

DESARROLLO DE MODELOS PREDICTIVOS
PARA LA CONCENTRACIÓN DE MATERIAL
PARTICULADO FINO EN ATMOSFERAS
URBANAS

por

Rodriguez Nuñez Martin

Presentado ante la

Facultad de Ciencias Exactas, Fisicas y Naturales

para obtener el grado de

Doctor en Ciencias de la Ingeniería

de la

Universidad Nacional de Córdoba

Octubre 2024

Directora: Hebe Alejandra Carreras



Esta obra está bajo una Licencia Creative Commons Atribución-CompartirIgual 4.0
Internacional.

Agradecimientos

Llega el tan ansiado día de defender mi tesis doctoral. Luego de muchos años de investigación, hay miles de agradecimientos que mencionar. En primer lugar, me siento profundamente afortunado y agradecido de haber nacido en la República Argentina. Le agradezco a mi país no solo por brindarme una educación gratuita de excelencia, sino además por la beca que me permitió dedicarme exclusivamente a mi investigación durante estos años de estudios doctorales.

En segundo lugar, el mayor agradecimiento lo debo tanto a la Dra. Mónica Balzarini como a la Dra. Hebe A. Carreras, mis directoras, quienes me iniciaron, guiaron y formaron en el ámbito de la investigación científica y los estudios de posgrado. Sus aportes fueron fundamentales, no solo para mi investigación, sino también para mi desarrollo personal. Asimismo, agradezco a mis compañeros de la cátedra de Química General y de la cátedra de Estadística y Biometría, quienes con sus comentarios y sugerencias resultaron ser un pilar fundamental para mi trabajo, además de brindarme grandes momentos de amistad, mates, cerveceadas y salidas.

Este trabajo no hubiera sido posible sin momentos de esparcimiento. Aquí quiero agradecer principalmente a mi familia: Ester, má, y Jose, pá, quienes siempre me apoyaron en lo que me hacía feliz, acompañándome en todo momento y brindándome un hogar cómodo para estudiar y trabajar. A mi hermano Juan Manuel, quien, sin dedicarse a la investigación, siempre me apoyó, acompañándome de cerca y preocupándose por mi salud mental. Y en especial a mi hermano Juan Cruz, cuya notable trayectoria en investigación y ciencia de datos siempre me incentivó a seguir

capacitándome y superándome. A mis abuelitos, que en paz descansen, y a todos mis familiares que confiaron en mí y me apoyaron. Agradezco a cada uno que me enseñó el camino de la perseverancia y la resiliencia, valores que inconscientemente estuvieron en mi mente a lo largo de estos años.

También quiero agradecer a mis grupos de amigos: la sub 19, los Outliers, los del cole y los de la vida. Aunque son grupos bien distintos, siempre estuvieron allí, brindándome toneladas de cariño y ayudándome a solucionar problemas o relajarme cuando era necesario. Con mates, juntadas y giras, hicieron este trayecto más ameno. No los nombraré uno por uno porque no terminaría más, pero cada uno sabe cuánto le agradezco. Además, a todas esas personas que la vida acerca y nos impulsan a ser mejores, aunque sea de forma efímera o duradera: gracias por ser esos “angelitos del universo”.

Gracias, gracias a todos. Les dedico este trabajo y esta etapa de mi vida.

Resumen

La Organización Mundial de la Salud (OMS) estima que nueve de cada diez personas respiran aire con altos niveles de contaminantes, lo que provoca la muerte de siete millones de personas cada año en todo el mundo. Entre estos contaminantes, el material particulado fino ($PM_{2.5}$), con un diámetro igual o inferior a 2.5 micrómetros, es particularmente nocivo por su capacidad para penetrar la barrera pulmonar y acceder al torrente sanguíneo. En países en vías de desarrollo, el monitoreo de la concentración de $PM_{2.5}$ mediante estaciones de referencia está limitado por sus elevados costos de adquisición y mantenimiento, lo que subraya la necesidad de métodos alternativos para monitorear la calidad del aire. Esta tesis aborda el desarrollo de un sistema de alerta de bajo costo ante condiciones adversas de calidad del aire, evaluando la capacidad predictiva de variables derivadas de información satelital de público acceso y algoritmos de aprendizaje computacional para pronosticar concentraciones de $PM_{2.5}$ en el área metropolitana de la provincia de Córdoba, Argentina. Además, se busca describir la variación espacio-temporal de este contaminante y analizar la contribución de variables derivadas de información satelital en sus patrones de variación. Para ello, se diseñaron sistemas de medición basados en sensores de bajo costo, se validaron contra una estación de referencia, y se recopilaron series temporales horarias durante un año en diez sitios con distintos usos de suelo. Además, se descargaron datos de 164 variables climáticas de la plataforma *ERA5* y de los satélites *MODIS*, *MERRA-2* y *VIIRS*. Se ajustaron diversos algoritmos de aprendizaje computacional, empleando distintas versiones de la base de datos reducidas mediante métodos de selección de

características (*featurewiz* y *Boruta*). Los sensores demostraron alta consistencia en sus mediciones. Se observó que las condiciones adversas de calidad del aire ocurrieron principalmente durante la noche y las primeras horas de la mañana en invierno. Aunque los niveles promedio de PM_{2.5} en Córdoba fueron menores en comparación con otras ciudades, se registraron aumentos estacionales significativos. Los modelos predictivos presentaron errores de 3.44, 2.60 y 6.17 $\frac{\mu\text{g}}{\text{m}^3}$ para los usos de suelo urbano, agrícola/área verde e industrial, respectivamente. En áreas agrícolas/verdes, los bosques aleatorios fueron los más precisos, mientras que las redes neuronales recurrentes destacaron en entornos urbanos e industriales. Las predicciones a cinco días tuvieron errores inferiores al 23% en todos los casos. Estos resultados indican que la integración de datos de concentración de PM_{2.5} provenientes de sistemas basados en sensores de bajo costo, datos satelitales de acceso público con algoritmos de aprendizaje automático/profundo constituye una herramienta valiosa para predecir la calidad del aire.

Palabras claves: contaminación atmosférica, monitoreo de calidad del aire, variables derivadas a partir de información satelital, sensores de bajo costo, aprendizaje automático, aprendizaje profundo.

The World Health Organization (WHO) estimates that nine out of ten people breathe air with high levels of pollutants, causing the death of seven million people worldwide each year. Among these pollutants, fine particulate matter (PM_{2.5}), with a diameter equal to or less than 2.5 micrometers, is particularly harmful due to its ability to penetrate the pulmonary barrier and enter the bloodstream. In developing countries, monitoring PM_{2.5} concentration using reference stations is limited by high acquisition and maintenance costs, highlighting the need for alternative methods to monitor air quality. This thesis addresses the development of a low-cost alert system for adverse air quality conditions, evaluating the predictive capacity of variables derived from publicly accessible satellite information and computational learning algorithms to forecast PM_{2.5} concentrations in the metropolitan area of the province

of Córdoba, Argentina. Additionally, it aims to describe the spatiotemporal variation of this pollutant and analyze the contribution of variables derived from satellite information to its variation patterns. For this purpose, measurement systems based on low-cost sensors were designed, validated against a reference station, and hourly time series data were collected over one year at ten sites with different land uses. Moreover, data from 164 climatic variables from the *ERA5* platform and the satellites *MODIS*, *MERRA-2*, and *VIIRS* were downloaded. Several computational learning algorithms were fitted, using different versions of the dataset reduced through feature selection methods (*featurewiz* and *Boruta*). The sensors demonstrated high consistency in their measurements. It was observed that adverse air quality conditions occurred mainly during the night and early morning in winter. Although average PM_{2.5} levels in Córdoba were lower compared to other cities, significant seasonal increases were recorded. Predictive models presented errors of 3.44, 2.60, and 6.17 $\frac{\mu\text{g}}{\text{m}^3}$ for urban, agricultural/green area, and industrial land uses, respectively. In agricultural/green areas, random forests were the most accurate, while recurrent neural networks stood out in urban and industrial environments. Five-day forecasts showed errors below 23 % in all cases. These results indicate that integrating PM_{2.5} concentration data from low-cost sensor systems, publicly accessible satellite data, and machine learning/deep learning algorithms constitutes a valuable tool for predicting air quality.

Keywords: air pollution, air quality monitoring, satellite-derived variables, low-cost sensors, machine learning, deep learning.

A Organizaçāo Mundial da Saúde (OMS) estima que nove em cada dez pessoas respiram ar com altos níveis de poluentes, causando a morte de sete milhões de pessoas em todo o mundo a cada ano. Entre esses poluentes, o material particulado fino (PM_{2.5}), com um diâmetro igual ou inferior a 2,5 micrômetros, é particularmente nocivo devido à sua capacidade de penetrar a barreira pulmonar e acessar a corrente sanguínea. Em países em desenvolvimento, o monitoramento da concentração de PM_{2.5} por meio

de estações de referência é limitado pelos altos custos de aquisição e manutenção, destacando a necessidade de métodos alternativos para monitorar a qualidade do ar. Esta tese aborda o desenvolvimento de um sistema de alerta de baixo custo para condições adversas de qualidade do ar, avaliando a capacidade preditiva de variáveis derivadas de informações de satélite de acesso público e algoritmos de aprendizado computacional para prever concentrações de PM_{2.5} na área metropolitana da província de Córdoba, Argentina. Além disso, busca descrever a variação espaço-temporal desse poluente e analisar a contribuição de variáveis derivadas de informações de satélite em seus padrões de variação. Para isso, foram desenvolvidos sistemas de medição baseados em sensores de baixo custo, validados contra uma estação de referência, e séries temporais horárias foram coletadas durante um ano em dez locais com diferentes usos do solo. Além disso, foram baixados dados de 164 variáveis climáticas da plataforma *ERA5* e dos satélites *MODIS*, *MERRA-2* e *VIIRS*. Diversos algoritmos de aprendizado computacional foram ajustados, utilizando diferentes versões do banco de dados reduzidas por meio de métodos de seleção de características (*featurewiz* e *Boruta*). Os sensores demonstraram alta consistência em suas medições. Observou-se que condições adversas de qualidade do ar ocorreram principalmente durante a noite e no início da manhã no inverno. Embora os níveis médios de PM_{2.5} em Córdoba fossem inferiores em comparação com outras cidades, aumentos sazonais significativos foram registrados. Os modelos preditivos apresentaram erros de 3,44, 2,60 e 6,17 $\frac{\mu g}{m^3}$ para os usos do solo urbano, agrícola/área verde e industrial, respectivamente. Em áreas agrícolas/verdes, as florestas aleatórias foram as mais precisas, enquanto as redes neurais recorrentes se destacaram em ambientes urbanos e industriais. As previsões de cinco dias apresentaram erros abaixo de 23 % em todos os casos. Esses resultados indicam que a integração de dados de concentração de PM_{2.5} provenientes de sistemas baseados em sensores de baixo custo, dados de satélite de acesso público e algoritmos de aprendizado de máquina/aprendizado profundo constitui uma ferramenta valiosa para prever a qualidade do ar.

Palavras-chave: poluição do ar, monitoramento da qualidade do ar, variáveis derivadas de informações de satélite, sensores de baixo custo, aprendizado de máquina, aprendizado profundo.

Índice general

Capítulo 1: Introducción General	1
1.1. Problema de investigación	1
1.2. Motivación	6
1.2.1. Objetivo general	8
1.2.2. Objetivos específicos	9
1.3. Desafíos metodológicos	9
Capítulo 2: Sistema de monitoreo basado en sensores de bajo costo .	13
2.1. Introducción	13
2.2. Materiales y métodos	15
2.2.1. Resumen gráfico del protocolo analítico de desarrollo y validación de sensores de monitoreo basados en tecnologías de bajo costo	15
2.2.2. Desarrollo del sistema de monitoreo basado en sensores de bajo costo	16
2.2.3. Base de datos	19
2.2.4. Protocolo estadístico	20
2.3. Resultados y discusión	22
2.3.1. Producto final	22
2.3.2. Consenso de medición de la concentración de PM _{2.5} entre los sistemas de monitoreo de bajo costo	24

2.3.3. Validación con respecto a la estación de referencia	29
2.3.4. Limitaciones del sensor de bajo costo	32
2.4. Conclusiones	32
Capítulo 3: Variación espacio temporal de la concentración de aerosoles	34
3.1. Introducción	34
3.2. Materiales y métodos	37
3.2.1. Resumen gráfico del protocolo analítico de caracterización de la variación espacio temporal de la concentración de PM _{2.5}	37
3.2.2. Área de estudio	38
3.2.3. Descripción de cada uno de los sitios de muestreo	39
3.2.4. Normativa	47
3.2.5. Preprocesamiento de datos	48
3.2.6. Protocolo estadístico	50
3.3. Resultados y discusión	52
3.3.1. Características temporales de la serie de PM _{2.5}	53
3.3.2. Análisis espacial de la concentración de PM _{2.5}	61
3.3.3. Análisis espacio-temporal por uso del suelo de PM _{2.5}	72
3.4. Conclusiones	83
Capítulo 4: Influencia de variables meteorológicas obtenidas de satélites	85
4.1. Introducción	85
4.2. Materiales y métodos	90
4.2.1. Resumen gráfico del protocolo de análisis de correlaciones entre series	90
4.2.2. Base de datos	91
4.2.3. Protocolo estadístico	106
4.3. Resultados	111

4.4. Discusión	139
4.5. Conclusiones	148
Capítulo 5: Modelado predictivo de la concentración de PM_{2.5}	150
5.1. Introducción	150
5.2. Materiales y métodos	154
5.2.1. Resumen gráfico del protocolo analítico de modelado predictivo	154
5.2.2. Base de datos	157
5.2.3. Protocolo estadístico	160
5.3. Resultados	175
5.3.1. Modelado de PM _{2.5} en entornos urbanos	175
5.3.2. Modelado de PM _{2.5} en entornos industriales	184
5.3.3. Modelado de PM _{2.5} en entornos agrícola y área verde	192
5.3.4. Evaluación de tiempos de cómputo	198
5.4. Discusión	200
5.5. Conclusiones	210
Comentarios finales	212
5.6. Repuesta a los objetivos	216
5.6.1. Objetivo general	216
5.6.2. Objetivos específicos	217
5.6.3. Conclusión Final	218
Futuras líneas de investigación	220
Anexo 1: Bases de datos	222
5.6.4. Base de datos original	222
5.6.5. Bases de datos reducidas	226
Anexo 2: Procedimiento de análisis de series temporales	231

Anexo 3: Descripción de variables	310
Anexo 4: Optimización de los modelos predictivos	416
Anexo 5: Variables no citadas en la literatura como relacionadas a la concentración de PM_{2.5}	422
Referencias bibliográficas	436

Capítulo 1

Introducción General

1.1. Problema de investigación

El crecimiento desmedido de la población y la ambición de la sociedad por mejorar su calidad de vida, han resultado en un creciente deterioro ambiental. La atmósfera no es ajena a este deterioro que acaba empeorando la calidad del aire, impactando de forma directa o indirecta en la calidad de vida de las personas.

