nuestro segundo caso de estudio, se contó solo con datos de 3 estaciones de monitoreo continuas inmersas en el tejido urbano las cuales midieron  $PM_{10}$ ,  $NO_2$  y CO. Utilizando la metodología propuesta se observaron concentraciones elevadas de  $PM_{10}$  (con una media de  $30.0 \mu g.m^{-3}$ ,  $28.5 \mu g.m^{-3}$  y  $29.3 \mu g.m^{-3}$  para Centenario, Córdoba y La Boca, respectivamente) para todo el período y la ocurrencia de eventos extremos alarmantes de CO (concentraciones superiores a 40 ppm). Al analizar las tendencias históricas, el  $NO_2$  disminuyó en todos los casos, no así el  $PM_{10}$  y el CO que en algunas estaciones la tendencia no fue significativa o aumentó con los años.

En base a los datos satelitales se confeccionó un modelo de  $PM_{10}$  para el Área Metropolitana de Buenos Aires que presentó un  $R^2 = 0.96$ , un RMSE de  $10.16 \mu g.m^{-3}$  y un MAPE del 24%. Analizando los errores, se observó que el modelo tiene un sesgo negativo, subestimando la situación de riesgo a la cual se encuentra expuesta la población. En base a nuestras estimaciones, la calidad del aire fue dañina en Buenos Aires para las personas sensibles en el período de estudio,

donde el riesgo relativo fue del 1.02 y la fracción atribuible de muerte entre el 1.5-2%. En este sentido, resulta alarmante la falta de un mayor control de la calidad del aire, en particular en los municipios en la zona norte y noroeste de la Ciudad de Buenos Aires.

Por último, en La Plata se observó que la falta de datos locales imposibilita el desarrollo de modelos espaciales mediante técnicas estadísticas. En base a esto, la aplicación de la metodología se limitó al análisis de los datos meteorológicos y a la validación de la simulación realizada para la Ciudad de Buenos Aires a 50 km de nuestro sitio de estudio. A su vez, al no existir ningún tipo de monitoreo continuo, la validación del modelo debió realizarse con mediciones integradas en el tiempo (monitoreo discreto). Con estas consideraciones, la validación del modelo dio un RMSE = 18.7  $\mu\text{g}\cdot\text{m}^{-3}$ , un MAPE = 41.5% y presentó una marcada subestimación de las concentraciones de  $\text{PM}_{10}$  modeladas.

Los mapas obtenidos para La Plata mostraron que la calidad del aire es dañina para las personas sensibles, generando una gran inquietud sobre las condiciones a las cuales está expuesta la población. En tal sentido, la falta de información es preocupante. El Partido de La Plata está situado en una zona donde la atmósfera tiene una baja capacidad de autolimpieza, principalmente durante el otoño y el invierno [Gassmann & Mazzeo, 2000; Ratto et al. 2018], y donde distintos estudios han mostrado efectos sobre la salud de la población [Wichmann et al., 2009].

Como una última cuestión, resulta importante remarcar que la escasez de información se produce en un momento en que las instituciones estatales requieren más que nunca información para comprender los efectos ambientales locales de la urbanización, el cambio ambiental global y los efectos de la vida urbana en el bienestar humano. Los recientes llamamientos a favor de una "ciencia de la urbanización" hacen hincapié en la necesidad de enfocar la escala de trabajo en lo local para atender los cambios que se producen por el creciente fenómeno de urbanización [Solecki et al., 2013]. En esta escala de trabajo, hemos visto que la teledetección puede proporcionar versátiles herramientas para su análisis, proveyendo una visión amplia de los niveles de los contaminantes en el aire urbano y su transporte en la atmósfera.

Sin embargo, si bien las concentraciones de algunos contaminantes pueden modelarse mediante el uso de variables medidas por sensores remotos u otras técnicas innovadoras, aún se precisan de las mediciones de campo para validar las simulaciones [Grey et al., 2013]. Mientras los parámetros de calidad de aire se confeccionen en función de las concentraciones máxicas de inmisión, no será factible el reemplazo de las mediciones locales por instrumental remoto, aunque puede complementar como se ha visto en este trabajo.

Esto pone en foco las implicancias que tienen las decisiones de gestión y planificación urbana en la salud de la población. Es imperativo que rechacemos la creencia de que los desafíos ambientales y de salud que enfrentamos son el resultado de elecciones personales llevadas de forma individual. Las opciones de estilo de vida no surgen en el vacío y son temas legítimos para el debate democrático y la acción gubernamental [World Health Organization, 2015]. Frente al cambio climático que nos acecha, se requiere un involucramiento de los gobiernos en conversaciones trans sectoriales, compartiendo conocimientos y perspectivas, y estableciendo agendas conjuntas que integren de forma vertical y horizontal a la ciudadanía [World Health Organization, 2016]. La gestión de la calidad del aire es un asunto de primordial e imperiosa importancia a nivel local y global y lograr que el diseño institucional sea el correcto es clave para el éxito.

En este marco, el diálogo efectivo entre la ciencia y la gestión debe ser un asunto prioritario. La ciencia deberá proporcionar herramientas innovadoras para avanzar en el nivel de conocimiento

de la atmósfera e identificar los problemas y soluciones que puedan surgir [Grey et al., 2013], pero es necesario dar un marco político que garantice los fondos económicos, de modo que la ciencia pueda reunir los datos y las pruebas necesarias. Tal como se mencionó, en ausencia de un conocimiento avanzado y una comprensión compartida, la complejidad alimenta las decisiones deficientes y la tensión asociada. La falta de información contribuye a que se generen percepciones erróneas con serias consecuencias para la población [Sadoff & Grey, 2005; Grey et al., 2013], que como se señaló anteriormente en esta tesis pueden estar directamente ligadas con padecimientos graves como el cáncer o la muerte.

Para contener y gestionar los riesgos relacionados con la mala calidad del aire, desde la escala local hasta la escala mundial, hace falta información. Este es un requisito previo para la gestión de los riesgos sociales, políticos y económicos en nuestro mundo cambiante. Y debe siempre contemplarse que los riesgos se acentúan cuando reina el desconocimiento.

Un último aspecto que quisiera rescatar es una reflexión sobre una deformación profesional que nos pone al filo de nuestra propia ética. Creo firmemente que no debe perderse de vista que, para la resolución de los problemas modernos como los que aquí se señalan, producto ineludible del sistema económico imperante, se requiere no sólo de la investigación y del desarrollo tecnológico, sino también de una revisión de los valores fundamentales de nuestras sociedades. Debe ser siempre nuestro mayor desafío diferenciar el consumo material y el crecimiento económico de la felicidad y el bienestar humano.

*“Tenemos que cambiarlo todo”*

Greta Thunberg

“La única forma de saber con certeza si existen, si se están generando, o si se están empeorando los problemas de la contaminación del aire es mediante la medición de los contaminantes.”

[Korc & Sáenz, 1999]

## Bibliografía

---

### A

- Abril, G. A., Diez, S. C., Pignata, M. L., & Britch, J. (2016). Particulate matter concentrations originating from industrial and urban sources: Validation of atmospheric dispersion modeling results. *Atmospheric Pollution Research*, 7(1), 180-189.
- Abrutzky, R., Dawidowski, L., Matus, P., & Lankao, P. R. (2012). Health effects of climate and air pollution in Buenos Aires: a first time series analysis. *Journal of Environmental Protection*, 3(3), 262.
- Achad, M., López, M. L., Ceppi, S., Palancar, G. G., Tirao, G., & Toselli, B. M. (2014). Assessment of fine and sub-micrometer aerosols at an urban environment of Argentina. *Atmospheric environment*, 92, 522-532.
- Akimoto, H. (2003). Global air quality and pollution. *Science*, 302(5651), 1716-1719.
- Aldabe, J., Elustondo, D., Santamaría, C., Lasheras, E., Pandolfi, M., Alastuey, A., Querol, X. and Santamaría, J.M., 2011. Chemical characterisation and source apportionment of PM2.5 and PM10 at rural, urban and traffic sites in Navarra (North of Spain). *Atmospheric Research*, 102(1-2), pp.191-205.
- Allende, D. G., Puliafito, S. E., Fernández, R. P., Castro, F. H. y Cremades, P. G. 2012. Herramientas para la evaluación integral de la calidad de aire en Argentina: modelos de escala urbana y regional. *Revista Proyecciones- UTN-FRBA* 10: 31 - 44.
- An, X.Q., Zuo, H.C., Chen, L.J., 2007. Atmospheric environmental capacity of SO<sub>2</sub> in winter over Lanzhou in China: a case study. *Adv. Atmos. Sci.* 24, 688e699
- Anderson, T. W. 1984. *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, New York.
- Ångström, A., 1929. On the atmospheric transmission of sun radiation and on dust in the air. *Geografiska Annaler* 11, 156-166.
- Anselin, L., 1995, Local indicators of spatial association—LISA. *Geographical Analysts*, 27, pp. 93–115.
- Arkouli, M., Ulke, A. G., Endlicher, W., Baumbach, G., Schultz, E., Vogt, U., M. Müller, L. Dawidowski, A. Faggi, U. Wolf-Benning & Scheffknecht, G. (2010). Distribution and temporal behavior of particulate matter over the urban area of Buenos Aires. *Atmospheric Pollution Research*, 1(1), 1-8.
- Arasa, R., Soler, M. R., Ortega, S., Olid, M., & Merino, M. (2010). A performance evaluation of MM5/MNEQA/CMAQ air quality modelling system to forecast ozone concentrations in Catalonia. *Tethys*, 7, 11-23.
- Arora, P., & Varshney, S. (2016). Analysis of k-means and k-medoids algorithm for big data. *Procedia Computer Science*, 78, 507-512.
- ARS (2015) Astillero Río Santiago, Información obtenida de: <http://www.astillero.gba.gov.ar>

Asociación de Fábricas Argentinas de Componentes, AFAC. 2018. Flota circulante en Argentina.

Austin, E., Coull, B. A., Zanobetti, A., & Koutrakis, P. (2013). A framework to spatially cluster air pollution monitoring sites in US based on the PM<sub>2.5</sub> composition. *Environment international*, 59, 244-254.

## B

Bakhtiarifar, M. H., Bashiri, M., & Amiri, A. (2017). Optimization of problems with multivariate multiple functional responses: A case study in air quality. *Communications in Statistics-Simulation and Computation*, 46(10), 8049-8063.

Baklanov, A., Molina, L. T., & Gauss, M. (2016). Megacities, air quality and climate. *Atmospheric Environment*, 126, 235-249.

Baldasano, J. M., Valera, E., & Jimenez, P. (2003). Air quality data from large cities. *Science of the Total Environment*, 307(1-3), 141-165.

Ballester, F., Corella, D., Perez-Hoyos, S., & Hervas, A. (1996). Air pollution and mortality in Valencia, Spain: a study using the APHEA methodology. *Journal of Epidemiology & Community Health*, 50(5), 527-533.

Ballester, F., Estarlich, M., Iñiguez, C., Llop, S., Ramón, R., Esplugues, A., Lacasaña, M. and Rebagliato, M., 2010. Air pollution exposure during pregnancy and reduced birth size: a prospective birth cohort study in Valencia, Spain. *Environmental Health*, 9(1), p.6.

Barros V, Menéndez A, Nagy G (2005) El Cambio Climático en el Río de La Plata, CIMA Textos del reporte técnico de los proyectos: Impactos del Cambio Global en las áreas costeras del Río de la Plata y Variabilidad hidrolimática del estuario del Río de la Plata: Influencia humana, ENSO y estado trófico. Proyecto "Assessments of Impacts and Adaptations to Climate Change (AIACC)", START-TWAS-UNEP, 1-204.

Baró Porras, F., Chaparro, L., Gómez Baggethun, E., Langemeyer, J., Nowak, D. J., & Terradas, J. (2014). Contribution of ecosystem services to air quality and climate change mitigation policies. *Ambio*, 43(4), 466-479.

Barros, V., Clarke, R., & Dias, P. S. (2006). Climate change in the La Plata basin. *Publication of the Inter-American Institute for Global Change Research (IAI), São José dos Campos, Brazil*.

Barros, V. R., Boninsegna, J. A., Camilloni, I. A., Chidiak, M., Magrín, G. O., & Rusticucci, M. (2015). Climate change in Argentina: trends, projections, impacts and adaptation. *Wiley Interdisciplinary Reviews: Climate Change*, 6(2), 151-169.

Baulant, A. L., Perron, G., Kleinpeter, J., Weber, C., Ranchin, T., & Wald, L. (2008). Adding virtual measuring stations to a network for urban air pollution mapping. *Environment international*, 34(5), 599-605.

Beck, U., and E. Beck-Gernsheim. 2002. Individualization: Institutionalized Individualism and its Social and Political Consequences. London: Sage

Beelen, R., Hoek, G., Vienneau, D., Eeftens, M., Dimakopoulou, K., Pedeli, X., Tsai, M.Y., Künzli, N., Schikowski, T., Marcon, A. and Eriksen, K.T., 2013. Development of NO<sub>2</sub> and NO<sub>x</sub> land use regression models for estimating air pollution exposure in 36 study areas in Europe–The ESCAPE project. *Atmospheric Environment*, 72, pp.10-23.

Belda, F., Meliá, J., & Segarra, D. (2001). Estudio climatológico en el sur de la comunidad valenciana utilizando un modelo digital del terreno (MDT).