Se considera contaminante atmosférico a cualquier sustancia emitida al aire desde una fuente biogénica, geogénica o antropogénica, que es ajena a la composición natural del aire, o que se presenta en concentraciones superiores a las naturales, y que puede poner en riesgo a los seres vivos en un corto o largo plazo (Fuzzi et al., 2015; Seinfeld & Pandis, 1998).

Uno de los contaminantes más preocupantes en atmósferas urbanas son los aerosoles atmosféricos, suspensiones relativamente estables de partículas sólidas o líquidas en un gas y que presentan un diámetro menor a $10 \mu m$. Distinto es a lo que refiere el término material particulado (PM), el cual define un conjunto de materia en estado sólido y/o líquido presente en suspensión en la atmósfera, excluyendo el agua pura (Meszaros et al., 1999; Salameh et al., 2015). Los aerosoles engloban tanto las partículas en

suspensión como la masa de aire en la que se encuentran (Finlayson-Pitts & Pitts Jr, 1999; VandenBoer et al., 2013). A pesar de que semánticamente estos términos no son equivalentes, son utilizados indistintamente en el lenguaje coloquial.

Las fuentes de emisión de partículas pueden clasificarse en antropogénicas y naturales, ya que los cambios en la calidad del aire no siempre están ligados a la intervención del ser humano. Se entiende por fuentes antropogénicas aquellas que resultan de la actividad humana, este grupo aglomera procesos industriales, mecánicos y de combustión, construcción y demolición, agricultura y ganadería, entre otros. Por otra parte, las emisiones naturales refieren a partículas generadas como consecuencia de procesos naturales como las erupciones volcánicas, incendios originados a partir de rayos, re suspensión de partículas por el viento, entre otras (Jiménez-Guerrero et al., 2008; Pey Betrán, 2008).

Las propiedades aerodinámicas engloban las características físicas de mayor relevancia para el PM. Estas no solo determinan el tipo de transporte y remoción en la atmósfera, si no que también su deposición en el sistema respiratorio, además de permitir identificar la fuente que les dio origen. El diámetro aerodinámico es la propiedad de mayor relevancia, que refiere al diámetro de una esfera de densidad unitaria ($1 \frac{kg}{cm^3}$) con las mismas características aerodinámicas que la partícula bajo consideración. Según el diámetro las partículas pueden clasificarse en tres grandes grupos: gruesas (PM_{10} : 2.5-10 μm), finas ($PM_{2.5}$: menor a 2.5 μm) y ultrafinas ($PM_{0.1}$: menor a 0.1 μm) (Gopalakrishnan et al., 2015; Kulkarni et al., 2011).

De acuerdo a su mecanismo de formación, las partículas pueden clasificarse en primarias y secundarias. Las partículas primarias son aquellas emitidas de forma directa a la atmósfera y que permanecen en ella de la misma forma en que fueron liberadas. Por otra parte, las partículas secundarias son aquellas que han estado sujetas a cambios químicos o bien son producto de reacciones entre dos o más contaminantes primarios en la atmósfera (Finlayson-Pitts & Pitts Jr, 1999; VandenBoer et al., 2013).

De acuerdo con los informes de la Organización Mundial de la Salud (OMS), el PM es considerado uno de los contaminantes atmosféricos más perjudiciales para la salud de las personas (WHO et al., 2021). Su presencia en la atmósfera impacta significativamente en el desarrollo de la sociedad dado que afecta múltiples componentes con los cuales el ser humano interactúa, además de su salud, se destaca el efecto que tiene en los ecosistemas y en clima.

La vía de ingreso de las partículas al cuerpo humano es a través del sistema respiratorio y su impacto en la salud depende de múltiples factores tales como el tamaño, componentes adsorbidos, composición química, condición sanitaria del individuo expuesto, intensidad y duración de la exposición. A medida que el tamaño de la partícula disminuye su peligrosidad aumenta dada su capacidad de penetrar más profundamente en el sistema respiratorio y su mayor facilidad para ser respiradas. Las partículas gruesas se depositan en la tráquea, bronquios y bronquiolos, mientras que las finas y ultrafinas son capaces de alcanzar los alvéolos pulmonares, donde pueden disolverse o depositarse (K.-H. Kim et al., 2015; Schlesinger, 2007). Además, su capacidad de generar efectos adversos a la salud también se encuentra ligado a los componentes adsorbidos que transporta, entre los más nocivos se destacan metales y microorganismos patógenos (Blanco Becerra, 2006; Nordberg et al., 2014). Por otra parte, se destaca su composición química, la cual determina su capacidad oxidante, así como el carácter cancerígeno y/o mutagénico de la partícula (IARC, 2004, 2010). Finalmente, los efectos de las partículas también varían con la intensidad y duración de la exposición a la que se ven sometidos los individuos, además de su condición sanitaria y edad. La población más susceptible de sufrir efectos adversos a la salud incluye a niños, embarazadas, personas mayores de 65 años y personas con enfermedades cardiovasculares y/o respiratorias (Sacks et al., 2011; Yan Zhao et al., 2020).

El impacto de los aerosoles en los ecosistemas deriva de la deposición en la superficie de las plantas, generando abrasión y calentamiento radiativo, conduciendo a una reduc-

ción de la capacidad fotosintética. Además, si los compuestos depositados son ácidos o alcalinos, pueden causar lesiones en la superficie de la hoja (Rai, 2016; Rai & Panda, 2014). Por otra parte, el PM depositado de forma directa en el suelo puede intervenir en el ciclado de nutrientes en los ecosistemas, afectando principalmente el ciclo del nitrógeno, dado su efecto en las bacterias y hongos de la rizósfera. Adicionalmente, los animales resultan afectados de una forma equivalente a la que es afectada la población humana, variando el impacto en función de la sensibilidad de cada especie.

La influencia de los aerosoles atmosféricos en el clima, deriva de su incidencia en el equilibrio térmico de la Tierra y el cambio climático global, la cual se halla vinculado a múltiples procesos. En primer lugar, resalta la influencia de los aerosoles en el balance radiativo de la Tierra, como consecuencia de la absorción y dispersión de la radiación solar incidente. En segundo lugar, destaca su participación en la formación de nubes y precipitaciones, derivada de su capacidad de actuar como núcleos de condensación. La influencia en la radiación se encuentra ligada a las propiedades fisicoquímicas de las partículas, dando lugar a efectos de enfriamiento y calentamiento de la atmósfera terrestre. Las partículas compuestas por sulfatos se caracterizan por generar un enfriamiento de la atmósfera, mientras que aquellas con una proporción notable de carbono en estado elemental y partículas de polvo tienen un efecto de calentamiento (S.-L. Chen et al., 2021). Por otra parte, el proceso por medio del cual la formación de nubes tiene influencia en el balance radiativo de la Tierra se debe a un aumento en el número de partículas, que determina un incremento del espesor óptico de las nubes, disminuyendo la radiación solar superficial neta incidente. Cabe destacar que las partículas más pequeñas disminuyen la eficiencia del proceso de precipitación, prolongando de este modo la vida media de las nubes (Unkašević et al., 2003). Además, los aerosoles atmosféricos antropogénicos como el hollín actúan afectando el contenido de agua de las nubes causando un forzamiento radiativo negativo (Solomon et al., 2007). Finalmente, dada la naturaleza de los aerosoles y su participación en un gran

número de reacciones secundarias, éstos pueden afectar la concentración y distribución de gases traza atmosféricos, alterando los ciclos de nitrógeno, azufre y oxidantes atmosféricos (Meszaros et al., 1999; Salameh et al., 2015).

El monitoreo de la calidad del aire es una herramienta fundamental a la hora de identificar y evaluar problemas de contaminación (OMS, 2021). No obstante, es preocupante que no exista una equitativa accesibilidad a sistemas de monitoreo en los distintos países del mundo. Esto ha generado una notoria brecha en el acceso a datos de calidad del aire entre países desarrollados y en vías de desarrollo. Esta tesis contribuye al desarrollo y validación de un sistema de monitoreo de bajo costo que permita alertar ante condiciones adversas de calidad de aire, permitiendo mitigar la escasa disponibilidad de datos sobre calidad de aire en países en desarrollo. Así, ciudades de países en vías de desarrollo podrán gestionar y controlar la calidad del aire, salvaguardando la salud pública.

La investigación sobre series temporales para el desarrollo de modelos predictivos en este estudio se centró en el PM_{2.5}, pero podría ser aplicable a cualquier otro contaminante, teniendo en cuenta las salvedades intrínsecas propias de cada uno de ellos. Para cumplir con el objetivo, fue necesario primero implementar un sistema de monitoreo de partículas basado en sensores de bajo costo que permitiese disponer de registros del contaminante a lo largo del tiempo. Una vez obtenidos estos datos, fue necesario considerar variables meteorológicas obtenidas a través de satélites, dada la estrecha relación existente entre la concentración de PM_{2.5} y los fenómenos atmosféricos (Jianjun He et al., 2017; Tiwari et al., 2017). La incorporación de estas variables como predictoras en nuestro sistema de alerta permitirá explicar las fluctuaciones en la concentración de PM_{2.5}, siguiendo la premisa de contar con un sistema de bajo costo económico.

Finalmente, se buscó interpretar la interrelación entre todas las variables predictoras por medio de algoritmos de inteligencia artificial. Estos algoritmos permiten obtener

predicciones futuras de la variable respuesta con un horizonte de predicción, en nuestro caso de cinco días de anticipación. La elección del horizonte de predicción se encuentra fundada en que la disponibilidad de las variables derivadas a partir de información satelital es gratuita solo cinco días previos al momento presente, de esta forma para poder obtener la predicción de la hora siguiente, es necesario emplear como predictoras los datos de la hora anterior con cinco días de antelación. Los algoritmos de inteligencia artificial entrenados captan la variabilidad de la concentración de PM_{2.5} respecto de las variables derivadas a partir de información satelital con cinco días de antelación, a pesar de esto como consecuencia de su disponibilidad, solo permitirá la predicción de la concentración de PM_{2.5} la hora siguiente. Por otra parte, si los datos de variables derivadas a partir de información satelital se obtuviesen en tiempo real, se permitiría obtener predicciones con cinco días de anticipación. Además, podría entrenarse otro algoritmo diferente considerando cualquier ventana de tiempo en función del objetivo deseado. La implementación de estos algoritmos permitirá capturar y explicar la variabilidad del contaminante en cuestión, prescindiendo así de los sistemas de medición en Tierra y posibilitando la realización de predicciones futuras de la concentración de PM_{2.5}. De esta forma, se pretende lograr el desarrollo de un sistema de alerta de bajo costo ante condiciones adversas de calidad del aire, cuya potencialidad deviene de la capacidad de alertar a la población y proteger la salud pública.

1.2. Motivación

Estudios previamente realizados en la ciudad de Córdoba han demostrado que se convive con elevadas concentraciones de PM, especialmente en la zona céntrica, cuya principal fuente de emisión se vincula al transporte (Amarillo et al., 2021; Carreras et al., 2006; Carreras & Pignata, 2001). También se ha demostrado la asociación entre los

niveles de PM y los riesgos de enfermedades respiratorias en la población de la ciudad de Córdoba (Busso et al., 2021; Carreras et al., 2009; Mateos et al., 2018). Además, la relación existente entre los niveles de PM y variables meteorológicas satelitales para el sitio de interés también ha sido foco de investigación (Amarillo et al., 2021; Della Ceca et al., 2018). Sin embargo, ninguno de estos estudios ha analizado la variable respuesta con una frecuencia de variación horaria, ni siguiendo un enfoque de series temporales que permita analizar la variabilidad conjunta de la concentración horaria de PM respecto de otras covariables.