Bellinger, C., Jabbar, M. S. M., Zaïane, O., & Osornio-Vargas, A. (2017). A systematic review of data mining and machine learning for air pollution epidemiology. *BMC public health*, 17(1), 907.

Bernard, P., Charafeddine, R., Frohlich, K. L., Daniel, M., Kestens, Y., & Potvin, L. (2007). Health inequalities and place: a theoretical conception of neighbourhood. *Social science & medicine*, 65(9), 1839-1852.

- Biancofiore, F., Busilacchio, M., Verdecchia, M., Tomassetti, B., Aruffo, E., Bianco, S., Di Tommaso, S., Colangeli, C., Rosatelli, G. and Di Carlo, P., 2017. Recursive neural network model for analysis and forecast of PM<sub>10</sub> and PM<sub>2.5</sub>. *Atmospheric Pollution Research*, 8(4), pp.652-659.
- Bianchi, B. V. M. (2009). Un caso de estudio en calidad de datos para ingeniería de software empírica. *Universidad de la República, Uruguay*.
- Binkowski, F. S., & Roselle, S. J. (2003). Models-3 Community Multiscale Air Quality (CMAQ) model aerosol component 1. Model description. *Journal of geophysical research: Atmospheres*, 108(D6).
- Bird, R. B., Stewart, W. E., & Lightfoot, E. N. (1992). *Fenómenos de transporte: un estudio sistemático de los fundamentos del transporte de materia, energía y cantidad de movimiento* (No. TP156 B5321).
- Bivand R. (2019). classInt: Choose Univariate Class Intervals. R package version 0.3-3. <https://CRAN.R-project.org/package=classInt>
- Bivand R, Tim Keitt and Barry Rowlingson (2019). rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.4-3. <https://CRAN.R-project.org/package=rgdal>
- Boersma, K. F., Eskes, H. J., Dirksen, R. J., van der A, R. J., Veefkind, J. P., Stammes, P., Huijnen, V., Kleipool, Q. L., Sneep, M., Claas, J., Leitão, J., Richter, A., Zhou, Y., and Brunner, D.: An improved tropospheric NO<sub>2</sub> column retrieval algorithm for the Ozone Monitoring Instrument, *Atmos. Meas. Tech.*, 4, 1905– 1928, doi:10.5194/amt-4-1905-2011, 2011
- Bogo, H., Otero, M., Castro, P., Ozafrán, M. J., Kreiner, A., Calvo, E. J., & Negri, R. M. (2003). Study of atmospheric particulate matter in Buenos Aires city. *Atmospheric Environment*, 37(8), 1135-1147.
- Boldo, E., & Querol, X. (2014). Nuevas políticas europeas de control de la calidad del aire: ¿ un paso adelante para la mejora de la salud pública?.
- Bösenberg, J., & Matthias, V. (2003). EARLINET: A European Aerosol Research Lidar Network to establish an aerosol climatology.
- Bostrom, A., Morgan, M. G., Fischhoff, B., & Read, D. (1994). What do people know about global climate change? 1. Mental models. *Risk Analysis*, 14(6), 959-970.
- Brasseur, G. P., & Prinn, R. G. (2003). *Atmospheric Chemistry in a Changing World: An Integration and Synthesis of a Decade of Tropospheric Chemistry Research; the International Global Atmospheric Chemistry Project of the International Geosphere-Biosphere Programme; with 22 Tables*. Springer Science & Business Media.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Brook, R.D., Rajagopalan, S., Pope III, C.A., Brook, J.R., Bhatnagar, A., Diez-Roux, A.V., Holguin, F., Hong, Y., Luepker, R.V., Mittleman, M.A. and Peters, A., 2010. Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association. *Circulation*, 121(21), pp.2331-2378.
- Brunekreef B, Janssen NA, de Hartog J, Harssema H, Knape M, van Vliet P. 1997. Air pollution from truck traffic and lung function in children living near motorways. *Epidemiology* 8:298–303
- Brunelli, U., Piazza, V., Pignato, L., Sorbello, F., & Vitabile, S. (2007). Two-days ahead prediction of daily maximum concentrations of SO<sub>2</sub>, O<sub>3</sub>, PM<sub>10</sub>, NO<sub>2</sub>, CO in the urban area of Palermo, Italy. *Atmospheric Environment*, 41(14), 2967-2995.
- Brulle, R. J. (2010). From environmental campaigns to advancing the public dialog: Environmental communication for civic engagement. *Environmental Communication*, 4(1), 82-98.
- Bucsela, E. J., Krotkov, N. A., Celarier, E. A., Lamsal, L. N., Swartz, W. H., Bhartia, P. K., Boersma, K. F., Veefkind, J. P., Gleason, J. F., and Pickering, K. E.: A new stratospheric and tropospheric NO<sub>2</sub> retrieval algorithm for nadir-viewing satellite instruments: applications to OMI, *Atmos. Meas. Tech.*, 6, 2607– 2626, doi:10.5194/amt-6-2607-2013, 2013.

Büttner, G., Feranec, J., Jaffrain, G., Mari, L., Maucha, G., & Soukup, T. (2004). The CORINE land cover 2000 project. *EARSeL eProceedings*, 3(3), 331-346.

Bytnerowicz, A., Arbaugh, M., Schilling, S., Fraczek, W., Alexander, D., & Dawson, P. (2007). Air pollution distribution patterns in the San Bernardino Mountains of southern California: a 40-year perspective. *The Scientific World Journal*, 7, 98-109.

## C

Cámara Argentina de Comercio (CAC), 2011. RELEVANCIA DEL SECTOR DE COMERCIO Y SERVICIOS EN LA ECONOMÍA ARGENTINA. Visto en [https://www.cac.com.ar/data/documentos/2\\_informe%20relevancia%20consolidado.pdf](https://www.cac.com.ar/data/documentos/2_informe%20relevancia%20consolidado.pdf) (Visitado 23-09-2019)

Carreras, H. A., & Pignata, M. L. (2002). Biomonitoring of heavy metals and air quality in Cordoba City, Argentina, using transplanted lichens. *Environmental Pollution*, 117(1), 77-87.

Carlsaw DC & Ropkins K (2012). "Openair — An R package for air quality data analysis." *Environmental Modelling & Software*, 27–28(0), pp. 52–61. ISSN 1364-8152, doi: [10.1016/j.envsoft.2011.09.008](https://doi.org/10.1016/j.envsoft.2011.09.008).

Cavallotto, J.L. 1995. Evolución Geomorfológica de la Llanura Costera del Margen Sur del Río de La Plata. Tesis Doctoral, Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata (inédita), 237 p., La Plata.

Chalvatzaki, E., Chatoutsidou, S. E., Lehtomäki, H., Almeida, S. M., Eleftheriadis, K., Hänninen, O., & Lazaridis, M. (2019). Characterization of Human Health Risks from Particulate Air Pollution in Selected European Cities. *Atmosphere*, 10(2), 96.

Chameides, W., & Walker, J. C. (1973). A photochemical theory of tropospheric ozone. *Journal of Geophysical Research*, 78(36), 8751-8760.

Chang, J.C., Hanna, S.R., 2004. Air quality model performance evaluation. *Meteorology and Atmospheric Physics* 87.

Channan, S., K. Collins, and W. R. Emanuel. 2014. Global mosaics of the standard MODIS land cover type data. University of Maryland and the Pacific Northwest National Laboratory, College Park, Maryland, USA.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. & Wirth, R. 2000. CRISP-DM 1.0 Step-by-Step Data Mining Guide. Technical report, CRISP-DM.

Chen, M., Wang, P., Chen, Q., Wu, J., & Chen, X. (2015). A clustering algorithm for sample data based on environmental pollution characteristics. *Atmospheric Environment*, 107, 194-203.

Chen, Y., Wang, L., Li, F., Du, B., Choo, K. K. R., Hassan, H., & Qin, W. (2017). Air quality data clustering using EPLS method. *Information Fusion*, 36, 225-232.

Chen, G., Wang, Y., Li, S., Cao, W., Ren, H., Knibbs, L.D., Abramson, M.J. and Guo, Y., (2018). Spatiotemporal patterns of PM10 concentrations over China during 2005–2016: a satellite-based estimation using the random forests approach. *Environmental pollution*, 242, pp.605-613.

Chin, M. (2005). The Goddard Chemistry Aerosol Radiation and Transport Model.

Cimorelli, A.J., Perry, S.G., Venkatram, A., Weil, J.C., Paine, R.J., Wilson, R.B., Lee, R.F., Peters, W.D. and Brode, R.W., 2005. AERMOD: A dispersion model for industrial source applications. Part I: General model formulation and boundary layer characterization. *Journal of applied meteorology*, 44(5), pp.682-693.

Ciudad de Buenos Aires, 2014. La Ciudad de Buenos Aires y la industria manufacturera. Informe de resultados. [https://www.estadisticaciudad.gob.ar/eyc/wp-content/uploads/2015/04/ir\\_2014\\_707.pdf](https://www.estadisticaciudad.gob.ar/eyc/wp-content/uploads/2015/04/ir_2014_707.pdf)

- Ciudad de Buenos Aires, 2019. Boletín Oficial de la Ciudad de Buenos Aires N° 5547 - 29/1/2019 . ANEXO Resolución 109/MEIGC/19 <https://documentosboletinoficial.buenosaires.gob.ar/publico/PE-RES-MEIGC-MEIGC-609-19-ANX.pdf>
- Climate Science Investigations. (2019). Energy: The Driver of Climate. Recuperado de <http://www.ces.fau.edu/nasa/module-2/how-greenhouse-effect-works.php>
- Cohen, M. A. (2007). Una nueva gestión ambiental: El riesgo y el principio precautorio. *espacio abierto*, 16(2), 209-222.
- Collett, R. S., & Oduyemi, K. (1997). Air quality modelling: a technical review of mathematical approaches. *Meteorological Applications*, 4(3), 235-246.
- Collini, E., Osoreo, M. S., Folch, A., Viramonte, J. G., Villarosa, G., & Salmuni, G. (2013). Volcanic ash forecast during the June 2011 Cordón Caulle eruption. *Natural hazards*, 66(2), 389-412.
- Colman-Lerner, A. Morales, M. Aguilar, P. Barrionuevo, J. Ditondo, A. Porta, L. Massolo y V. Dodero, «Calidad Del Aire En Áreas Urbanas E Industriales En La Provincia De Buenos Aires. Estudio Comparativo Entre La Plata Y Bahía Blanca,» de Contribuciones de la III Reunión Anual PROIMCA, Buenos Aires, 2011.
- Colman-Lerner, JE. Análisis y mitigación/remoción de Material Particulado (PM) y Compuestos Orgánicos Volátiles (COVs) y semivolátiles (COSVs). Tesis Doctoral, Facultad de Ciencias Exactas, UNLP. 2013
- Colman Lerner, J. E. C., Kohajda, T., Aguilar, M. E., Massolo, L. A., Sánchez, E. Y., Porta, A. A. & Mueller, A. (2014). Improvement of health risk factors after reduction of VOC concentrations in industrial and urban areas. *Environmental Science and Pollution Research*, 21(16), 9676-9688.
- Colman Lerner JE., A. Morales, M. Aguilar, D. Giuliani, J. Ditondo, V.I. Dodero, L. Massolo, EY. Sánchez, N. Matamoros, A. Porta (2014a). "The effect of Air Pollution on Children's Health: a Comparative Study between La Plata and Bahía Blanca, Buenos Aires Province, Argentina". In: WIT Transactions on Ecology and the Environment, Vol. 181, Environmental Impact II, C.A. Brebbia, G. Passerini, Eds. 659-670. Print ISBN: 978-1- 84564-762- 9. 768 pp.
- Colman Lerner JE., T. Kohajda, ME. Aguilar, LA. Massolo, EY. Sánchez, AA. Porta, P. Opitz, G. Wichmann, O. Herbarth, A. Mueller (2014b). "Improvement of health risk factors after reduction of VOC concentrations in industrial and urban areas". *Environmental Science and Pollution Research International*, 21: 9676-9688.
- Conselleria de Agricultura, Desarrollo Rural, Emergencia Climática y Transición Ecológica (2017a). Evaluación de la Calidad del Aire en la Comunidad Valenciana: U. Móvil La Pobla de Vallbona-La Vereda Zona ES1007: Turia (A. Costera) 28/03/17-28/05/17. Recuperado de <http://www.agroambient.gva.es/documents/20549779/164404226/LA+POBLA+DE+VALLBONA-LA+VEREDA/943a4ab9-0377-48eb-8a33-3dbd8c45c3d9;jsessionid=F8ECA8890131F12B24CDB5AEAE7A3DC>
- Conselleria de Agricultura, Desarrollo Rural, Emergencia Climática y Transición Ecológica. (2017b). Evaluación de la calidad del aire en la Comunidad Valenciana Zona ES1007: Turia (A. Costera) año 2017. Recuperado de <http://www.agroambient.gva.es/documents/20549779/164404040/ZONA+ES1007.+TURIA.+%C3%80REA+COSTERA/od1528fc-0aa4-421b-a520-5b849f96b2ce;jsessionid=D42A843528C4D38E32C6A41BA71B6875>
- Cowie, C.T., Garden, F., Jegasothy, E., Knibbs, L.D., Hanigan, I., Morley, D., Hansell, A., Hoek, G. and Marks, G.B., 2019. Comparison of model estimates from an intra-city land use regression model with a national satellite-LUR and a regional Bayesian Maximum Entropy model, in estimating NO<sub>2</sub> for a birth cohort in Sydney, Australia. *Environmental research*, 174, pp.24-34. Cressie, N. 1991. *Statistics for Spatial Data*, 1 edn, John Wiley & Sons.
- Cressie, N. 1989. Geostatistics. *The American Statistician*. 43(4): 611(23).
- Crutzen, P. J. (1974). Photochemical reactions initiated by and influencing ozone in unpolluted tropospheric air. *Tellus*, 26(1-2), 47-57.
- Cullen, A. and Frey, H. (1999). *Probabilistic Techniques in Exposure Assessment*. Plenum Publishing Co., 1st edition.