Actualmente existe abundante información respecto de los efectos nocivos del PM en la salud del hombre, su composición, fuentes de emisión, relación con otras covariables y su modelado predictivo. A pesar de esto, la mayoría de los estudios se han realizado en países de Europa y América del Norte. En los últimos años ha ocurrido una creciente contribución a esta área en el resto del mundo, permitiendo comparar diferentes situaciones en cuanto a niveles y tipos de emisión, así como respecto de las condiciones socioeconómicas de la población (Amegbor et al., 2023 ; Bonilla et al., 2023; De Oliveira et al., 2023; Imane et al., 2022 ; Kurtz et al., 2022 ; Mlambo et al., 2023 ; Thao et al., 2022; Z. Yang et al., 2022). Dado que las condiciones meteorológicas, topográficas y las actividades antropogénicas en un sitio específico determinan los patrones de emisión de PM, se ve justificada la realización de estudios de monitoreo *in situ*. Hasta el día de hoy, en la ciudad de Córdoba, se han desarrollado estudios que demuestran los impactos en la salud de este contaminante, además de su relación con covariables y la identificación de su fuente de emisión dada su composición (Amarillo et al., 2021; Busso et al., 2021; Mateos et al., 2018, 2019). Sin embargo, ningún estudio realizado en el país ha tenido por objetivo el desarrollo de un modelo predictivo de contaminación de aire en pos de generar un sistema de alerta temprana ante condiciones adversas de calidad de aire implementando técnicas de inteligencia artificial para la concentración horaria de PM_{2.5}.

Contar con información sobre la variabilidad del PM y su relación con variables meteorológicas en diversas escalas temporales en distintos sitios del área metropolitana de la provincia de Córdoba, permitirá caracterizar tanto espacial como temporalmente este contaminante. Además, estos resultados son el punto de partida para el desarrollo de modelos estadísticos de elevada capacidad predictiva, ya que no solo aprovechan la variabilidad respecto de covariables derivadas a partir de información satelital, sino que también lo hacen respecto de información espacial y temporal derivada del análisis exploratorio. Esto permitirá la creación de un sistema de alerta cuya capacidad predictiva se vea potenciada, permitiendo así advertir ante condiciones adversas de calidad de aire y reforzando los pilares en los que se sustenta la salud pública, que garanticen una mejor calidad de vida a la población. Finalmente, la disponibilidad de nuevos datos paliará la falta de acceso a información sobre contaminación de aire en la provincia.

La inaccesibilidad a sistemas de monitoreo de referencia en países cuyos ingresos son limitados genera una ausencia de control y gestión de la calidad del aire. La capacidad de generar redes de vigilancia de la calidad del aire basadas en equipos certificados de referencia es exclusivo de países desarrollados económicamente dados sus elevados costes de adquisición y mantenimiento. Esto es prohibitivo para países en vías de desarrollo que dejan la gestión y control de calidad de aire de lado al ser una necesidad secundaria.

1.2.1. Objetivo general

Desarrollar un modelo predictivo de concentración de aerosoles para el área metropolitana de la provincia de Córdoba a partir de datos relevados por sensores de bajo costo, en función de variables derivadas a partir de información satelital de acceso público, variables meteorológicas, variables temporales y variables relacionadas al uso del suelo, y validar el empleo de sensores de bajo costo para determinar partículas

finas en áreas urbanas.

1.2.2. Objetivos específicos

1. Describir la variación espacio temporal de la concentración de aerosoles colectados con sensores de bajo costo en zonas del Área Metropolitana de Córdoba, Argentina.
2. Evaluar la utilidad de sensores de bajo costo para el estudio de aerosoles en zonas urbanas del Área Metropolitana de Córdoba, Argentina.
3. Analizar la influencia de distintos tipos de variables obtenidas a través de satélites, entre ellas meteorológicas, en los patrones de variación temporal de aerosoles en el Área Metropolitana de Córdoba, Argentina.
4. Validar el empleo de productos satelitales, tales como AOD (del inglés *Aerosol Optical Depth*), para estimar la concentración de material particulado en el Área Metropolitana de Córdoba, Argentina, a través de análisis estadísticos.
5. Proponer un modelo predictivo que incluya variables de uso del suelo, satelitales y temporales para la estimación de la concentración de aerosoles en el Área Metropolitana de Córdoba, Argentina.

1.3. Desafíos metodológicos

Los procesos de contaminación atmosférica varían de manera continua en el tiempo y en el espacio. Por este motivo, resulta de interés predecir la concentración de contaminantes atmosféricos referenciados a un determinado momento de tiempo y ubicación geográfica. Los datos que alimentan estos análisis son series temporales condicionadas a una locación espacial, o bien una colección de observaciones recogidas secuencialmente en el tiempo georreferenciadas a un sitio en particular. Cabe destacar

que la variabilidad espacio temporal de los contaminantes no se encuentra determinada solo por las características espaciales y temporales de los datos recopilados, sino por múltiples covariables que también varían espacial y temporalmente, en este caso variables meteorológicas.

En parte, la novedad de la investigación recae en el análisis de series temporales de PM_{2.5} cuya frecuencia de observación es horaria, es decir, que los datos se encuentran equiespaciados por un intervalo de tiempo igual a una hora, algo nunca antes realizado para la provincia de Córdoba, Argentina. El primer impedimento a esta investigación es la ausencia de un sistema de medición y seguimiento de este tipo de contaminante en la escala temporal de interés. Como consecuencia, para poder recopilar los datos, fue necesario el emplear sensores que permitiesen monitorear y recopilar datos de concentración de PM_{2.5} en la frecuencia temporal de interés. Estos dispositivos debieron ser diseñados y ensamblados desde cero, atravesando todas las etapas transversales en el desarrollo de un equipo de monitoreo, desde la creación del algoritmo y el software, hasta el ensamblado de los prototipos finales y su validación con respecto a un sistema de referencia certificado.

Una vez obtenidos los datos del contaminante de interés debió procederse con una exhaustiva limpieza y pre tratamiento de los datos, de acuerdo a las características espaciales y temporales particulares de cada uno. Por este motivo fue inevitable un análisis condicionado tanto por ubicación geografía como por momento de tiempo, garantizando una limpieza exclusiva para cada uno de los sitios y períodos de tiempo considerados. Como consecuencia fue necesario el estudio de la variabilidad explicada por cada una de las variables temporales derivadas, con el objetivo de determinar el porcentaje de variabilidad representado por cada una de ellas.

El estudio de series temporales tiene como característica principal la dependencia que vincula observaciones contiguas, que puede verse influenciada por comportamientos estacionales. Es importante determinar cuáles son los factores temporales que

desencadenan la estacionalidad que influencia la serie temporal, para así contar con las herramientas necesarias para explicar su variabilidad. Además, las características propias de las series temporales son las que nos permitirán encontrar relaciones entre las distintas ubicaciones espaciales donde los datos fueron recopilados. Por otra parte, determinar estas características de la serie temporal es necesario para poder entrenar un algoritmo predictivo que logre acertadas predicciones de la concentración de PM_{2.5}.

Como se mencionó anteriormente, no basta solo con variables derivadas de las características espaciales y temporales para explicar la variabilidad del contaminante en cuestión. Por este motivo, fue necesaria la inclusión de covariables que fluctúasen conjuntamente con la concentración de PM_{2.5} durante el período de muestreo, en la misma frecuencia temporal y ubicación geográfica en que los datos fueron recopilados por los sensores desarrollados. El modelado estadístico busca captar la forma en que se desencadena la variabilidad conjunta de la variable respuesta (concentración de PM_{2.5}) con respecto a las ingresadas como predictoras, en función de los datos de entrenamiento, con el objetivo de lograr generar predicciones en datos donde no se cuenta con la respuesta. El análisis de variabilidad conjunto de la respuesta en función de las covariables y la incorporación de mecanismos de selección de características permitirá identificar un subconjunto de datos que maximice la explicación de la forma en que fluctúa la concentración de PM_{2.5}, resultando en una reducción de la dimensionalidad y disminuyendo la complejidad. Además, este análisis se complementa con la creación de nuevas características derivadas de los datos existentes, algo que permitirá la introducción de covariables que garanticen una mejor representación de la variable respuesta. Finalmente, es fundamental el estudio de múltiples algoritmos y modelos estadísticos, con el objetivo de determinar cuál de ellos es el que mejor ajusta a los datos, logrando captar la variabilidad presente y garantizando predicciones de elevada exactitud.

A modo de resumen y de forma unificada, el objeto de este trabajo es proveer

las herramientas necesarias para la creación de un sistema de alerta ante condiciones adversas de calidad de aire basado en tecnologías de bajo costo. En primer lugar fue necesario contar con un sistema de monitoreo de bajo costo y validarla con respecto a una estación de referencia. Posteriormente es necesario recopilar datos de concentración de PM_{2.5}, además de buscar covariables que permitan explicar su variabilidad. Una vez obtenidos los datos, es necesario llevar a cabo el modelado predictivo, modelado que busca inferir la forma en que se vinculan las covariables y la concentración de PM_{2.5}, con el objetivo de explicar sus fluctuaciones y obtener predicciones acertadas. La principal aplicación de estos modelos recae en alertar a la población ante condiciones adversas de calidad de aire en ciudades de países en vías de desarrollo con un horizonte de predicción de cinco días, permitiendo la accesibilidad a datos de calidad de aire y preservando la salud pública.

Capítulo 2

Sistema de monitoreo basado en sensores de bajo costo

2.1. Introducción

Solo países desarrollados económicamente tienen la capacidad de generar redes de vigilancia de la calidad del aire basadas en equipos certificados de referencia. Un país que marca tendencia en cuanto al seguimiento y control del PM es Estados Unidos, donde se lleva a cabo una vigilancia exhaustiva de este contaminante a nivel del suelo por medio del Método Federal de Referencia (FRM) y los Métodos Federales Equivalentes (FEM) (Noble et al., 2001). Cuenta con aproximadamente 600 estaciones de referencia en todo su territorio, que además se complementan con mediciones satelitales y otros métodos alternativos de seguimiento para lograr extrapolar los resultados a toda la nación (Christopher & Gupta, 2010). A pesar de marcar tendencia en el control y gestión de la calidad del aire, establece protocolos y normativas que responden a la realidad de un país con una fuerte economía, y produce tecnologías cuya accesibilidad requiere elevados presupuestos. Estos estándares entonces representan metas inalcanzables para naciones en vías de desarrollo, motivo por el cual se hace

necesario migrar hacia alternativas de bajo costo para la gestión y control de la calidad del aire.

En esta tesis se armaron instrumentos de bajo costo empleando un sensor láser de medición de partículas que sigue el principio de funcionamiento del fotómetro de dispersión, lo cual proporciona una solución económica para el monitoreo y control de la concentración de PM_{2.5}. Sus principales ventajas frente a los FRM y los FEM incluyen mediciones en tiempo real de los valores de concentración, además de su portabilidad y rentabilidad. Estas ventajas son a expensas de la exactitud en la medición e información acerca de la composición e individualidad de la partícula. A efectos de esta tesis, se entiende por sensores de bajo costo a aquellos que pueden ser adquiridos a menos de 100 dólares americanos.

El monitoreo de la calidad del aire es una actividad fundamental a la hora de identificar y evaluar problemas de contaminación (OMS, 2021). Esta actividad combinada con herramientas como los modelos predictivos e inventarios de emisión, forman parte integral de los sistemas de gestión y control de la calidad del aire. El propósito principal de estos sistemas recae en la provisión de información a científicos, legisladores y planificadores, fomentando la toma de decisiones para la gestión, mejora del ambiente y preservación de la salud pública.

El objetivo del presente capítulo es evaluar la utilidad de sensores de bajo costo para el monitoreo de la concentración de PM_{2.5} en el área metropolitana de la provincia de Córdoba, Argentina. Para abarcar este objetivo fue necesario el diseño y desarrollo de sistemas de monitoreo basado en sensores de bajo costo que permitiesen recopilar datos de concentración de PM_{2.5}. Además, resultó indispensable determinar la integridad de las mediciones realizadas, no solo la estabilidad entre sensores, si no que también su validación respecto a una estación de referencia.

2.2. Materiales y métodos

2.2.1. Resumen gráfico del protocolo analítico de desarrollo y validación de sensores de monitoreo basados en tecnologías de bajo costo

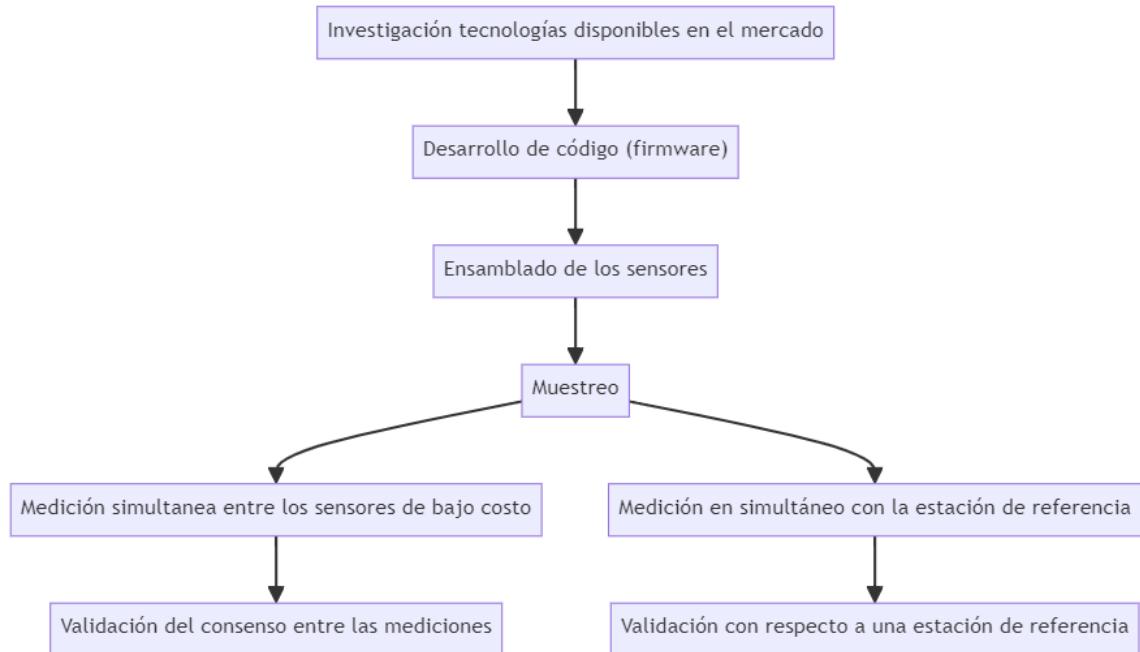


Figura 2.1: Resumen gráfico del protocolo analítico desarrollado en el capítulo 2.

La Figura 2.1 presenta el protocolo analítico desarrollado en el capítulo 2, ofreciendo una visión ordenada de las distintas actividades llevadas a cabo para alcanzar el objetivo propuesto. En primer lugar, se destaca la intensa investigación de tecnologías disponibles en el mercado referidas a sensores de bajo costo. A continuación, se ilustra el procedimiento de desarrollo del *firmware*, el software embebido en la memoria de lectura de un dispositivo que se encarga de proporcionar las instrucciones sobre el comportamiento del sistema. Posteriormente se hace referencia al armado y ensamblado de los sensores de bajo costo, específicamente diseñado para la medición

de concentración de PM_{2.5}. Posteriormente se hace referencia a la evaluación de la calidad del dato recopilado, la cual se abordó en dos etapas. La primera se centró en determinar el consenso de las mediciones obtenidas por los diez sistemas de bajo costo ensambladas. Por otra parte, se realizó una comparación de las mediciones obtenidas por un sensor de bajo costo contra las de una estación de referencia, evaluando su fiabilidad y precisión. A pesar del orden en que se menciona el procedimiento, en primer lugar fue realizada la comparación contra una estación de referencia y luego la medición del consenso entre los sistemas de medición de bajo costo, al tratarse de dos procedimientos totalmente independientes decide mantenerse la estructura dual del procedimiento.

2.2.2. Desarrollo del sistema de monitoreo basado en sensores de bajo costo

En la presente tesis, se define un sistema de monitoreo como un conjunto de sensores ensamblados, operando mediante una placa de microcontrolador, junto con el código de base o *firmware* que permite la intercomunicación y protocolo de funcionamiento. Estos sistemas, al basarse en sensores de bajo costo, se unifican bajo el término “sistema de monitoreo de bajo costo”. Este sistema permite la medición de diversas variables utilizando tecnologías económicas. El sistema de monitoreo de bajo costo resultante quedó conformado por los siguientes componentes:

1. Sensor PMS 1003 Plantower.
2. Sensor RTC-DS3231 - con EEPROM I2C.
3. Sensor DHT22 - Módulo Sensor Digital de Temperatura y Humedad.
4. Arduino UNO R3 SMD (CH340G) comp / ATmega328 16MHz.
5. Micro SD Card - Conexión SPI y Reg de Tensión.
6. Protoshield para Arduino uno shield y proto 170p uno electro.

7. Fuente Arduino 12V 1,25 A.

En el marco del estudio de sensores de bajo costo para la medición de PM, se llevó a cabo una investigación de las distintas tecnologías disponibles en el mercado. Luego de analizar diferentes aspectos, como precisión de medición, costo, funcionamiento, tamaño, sensibilidad e interfaz de programación, se seleccionó el sensor PMS 1003 como la opción más viable. Este es un sensor de calidad del aire basado en el principio de dispersión láser, desarrollado por la compañía Plantower, una industria de origen chino con sede en Shenzhen y oficinas en Pekín. A partir de un haz láser de 660 nm este sensor mide la concentración de PM_{2.5}, un algoritmo de transformación propio del sensor permite la estimación en las fracciones PM_{1.0}, y PM₁₀. El fabricante reporta una precisión de $\pm 10 \frac{\mu g}{m^3}$ para el PM_{2.5} en el rango de 0 a 500 $\pm 10 \frac{\mu g}{m^3}$, mientras que para el rango de 500 a 1000 $\frac{\mu g}{m^3}$ el error se estima como $\pm 10\% m.v.$, lo que significa que el error puede ser de hasta el 10 % del valor medido en ese rango de concentración. Es importante tener en cuenta estas especificaciones a la hora de utilizar el sensor, con el objetivo de lograr una correcta interpretación de los datos.

Para garantizar el correcto funcionamiento del sensor, se procedió a desarrollar y programar el código (*firmware*) que permitiría la conexión y comunicación entre los diferentes componentes. Se decidió utilizar la plataforma de desarrollo Arduino. Esta es una plataforma de electrónica abierta para la creación de prototipos basada en software y hardware flexibles, diseñada con el objetivo de facilitar la creación de entornos y objetos interactivos. La tarjeta de Arduino que se empleo fue la UNO R3 SMD, esta es una tarjeta que reduce su costo dado que contiene el chip adaptador USB – SERIAL CH340, a diferencia de otras tarjetas más costosas que contienen el FT232. Otra característica que abarata su costo es que emplea ATmega328 en su versión SMD o de montaje superficial. En resumen, se trata de una placa electrónica basada en el microcontrolador ATmega328 que cuenta con 14 pines de entrada y salida digitales (de los cuales 6 pueden ser utilizados como salidas PWM) y 6 entradas

analógicas, un conector USB, uno de alimentación y un botón de reinicio. Junto con la placa seleccionada se incorporó una protoshield para Arduino uno junto con el proto uno electro, permiten duplicar el número de conexiones que podemos realizar en la placa, otorgando mayor espacio para realizar las conexiones. La protoshield, es una placa de expansión analógica y digital que permite la interconexión de múltiples periféricos al estandarizar la disposición de los pines. Esta se acopla al Arduino ampliando las capacidades de la placa de base.

El chip RTC DS3231 en combinación con la EEPROM I2C proporcionan un reloj en tiempo real (RTC) que permite registrar la fecha y hora en que se realizan las mediciones. Además, cuenta con una interfaz I2C, un oscilador integrado con compensación de temperatura de cristal (TCXO) y un cristal, que mejoran su precisión y la estabilidad del reloj.

El sensor DHT22 es un sensor digital de temperatura y humedad relativa de bajo costo, que permite detectar cualquier anomalía en las condiciones ambientales durante el funcionamiento del sensor. A partir de los datos registrados por este sensor podemos cerciorarnos de las condiciones de funcionamiento, con el objetivo de determinar la calidad de los datos recopilados. Este módulo utiliza un sensor capacitivo de humedad y un termistor que permite medir el aire circundante. Es importante asegurarse de que las condiciones en las que se realiza la medición sean estables y adecuadas para obtener datos confiables.

El sensor también incorporó un módulo micro SD, permitiendo incorporar una memoria micro SD para almacenar los datos, dada su gran capacidad de almacenamiento en comparación con su reducido tamaño físico lo vuelven la opción más viable para este tipo de sensores. De esta forma, son una alternativa sumamente compatible con la plataforma Arduino, sobre todo cuando se requiere almacenar grandes cantidades de información en productos de reducido tamaño.

Finalmente, la fuente de alimentación de 12 V es necesaria para garantizar el

funcionamiento continuo del sistema de medición. Esta se conecta a una toma eléctrica de pared permitiendo la provisión de energía de manera continua al instrumento.

2.2.3. Base de datos

Se generó una base de datos de mediciones simultáneas entre los diez sistemas ensambladas. Estas mediciones fueron llevadas a cabo durante un período de 14 días, entre el 05-07-2022 y el 21-07-2022, en la terraza de la Facultad de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de Córdoba, Argentina.



Figura 2.2: Medición simultánea de los sistemas de monitoreo de bajo costo.

Por otra parte, para validar los datos de los sensores de bajo costo con datos

obtenidos con un equipo de referencia, se obtuvieron datos de la estación de monitoreo instalada por el Centro de Investigación y Química Ambiental (CIQA) en el centro de la ciudad, en la Casona Municipal ubicada en la avenida General Paz y La Rioja. Esta estación de monitoreo cuenta con un equipo atenuador beta (Instrumento FEM) aprobado por la Agencia de Protección Ambiental de Estados Unidos (USEPA), que registra concentración de PM₁₀. Los datos fueron recopilados desde el 28-10-2021 hasta el 05-11-2021.

Junto a la estación de monitoreo se colocó un sensor y se compraron datos de PM₁₀ de la estación de referencia con los valores de PM₁₀ estimados por el sensor de bajo costo a partir de PM_{2.5}.

2.2.4. Protocolo estadístico

Validación del consenso entre las mediciones de los sistemas de monitoreo de bajo costo

La determinación del consenso entre las mediciones de los sistemas de monitoreo de bajo costo fue abordada por diversas estrategias. En primer lugar, se llevó a cabo el cálculo y comparación de medidas descriptivas a partir de las series temporales recopiladas. Estas medidas incluyeron la media, la desviación estándar, el primer cuartil, la mediana, el tercer cuartil, el mínimo y el máximo de la medición de PM_{2.5}, con frecuencia horaria. Esta comparación permitió determinar si existían discrepancias significativas entre los distintos sensores.

En segundo lugar, se realizaron gráficos exploratorios de las series temporales recopiladas por los sensores. Estos fueron comparados con el objetivo de detectar patrones comunes en los datos y evaluar si los sensores mostraban tendencias similares. Además, se buscó detectar anomalías o fluctuaciones inusuales en los datos.

Finalmente, se implementó el coeficiente de correlación de Pearson para determinar la semejanza existente entre los datos. Este coeficiente de correlación es una medida

estadística que indica el grado de asociación lineal entre dos variables, en este caso, las mediciones de concentración de PM_{2.5} entre los sensores de bajo costo de a pares.

Comparación con una estación de referencia

La comparación de los registros de los sistemas de monitoreo de bajo costo con la estación de referencia fue abordada a partir de diversas estrategias. En primer lugar, se procedió a visualizar las series temporales correspondientes a ambas fuentes de datos de forma conjunta. Esta visualización permitió detectar eventuales discrepancias, así como también el grado de similitud entre las mediciones.

Para mejorar el contraste, se implementó un gráfico de similitud que permitió representar los datos de los sensores en ambos ejes cartesianos, conjuntamente con la recta de similitud ubicada a 45°. La finalidad de este gráfico fue facilitar la identificación de patrones de similitud o discrepancia entre las mediciones. En particular, si los datos de los sensores y la estación de referencia mostraran coincidencia perfecta, estos se posicionarían sobre la línea diagonal.

Finalmente, se procedió a calcular el coeficiente de correlación lineal de Pearson, que permitió obtener una medida cuantitativa de la similitud existente entre los datos de los sistemas de monitoreo de bajo costo y los de la estación de referencia. Este coeficiente proporciona información acerca del grado de asociación lineal entre las dos variables.

2.3. Resultados y discusión

2.3.1. Producto final

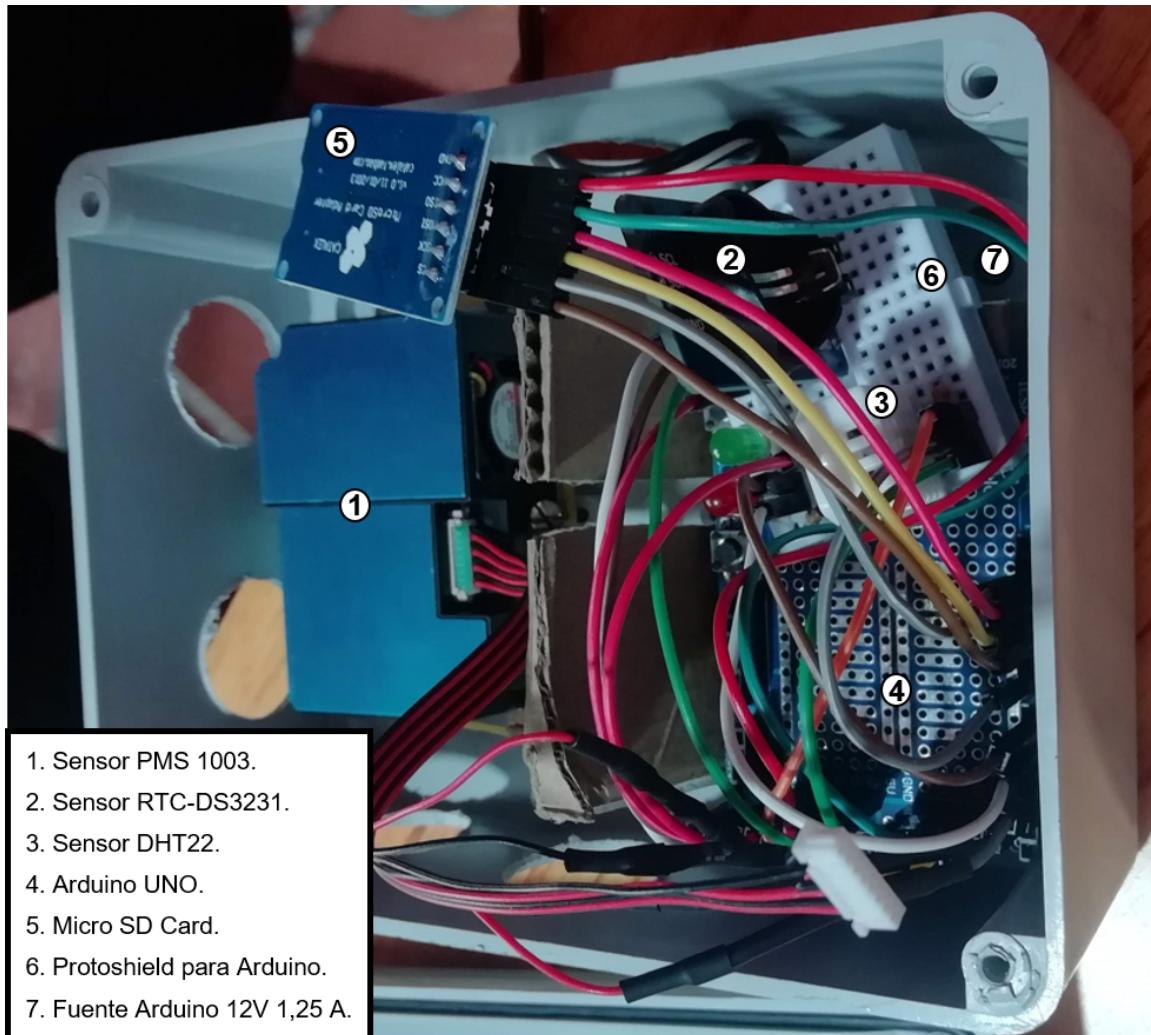


Figura 2.3: Sistema de monitoreo basado en tecnologías de bajo costo ensamblado.