Csépe, Z., Makra, L., Voukantsis, D., Matyasovszky, I., Tusnányi, G., Karatzas, K., & Thibaudon, M. (2014). Predicting daily ragweed pollen concentrations using computational intelligence techniques over two heavily polluted areas in Europe. *Science of the Total Environment*, 476, 542-552.

## D

D'Angiola, A., Dawidowski, L. E., Gómez, D. R., & Osses, M. (2010). On-road traffic emissions in a megacity. *Atmospheric Environment*, 44(4), 483-493.

Dadson, S. J. (2017). *Statistical Analysis of Geographical Data: An Introduction*. John Wiley & Sons.

Davis, D. L. (1997). Short-term improvements in public health from global-climate policies on fossil-fuel combustion: an interim report. *The Lancet*, 350(9088), 1341-1349.

de Hoogh, K., Korek, M., Vienneau, D., Keuken, M., Kukkonen, J., Nieuwenhuijsen, M.J., Badaloni, C., Beelen, R., Bolignano, A., Cesaroni, G. and Pradas, M.C., 2014. Comparing land use regression and dispersion modelling to assess residential exposure to ambient air pollution for epidemiological studies. *Environment international*, 73, pp.382-392.

De Leeuw, G., Andreas, E. L., Anguelova, M. D., Fairall, C. W., Lewis, E. R., O'Dowd, C., ... & Schwartz, S. E. (2011). Production flux of sea spray aerosol. *Reviews of Geophysics*, 49(2).

Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae et al. (2011). *The ERA-Interim reanalysis: Configuration and performance of the data assimilation system*. Quarterly Journal of the Royal Meteorological Society, 137(656), 553--597, doi: 10.1002/qj.828

Deguen, S., & Zmirou-Navier, D. (2010). Social inequalities resulting from health risks related to ambient air quality—a European review. *European Journal of Public Health*, 20(1), 27-35.

Derwent, D., Fraser, A., Abbott, J., Willis, P., Murrells, T., 2010. Evaluating the performance of air quality models (No. Issue 3). Department for Environment and Rural Affairs.

Desarkar, A., & Das, A. (2018). Implementing decision tree in air pollution reduction framework. In *Smart Computing and Informatics* (pp. 105-113). Springer, Singapore.

Díaz, J., Linares, C., Carmona, R., Russo, A., Ortiz, C., Salvador, P., & Trigo, R. M. (2017). Saharan dust intrusions in Spain: health impacts and associated synoptic conditions. *Environmental research*, 156, 455-467.

Dickerson, R., Kondragunta, S., Stenchikov, G., Civerolo, K., Doddridge, B., and Holben, B.: The impact of aerosols on solar ultraviolet radiation and photochemical smog, *Science*, 278, 827–830, 1997.

Dincer, N. G., & Akkuş, Ö. (2018). A new fuzzy time series model based on robust clustering for forecasting of air pollution. *Ecological Informatics*, 43, 157-164.

Diner, D.J.; Abdou, W.; Ackerman, T.; Crean, K.; Gordon, H.; Kahn, R.; Martonchik, J.; McMuldloch, S.; Paradise, S.; Pinty, B.; et al. *MISR Level 2 Aerosol Retrieval Algorithm Theoretical Basis*; NASA: Greenbelt, MD, USA, 1999; JPL D11400.

Directive 2008/50/EC, Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on Ambient Air Quality and Cleaner Air for Europe. Annex I. [WWW Document] URL <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32008L0050>.

Dirección General de Estadísticas y Censos, 2016. Estadísticas vitales. Ministerio de Hacienda GCBA. Consultado última vez 1/10/2019 en [https://www.estadisticaciudad.gob.ar/eyc/wp-content/uploads/2016/05/poblacion\\_2016\\_023.pdf](https://www.estadisticaciudad.gob.ar/eyc/wp-content/uploads/2016/05/poblacion_2016_023.pdf)

Domańska, D., & Łukasik, S. (2016). Handling high-dimensional data in air pollution forecasting tasks. *Ecological Informatics*, 34, 70-91.

- Dominici, F., McDermott, A., Zeger, S. L., & Samet, J. M. (2002). On the use of generalized additive models in time-series studies of air pollution and health. *American journal of epidemiology*, 156(3), 193-203.
- Dorling, S. R., Davies, T. D., & Pierce, C. E. (1992). Cluster analysis: A technique for estimating the synoptic meteorological controls on air and precipitation chemistry—Method and applications. *Atmospheric Environment. Part A. General Topics*, 26(14), 2575-2581.
- Downey, A. (2018). *Think complexity: complexity science and computational modeling*. " O'Reilly Media, Inc."
- Duboue, M. (1978). Pollution roses: a simple way of interpreting the data obtained by air pollution measurement systems in the proximity of refineries. In *Studies in Environmental Science (Vol. 1, pp. 133-136)*. Elsevier.

## E

- Eastoe, E. F. (2009). A hierarchical model for non-stationary multivariate extremes: a case study of surface-level ozone and NOx data in the UK. *Environmetrics: The official journal of the International Environmetrics Society*, 20(4), 428-444.
- EEA, European Environment Agency (2017). Air quality in Europe 2017. ISBN 978-92-9213-921-6
- Elangasinghe, M. A., Singhal, N., Dirks, K. N., Salmond, J. A., & Samarasinghe, S. (2014 a). Complex time series analysis of PM10 and PM2.5 for a coastal site using artificial neural network modelling and k-means clustering. *Atmospheric Environment*, 94, 106-116.
- Engel-Cox, J., Oanh, N. T. K., van Donkelaar, A., Martin, R. V., & Zell, E. (2013). Toward the next generation of air quality monitoring: Particulate Matter. *Atmospheric Environment*, 80, 584-590.
- Engel-Cox, J. A., Hoff, R. M., & Haymet, A. D. J. (2004). Recommendations on the use of satellite remote-sensing data for urban air quality. *Journal of the Air & Waste Management Association*, 54(11), 1360-1371.
- El-Harbawi, M. (2013). Air quality modelling, simulation, and computational methods: a review. *Environmental Reviews*, 21(3), 149-179.
- Engelstaedter, S., & Washington, R. (2007). Atmospheric controls on the annual cycle of North African dust. *Journal of Geophysical Research: Atmospheres*, 112(D3).
- Environmental Protection Agency (2006), "Data Quality Assessment: A Reviewer's Guide". EPA/240/B-06/002
- ENVIRON, O. (2014). User's Guide to the Comprehensive Air Quality Model with Extensions Version 5.40. *ENVIRON International Corporation, Novato, CA. Available at: www.camx.com.*
- Eritja, M. C. (2007). El mecanismo de cumplimiento del Protocolo de Kioto: un nuevo paso en aras al control de cumplimiento de los acuerdos internacionales ambientales. *Revista electrónica de estudios internacionales (REEI)*, (14), 11.
- Estrabou, C., Filippini, E., Soria, J. P., Schelotto, G., & Rodriguez, J. M. (2011). Air quality monitoring system using lichens as bioindicators in Central Argentina. *Environmental monitoring and assessment*, 182(1), 375-383.
- European Environment Agency (2016). Air quality in Europe — 2016 report. EEA Report No 28/2016. <https://www.eea.europa.eu/.../air-quality- in-europe-2016/download>
- Evans, G. W., & Kantrowitz, E. (2002). Socioeconomic status and health: the potential role of environmental risk exposure. *Annual review of public health*, 23(1), 303-331.

## F

- Fan, J., & Yao, Q. (2008). *Nonlinear time series: nonparametric and parametric methods*. Springer Science & Business Media.
- Fayyad, U. M., Piatetsky-Shapiro, G. & Smyth, P. 1996a. From data mining to knowledge discovery: an overview, *Advances in Knowledge Discovery and Data Mining*, 1–34. American Association for Artificial Intelligence.
- Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. 1996b. The KDD PROCESS for extracting useful knowledge from volumes of data. *Communication of the ACM* 39, 27–34.
- Fayyad, U., Piatetsky-Shapiro, G., Smith, P. & Uthurusamy, R. 1996c. *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press.
- Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., & Herrera, F. (2018). Introduction to KDD and Data Science. In *Learning from Imbalanced Data Sets* (pp. 1-17). Springer, Cham.
- Fidalgo, F. y Martínez, O.R. 1983. Algunas características geomorfológicas dentro del Partido de La Plata (Buenos Aires). *Revista de la Asociación Geológica Argentina* 38: 263-279.
- Finlayson-Pitts, B. J., & Pitts Jr, J. N. (1999). *Chemistry of the upper and lower atmosphere: theory, experiments, and applications*. Elsevier.
- Formenti, P., D'Anna, B., Flamant, C., Mallet, M., Piketh, S.J., Schepanski, K., Waquet, F., Auriol, F., Brogniez, G., Burnet, F. and Chaboureaud, J.P., 2019. The Aerosols, Radiation and Clouds in southern Africa (AEROCLO-SA) field campaign in Namibia: overview, illustrative observations and way forward. *Bulletin of the American Meteorological Society*, (2019).
- Frenguelli, J. (1950). Rasgos generales de la morfología y geología de la provincia de Buenos Aires. Provincia de Buenos Aires, Ministerio de Obras Publicas de la Provincia de Buenos Aires. *Lab. Ensayos Materiales Invest. Tecnol.(LEMIT) Ser. II, 33*, 1-72.
- Fucks, E. E., D'amico, G. M., Pisano, M. F., & Nuccetelli, G. (2017). Evolución geomorfológica de la región del Gran La Plata y su relación con eventos catastróficos.
- Funtowicz, S.O. & Ravetz, J.R. (1991). A new scientific methodology for global environmental issues. *Ecological economics: The science and management of sustainability*, 10, p.137.
- Funtowicz, S. O., & Ravetz, J. R. (2000). *La ciencia posnormal: ciencia con la gente* (Vol. 160). Icaria editorial.

## G

- García, S., Ramírez, S., Luengo, J., & Herrera, F. C. (2016). Big Data: Preprocesamiento y calidad de datos. *University of Granada*, 18-20.
- García Ferreyra, M. F. (2013). Obtención de mapas de calidad de aire, a través de la implementación y primera aplicación del modelo de transporte químico CHIMERE sobre Argentina. Tesis de Maestría en Aplicaciones Espaciales de Alerta y Respuesta Temprana a Emergencias. Facultad de Matemática, Astronomía y Física, Universidad Nacional de Córdoba.
- García Ferreyra, M. F., Curci, G., & Lanfri, M. (2016). First Implementation of the WRF-CHIMERE-EDGAR Modeling System Over Argentina. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(12), 5304-5314.

- Gaitán Varón, M. & Cárdenas Ruiz, P. A., 2017. Guía para la elaboración de inventarios de emisiones atmosféricas [Recurso electrónico] - Dirección de Asuntos Ambientales, Sectorial y Urbana. -- Bogotá, D.C.: Colombia. Ministerio de Ambiente y Desarrollo Sostenible, 2017. 92 p. ISBN: ISBN: 978-958-8901-64-0
- Gassó, S., & Stein, A. F. (2007). Does dust from Patagonia reach the sub-Antarctic Atlantic Ocean?. *Geophysical Research Letters*, 34(1).
- Gassó, S., Stein, A., Marino, F., Castellano, E., Udisti, R., & Ceratto, J. (2010). A combined observational and modeling approach to study modern dust transport from the Patagonia desert to East Antarctica.
- Garaga, R., Sahu, S. K., & Kota, S. H. (2018). A review of air quality modeling studies in India: local and regional scale. *Current Pollution Reports*, 1-15.
- Garratt, J. R. (1994). The atmospheric boundary layer. *Earth-Science Reviews*, 37(1-2), 89-134.
- Garreaud, R. D., & Falvey, M. (2009). The coastal winds off western subtropical South America in future climate scenarios. *International Journal of Climatology*, 29(4), 543-554.
- Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R. and Wargan, K. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, 30(14), pp.5419-5454.
- Georgakopoulos, D., Hornick, M., & Sheth, A. (1995). *An overview of workflow management: From process modeling to workflow automation infrastructure. Distributed and Parallel Databases*, 3(2), 119-153. doi:10.1007/bf01277643
- Gibson, N. (1990). The right to a clean environment. *Sask. L. Rev.*, 54, 5.
- Ginoux, P., Chin, M., Tegen, I., Prospero, J. M., Holben, B., Dubovik, O., & Lin, S. J. (2001). Sources and distributions of dust aerosols simulated with the GOCART model. *Journal of Geophysical Research: Atmospheres*, 106(D17), 20255-20273.
- Gilks, W. R., Richardson S., & Spiegelhalter D. (1995). *Markov chain Monte Carlo in practice*. Chapman and Hall/CRC.
- Giuliani, D., Orte, M., Martins, E., Matamoros, N., Colman, J., & Porta, A. (2017, September). Caracterización química y espectroscópica de material particulado en aire extramuros en la ciudad de La Plata y alrededores. In *IV Congreso Internacional Científico y Tecnológico-CONCYT 2017*.
- Goldberg, D. L., Gupta, P., Wang, K., Jena, C., Zhang, Y., Lu, Z., & Streets, D. G. (2019). Using gap-filled MAIAC AOD and WRF-Chem to estimate daily PM<sub>2.5</sub> concentrations at 1 km resolution in the Eastern United States. *Atmospheric environment*, 199, 443-452.
- Gómez-Losada, Á. (2017). Clustering Air Monitoring Stations According to Background and Ambient Pollution Using Hidden Markov Models and Multidimensional Scaling. In *Data Science* (pp. 123-132). Springer, Cham.
- Gong, B., & Ordieres-Meré, J. (2016). Prediction of daily maximum ozone threshold exceedances by preprocessing and ensemble artificial intelligence techniques: Case study of Hong Kong. *Environmental Modelling & Software*, 84, 290-303.
- Goodchild, MF (1987). A spatial analytical perspective on geographical information systems. *International journal of geographical information system*, 1(4), 327-334.
- Goodchild MF (2007) Citizens as sensors: The world of volunteered geography. *GeoJournal* 69:211-221.
- Gorman, D., Drewry, A., Huang, Y. L., & Sames, C. (2003). The clinical toxicology of carbon monoxide. *Toxicology*, 187(1), 25-38.
- Gramsch, E., Cereceda-Balic, F., Oyola, P., & Von Baer, D. (2006). Examination of pollution trends in Santiago de Chile with cluster analysis of PM<sub>10</sub> and ozone data. *Atmospheric environment*, 40(28), 5464-5475.

- Grimm, N. B., Faeth, S. H., Golubiewski, N. E., Redman, C. L., Wu, J., Bai, X., & Briggs, J. M. (2008). Global change and the ecology of cities. *science*, 319(5864), 756-760.
- Guo H, Lee SC, Chan LY, Li WM (2004). Risk assessment of exposure to volatile organic compounds in different indoor environments. *Environ Res.* 94(1):57-66.
- Gupta, P., Christopher, S.A. (2009). Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach. *J. Geophys. Res.* 114 (D20), D20205.
- Gupta, P.; Remer, L.A.; Levy, R.C.; Mattoo, S (2018). Validation of MODIS 3 km land aerosol optical depth from NASA's EOS Terra and Aqua missions. *Atmos. Meas. Tech.*, 11, 3145.
- Gurjar, B. R., Jain, A., Sharma, A., Agarwal, A., Gupta, P., Nagpure, A. S., & Lelieveld, J. (2010). Human health risks in megacities due to air pollution. *Atmospheric Environment*, 44(36), 4606-4613.

## H

- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- Hancock, J. (2003). *Environmental human rights: Power, ethics, and law*. London, ENG: Ashgate.
- Hamra, G.B., Guha, N., Cohen, A., Laden, F., Raaschou-Nielsen, O., Samet, J.M., Vineis, P., Forastiere, F., Saldiva, P., Yorifuji, T. and Loomis, D., 2014. Outdoor particulate matter exposure and lung cancer: a systematic review and meta-analysis. *Environmental health perspectives*.
- Harrop, O. (2018). *Air quality assessment and management: A practical guide*. CRC Press.
- Hashim, J. H., Pillay, M. S., Hashim, Z., Shamsudin, S. B., Sinha, K., Zulkifli, Z. H., & Lian, K. S. (2004). A study of health impact and risk assessment of urban air pollution in the Klang Valley, Malaysia. *A research project report submitted by UKM Pakarunding Sdn. Bhd. Universiti Kebangsaan Malaysia to WHO-Western Pacific Regional Office Available at: [http://www.wpro.who.int/environmental\\_health/documents/docs/UKM\\_Report.pdf](http://www.wpro.who.int/environmental_health/documents/docs/UKM_Report.pdf)*.
- Hasenfratz, D., Saukh, O., Walser, C., Hueglin, C., Fierz, M., Arn, T., Beutel, J. and Thiele, L., (2015). Deriving high-resolution urban air pollution maps using mobile sensor nodes. *Pervasive and Mobile Computing*, 16, pp.268-285.
- Hastie, T, Tibshirani, R, Friedman, J (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York: Springer.
- He, J., Gong, S., Yu, Y., Yu, L., Wu, L., Mao, H., Song, C., Zhao, S., Liu, H., Li, X. and Li, R., 2017. Air pollution characteristics and their relation to meteorological conditions during 2014–2015 in major Chinese cities. *Environmental pollution*, 223, pp.484-496.
- Heal, M. R., Kumar, P., & Harrison, R. M. (2012). Particles, air quality, policy and health. *Chemical Society Reviews*, 41(19), 6606-6630.
- Hemminki, K., & Pershagen, G. (1994). Cancer risk of air pollution: epidemiological evidence. *Environmental Health Perspectives*, 102(suppl 4), 187-192.
- Henderson, S., Beckerman, B., Jerrett, M., Brauer, M., 2007. Application of land use regression to estimate long-term concentrations of traffic related nitrogen oxides and fine particulate matter. *Environ. Sci. Technol.* 41, 2422–2428.
- Henschel, S., Le Tertre, A., Atkinson, R.W., Querol, X., Pandolfi, M., Zeka, A., Haluza, D., Analitis, A., Katsouyanni, K., Bouland, C. and Pascal, M., 2015. Trends of nitrogen oxides in ambient air in nine European cities between 1999 and 2010. *Atmospheric environment*, 117, pp.234-241.

- Héroux, M.E., Anderson, H.R., Atkinson, R., Brunekreef, B., Cohen, A., Forastiere, F., Hurley, F., Katsouyanni, K., Krewski, D., Krzyzanowski, M. and Künzli, N., 2015. Quantifying the health impacts of ambient air pollutants: recommendations of a WHO/Europe project. *International journal of public health*, 60(5), pp.619-627.
- Hijmans RJ (2019). raster: Geographic Data Analysis and Modeling. R package version 2.8-19. <https://CRAN.R-project.org/package=raster>
- Ho, T.K. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal.* 1998, 20, 832–844.
- Hoek, G., Beelen, R., De Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., & Briggs, D. (2008). A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric environment*, 42(33), 7561-7578.
- Hoek, G., Eeftens, M., Beelen, R., Fischer, P., Brunekreef, B., Boersma, K. F., & Veeffkind, P. (2015). Satellite NO<sub>2</sub> data improve national land use regression models for ambient NO<sub>2</sub> in a small densely populated country. *Atmospheric environment*, 105, 173-180.
- Holben, B.N., Eck, T.F., Slutsker, I., Tanre, D., Buis, J.P., Setzer, A., Vermote, E., Reagan, J.A., Kaufman, Y.J., Nakajima, T. and Lavenue, F., 1998. AERONET—A federated instrument network and data archive for aerosol characterization. *Remote sensing of environment*, 66(1), pp.1-16.
- Hooyberghs, J., Mensink, C., Dumont, G., Fierens, F., & Bresseur, O. (2005). A neural network forecast for daily average PM<sub>10</sub> concentrations in Belgium. *Atmospheric Environment*, 39(18), 3279-3289.
- Hu, X., Belle, J. H., Meng, X., Wildani, A., Waller, L. A., Strickland, M. J., & Liu, Y. (2017). Estimating PM<sub>2.5</sub> concentrations in the conterminous United States using the random forest approach. *Environmental science & technology*, 51(12), 6936-6944.
- Huang, K., Xiao, Q., Meng, X., Geng, G., Wang, Y., Lyapustin, A., Gu, D. and Liu, Y., 2018. Predicting monthly high-resolution PM<sub>2.5</sub> concentrations with random forest model in the North China Plain. *Environmental pollution*, 242, pp.675-683.
- Huth, R., 1996. An intercomparison of computer-assisted circulations classification methods. *Int. J. Climatol.* 16, 893e922
- Hsu, A., Reuben, A., Shindell, D., de Sherbinin, A., & Levy, M. (2013). Toward the next generation of air quality monitoring indicators. *Atmospheric environment*, 80, 561-570.
- Hsu, N.C., Jeong, M.J., Bettenhausen, C., Sayer, A.M., Hansell, R., Seftor, C.S., Huang, J. and Tsay, S.C., (2013). Enhanced Deep Blue aerosol retrieval algorithm: The second generation. *Journal of Geophysical Research: Atmospheres*, 118(16), pp.9296-9315.

## I

- Ichoku, C., Chu, D.A., Mattoo, S., Kaufman, Y.J., Remer, L.A., Tanré, D., Slutsker, I. and Holben, B.N. (2002). A spatio-temporal approach for global validation and analysis of MODIS aerosol products. *Geophysical Research Letters*, 29(12), pp.MOD1-1.
- Ichoku, C., Kaufman, Y. J., Remer, L. A., & Levy, R. (2004). Global aerosol remote sensing from MODIS. *Advances in Space Research*, 34(4), 820-827.
- INDEC (Instituto Nacional de Estadísticas y Censos). Censo Nacional de Población, Hogares y Viviendas 2010. <http://www.censo2010.indec.gov.ar/>. (2011).
- Instituto Nacional de Estadísticas y Censo (INDEC). 2010. en línea. Censo Nacional de Población, Hogares y Viviendas. Disponible en: [www.censo2010.indec.gov.ar](http://www.censo2010.indec.gov.ar)
- IPCS. Environmental Health criteria 214: Human Exposure Assessment. World Health Organization. Genova, 2000.

ISO/IEC. 25012:2008 – software engineering – software product quality requirements and evaluation (square) – data quality model. Technical report, ISO/IEC, 2008

Irwin, J.S., Civerolo, K., Hogrefe, C., Appel, W., Foley, K., Swall, J., 2008. A procedure for inter-comparing the skill of regional-scale air quality model simulations of daily maximum 8-h ozone concentrations. *Atmospheric Environment* 42, 5403e5412.

## J

Jackson, J.M.; Liu, H.; Laszlo, I.; Kondragunta, S.; Remer, L.A.; Huang, J.; Huang, H.C. Suomi-NPP VIIRS aerosol algorithms and data products. *J. Geophys. Res. Atmos.* **2013**, 118.

Jacobson, M. Z. (2002). *Atmospheric pollution: history, science, and regulation*. Cambridge University Press.

Jacobson, M. Z. (2005). *Fundamentals of atmospheric modeling*. Cambridge university press.

Janssen, S., Dumont, G., Fierens, F., & Mensink, C. (2008). Spatial interpolation of air pollution measurements using CORINE land cover data. *Atmospheric Environment*, 42(20), 4884-4903.

Jarvis A., H.I. Reuter, A. Nelson, E. Guevara, 2008, Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT), available from <http://srtm.csi.cgiar.org>.

Jenks, George F. (1967): "*The Data Model Concept in Statistical Mapping*", en *International Yearbook of Cartography* n° 7: pp. 186-190.

Jiang, Z., Mao, B., Meng, X., Du, X., Liu, S., & Li, S. (2010, August). An air quality forecast model based on the BP neural network of the samples self-organization clustering. In *2010 Sixth International Conference on Natural Computation* (Vol. 3, pp. 1523-1527). IEEE.

Jiang, P., Dong, Q., & Li, P. (2017). A novel hybrid strategy for PM<sub>2.5</sub> concentration analysis and prediction. *Journal of environmental management*, 196, 443-457.

Johnson, M. S., Meskhidze, N., Kiliyanpilakkil, V. P., & Gassó, S. (2011). Understanding the transport of Patagonian dust and its influence on marine biological activity in the South Atlantic Ocean. *Atmospheric Chemistry & Physics*, 11(6).

Just, A. C., Wright, R. O., Schwartz, J., Coull, B. A., Baccarelli, A. A., Tellez-Rojo, M. M., ... & Kloog, I. (2015). Using high-resolution satellite aerosol optical depth to estimate daily PM<sub>2.5</sub> geographical distribution in Mexico City. *Environmental science & technology*, 49(14), 8576-8584.

## K

Kalabokas, P. D., Bartzis, J. G., & Papagiannakopoulos, P. (2002). Atmospheric levels of nitrogen oxides at a Greek oil refinery compared with the urban measurements in Athens. *Water, Air and Soil Pollution: Focus*, 2(5-6), 703-716.

Kalnay, E., & Cai, M. (2003). Impact of urbanization and land-use change on climate. *Nature*, 423(6939), 528.

Kam, W., Liacos, J. W., Schauer, J. J., Delfino, R. J., & Sioutas, C. (2012). Size-segregated composition of particulate matter (PM) in major roadways and surface streets. *Atmospheric environment*, 55, 90-97.

Kampa, M., & Castanas, E. (2008). Human health effects of air pollution. *Environmental pollution*, 151(2), 362-367.

Kasperson, R. E., Renn, O., Slovic, P., Brown, H. S., Emel, J., Goble, R., ... & Ratick, S. (1988). The social amplification of risk: A conceptual framework. *Risk analysis*, 8(2), 177-187.