La Figura 2.3 presenta una vista detallada de uno de los sistemas de monitoreo de bajo costo en su última versión, donde pueden observarse los diferentes componentes, así como las interconexiones que permiten la comunicación entre ellos.

El sensor de mayor tamaño, localizado en el frente del sistema es el sensor PMS 1003, responsable de la medición la concentración de PM. Por este motivo es que tanto

en su frente como en sus alrededores se realizaron perforaciones que garantizan el flujo de aire.

Por otra parte, puede apreciarse como el resto de los dispositivos se encuentran protegidos y aislados del exterior para garantizar su integridad y correcto funcionamiento.



Figura 2.4: Sistema de monitoreo de bajo costo en funcionamiento durante el muestreo en La Reserva Natural Urbana General San Martín de la Provincia de Córdoba.

La Figura 2.4 muestra el sistema de monitoreo de bajo costo dispuesto en La Reserva Natural Urbana General San Martín. En la imagen, puede apreciarse cómo el sensor se encuentra instalado en el exterior del edificio, pero protegido de la lluvia, lo que garantiza su integridad y correcto funcionamiento.

2.3.2. Consenso de medición de la concentración de PM_{2.5} entre los sistemas de monitoreo de bajo costo

Tabla 2.1: Medidas descriptivas de resumen de los distintos sensores.

Sensor	Media	D.E.	Q1	Mediana	Q3	Mínimo	Máximo
1	14.07	13.90	3.97	9.84	19.39	0.10	70.85
2	15.09	14.08	4.41	11.4	20.58	0.10	72.13
3	13.96	13.47	4.07	9.99	19.26	0.01	68.82
4	14.95	14.37	4.41	10.7	20.87	0.02	72.28
5	14.53	13.35	4.59	10.8	19.82	0.16	69.80
6	13.52	13.44	3.6	9.4	18.76	0.02	68.25
7	14.08	13.85	3.79	10.1	19.64	0.05	69.99
8	13.05	12.77	3.66	9.22	17.79	0.08	64.73
9	14.86	13.92	4.58	10.9	20.48	0.11	71.40
10	12.42	12.12	3.56	8.73	17.12	0.11	62.56
General	14.05	13.55	4.07	10.05	19.38	0.08	69.08

Sensor, media, desvío estándar (D.E.), percentil Q1, mediana, percentil Q3, mínimo y máximo.

Durante el periodo de calibración se obtuvieron un total de 3360 datos horarios, a partir de los cuales se calcularon medidas de resumen (Tabla 2.1). La media general obtenida durante este muestreo fue de $14.05 \frac{\mu g}{m^3}$, con un desvío estándar promedio de $13.55 \frac{\mu g}{m^3}$, un valor mínimo promedio registrado de $0.08 \frac{\mu g}{m^3}$ y un valor máximo de $69.08 \frac{\mu g}{m^3}$.

Puede apreciarse que el sensor 10 es el que más se aleja de la media general, con una diferencia promedio de $1.63 \frac{\mu g}{m^3}$. Sin embargo, esta diferencia es baja en magnitud con respecto al desvío estándar promedio, lo que indica que incluso el sensor más deficiente en medir se desvía poco de las mediciones generales frente al amplio rango

de variación de la variable en cuestión. Además, puede notarse que la incapacidad del sensor 10 en la medición se debe a una subestimación de los valores de concentración de PM_{2.5}, ya que todas sus medidas descriptivas se encuentran por debajo del promedio general obtenido.

La Tabla 2.1 refleja la consistencia existente entre las mediciones realizadas por los distintos sensores a lo largo del tiempo. El rango de variación de las medidas descriptivas obtenidos para los distintos sensores respecto a la media general es acotado, lo cual demuestra que no existen desvíos respecto a las medidas descriptivas generales que indiquen un funcionamiento atípico en ninguno de los sensores.

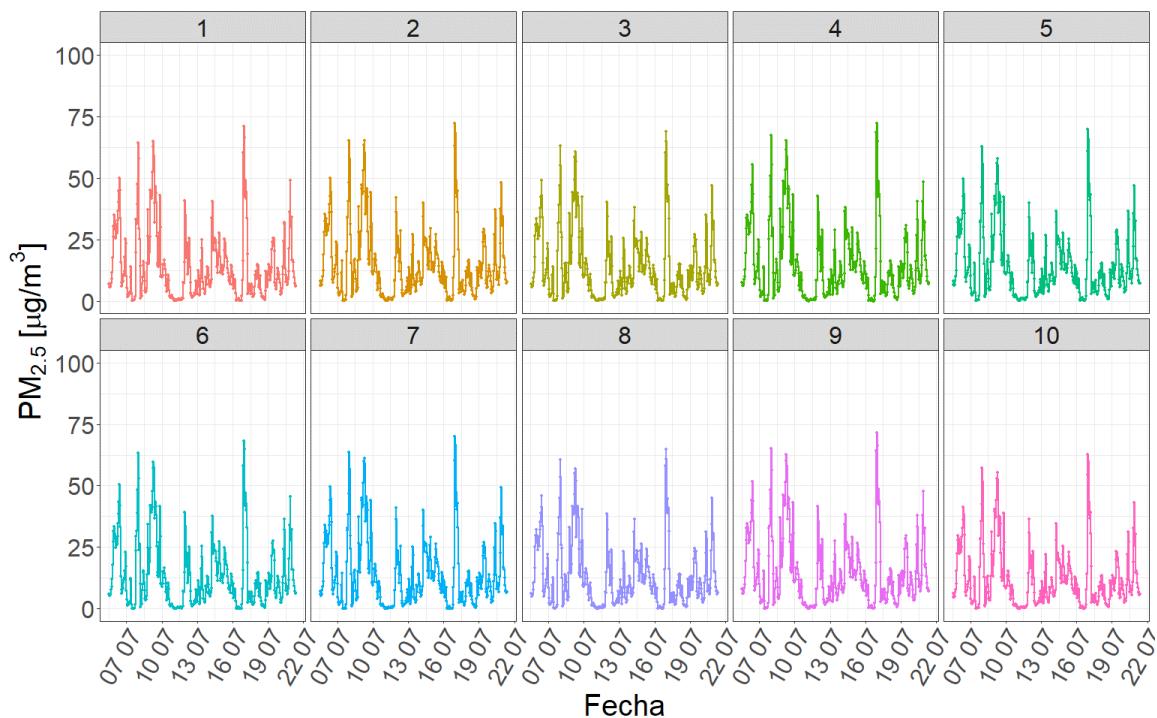


Figura 2.5: Series de tiempo de concentración de PM_{2.5} generadas durante la medición conjunta de los sistemas de monitoreo de bajo costo.

Todas las series de tiempo exhiben un patrón general común, identificando de manera similar los momentos en que ocurren picos de concentración y baches mínimos (Figura 2.5). Es interesante destacar que el pico máximo de concentración y el mínimo fueron registrados de forma simultánea por todos los sensores en el día 2022-07-16 a

las 23:00:00 y 15:00:00, respectivamente. Esto demuestra una notable consistencia en las mediciones de los sensores, que consiguen reflejar de manera similar las variaciones en la concentración.

Además, puede observarse que a pesar de existir una leve variación en la magnitud de las mediciones registradas en un mismo momento, los valores muestran un consenso y reflejan el patrón de variación de PM_{2.5}. En otras palabras, los sensores de bajo costo presentan una capacidad consistente para detectar cambios en la concentración de PM_{2.5} a lo largo del tiempo, lo que sugiere que son una valiosa herramienta para la monitorización continua de la calidad del aire.

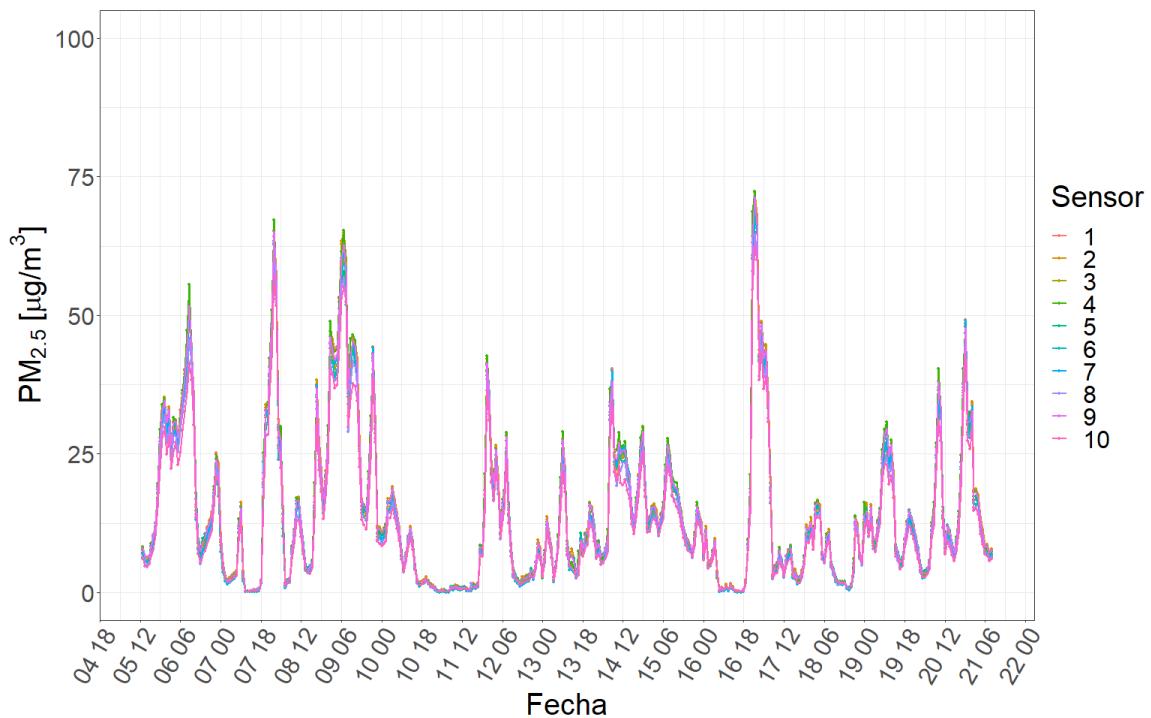


Figura 2.6: Series de tiempo de concentración de PM_{2.5} generadas durante la medición conjunta de los sistemas de monitoreo de bajo costo superpuestas.

En la Figura 2.6 puede notarse cómo a pesar de que las mediciones de los sensores a lo largo del tiempo siguen un patrón general común, pueden variar en magnitud en un mismo instante, dando lugar a pequeñas desviaciones en los registros. A pesar de esto, puede apreciarse que estas diferencias son bajas en comparación con el rango de

variación de los valores que toma la variable de interés.

Para evaluar esta deficiencia en el consenso de los sensores, se calculó la métrica de MAD (desvío absoluto medio), que refleja la media de la suma de las diferencias en valor absoluto de las mediciones entre los sensores, otorgando un resultado máximo de 2.66 entre los sensores. Este valor indica que los dos sensores que mayor discrepancia reflejaron en sus mediciones (sensor 2 y sensor 10) tienen una diferencia promedio en valor absoluto de $2.66 \frac{\mu g}{m^3}$ en sus mediciones a lo largo del tiempo. Este valor representa el 4 % del rango de variación promedio entre las mediciones de todos los sensores, el cual es de $69 \frac{\mu g}{m^3}$ (Tabla 2.1). Además, la media del MAD para todos los sistemas de medición de bajo costo considerados arrojó un valor de $1.17 \frac{\mu g}{m^3}$, demostrando que esta es la diferencia promedio en valor absoluto entre las mediciones de concentración de PM_{2.5} de los sistemas de monitoreo de bajo costo a lo largo del tiempo.

De esta manera, puede concluirse que todos los sensores logran capturar el mismo patrón de variación de PM_{2.5} en el tiempo, identificando de manera similar los baches mínimos y los picos máximos. Además, a pesar de que existen variaciones en los valores registrados, estas son bajas en comparación con el rango de variación de la variable de interés.