- Kelling, S, Hochachka, W, Fink, D (2009) Data-intensive Science: A new paradigm for biodiversity studies. *BioScience* 59(7): 613–620.
- Kikuchi, M.; Murakami, H.; Suzuki, K.; Nagao, T.M.; Higurashi, A. improved hourly estimates of aerosol optical thickness using spatiotemporal variability derived from Himawari-8 geostationary satellite. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3442–3455.
- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big data & society*, *1*(1), 2053951714528481.
- Knafllic, C. N. (2015). *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons.
- Krotkov, N.A., McLinden, C.A., Li, C., Lamsal, L.N., Celarier, E.A., Marchenko, S.V., et al., 2016. Aura OMI observations of regional SO<sub>2</sub> and NO<sub>2</sub> pollution changes from 2005 to 2015. *Atmos. Chem. Phys.* *16*, 4605–4629
- Kruse, E., Sarandón, R. y Gaspari, F. 2014. Impacto del cambio climático en el Gran La Plata. Editorial Universidad Nacional de La Plata, 105 p., La Plata.
- Krzyzanowski, M., & Cohen, A. (2008). Update of WHO air quality guidelines. *Air Quality, Atmosphere & Health*, *1*(1), 7-13.
- Kuhn M. Contributions from J Wing, S Weston, A Williams, C Keefer, A Engelhardt, T Cooper, Z Mayer, B Kenkel, the R Core Team, M Benesty, R Lescarbeau, A Ziem, L Scrucca, Y Tang, Can Candan and T Hunt. (2019). caret: Classification and Regression Training. R package version 6.0-84. <https://CRAN.R-project.org/package=caret>
- Kumar, P., Morawska, L., Martani, C., Biskos, G., Neophytou, M., Di Sabatino, S., ... & Britter, R. (2015). The rise of low-cost sensing for managing air pollution in cities. *Environment international*, *75*, 199-205.

## L

- Lamb, R. G., & Seinfeld, J. H. (1973). Mathematical modeling of urban air pollution. General theory. *Environmental Science & Technology*, *7*(3), 253-261.
- Laughner, J. L., Zare, A., & Cohen, R. C. (2016). Effects of daily meteorology on the interpretation of space-based remote sensing of NO<sub>2</sub>. *Atmospheric Chemistry and Physics*, *16*(23), 15247-15264.
- Laughner, J. L., Zare, A., & Cohen, R. C. (2016). Effects of daily meteorology on the interpretation of space-based remote sensing of NO<sub>2</sub>. *Atmospheric Chemistry and Physics*, *16*(23), 15247-15264.
- Lee, H.J., Liu, Y., Coull, B.A., Schwartz, J., Koutrakis, P., 2011. A novel calibration approach of MODIS AOD data to predict PM<sub>2.5</sub> concentrations. *Atmos. Chem. Phys.* *11*, 7991e8002
- Leelőssy, Á., Molnár, F., Izsák, F., Havasi, Á., Lagzi, I., & Mészáros, R. (2014). Dispersion modeling of air pollutants in the atmosphere: a review. *Open Geosciences*, *6*(3), 257-278.
- Leroy, M.; Deuzé, J.L.; Bréon, F.M.; Hautecoeur, O.; Herman, M.; Buriez, J.C.; Tanré, D.; Bouffies, S.; Chazette, P.; Roujean, J.L. Retrieval of atmospheric properties and surface bidirectional reflectances over land from POLDER/ADEOS. *J. Geophys. Res. Atmos.* 1997, *102*, 17023–17037.
- Leung, Y., Leung, K. S., Wong, M. H., Mak, T., Cheung, K. Y., Lo, L. Y., ... & Dong, Y. L. (2018). An integrated web-based air pollution decision support system—a prototype. *International Journal of Geographical Information Science*, 1-28.
- Levelt, P.F., Joiner, J., Tamminen, J., Veefkind, J.P., Bhartia, P.K., Zweers, D.C.S., et al., 2018. The Ozone Monitoring Instrument: overview of 14 years in space. *Atmos. Chem. Phys.* *18*, 5699–5745.
- Levy, R. J. (2015). Carbon monoxide pollution and neurodevelopment: a public health concern. *Neurotoxicology and teratology*, *49*, 31-40.



- Levy, J. I., Greco, S. L., & Spengler, J. D. (2002). The importance of population susceptibility for air pollution risk assessment: a case study of power plants near Washington, DC. *Environmental Health Perspectives*, 110(12), 1253-1260.
- Li, F., Ginoux, P., & Ramaswamy, V. (2010). Transport of Patagonian dust to Antarctica. *Journal of Geophysical Research: Atmospheres*, 115(D18)
- Li, C., Hsu, N.C., Tsay, S.-C., 2011. A study on the potential applications of satellite data in air quality monitoring and forecasting. *Atmos. Environ.* 45 (22), 3663e3675.
- Li, Y., Xue, Y., Guang, J., Mei, L., She, L., Fan, C., & Chen, G. (2017, July). Estimating ground-level PM 2.5 concentration in Beijing using BP ANN model from satellite data. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 4870-4873). IEEE.
- Li, Y., Xue, Y., Guang, J., She, L., Fan, C., & Chen, G. (2018). Ground-Level PM<sub>2.5</sub> Concentration Estimation from Satellite Data in the Beijing Area Using a Specific Particle Swarm Extinction Mass Conversion Algorithm. *Remote Sensing*, 10(12), 1906.
- Li, Z., Roy, D. P., Zhang, H. K., Vermote, E. F., & Huang, H. (2019). Evaluation of Landsat-8 and Sentinel-2A Aerosol Optical Depth Retrievals across Chinese Cities and Implications for Medium Spatial Resolution Urban Aerosol Monitoring. *Remote Sensing*, 11(2), 122.
- Licursi, M., Sierra, M. V., & Gómez, N. (2006). Diatom assemblages from a turbid coastal plain estuary: Río de la Plata (South America). *Journal of Marine systems*, 62(1-2), 35-45.
- Lijteroff, R., Cortinez, V., & Raba, J. (1999). Urban development and air quality in San Luis City, Argentina. *Environmental monitoring and assessment*, 57(2), 169-182.
- Lillesand, T., Kiefer, R. W., & Chipman, J. (2015). *Remote sensing and image interpretation*. John Wiley & Sons.
- Limaye, V., Knowlton, K., Sarkar, S., Ganguly, P., Pingle, S., Dutta, P., ... & Kakkad, K. (2018). Development of Ahmedabad's Air Information and Response (AIR) Plan to Protect Public Health. *International journal of environmental research and public health*, 15(7), 1460.
- Lin, S.-J., 2004. A vertically Lagrangian finite-volume dynamical core for global models. *Mon. Wea. Rev.*, 132, 2293–2307.
- Lippmann Morton (2014). Toxicological and epidemiological studies of cardiovascular effects of ambient air fine particulate matter (PM<sub>2.5</sub>) and its chemical components: coherence and public health implications. *CritRev Toxicol.*, 44(4): 299-347.
- Liu, C.W., Lin, K.H., Kuo, Y.M., 2003. Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Science of Total Environment* 313 (1–3), 77–89
- Longhurst, J. W. S., Beattie, C. I., Chatterton, T. J., Hayes, E. T., Leksmono, N. S., & Woodfield, N. K. (2006). Local air quality management as a risk management process: Assessing, managing and remediating the risk of exceeding an air quality objective in Great Britain. *Environment international*, 32(8), 934-947.
- Loomis D, Grosse Y, Lauby-Secretan B, et al; International Agency for Research on Cancer Monograph Working Group, IARC. The carcinogenicity of outdoor air pollution. *Lancet Oncol* 2013; 14: 1262–1263.
- López, M. L., Ceppi, S., Palancar, G. G., Olcese, L. E., Tirao, G., & Toselli, B. M. (2011). Elemental concentration and source identification of PM<sub>10</sub> and PM<sub>2.5</sub> by SR-XRF in Córdoba City, Argentina. *Atmospheric environment*, 45(31), 5450-5457.
- Lorenz, E. N. The Problem of Deducing the Climate from the Governing Equations. *Telus*. 1964, 16: 1-11

- Lu F., Xu D., Cheng Y., Dong S., Guo C., Jiang X., Zheng X., 2015. Systematic review and meta-analysis of the adverse health effects of ambient PM<sub>2.5</sub> and PM<sub>10</sub> pollution in the Chinese population. *Environ Res.*, 136:196-204.
- Luecken, D. J., Hutzell, W. T., & Gipson, G. L. (2006). Development and analysis of air quality modeling simulations for hazardous air pollutants. *Atmospheric Environment*, 40(26), 5087-5096.
- Lyapustin, A., Wang, Y., Laszlo, I., Kahn, R., Korkin, S., Remer, L., ... & Reid, J. S. (2011). Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm. *Journal of Geophysical Research: Atmospheres*, 116(D3).
- Lyapustin, A.; Wang, Y.; Korkin, S.; Huang (2018), D. MODIS Collection 6 MAIAC algorithm. *Atmos. Meas. Tech.*, 11, 5741-5765
- Lyapustin, A., Wang, Y. (2018). MCD19A2 MODIS/Terra+Aqua Land Aerosol Optical Depth Daily L2G Global 1km SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC. doi: 10.5067/MODIS/MCD19A2.006

## M

- Ma, K. L., Liao, I., Frazier, J., Hauser, H., & Kostis, H. N. (2012). Scientific storytelling using visualization. *IEEE Computer Graphics and Applications*, 32(1), 12-19.
- Mahlman, J. D. (1997). Dynamics of transport processes in the upper troposphere. *Science*, 276(5315), 1079-1083.
- Mao, X., Shen, T., & Feng, X. (2017). *Prediction of hourly ground-level PM 2.5 concentrations 3 days in advance using neural networks with satellite data in eastern China. Atmospheric Pollution Research*, 8(6), 1005-1015. doi:10.1016/j.apr.2017.04.002
- Martínez, J., Saavedra, Á., García-Nieto, P. J., Piñeiro, J. I., Iglesias, C., Taboada, J., ... & Pastor, J. (2014). Air quality parameters outliers detection using functional data analysis in the Langreo urban area (Northern Spain). *Applied Mathematics and Computation*, 241, 1-10.
- Masiol, M., Squizzato, S., Formenton, G., Harrison, R. M., & Agostinelli, C. (2017). Air quality across a European hotspot: Spatial gradients, seasonality, diurnal cycles and trends in the Veneto region, NE Italy. *Science of The Total Environment*, 576, 210-224.
- Massolo L., Rehwagen M., Porta A., Herbarth O., Ronco A., Müller A (2010). "Indoor-outdoor distribution and risk assessment of volatile organic compounds in the atmosphere of industrial and urban areas". *Environmental Toxicology*, 25(4):339-49.
- Massolo, L., Rehwagen, M., Porta, A., Ronco, A., Herbarth, O., Mueller, A. "Indoor outdoor distribution and risk assessment of volatile organic compounds in the atmosphere of industrial and urban areas", *Environmental Toxicology*. Vol. 25, pp: 339-349 (2010)
- Mavko, M. E., Tang, B., & George, L. A. (2008). A sub-neighborhood scale land use regression model for predicting NO<sub>2</sub>. *Science of the Total Environment*, 398(1), 68-75.
- Mavroidis, I., & Ilia, M. (2012). Trends of NO<sub>x</sub>, NO<sub>2</sub> and O<sub>3</sub> concentrations at three different types of air quality monitoring stations in Athens, Greece. *Atmospheric environment*, 63, 135-147.
- Mayer, H. (1999). Air pollution in cities. *Atmospheric environment*, 33(24-25), 4029-4037.
- Mazzeo, N. A., Venegas, L. E., & Choren, H. (2005). Analysis of NO, NO<sub>2</sub>, O<sub>3</sub> and NO<sub>x</sub> concentrations measured at a green area of Buenos Aires City during wintertime. *Atmospheric Environment*, 39(17), 3055-3068.
- McCollister, G. M., & Wilson, K. R. (1975). Linear stochastic models for forecasting daily maxima and hourly concentrations of air pollutants. *Atmospheric Environment (1967)*, 9(4), 417-423.

- McConnell, J. R., Aristarain, A. J., Banta, J. R., Edwards, P. R., & Simões, J. C. (2007). 20th-Century doubling in dust archived in an Antarctic Peninsula ice core parallels climate change and desertification in South America. *Proceedings of the National Academy of Sciences*, 104(14), 5743-5748.
- McKinney, M. L. (2002). Urbanization, Biodiversity, and Conservation. The impacts of urbanization on native species are poorly studied, but educating a highly urbanized human population about these impacts can greatly improve species conservation in all ecosystems. *Bioscience*, 52(10), 883-890.
- Mei, L.; Rozanov, V.; Vountas, M.; Burrows, J.P.; Levy, R.C.; Lotz, W. Retrieval of aerosol optical properties using MERIS observations: Algorithm and some first results. *Remote Sens. Environ.* **2017**, 197, 125–140.
- Mendiguchía, C., Moreno, C., Galindo-Riaño, M.D., García-Vargas, M., 2004. Using chemometric tools to assess anthropogenic effects in river water: a case study: Guadalquivir River (Spain). *Analytica Chimica Acta* 515 (1), 143–149.
- Meng, X., Fu, Q., Ma, Z., Chen, L., Zou, B., Zhang, Y., ... & Liu, Y. (2016). Estimating ground-level PM<sub>10</sub> in a Chinese city by combining satellite data, meteorological information and a land use regression model. *Environmental Pollution*, 208, 177-184.
- Middleton, D. R. (1998). A new box model to forecast urban air quality: BOXURB. *Environmental Monitoring and Assessment*, 52(1-2), 315-335.
- Millán, M., Estrela, M. J., & Badenas, C. (1997). Situaciones sinópticas asociadas a la dinámica de incendios forestales en la Comunidad Valenciana. *Ecología*, (11), 3-29.
- Ministerio para la Transición Ecológica (2018). Evaluación de la Calidad del Aire en España 2018. Consultado en [https://www.miteco.gob.es/es/calidad-y-evaluacion-ambiental/temas/atmosfera-y-calidad-del-aire/informeevaluacioncalidadairespana2018\\_tcm30-498764.pdf](https://www.miteco.gob.es/es/calidad-y-evaluacion-ambiental/temas/atmosfera-y-calidad-del-aire/informeevaluacioncalidadairespana2018_tcm30-498764.pdf). Último acceso 20-09-2019
- Mintz, D. (2012). Technical assistance document for the reporting of daily air quality-the air quality index (aqi): Us environmental protection agency. *Office of Air Quality Planning and Standards*, 1-26.
- Mitchell, G., Norman, P., & Mullin, K. (2015). Who benefits from environmental policy? An environmental justice analysis of air quality change in Britain, 2001–2011. *Environmental Research Letters*, 10(10), 105009.
- Molina, M. J., & Molina, L. T. (2004). Megacities and atmospheric pollution. *Journal of the Air & Waste Management Association*, 54(6), 644-680.
- Monks, P. S., Granier, C., Fuzzi, S., Stohl, A., Williams, M. L., Akimoto, H., ... & Blake, N. (2009). Atmospheric composition change—global and regional air quality. *Atmospheric environment*, 43(33), 5268-5350.
- Monteiro, A., Vautard, R., Borrego, C., & Miranda, A. I. (2005). Long-term simulations of photo oxidant pollution over Portugal using the CHIMERE model. *Atmospheric Environment*, 39(17), 3089-3101.
- Moran, P. A. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(2), 243-251.

## N

- Næss, Ø., Piro, F. N., Nafstad, P., Smith, G. D., & Leyland, A. H. (2007). Air pollution, social deprivation, and mortality: a multilevel cohort study. *Epidemiology*, 18(6), 686-694.
- NASA. (2019, 6 noviembre). Validation Method. Recuperado 25 noviembre, 2019, de <https://darktarget.gsfc.nasa.gov/validation>
- Natenzon, C; Marlenko, N.; González, S.; Ríos, D.; Murgida, A; Meconi G; Calvo, A. 2003. "Las dimensiones del riesgo en ámbitos urbanos. Catástrofes en el Área Metropolitana de Buenos Aires". En: ALESSANDRI CARLOS, Ana Fani;

Roberto BERTONCELLO. *Procesos territoriales en Argentina y Brasil*. Buenos Aires: Universidad de Buenos Aires/ Universidad de San Pablo. pp. 255-276.

Ni, X., Cao, C., Zhou, Y., Cui, X., & P Singh, R. (2018). Spatio-Temporal Pattern Estimation of PM<sub>2.5</sub> in Beijing-Tianjin-Hebei Region Based on MODIS AOD and Meteorological Data Using the Back Propagation Neural Network. *Atmosphere*, 9(3), 105.

Niska, H., Rantamäki, M., Hiltunen, T., Karppinen, A., Kukkonen, J., Ruuskanen, J., & Kolehmainen, M. (2005). Evaluation of an integrated modelling system containing a multi-layer perceptron model and the numerical weather prediction model HIRLAM for the forecasting of urban airborne pollutant concentrations. *Atmospheric Environment*, 39(35), 6524-6536.

NTP (National Toxicology Program). Report on Carcinogens, Thirteenth Edition. Research Triangle Park, NC: U.S. Department of Health and Human Services, Public Health Service. 2014

## O

Olcese, L. E., & Toselli, B. M. (1997). Effects of meteorology and land use on ambient measurements of primary pollutants in Córdoba city, Argentina. *Meteorology and Atmospheric Physics*, 62(3), 241-248.

Olsthoorn, X., Amann, M., Bartonova, A., Clench-Aas, J., Cofala, J., Dorland, K., ... & Larssen, S. (1999). Cost benefit analysis of European air quality targets for sulphur dioxide, nitrogen dioxide and fine and suspended particulate matter in cities. *Environmental and resource economics*, 14(3), 333-351.

Oliver, M.A., (1990) KRIGING, A method of interpolations for geographical information systems. *Int. J. Geographical Information Systems*. Vol. 4(3): 313-332.

Oliver, M. A. "Kriging: A Method of Interpolation for Geographical Information Systems". *International Journal of Geographic Information Systems* 4: 313-332. 1990.

Olvera-García, M. Á., Carbajal-Hernández, J. J., Sánchez-Fernández, L. P., & Hernández-Bautista, I. (2016). Air quality assessment using a weighted Fuzzy Inference System. *Ecological informatics*, 33, 57-74.

O'Neill, M. S., Jerrett, M., Kawachi, I., Levy, J. I., Cohen, A. J., Gouveia, N., ... & Workshop on Air Pollution and Socioeconomic Conditions. (2003). Health, wealth, and air pollution: advancing theory and methods. *Environmental health perspectives*, 111(16), 1861-1870.

Orona, N. S., Ferraro, S. A., Astort, F., Morales, C., Brites, F., Boero, L., ... & Tasat, D. R. (2016). Acute exposure to Buenos Aires air particles (UAP-BA) induces local and systemic inflammatory response in middle-aged mice: A time course study. *Environmental pollution*, 208, 261-270.

Orte, M. A., Coman Lerner, J., Gutiérrez, M., Elordi, L., Matamoros, N., Reyna Almandos, J., & Porta, A. (2015). Estudio de hidrocarburos aromáticos policíclicos asociados al material particulado y en fase gaseosa en la ciudad de La Plata y alrededores. *Libro de Actas de PROIMCA*.

Otero L. A.; P. R. Ristori; E. Pawelko; J. Pallotta; R. D'Elia; E. Quel. Biomass Burning Aerosol Detection over Buenos Aires City, August 2009. *Journal of Physics: Conference Series*.: IOP Publishing (doi:10.1088/1742-6596/274/1/012092), vol.274 n°1, ISSN 1742-6596 [http://iopscience.iop.org/1742-6596/274/1/012092] (2011).

Otero L. A., P. Ristori, J. V. Pallotta, E. Pawelko, E. Quel. (2012). The Puyehue-Cordón Caulle Volcanic Eruption, June 2011: Water Vapor, Atmospheric Boundary Layer and Aerosol Temporal Evolution in Buenos Aires, Argentina. *Revista Boliviana de Física*. [online]. 2012, vol.21, no.21

Organización Mundial de la Salud (2006). Guías de calidad del aire de la OMS relativas al material particulado, el ozono, el dióxido de nitrógeno y el dióxido de azufre. Actualización mundial 2005. Recuperado de [https://www.who.int/phe/health\\_topics/AQG\\_spanish.pdf](https://www.who.int/phe/health_topics/AQG_spanish.pdf)

- Ortegón, E. (2007). *Manual de prospectiva y decisión estratégica: bases teóricas e instrumentos para América Latina y el Caribe* (Vol. 51). United Nations Publications.
- Oshima, N., Kato, H., & Kadokura, S. (2002). An application of statistical downscaling to estimate surface air temperature in Japan. *Journal of Geophysical Research: Atmospheres*, 107(D10), ACL-14.
- Osseiran, N., & Chriscaden, K. (2016). Air pollution levels rising in many of the world's poorest cities. World Health Organization
- Ostro, B., & World Health Organization (2004). Outdoor air pollution: assessing the environmental burden of disease at national and local levels.

## P

- Paatero, P., Tapper, U., 1994. Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* 5, 111–126.
- Paciorek, C.J., Liu, Y., Moreno-Macias, H., Kondragunta, S., 2008. Spatiotemporal associations between GOES aerosol optical depth retrievals and ground-level PM<sub>2.5</sub>. *Environ. Sci. Technol.* 42, 5800e5806.
- Pacyna, J. M., & Graedel, T. E. (1995). Atmospheric emissions inventories: status and prospects. *Annual review of energy and the environment*, 20(1), 265-300.
- Patier, R. F. (2015). La evaluación de la calidad del aire en el marco de la Unión Europea y su repercusión en España. *Revista de Salud Ambiental*, 15, 61-64.
- Paul, M. J., & Meyer, J. L. (2001). Streams in the urban landscape. *Annual review of Ecology and Systematics*, 32(1), 333-365.
- Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and earth system sciences discussions*, 4(2), 439-473.
- Peled, R. (2011). Air pollution exposure: Who is at high risk? *Atmospheric Environment*, 45(10), 1781-1785.
- Petkova, E. P., Jack, D. W., Volavka-Close, N. H., & Kinney, P. L. (2013). Particulate matter pollution in African cities. *Air Quality, Atmosphere & Health*, 6(3), 603-614.
- Petrenko, M., Ichoku, C., & Leptoukh, G. (2012). Multi-sensor aerosol products sampling system (MAPSS). *Atmospheric Measurement Techniques*, 5(5), 913-926.
- Pfender, W., Graw, R., Bradley, W., Carney, M., & Maxwell, L. (2006). Use of a complex air pollution model to estimate dispersal and deposition of grass stem rust urediniospores at landscape scale. *Agricultural and forest meteorology*, 139(1-2), 138-153.
- Pires, J.C.M., Sousa, S.I.V., Pereira, M.C., Alvim-Ferraz, M.C.M., Martins, F.G., 2008a. Management of air quality monitoring using principal component and cluster analysis – part I: SO<sub>2</sub> and PM<sub>10</sub>. *Atmospheric Environment* 42 (6), 1249–1260.
- Pires, J.C.M., Sousa, S.I.V., Pereira, M.C., Alvim-Ferraz, M.C.M., Martins, F.G., 2008b. Management of air quality monitoring using principal component and cluster analysis – part II: CO, NO<sub>2</sub> and O<sub>3</sub>. *Atmospheric Environment* 42 (6), 1261–1274
- Pires, J. C. M., Pereira, M. C., Alvim-Ferraz, M. C. M., & Martins, F. G. (2009). Identification of redundant air quality measurements through the use of principal component analysis. *Atmospheric environment*, 43(25), 3837-3842.

- PNUMA/OMS 2002 Manuales de Metodología de GEMS/Aire. Volumen 1. Aseguramiento de la calidad en el monitoreo de la calidad del aire urbano Traducción del Centro Panamericano de Ingeniería Sanitaria y Ciencias del Ambiente (OPS/CEPIS) On line: [http://www.bvsde.paho.org/bvsci/e/fulltext/gemsaire/volumen\\_1.pdf](http://www.bvsde.paho.org/bvsci/e/fulltext/gemsaire/volumen_1.pdf)
- Podobnik, B., Grosse, I., Horvatić, D., Ilic, S., Ivanov, P. C., & Stanley, H. E. (2009). Quantifying cross-correlations using local and global detrending approaches. *The European Physical Journal B*, *71*(2), 243.
- Pöschl, U. (2005). Atmospheric aerosols: composition, transformation, climate and health effects. *Angewandte Chemie International Edition*, *44*(46), 7520-7540.
- Prasad, A. K., Singh, S., Chauhan, S. S., Srivastava, M. K., Singh, R. P., & Singh, R. (2007). Aerosol radiative forcing over the Indo-Gangetic plains during major dust storms. *Atmospheric Environment*, *41*(29), 6289-6301.
- Pressman, R. S. 2005. Software Engineering: A Practitioner's Approach, 6th edition. McGraw-Hill Science
- Pi, D. J. D., Garay, M. J., Seidel, F. C., Bruegge, C. J., Jovanovic, V. M., Bull, M. A., ... & Xu, F. (2017). AirMSPI Data Quality Statement: ORACLES Campaign.
- Pires, J. C. M., Sousa, S. I. V., Pereira, M. C., Alvim-Ferraz, M. C. M., & Martins, F. G. (2008). Management of air quality monitoring using principal component and cluster analysis—Part II: CO, NO<sub>2</sub> and O<sub>3</sub>. *Atmospheric Environment*, *42*(6), 1261-1274.
- Pizarro Hofer, R. (2001). *La vulnerabilidad social y sus desafíos: una mirada desde América Latina*. CEPAL.
- Pope 3rd, C. A. (2000). Epidemiology of fine particulate air pollution and human health: biologic mechanisms and who's at risk?. *Environmental health perspectives*, *108*(suppl 4), 713-723.
- Pope III, C. A., Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K., & Thurston, G. D. (2002). Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *Jama*, *287*(9), 1132-1141.
- Puliafito, E., Gantuz, M., & Puliafito, J. L. (2010). Characterizing mobile emissions by on-board measurements. *International Journal for Applied Environmental Studies*, *5*(2), 297-316.
- Puliafito, S. E., Allende, D., Fernández, R., Castro, F., & Cremades, P. (2011). New approaches for urban and regional air pollution modelling and management. In *Advanced Air Pollution*. InTech.
- Puliafito, S. E., Allende, D., Pinto, S., & Castesana, P. (2015). High resolution inventory of GHG emissions of the road transport sector in Argentina. *Atmospheric Environment*, *101*, 303-311.
- Prüss-Üstün, A., Wolf, J., Corvalán, C., Bos, R., & Neira, M. (2016). *Preventing disease through healthy environments: a global assessment of the burden of disease from environmental risks*. World Health Organization.