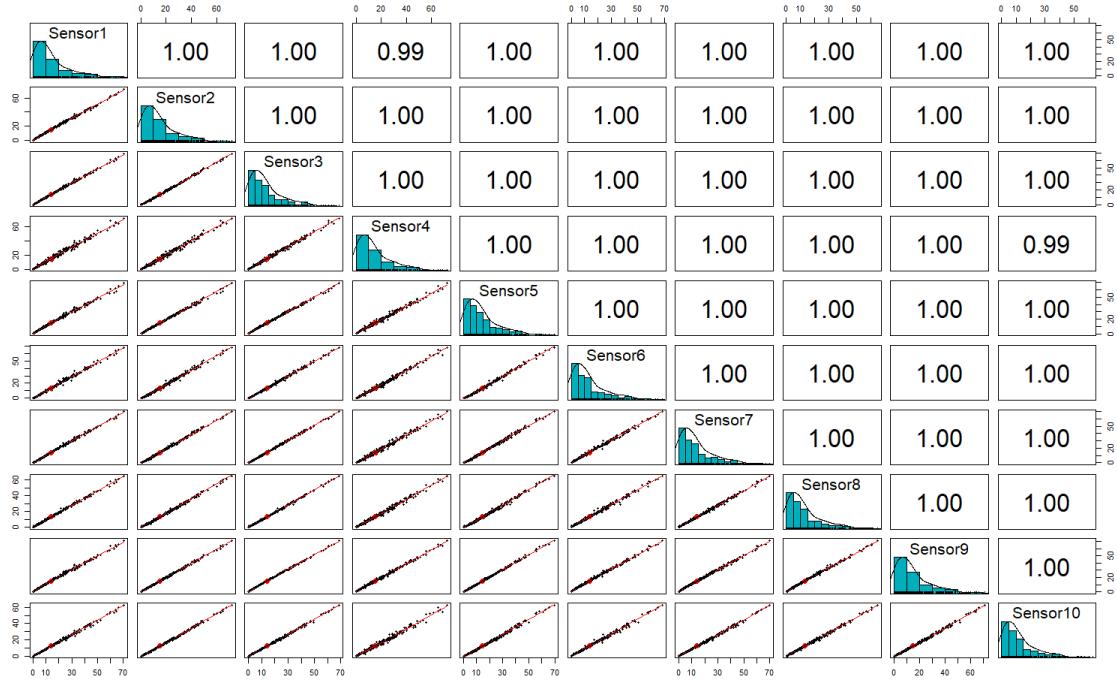
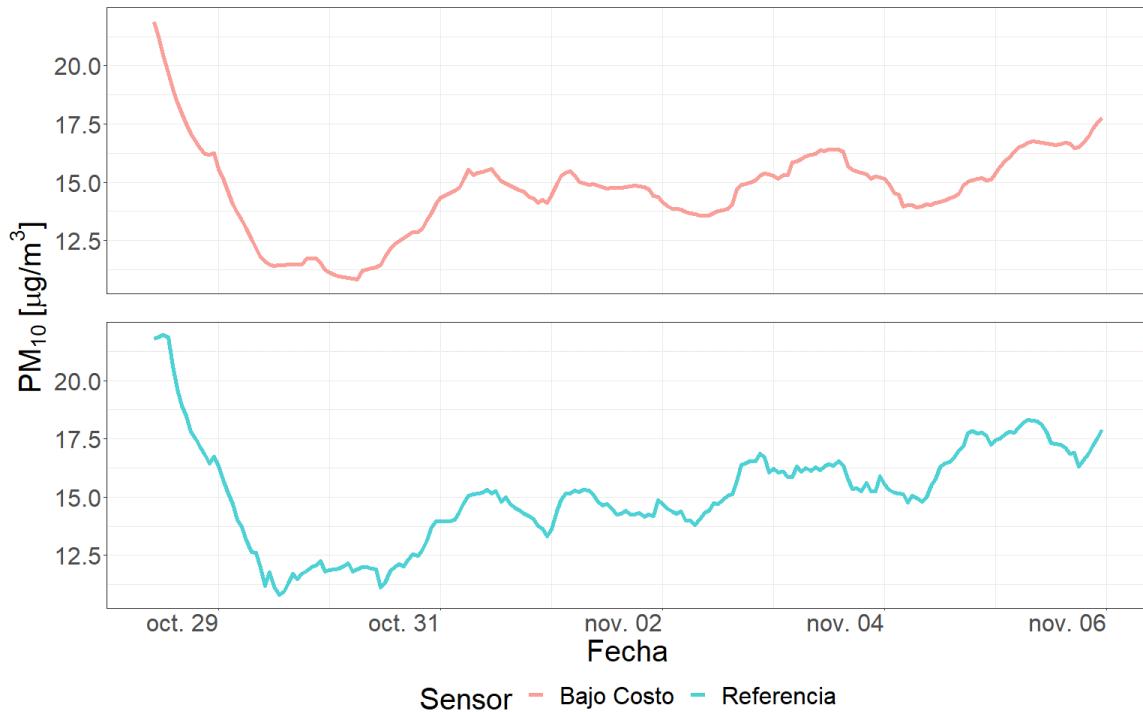


Figura 2.7: Relación entre las mediciones de concentración de PM_{2.5} de los distintos sistemas de monitoreo de bajo costo, su distribución y coeficientes de correlación.

La Figura 2.7 confirma el consenso entre las mediciones de los distintos sensores. El diagrama de puntos generado por la combinación de cada par de sensores muestra una estrecha relación con la recta de similitud de 45 grados, demostrando una alta coincidencia en las mediciones a lo largo del tiempo. Además, los coeficientes de correlación calculados para cada par de sensores son siempre mayores a 0.99, indicando una correlación lineal muy fuerte entre las mediciones. En consecuencia, podemos concluir que los sistemas de monitoreo de bajo costo son capaces de medir con precisión el PM_{2.5} en el aire, y que existe un alto grado de acuerdo entre las mediciones obtenidas por cada uno de ellos.

2.3.3. Validación con respecto a la estación de referencia



En la Figura 2.8 puede apreciarse que la tendencia general de las mediciones es similar para ambos instrumentos de medición, ya que ambas series muestran fluctuaciones similares, demostrando acuerdo entre sus registros. A partir de estos resultados, podría establecerse en primera instancia, que las mediciones de concentración de PM_{2.5} registradas por los sensores de bajo costo permitirían representar en términos generales los datos obtenidos por una estación de referencia.

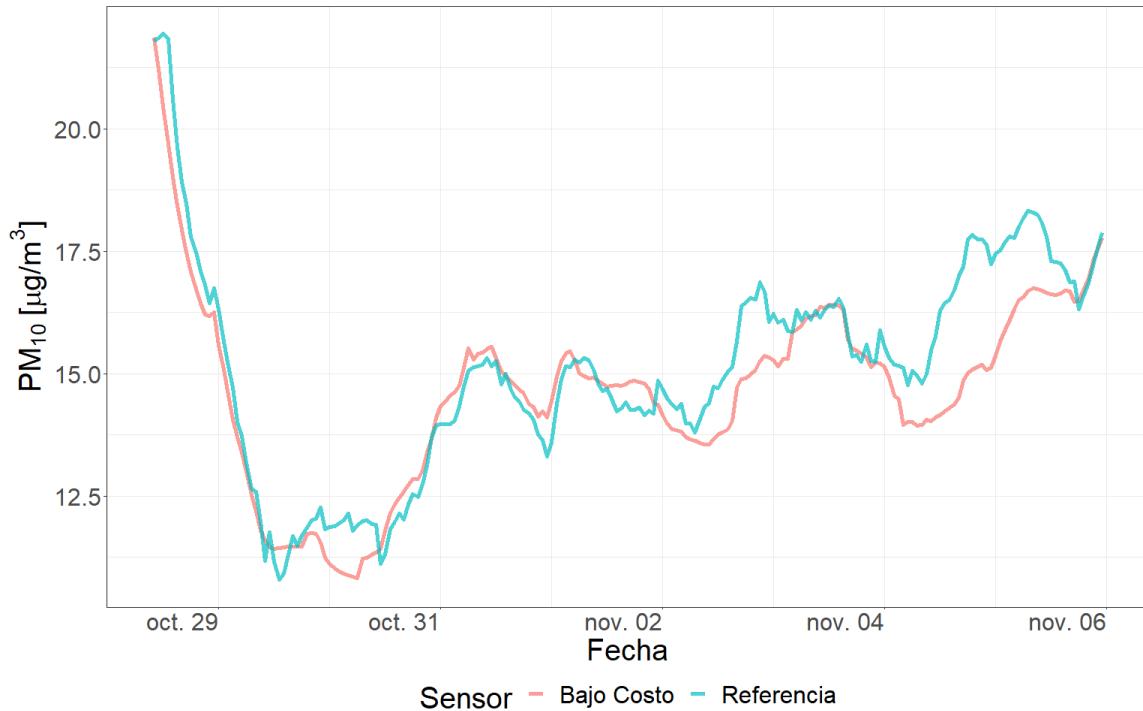


Figura 2.9: Series de tiempo generadas durante la medición simultánea de la estación de referencia y el sistema de monitoreo de bajo costo superpuestas.

Ambos equipos de medición registraron el mismo patrón de variación de la concentración de PM_{2.5} (Figura 2.9). Sin embargo, existen sucesivos eventos en los que las mediciones de la estación de referencia se encuentran por encima de las mediciones registradas por los sistemas de monitoreo de bajo costo. Esto indicaría que las mediciones de los sistemas de monitoreo de bajo costo podrían subestimar la concentración de PM.

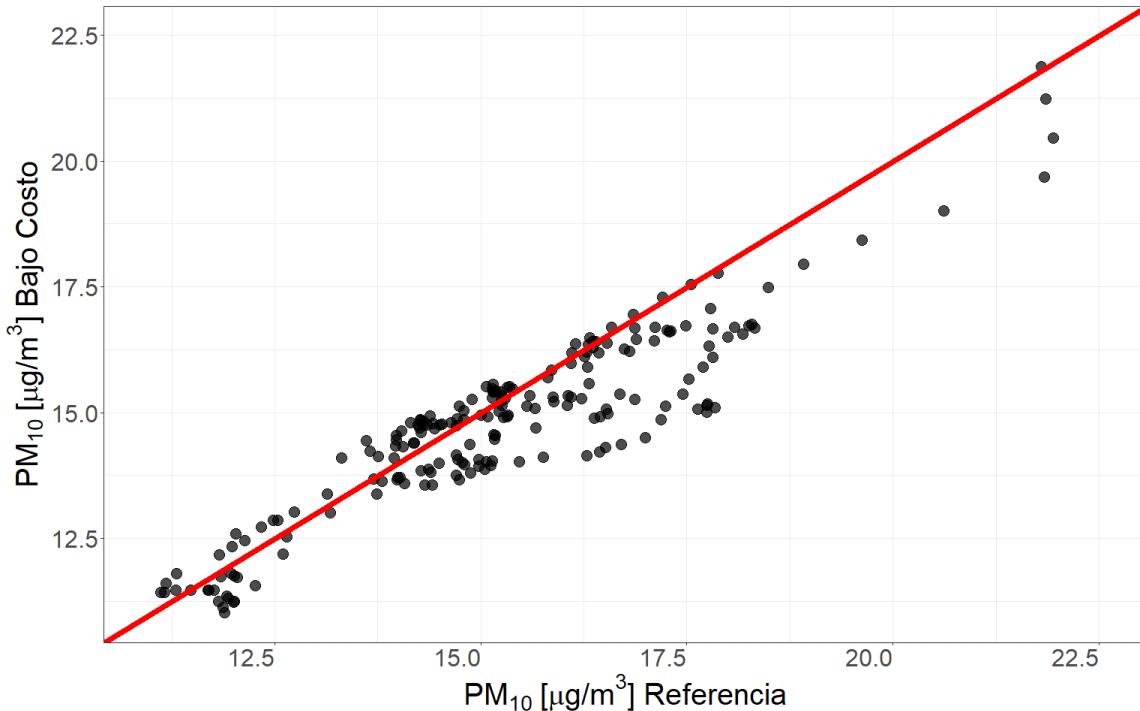


Figura 2.10: Diagrama de puntos de las mediciones registradas por la estación de referencia y el sistema de monitoreo de bajo costo, en rojo puede apreciarse la recta de similitud a 45 grados.

Las mediciones de ambos equipos junto con la recta de similitud a 45 grados demuestran una fuerte coincidencia entre las mediciones (Figura 2.10). La imagen permite apreciar una alta densidad de puntos que se sitúan por encima de la recta, indicando una completa concordancia entre los datos de ambas estaciones. Sin embargo, también puede apreciarse un conjunto de puntos que se alejan de la recta de similitud y se ubican por encima de la misma, sugiriendo una subestimación de la concentración de PM por parte de los sistemas de monitoreo de bajo costo en el rango de valores entre 13 y 17 $\frac{\mu g}{m^3}$. De todas maneras, la mayor densidad de puntos se observa cercana a la recta de similitud, demostrando consistencia y concordancia entre sus mediciones.

Esta fuerte coincidencia entre las mediciones de ambas estaciones además se ve reflejada en el valor que otorgó el coeficiente de correlación lineal de Pearson. El coeficiente de 0.93 permite concluir que, a pesar de la subestimación de los sistemas de bajo costo, existe una notable similitud entre ambos sistemas de medición.

Esto significa que, aunque los valores registrados por el sensor de bajo costo puedan ser ligeramente inferiores que los obtenidos por la estación de referencia, ambas fuentes de datos presentan patrones similares de variación de la concentración de PM. Por lo tanto, el uso de sistemas de monitoreo basado en sensores de bajo costo podría ser una alternativa viable y económica para la medición de la concentración de PM en el aire, especialmente en áreas donde no se dispone de sistemas de medición de referencia o donde se requiera una mayor densidad de instrumentos en la ciudad.