## R

- R Core Team (2013). R: A language and environment for statistical computing.
- Rasmussen, R. A. (1972). What do the hydrocarbons from trees contribute to air pollution?. *Journal of the Air Pollution Control Association*, *22*(7), 537-543.
- Ratto, G., Videla, F., Almandos, J. R., Maronna, R., & Schinca, D. (2006). Study of meteorological aspects and urban concentration of SO<sub>2</sub> in atmospheric environment of La Plata, Argentina. *Environmental monitoring and assessment*, *121*(1-3), 327-342.
- Ratto, G. (2016). *Estudio de parámetros ambientales utilizando técnicas espectroscópicas, datos meteorológicos y métodos estadísticos* (Doctoral dissertation, Universidad Nacional de La Plata).

- Reckziegel, F., Bustos, E., Mingari, L., B´aez, W., Villarosa, G., Folch, A., Collini, E., Viramonte, J., Romero, J., Osoros, S., Forecasting volcanic ash dispersal and coeval resuspension during the April–May 2015 Calbuco eruption, *Journal of Volcanology and Geothermal Research* (2016), doi: 10.1016/j.jvolgeores.2016.04.03
- Reichert, E. (2003). *Social work and human rights: A foundation for policy and practice*. New York, NY: Columbia University.
- Represa, S., Mellado, D., Bali, L., Lerner, C., Esteban, J., Sánchez, Y., & Porta, A. A. (2016). Aplicación de tecnologías de sistemas de información geográfica (SIG) para la estimación del riesgo por exposición a compuestos orgánicos volátiles (COVs). In *III Congreso de la Sociedad de Análisis de Riesgo Latinoamericana (SRA-LA) (Brasil, mayo de 2016)*.
- Resquin, M. D., Santágata, D., Gallardo, L., Gómez, D., Rössler, C., & Dawidowski, L. (2018). Local and remote black carbon sources in the Metropolitan Area of Buenos Aires. *Atmospheric environment*, 182, 105-114.
- Richter, A. and Wagner, T.: The Use of UV, Visible and Near IR Solar Back Scattered Radiation to Determine Trace Gases, in: *The Remote Sensing of Tropospheric Composition from Space*, edited by: Burrows, J., Platt, U., and Borrell, P., Springer, New York, 2011
- Ristori, P., Otero, L., Fochesatto, J., Flamant, P. H., Wolfram, E., Quel, E., ... & Holben, B. (2003). Aerosol optical properties measured in Argentina: wavelength dependence and variability based on sun photometer measurements. *Optics and lasers in engineering*, 40(1-2), 91-104.
- Ristori P., L. Otero, E. Pawelko, J. Pallotta. (2010). Aerosol and atmospheric boundary layer temporal evolution in Buenos Aires, Argentina during May 12, 2006. *Proceedings of the 25th International Laser Radar Conference*, 5–9 July 2010, St.-Petersburg, Russia, Publishing House of IAO SB RAS, 2010. ISBN 978-5-94458-109-9. Vol. I, 684-686 (2010).
- Rodríguez JD, Perez A, Lozano JA (2010) Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Trans Pattern Anal Mach Intell* 32:569–575
- Rodríguez, J. H., Weller, S. B., Wannaz, E. D., Klumpp, A., & Pignata, M. L. (2011). Air quality biomonitoring in agricultural areas nearby to urban and industrial emission sources in Córdoba province, Argentina, employing the bioindicator *Tillandsia capillaris*. *Ecological indicators*, 11(6), 1673-1680.
- Rückerl R., A. Schneider, S. Breitner, J. Cyrys, A. Peters (2011). Health effects of particulate air pollution: a review of epidemiological evidence. *Inhalation Toxicology*, 23(10):555–592.
- Ruiz, S., Fernández-Olmo, I., & Irabien, Á. (2014). Discussion on graphical methods to identify point sources from wind and particulate matter-bound metal data. *Urban Climate*, 10, 671-681.
- Russell, N., Ter Hofstede, A. H., Edmond, D., & van der Aalst, W. M. (2005, October). Workflow data patterns: Identification, representation and tool support. In *International Conference on Conceptual Modeling* (pp. 353-368). Springer, Berlin, Heidelberg.
- Russo, A., Lind, P. G., Raischel, F., Trigo, R., & Mendes, M. (2015). Neural network forecast of daily pollution concentration using optimal meteorological data at synoptic and local scales. *Atmospheric Pollution Research*, 6(3), 540-549.

## S

- Sadat, Y. K., Nikaein, T., & Karimipour, F. (2015). Fuzzy spatial association rule mining to analyze the effect of environmental variables on the risk of allergic asthma prevalence. *Geodesy and Cartography*, 41(2), 101-112.
- Saisana, M., Dubois, G., Chaloulakou, A., & Spyrellis, N. (2004). Classification criteria and probability risk maps: Limitations and perspectives.

- Salako, G. O., & Hopke, P. K. (2012). Impact of percentile computation method on PM 24-h air quality standard. *Journal of environmental management*, 107, 110-113.
- Sammarco, M., Tse, R., Pau, G., & Marfia, G. (2017). Using geosocial search for urban air pollution monitoring. *Pervasive and Mobile Computing*, 35, 15-31.
- Samper, F.J. & J. Carrera 1990. *Geoestadística. Aplicaciones a la Hidrogeología Subterránea*. Centro Internacional de Métodos Numéricos en Ingeniería. Universitat Politècnica de Catalunya. Barcelona.
- Schikowski, T., Krämer, U., Kuhlbusch, T. A., & Hoffmann, B. (2017). Land use regression modeling of oxidative potential of fine particles, NO<sub>2</sub>, PM<sub>2.5</sub> mass and association to type two diabetes mellitus. *Atmospheric Environment*, 171, 181-190.
- Scire, J. S., Strimaitis, D. G., & Yamartino, R. J. (2000). A user's guide for the CALPUFF dispersion model. *Earth Tech, Inc. Concord, MA*, 10.
- Seinfeld, J. H., & Pandis, S. N. (2016). *Atmospheric chemistry and physics: from air pollution to climate change*. John Wiley & Sons.
- Sekar, C., Gurjar, B. R., Ojha, C. S. P., & Goyal, M. K. (2015). Potential assessment of neural network and decision tree algorithms for forecasting ambient PM 2.5 and CO concentrations: Case study. *Journal of Hazardous, Toxic, and Radioactive Waste*, 20(4), A5015001.
- Sen, Pranab Kumar. "Estimates of the regression coefficient based on Kendall's tau." *Journal of the American Statistical Association* 63.324 (1968): 1379-1389.
- Shah, A. S., Langrish, J. P., Nair, H., McAllister, D. A., Hunter, A. L., Donaldson, K., ... & Mills, N. L. (2013). Global association of air pollution and heart failure: a systematic review and meta-analysis. *The Lancet*, 382(9897), 1039-1048
- Shah, T., Rabhi, F., & Ray, P. (2015). Investigating an ontology-based approach for Big Data analysis of inter-dependent medical and oral health conditions. *Cluster Computing*, 18(1), 351-367
- Shapiro, S. S., & Francia, R. S. (1972). An approximate analysis of variance test for normality. *Journal of the American Statistical Association*, 67(337), 215-216.
- Shi, J. P., & Harrison, R. M. (1997). Regression modelling of hourly NO<sub>x</sub> and NO<sub>2</sub> concentrations in urban air in London. *Atmospheric Environment*, 31(24), 4081-4094.
- Shimodaira, H. (2004). Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling. *The Annals of Statistics*, 32(6), 2616-2641.
- Shtein, A., Karnieli, A., Katra, I., Raz, R., Levy, I., Lyapustin, A., ... & Kloog, I. (2018). Estimating daily and intra-daily PM<sub>10</sub> and PM<sub>2.5</sub> in Israel using a spatio-temporal hybrid modeling approach. *Atmospheric Environment*, 191, 142-152.
- Silva, C., & Quiroz, A. (2003). Optimization of the atmospheric pollution monitoring network at Santiago de Chile. *Atmospheric Environment*, 37(17), 2337-2345.
- Simonds, M., Xiao, H., & Levine, S. P. (1994). Optical remote sensing for air pollutants-review. *American Industrial Hygiene Association Journal*, 55(10), 953-965.
- Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263-286.
- Skamarock, W. C., & Klemp, J. B. (2008). A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *Journal of computational physics*, 227(7), 3465-3485.
- Smith, M., Goodchild, M., & Longley, P (2015). *Univariate classification schemes en Geospatial Analysis*.



- Smithline, H. A., Ward, K. R., Chiulli, D. A., Blake, H. C., & Rivers, E. P. (2003). Whole body oxygen consumption and critical oxygen delivery in response to prolonged and severe carbon monoxide poisoning. *Resuscitation*, *56*(1), 97-104.
- Soares, A., & Pereira, M. J. (2007). Space–time modelling of air quality for environmental-risk maps: A case study in South Portugal. *Computers & Geosciences*, *33*(10), 1327-1336.
- Solecki, W., Seto, K. C., & Marcotullio, P. J. (2013). It's time for an urbanization science. *Environment: science and policy for sustainable development*, *55*(1), 12-17.
- Soler, Y. (2017). Teorías sobre los sistemas complejos. *Administración & Desarrollo*, *47*(2), 52-69.
- Sosa, B. S., Porta, A., Lerner, J. E. C., Noriega, R. B., & Massolo, L. (2017). Human health risk due to variations in PM 10-PM 2.5 and associated PAHs levels. *Atmospheric Environment*, *160*, 27-35.
- Sørensen, M. P. (2018). Ulrich Beck: exploring and contesting risk. *Journal of Risk Research*, *21*(1), 6-16.
- Sprenst, P., & Smeeton, N. C. (2016). *Applied nonparametric statistical methods*. CRC Press.
- Stafoggia, M., Schwartz, J., Badaloni, C., Bellander, T., Alessandrini, E., Cattani, G., ... & Sorek-Hamer, M. (2017). Estimation of daily PM 10 concentrations in Italy (2006–2012) using finely resolved satellite data, land use variables and meteorology. *Environment international*, *99*, 234-244.
- Stafoggia, M., Bellander, T., Bucci, S., Davoli, M., de Hoogh, K., De'Donato, F., ... & Scortichini, M. (2019). Estimation of daily PM10 and PM2.5 concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model. *Environment international*, *124*, 170-179.
- Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J., Cohen, M. D., & Ngan, F. (2015). NOAA's HYSPLIT atmospheric transport and dispersion modeling system. *Bulletin of the American Meteorological Society*, *96*(12), 2059-2077.
- Stern, N. (2007). *The Economics of Climate Change: The Stern Review*. Cambridge, UK: Cambridge University Press.
- Stieb, D. M., Burnett, R. T., Smith-Doiron, M., Brion, O., Shin, H. H., & Economou, V. (2008). A new multipollutant, no-threshold air quality health index based on short-term associations observed in daily time-series analyses. *Journal of the Air & Waste Management Association*, *58*(3), 435-450.
- Stocker, T. F., Qin, D., Plattner, G. K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., and Bex, V.: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Computational Geometry, 18, 95–123, 2013.
- Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the ACM*, *40*(5), 103-110.
- Sugimoto, N., Nishizawa, T., Shimizu, A., Matsui, I., & Jin, Y. (2014, November). Characterization of aerosols in East Asia with the Asian dust and aerosol lidar observation network (AD-Net). In *Lidar Remote Sensing for Environmental Monitoring XIV* (Vol. 9262, p. 92620K). International Society for Optics and Photonics.
- Sulemana, I. (2012). Assessing Over-aged Car Legislation as an Environmental Policy Law in Ghana. *International Journal of Business and Social Science*, *3*(20).
- Sun, W., Zhang, H., Palazoglu, A., Singh, A., Zhang, W., & Liu, S. (2013). Prediction of 24-hour-average PM2.5 concentrations using a hidden Markov model with different emission distributions in Northern California. *Science of the total environment*, *443*, 93-103.

## T

- Tai, A. P., Mickley, L. J., Jacob, D. J., Leibensperger, E. M., Zhang, L., Fisher, J. A., & Pye, H. O. T. (2012). Meteorological modes of variability for fine particulate matter (PM 2.5) air quality in the United States: implications for PM 2.5 sensitivity to climate change. *Atmospheric Chemistry and Physics*, *12*(6), 3131-3145.

- Tansley, S., & Tolle, K. M. (2009). *The fourth paradigm: data-intensive scientific discovery* (Vol. 1). A. J. Hey (Ed.). Redmond, WA: Microsoft research.
- Teillet, P. M., & Fedosejevs, G. (1995). On the dark target approach to atmospheric correction of remotely sensed data. *Canadian Journal of Remote Sensing*, 21(4), 374-387.
- Tenías, J. M., Ballester, F., & Rivera, M. L. (1998). Association between hospital emergency visits for asthma and air pollution in Valencia, Spain. *Occupational and environmental medicine*, 55(8), 541-547.
- Thurston, G.D., Spengler, J.D., 1985. A quantitative assessment of source contributions to inhalable particulate matter pollution in metropolitan Boston. *Atmospheric Environment* 19, 9–25
- Tian, H., Liu, K., Hao, J., Wang, Y., Gao, J., Qiu, P. and Zhu, C. (2013). Nitrogen oxides emissions from thermal power plants in China: Current status and future predictions. *Environ. Sci. Technol.* 47: 11350–11357.
- Tombesi, N., Pozo, K., & Harner, T. (2014). Persistent organic pollutants (POPs) in the atmosphere of agricultural and urban areas in the province of Buenos Aires in Argentina using PUF disk passive air samplers. *Atmospheric Pollution Research*, 5(2), 170-178.
- Torres, O.; Bhartia, P.K.; Herman, J.R.; Sinyuk, A.; Ginoux, P.; Holben, B (2002). A long-term record of aerosol optical depth from TOMS observations and comparison to AERONET measurements. *J. Atmos. Sci.*, 59, 398–413.
- Torres, O., Tanskanen, A., Veihelmann, B., Ahn, C., Braak, R., Bhartia, P. K., ... & Levelt, P. (2007). Aerosols and surface UV products from Ozone Monitoring Instrument observations: An overview. *Journal of Geophysical Research: Atmospheres*, 112(D24).

## U

- Ulke, A. G., & Mazzeo, N. A. (1998). Climatological aspects of the daytime mixing height in Buenos Aires city, Argentina. *Atmospheric Environment*, 32(9), 1615-1622.
- Ulke, A. G., Brizuela, M. M. T., Raga, G. B., & Baumgardner, D. (2016). Aerosol properties and meteorological conditions in the city of Buenos Aires, Argentina, during the resuspension of volcanic ash from the Puyehue-Cordón Caulle eruption. *Natural Hazards & Earth System Sciences*, 16(9).

## V

- Vahlsing, C., & Smith, K. R. (2012). Global review of national ambient air quality standards for PM 10 and SO 2 (24 h). *Air Quality, Atmosphere & Health*, 5(4), 393-399.
- van Donkelaar, A., Martin, R. V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., & Villeneuve, P. J. (2010). Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application. *Environmental health perspectives*, 118(6), 847-855.
- Van den Elshout, S., Léger, K., & Heich, H. (2014). CAQI common air quality index—Update with PM2. 5 and sensitivity analysis. *Science of the Total Environment*, 488, 461-468.
- Vardoulakis, S., Valiantis, M., Milner, J., & ApSimon, H. (2007). Operational air pollution modelling in the UK—Street canyon applications and challenges. *Atmospheric Environment*, 41(22), 4622-4637.
- Vasconcellos PC., DZ. Souza, SG. Ávila, MP. Araújo, E. Naoto, KH. Nascimento, FS. Cavalcante, M. Dos Santos, P.Smichowski, E.Behrentz (2011). Comparative study of the atmospheric chemical composition of three South American cities. *Atmospheric Environment*, 45: 5770-5777.

- Venegas, L. E., & Mazzeo, N. A. (1999). Atmospheric stagnation, recirculation and ventilation potential of several sites in Argentina. *Atmospheric Research*, 52(1), 43-57.
- Venegas, L. E., & Mazzeo, N. A. (2006). Modelling of urban background pollution in Buenos Aires City (Argentina). *Environmental Modelling & Software*, 21(4), 577-586.
- Voorhees, A. S. (2005). Benefits analysis of particulate matter control programs—a case study of Tokyo. *Journal of Risk Research*, 8(4), 311-329.
- Voukantsis, D., Karatzas, K., Kukkonen, J., Räsänen, T., Karppinen, A., & Kolehmainen, M. (2011). Intercomparison of air quality data using principal component analysis, and forecasting of PM<sub>10</sub> and PM<sub>2.5</sub> concentrations using artificial neural networks, in Thessaloniki and Helsinki. *Science of the Total Environment*, 409(7), 1266-1276.

## W

- Wackernagel, H. 1995. *Multivariate Geostatistics. An Introduction with Applications*. Springer-Verlag, Berlin
- Wang, J., Zhang, X., Guo, Z., & Lu, H. (2017b). Developing an early-warning system for air quality prediction and assessment of cities in China. *Expert Systems with Applications*, 84, 102-116.
- Wang, J., & Song, G. (2018). A deep spatial-temporal ensemble model for air quality prediction. *Neurocomputing*, 314, 198-206.
- Wang, D., Wei, S., Luo, H., Yue, C., & Grunder, O. (2017 a). A novel hybrid model for air quality index forecasting based on two-phase decomposition technique and modified extreme learning machine. *Science of The Total Environment*, 580, 719-733.
- Wang, L., Zhong, B., Vardoulakis, S., Zhang, F., Pilot, E., Li, Y., ... & Krafft, T. (2016). Air quality strategies on public health and health equity in Europe—a systematic review. *International journal of environmental research and public health*, 13(12), 1196.
- Wang, Z., King, E., Smith, G., Bellgard, M., Broomhall, M., Chedzey, H., ... & Schibeci, D. (2011). RS-YABI: A workflow system for Remote Sensing Processing in AusCover.
- Watson, A., Bates, R., and Kennedy, D. (1988). *Air Pollution, the Automobile, and Public Health*. National Academies Press
- Watson, J.G., Chow, J.C., Fujita, E.M., 2001. Review of volatile organic compounds source apportionment by chemical mass balance. *Atmospheric Environment* 35, 1567–1584
- Welton, E. J., Campbell, J. R., Berkoff, T. A., Valencia, S., Spinhirne, J. D., Holben, B., ... & Schmid, B. (2006). The NASA Micro-Pulse Lidar Network (MPLNET): an overview and recent results. *Opt. Pur. Apl*, 39, 67-74.
- Wentz, E., Anderson, S., Fragkias, M., Netzband, M., Mesev, V., Myint, S., ... & Seto, K. (2014). Supporting global environmental change research: A review of trends and knowledge gaps in urban remote sensing. *Remote Sensing*, 6(5), 3879-3905.
- Wichmann FA., LE. Busi, NF. Cianni, L. Massolo, A. Müller, A. Porta, PD. Sly (2009). “Increased asthma and respiratory symptoms in children exposed to petrochemical pollution”. *Journal of Allergy and Clinical Immunology*, 12 (3): 632-638.
- Wickham, H (2007). Reshaping Data with the reshape Package. *Journal of Statistical Software*, 21(12), 1-20. URL <http://www.jstatsoft.org/v21/i12/>.
- Wickham H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wickham, H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

- Wickham H, R François, L Henry and K Müller (2019). dplyr: A Grammar of Data Manipulation. R package version 0.8.0.1. <https://CRAN.R-project.org/package=dplyr>
- Winiwarter, W., & Klimont, Z. (2011). The role of N-gases (N<sub>2</sub>O, NO<sub>x</sub>, NH<sub>3</sub>) in cost-effective strategies to reduce greenhouse gas emissions and air pollution in Europe. *Current Opinion in Environmental Sustainability*, 3(5), 438-445.
- Wofsy, S.C., Afshar, S., Allen, H.M., APEL, E., Asher, E.C., Barletta, B., Bent, J., Bian, H., Biggs, B.C., Blake, D.R. and Blake, N., 2018. ATom: Merged atmospheric chemistry, trace gases, and aerosols. *ORNL DAAC*.
- World Air Quality Index project. (2015, 16 mayo). World Health Organization: 2014 Air Pollution Ranking. Recuperado 22 noviembre, 2019, de <https://aqicn.org/faq/2015-05-16/world-health-organization-2014-air-pollution-ranking/>
- World Health Organization (1999), Environmental Health Criteria, No. 213: Carbon Monoxide (second edition), ISBN 92 4 157213 2 <http://www.inchem.org/documents/ehc/ehc/ehc213.htm> Checked July 2018
- World Health Organization (2000) Air quality guidelines for Europe; second edition Copenhagen, WHO Regional Office for Europe, 2000 (WHO regional publications. European series; No 91), (full background material available on [http://www.euro.who.int/air/activities/20050223\\_4](http://www.euro.who.int/air/activities/20050223_4))
- World Health Organization, 2005. Air Quality Guidelines for Particulate Matter, Ozone, Nitrogen Dioxide and Sulfur Dioxide. World Health Organization, Geneva, Switzerland.
- World Health Organization. (2006). *Air quality guidelines: global update 2005: particulate matter, ozone, nitrogen dioxide, and sulfur dioxide*. World Health Organization.
- World Health Organization (2013). Review of evidence on health aspects of air pollution – REVIHAAP Project. WHO European Centre for Environment and Health, Bonn, WHO 2013. 309 pp.
- World Health Organization. Global Health Observatory (GHO) data: urban population growth. Geneva, Switzerland: WHO, 2015. [http://www.who.int/gho/urban\\_health/situation\\_trends/urban\\_population\\_growth\\_text/en/](http://www.who.int/gho/urban_health/situation_trends/urban_population_growth_text/en/) Accessed March 2015.
- World Health Organization (2016 a). Global report on urban health: equitable, healthier cities for sustainable development.
- World Health Organization (2016 b). Health as the pulse of the new urban agenda: United Nations conference on housing and sustainable urban development, Quito, October 2016.
- World Health Organization Ambient (Outdoor) Air Quality and Health, 2018. Available online: [https://www.who.int/newsroom/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/newsroom/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health) (accessed on 11 February 2019).
- World Health Organization (2019). Ambient Air Pollution Database. Geneva, Switzerland: WHO, 2019. [http://www.who.int/phe/health\\_topics/outdoorair/databases/cities/en/](http://www.who.int/phe/health_topics/outdoorair/databases/cities/en/) Accessed July 2019.

## X

- Xie, Y., Zhao, L., Xue, J., Gao, H.O., Li, H., Jiang, R., Qiu, X. and Zhang, S., 2018. Methods for defining the scopes and priorities for joint prevention and control of air pollution regions based on data-mining technologies. *Journal of cleaner production*, 185, pp.912-921.

## Y

- Yang, Q. & Wu, X. 2006. 10 challenging problems in data mining research. *International Journal of Information Technology and Decision Making* 5(4), 597–604.
- Yang, F., Tan, J., Zhao, Q., Du, Z., He, K., Ma, Y., Duan, F. and Chen, G.J.A.C., 2011. Characteristics of PM 2.5 speciation in representative megacities and across China. *Atmospheric Chemistry and Physics*, 11(11), pp.5207-5219.
- Yap, X.Q., Hashim, M., 2013. A robust calibration approach for PM10 prediction from MODIS aerosol optical depth. *Atmos. Chem. Phys.* 13, 3517e3526.
- Yeganeh, B., Hewson, M. G., Clifford, S., Knibbs, L. D., & Morawska, L. (2017). A satellite-based model for estimating PM2.5 concentration in a sparsely populated environment using soft computing techniques. *Environmental Modelling & Software*, 88, 84-92.
- Yeh, S., & Small, M. J. (2002). Incorporating exposure models in probabilistic assessment of the risks of premature mortality from particulate matter. *Journal of Exposure Science and Environmental Epidemiology*, 12(6), 389.
- Yidana, S.M., Ophori, D., Banoeng-Yakubo, B., 2008. A multivariate statistical analysis of surface water chemistry data at the Ankobra Basin, Ghana. *Journal of Environmental Management* 86, 80–87.
- Young, M. T., Bechle, M. J., Sampson, P. D., Szpiro, A. A., Marshall, J. D., Sheppard, L., & Kaufman, J. D. (2016). Satellite-based NO<sub>2</sub> and model validation in a national prediction model based on universal Kriging and land-use regression. *Environmental science & technology*, 50(7), 3686-3694.

## Z

- Zanobetti, A., & Schwartz, J. (2009). The effect of fine and coarse particulate air pollution on mortality: a national analysis. *Environmental health perspectives*, 117(6), 898.
- Zhang, C., Ni, Z., & Ni, L. (2015). Multifractal detrended cross-correlation analysis between PM2.5 and meteorological factors. *Physica A: Statistical Mechanics and its Applications*, 438, 114-123.
- Zhang, N. N., Ma, F., Qin, C. B., & Li, Y. F. (2018). Spatiotemporal trends in PM2.5 levels from 2013 to 2017 and regional demarcations for joint prevention and control of atmospheric pollution in China. *Chemosphere*, 210, 1176-1184.
- Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., & Baklanov, A. (2012). Real-time air quality forecasting, part I: History, techniques, and current status. *Atmospheric Environment*, 60, 632-655.
- Zhao, C., Liu, Z., Wang, Q., Ban, J., Chen, N. X., & Li, T. (2019). High-resolution daily AOD estimated to full coverage using the random forest model approach in the Beijing-Tianjin-Hebei region. *Atmospheric Environment*, 203, 70-78.
- Zhou, S., Deng, Q., & Liu, W. (2012). Extreme air pollution events: Modeling and prediction. *Journal of Central South University*, 19, 1668–1672.
- Zeng, X., Pielke, R. A., & Eykholt, R. (1993). Chaos theory and its applications to the atmosphere. *Bulletin of the American Meteorological Society*, 74(4), 631-644.
- Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009 a). *Mixed effects models and extensions in ecology with R*. Springer Science & Business Media.
- Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2009 b). Mixed effects modelling for nested data. In *Mixed effects models and extensions in ecology with R* (pp. 101-142). Springer, New York, NY.