2.3.4. Limitaciones del sensor de bajo costo

El sensor PMS 1003 de Plantower, al igual que muchos otros sensores de bajo costo, puede enfrentar dificultades al medir las concentraciones de PM_{2.5} en condiciones de elevada humedad relativa. En el estudio de Badura et al. (2018), han señalado que estos sensores pueden registrar micro gotas de agua como partículas de PM_{2.5}, especialmente cuando la humedad relativa supera el 80 % (Badura et al., 2018). Este fenómeno puede llevar a una sobreestimación de las concentraciones de PM_{2.5} bajo condiciones de alta humedad. A pesar de la coincidencia descripta entre los múltiples sensores de bajo costo, e inclusive con la estación de referencia, es crucial reconocer que las lecturas pueden desviarse describiendo errores de hasta un 50 % en las lecturas, en ambientes con alta humedad relativa (Jayaratne et al., 2018; Sayahi et al., 2019). Por este motivo, es esencial considerar estas variaciones al analizar y comparar datos de calidad del aire obtenidos con estos dispositivos.

2.4. Conclusiones

El ensamblaje de los sistemas de monitoreo de bajo costo implicó procedimientos específicos que aseguraron una integración efectiva y funcional. Para la conexión de los sensores a la placa Arduino, se utilizaron cables pelados que se soldaron empleando

mercurio previamente calentado hasta alcanzar estado líquido, lo que garantizó uniones firmes y conductoras al solidificarse. La carga del *firmware* se realizó directamente desde una computadora a la placa Arduino mediante una conexión USB. Este meticuloso proceso de ensamblaje contribuyó a una configuración robusta y fiable del sistema de monitoreo.

Los sistemas de monitoreo basados en sensores de bajo costo presentaron consistencia en las mediciones de concentración de PM_{2.5}, demostrando patrones similares independientemente del sistema utilizado. Asimismo, los datos registrados por los sistemas de bajo costo son también consistentes con los datos recopilados por la estación de referencia de la ciudad, indicando que estos sistemas proporcionan mediciones fiables de la concentración de PM. Esta consistencia de datos entre sistemas entre sí y entre los sistemas y la estación de referencia, sugieren que los sistemas de bajo costo pueden ser empleados para monitorear la concentración de PM_{2.5} en zonas urbanas del Área Metropolitana de Córdoba, constituyendo una alternativa viable y económica para la medición de la calidad del aire, y por lo tanto para la gestión y el control de la calidad del aire de la ciudad. Estos sistemas son el paso inicial para el desarrollo de sistemas de alerta temprana ante condiciones adversas de calidad del aire, paliando la falta de datos de contaminantes y fortaleciendo los pilares sobre los cuales recae la salud publica.

Capítulo 3

Variación espacio temporal de la concentración de aerosoles

3.1. Introducción

Las partículas en suspensión producidas principalmente por vehículos, industrias y reacciones fotoquímicas son uno de los contaminantes más preocupantes en entornos urbanos (Seinfeld & Pandis, 1998). Estas partículas no solo afectan al medio ambiente, sino que también son nocivas para la salud humana. En efecto, las partículas de $2.5 \mu m$ de diámetro aerodinámico o menos ($PM_{2.5}$) tienen la capacidad de penetrar profundamente en los pulmones, acumulándose en los alvéolos, (Xing et al., 2016) o incorporándose directamente al torrente sanguíneo alcanzando otros órganos. Dado que las partículas más pequeñas tienen una mayor superficie efectiva en comparación con las de mayor tamaño, su capacidad de adsorber compuestos tóxicos aumenta, al igual que su peligrosidad. A pesar de ello, actualmente la ciudad de Córdoba no cuenta con un sistema de monitoreo que informe a la población acerca de sus concentraciones. Sin embargo, gracias a la investigación científica local se dispone de numerosos antecedentes en esta ciudad, los cuales además de poner en evidencia

la concentración de contaminantes, su variación y composición, también proveen información respecto de sus impactos en la salud humana (Amarillo et al., 2021; Busso et al., 2021; Mateos et al., 2018, 2019).

Los estándares de emisión que se emplean en la provincia de Córdoba en cuanto a la concentración del PM son similares a los que emplea la Agencia de Protección Ambiental (EPA – EEUU), basándose en la Norma Nacional de Calidad del Aire Ambiente (del inglés, NAAQS) (EPA, 2010). Sin embargo, los lineamientos más recientes sugeridos por la OMS son más restrictivos y es importante tenerlos presentes al analizar los potenciales efectos en la salud de la concentración de PM_{2.5} (OMS, 2021).

Está comprobado que el PM muestra importantes variaciones estacionales, dado a su fuerte dependencia de las condiciones ambientales. Su concentración varía considerablemente en el espacio y el tiempo en distintos entornos (Hofman et al., 2016; P. Kumar et al., 2011; Peters et al., 2014; Van Poppel et al., 2013). En entornos urbanos la variación espacial se encuentra principalmente asociada a la intensidad del tráfico, la topología y la distancia a las fuentes de emisión; Mientras que la variación temporal está asociada al momento en el año, condiciones meteorológicas, dinámica del tráfico y fluctuaciones ambientales (Apparicio et al., 2016; Van den Bossche et al., 2015). Generalmente, su concentración es mayor durante períodos de bajas temperaturas, tal como fue informado en ciudades de Austria (Gomišček et al., 2004) y Grecia (Chaloulakou et al., 2003). De manera similar en Córdoba, la mayor concentración de partículas se produce en la época invernal dado las frecuentes inversiones térmicas que favorecen la concentración y permanencia de contaminantes en las capas inferiores de la atmósfera, particularmente durante la noche y primeras horas de la mañana (Amarillo et al., 2021; Mateos et al., 2019; Olcese & Toselli, 1997; Stein & Toselli, 1996).

Los datos que sustentan el análisis espacio temporal desarrollado en la presente

tesis son series temporales condicionadas a una localización específica. Una serie temporal se define como una secuencia de valores observados a lo largo del tiempo ordenados cronológicamente, en los cuales se destaca una estructura de dependencia entre observaciones (Chatfield, 2000), su condicionante de georreferenciación deriva del aspecto espacial. Si conociendo los valores pasados de la serie no es posible predecir con total certeza el próximo valor de la variable, decimos que la serie es no determinista o aleatoria (Hamilton, 2020; Scotto, 1998). Los objetivos del análisis de series de tiempo son diversos, pudiendo destacarse la predicción, la clasificación y la simulación de nuevos datos (Chatfield et al., 1994; Peña, 2005).

El principal objetivo del presente capítulo es describir la variabilidad tanto espacial como temporal, de la concentración de PM_{2.5} en distintas zonas del área metropolitana de la provincia de Córdoba. De esta forma, se describen las principales características de las series disponibles y se identifican sus patrones de variación espacial y temporal. La finalidad de caracterizar tanto espacial como temporalmente la concentración de PM_{2.5} reside en contemplar todas las características derivadas a la hora de desarrollar el modelado predictivo (capítulo 5), a fin de obtener resultados de elevada capacidad predictiva.

3.2. Materiales y métodos

3.2.1. Resumen gráfico del protocolo analítico de caracterización de la variación espacio temporal de la concentración de PM_{2.5}

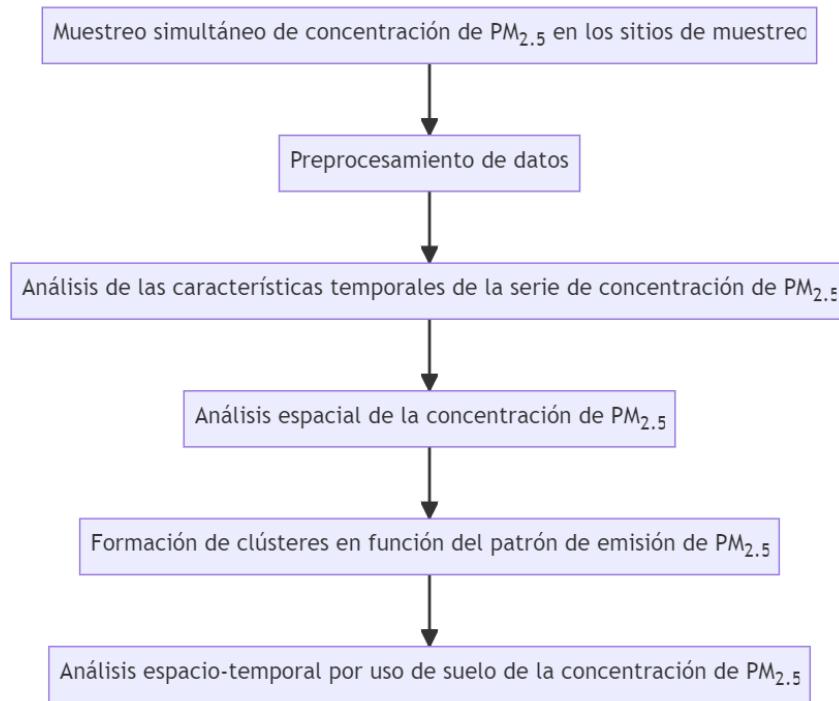


Figura 3.1: Resumen gráfico del protocolo analítico desarrollado en el capítulo 3.

La Figura 3.1 ilustra el protocolo analítico desarrollado en este capítulo con la finalidad de caracterizar la variación espacio temporal de la concentración de PM_{2.5} en zonas del Área Metropolitana de Córdoba, Argentina. En primer lugar, se desarrolló un muestreo simultáneo en el tiempo en los sitios con distinto uso del suelo seleccionados dentro del área de interés. Posteriormente, se llevó a cabo un análisis exploratorio de datos, el cual constó de dos etapas, una vinculada a la variabilidad temporal y otra a la espacial. Además, se desarrolló un estudio de la similitud existente en el patrón de emisión de PM_{2.5} a lo largo del periodo de muestreo entre los distintos sitios, el cual

terminó en la conformación de grupos homogéneos de los sitios en estudio. Finalmente, y a partir del agrupamiento realizado, se concluyó con un análisis que vinculó tanto la variabilidad espacial, como la temporal.

3.2.2. Área de estudio

La ciudad de Córdoba es la capital de la Provincia de Córdoba y la segunda ciudad más importante del país en cuanto a su población, con aproximadamente 1,3 millones de habitantes y una superficie de 576 km² (INDEC, 2010). Su centro urbano está ubicado a 470 metros sobre el nivel del mar y, dado que la ciudad se encuentra rodeada por un cordón montañoso, durante el invierno ocurren eventos de inversión térmica que producen el estancamiento de los contaminantes del aire en una fina capa (200 metros) sobre la superficie (Stein & Toselli, 1996 ; Olcese & Toselli, 2002). Estudios realizados sobre las fuentes de aerosoles dentro de la ciudad revelan que la principal es el tránsito (85 % de la emisión total de partículas), seguido por la resuspensión del polvo de calles y edificios en construcción (Stein & Toselli, 1996 ; Achad et al., 2013). Una gran variedad de industrias se ubican en la periferia de la ciudad, incluyendo metalmecánicas, alimentarias, metalúrgicas, entre otras. El importante crecimiento de la flota vehicular circulante, la falta de control de emisiones y el escaso uso de convertidores catalíticos, ha producido un incremento considerable en las emisiones de partículas (López et al., 2011). A pesar del crecimiento sostenido de la ciudad en las últimas décadas, el microcentro sigue siendo el lugar donde se concentran las actividades económicas, culturales, educativas, financieras, de prestación de servicios y administrativas, lo cual determina una gran demanda de acceso vehicular y transporte público de pasajeros. Se estima que de los aproximadamente 150 millones de pasajeros anuales que transporta el servicio público de transporte, más del 90 % desciende/asciende en el centro (Martin, 2005; Serio et al., 2020).

Para la construcción de la base de datos de PM_{2.5} se llevaron a cabo muestreos

simultáneos de concentración en diez sitios seleccionados, utilizando los equipos de medición presentados en el capítulo anterior, sobre distintos usos de suelo. Cada sitio se monitoreó durante un periodo de un año y tres meses, desde el 30/03/2021 hasta el 01/06/2022. En total, se recopilaron alrededor de 70.000 datos horarios.

Cabe destacar que los datos fueron tomados cada 10 segundos, frecuencia temporal que facilitó la limpieza y procesamiento de los mismos, así como la obtención de medias horarias que representan de mejor forma la tendencia a largo plazo de las series.

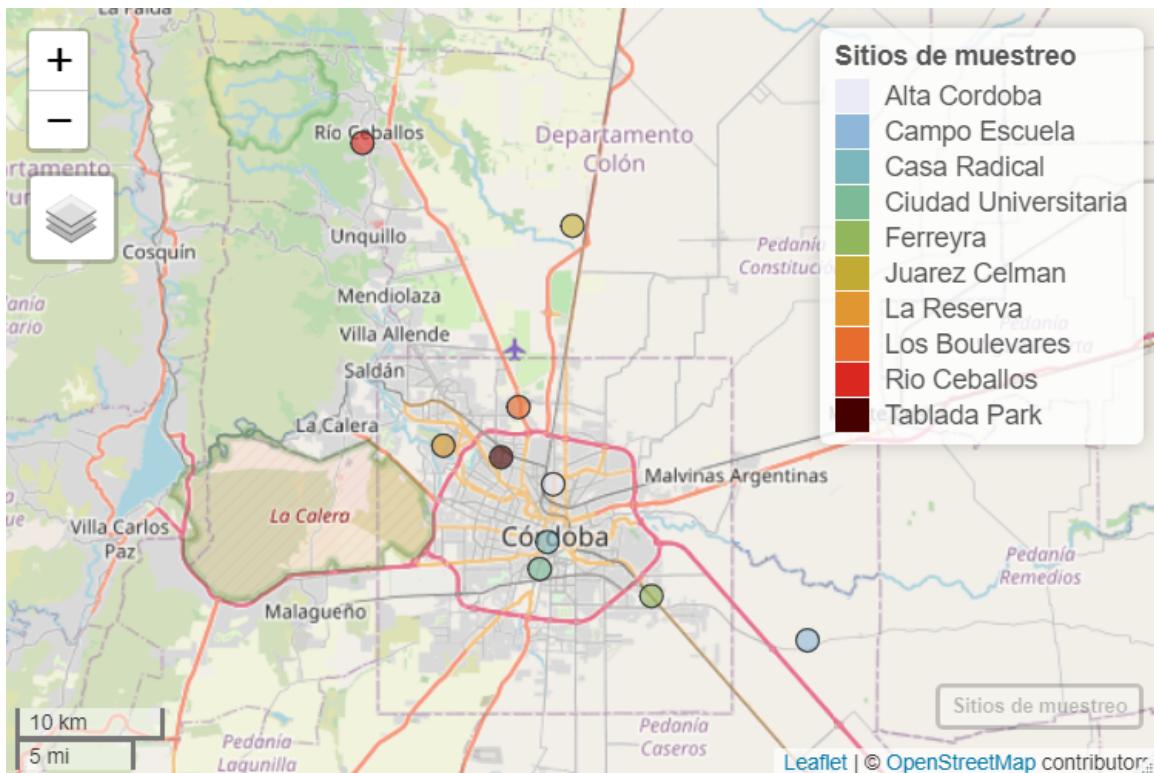


Figura 3.2: Ubicación específica de los sitios de muestreo dentro del área metropolitana de la provincia de Córdoba.

3.2.3. Descripción de cada uno de los sitios de muestreo

Incorporando la interpretación de los inventarios de emisión de PM_{2.5} publicados por Enrique Puliafito para la Argentina (Puliafito, 2009) y una comparación con la clasificación de uso del suelo de zonas climáticas locales (LCZ) (Stewart & Oke, 2012). La clasificación LCZ (*Land Cover Classification for Urban Zones*) es un sistema

exhaustivo diseñado para organizar y describir de manera sistemática el uso del suelo. Cada código LCZ sigue una estructura jerárquica que comienza con un número que representa una categoría principal, abarcando grupos generales de tipos de uso del suelo. Estos números principales se detallan aún más a través de letras de subcategoría, ofreciendo información adicional sobre las características específicas del uso del suelo dentro de esa categoría principal. Esta estructura proporciona una clasificación detallada y precisa de los entornos urbanos, lo que resulta fundamental para una variedad de aplicaciones, desde la planificación urbana hasta la gestión ambiental y el estudio del clima urbano. A continuación, se presenta la descripción detallada de cada uno de los sitios de muestreo.

Tabla 3.1: Descripción sitios de muestreo.

Sitio	Latitud	Longitud	Lugar de emplazamiento
Ferreyra	-31.453423	-64.114533	Techo primer piso
Ciudad Universitaria	-31.436398	-64.194059	Techo primer piso
Los Boulevares	-31.338636	-64.210035	Techo primer piso
Casa Radical	-31.420605	-64.189325	Techo primer piso
Tablada Park	-31.369442	-64.221829	Techo primer piso
Alta Córdoba	-31.385591	-64.184236	Techo primer piso
La Reserva	-31.362222	-64.263194	Techo primer piso
Río Ceballos	-31.177749	-64.321336	Techo primer piso
Campo Escuela	-31.480386	-64.003523	Techo primer piso
Juárez Celman	-31.227705	-64.170459	Techo primer piso

Sitio de muestreo, latitud del punto, longitud del punto, lugar de emplazamiento del sistema de monitoreo de bajo costo en el sitio.

1. Ferreyra: la Empresa Tubos Trans-Electric se ubica en barrio Ferreyra, polo industrial al sureste de la Ciudad, y se dedica a la producción de bienes y servicios

exportables relacionados con el rubro de la energía eléctrica, focalizándose en la producción de transformadores de distribución. La mayoría de las edificaciones vecinas son empresas, las más cercanas se dedican al rubro de la metalurgia, producción de envases de plástico, producción de piezas de automóviles, logística y reparación de vehículos particulares. Además, esta empresa se encuentra situada sobre la avenida Amadeo Sabattini, que es uno de los principales accesos a la Avenida de Circunvalación de Córdoba, facilitando el acceso de camiones y despacho de mercaderías; esta avenida se caracteriza por tener un tráfico vehicular moderado y constante a lo largo de las horas del día. La cobertura de vegetación en la zona es muy escasa. Las principales fuentes de emisión detectadas en el sitio son industrias, tráfico vehicular, residencial y trenes (Puliafito, 2009). La estación de trenes corresponde al Ferrocarril General Bartolomé Mitre de la red ferroviaria argentina, se encuentra actualmente sin operaciones, sin embargo por sus vías transita el servicio Retiro-Córdoba de la empresa estatal Trenes Argentinos Operadora Ferroviaria, aunque no hace parada en esta. El servicio sale desde Retiro San Martín los días miércoles y domingos a las 11:30 horas con llegada Córdoba a las 8:34 horas del día posterior. Dentro de la clasificación LCZ, este sitio sería clasificado como LCZ8₁₀, una combinación de las categorías *large low-rise* y *heavy industry*, donde la primera hace referencia a un sitio con una disposición abierta de grandes edificios de baja y mediana altura (1 a 3 pisos), cuya cubierta del suelo se encuentra mayoritariamente pavimentada, con pocos o ningún árbol y donde los materiales de construcción empleados son acero, hormigón, metal y piedra. Mientras que la segunda y menos importante categoría muestra el estilo industrial del sitio, aunque no es una industria pesada debido a la baja densidad de industrias con respecto a lo que establece la clasificación de usos de suelo.

2. Ciudad Universitaria: la Facultad de Ciencias Exactas, Físicas y Naturales

de la Universidad Nacional de Córdoba se encuentra ubicada en la Ciudad Universitaria, en el centro sur de la Ciudad, sitio donde se concentran las principales dependencias de toda la universidad. Su principal característica es el gran número de personas que frecuentan diariamente la zona usando todo tipo de medios de transporte como motocicletas, automóviles y colectivos, además de otros que no generan emisiones como lo son a pie y en bicicleta. Es una zona con una gran cantidad de espacios verdes donde la cobertura del suelo es compartida por la vegetación, el pavimento y la edificación. La principal fuente de emisión detectada en el sitio es el tránsito vehicular (Puliafito, 2009). Dentro de la clasificación LCZ, este sitio se clasificaría como LCZ5₈, una combinación de las categorías *open midrise* y *large low-rise*, la primera hace referencia a una disposición abierta de edificios de mediana altura (3-9 plantas), donde abunda la cubierta de suelo permeable (plantas bajas, árboles dispersos) y cuyos materiales de construcción son principalmente hormigón, acero, piedra y vidrio. Mientras que la segunda y menos importante categoría hace referencia a un uso del suelo urbano con un notable tráfico de vehículos.

3. Los Boulevares: esta empresa se ubica en barrio Los Boulevares al norte de la Ciudad y se dedica a la mecánica y puesta a punto de camiones. La mayoría de las edificaciones vecinas cercanas son industrias, del rubro de la electrónica, logística, automóviles, productos de limpieza y productos lácteos. Además, este establecimiento se encuentra situado muy cerca de la Avenida La Voz del Interior (E53) que conecta la Avenida de Circunvalación de Córdoba con otras aglomeraciones urbanas como Villa Allende, Mendiola y Río Ceballos, por lo que tiene un tráfico vehicular moderado y constante a lo largo del día. La cobertura de vegetación en la zona es muy escasa con una gran densidad de industrias en sus alrededores. Las principales fuentes de emisión son las industrias, el tráfico vehicular, emisiones residenciales, el tránsito aéreo, emisiones

provenientes de rutas y actividades comerciales (Puliafito, 2009). Dentro de la clasificación LCZ, este sitio se clasificaría como LCZ8₁₀, una combinación de las categorías *large low-rise* y *heavy industry*, donde la primera categoría hace referencia a un sitio con una disposición abierta de grandes edificios de baja y mediana altura (1 a 3 pisos), cuya cubierta del suelo se encuentra mayoritariamente pavimentada, con pocos o ningún árbol y donde principalmente los materiales de construcción empleados son acero, hormigón, metal y piedra. Mientras que la segunda categoría muestra el estilo industrial del sitio.

4. Casa Radical: la Casa Radical, ubicada en el barrio Nueva Córdoba, en el centro de la Ciudad, es la sede del partido Unión Cívica Radical en Córdoba y por lo tanto aquí se desarrollan actividades políticas. En sus alrededores se encuentran gran cantidad de edificios residenciales con numerosos pisos. Además, este sitio se encuentra frente al Boulevard San Juan, una de las calles más transitadas de la ciudad, donde además de edificios residenciales hay numerosas empresas de servicios y locales comerciales. La vegetación que caracteriza la zona es escasa y las principales fuentes de emisión detectadas son el tráfico vehicular, la residencial y el comercio (Puliafito, 2009). Dentro de la clasificación LCZ, este sitio se clasificaría como LCZ2₈, una combinación de las categorías *compact midrise* y *large low-rise*, donde la primera hace referencia a un sitio que posee una mezcla densa de edificios de mediana altura (3-9 pisos), con muy poca vegetación (pocos o ningún árbol), donde la superficie se encuentra mayoritariamente pavimentada y cuyos materiales de construcción son principalmente roca, ladrillo, teja y hormigón. Mientras que la segunda categoría hace referencia a la disposición abierta de edificios que existe.
5. Tablada Park: se trata de una residencia familiar ubicado en el barrio Tablada Park, en la zona noroeste de la Ciudad Capital. Este sitio se encuentra muy

cercano a la avenida Rodríguez Del Busto, con numerosos locales comerciales y supermercados y posee un alto tránsito vehicular. Los alrededores del sitio se caracterizan por ser viviendas residenciales y comercios. Las principales fuentes de emisión detectadas en el sitio son el tránsito vehicular, la residencial y el comercio (Puliafito, 2009), además la vegetación en la zona es moderada. Dentro de la clasificación LCZ, este sitio se clasificaría como LCZ3₆, una combinación de las categorías *compact low rise* y *open low rise*, donde la primera hace referencia a una mezcla densa de edificios de baja altura (1-3 pisos), cuya superficie se encuentra mayoritariamente pavimentada con pocos o ningún árbol y donde los principales materiales de construcción son roca, ladrillo, teja y hormigón. Mientras que la segunda categoría, hace referencia a la presencia de árboles y plantas que caracterizan la zona.

6. Alta Córdoba: se trata de una residencia familiar ubicada en el barrio residencial Alta Córdoba, en la zona centro norte de la Ciudad. Este sitio se encuentra sobre la calle Mariano Fragueiro que aglomera los principales comercios de la zona y posee un alto tránsito vehicular. Los alrededores del sitio se caracterizan por ser casas de familia y locales comerciales. Las principales fuentes de emisión detectadas en el sitio son el tránsito vehicular, la residencial, el comercio y los trenes (Puliafito, 2009), además la vegetación en la zona es baja. Dentro de la clasificación LCZ, este sitio se clasificaría como LCZ3, una combinación de las categorías *compact low rise* y *open low rise*, la cual hace referencia a una mezcla densa de edificios de baja altura (1-3 pisos), cuya superficie se encuentra mayoritariamente pavimentada con pocos o ningún árbol y donde los principales materiales de construcción son piedra, ladrillo, teja y hormigón.
7. La Reserva: la Reserva Natural Urbana San Martín se ubica en la zona noreste de la Ciudad, al margen de la Avenida de Circunvalación de Córdoba; su actividad

principal es proteger y preservar alrededor de 115 hectáreas de bosque nativo, además de concientizar acerca de la protección del medio ambiente. A pesar de ser un área verde donde predomina la cobertura de vegetación, existe una gran cantidad de barrios residenciales rodeándola, generando aportes de contaminantes provenientes del tránsito vehicular y residenciales. Además, este sitio cuenta con un camping que promueve las actividades recreativas al aire libre y que es frecuentado por un gran número de personas principalmente durante los fines de semana. Las principales fuentes de emisión detectadas en el sitio son el tránsito vehicular, y la residencial (Puliafito, 2009). Dentro de la clasificación LCZ, este sitio se clasificaría como LCZB_{A9}, una combinación de las categorías *scattered trees*, *dense trees* y *sparsely built*, donde la primera hace referencia a un sitio con un paisaje ligeramente arbolado de especímenes de hoja caduca y/o de hoja perenne, cuya cobertura del suelo es mayoritariamente permeable (plantas bajas) y su función principal es la de una zona de bosque natural, cultivo de árboles o parque urbano. Mientras que la segunda categoría busca dar información acerca de la gran densidad de árboles que existe en los alrededores y la tercer categoría hace referencia a la disposición dispersa de construcciones de pequeño y mediano tamaño.

8. Río Ceballos: Se trata de un hogar familiar ubicado en las afueras de la Ciudad de Río Ceballos, que se caracteriza por tener un uso del suelo de área verde. Este sitio se encuentra ubicado en una zona de sierras donde predomina la cobertura de vegetación y los principales accesos son calles no pavimentadas. Es un sitio prácticamente prístino, donde las principales fuentes de PM pueden ser consideradas la resuspensión de polvo de las calles e incendios generados como consecuencia de la quema ya sea de residuos o bien de biomasa. Dentro de la clasificación LCZ, este sitio se clasificaría como LCZA_{B9}, una combinación de las categorías *dense trees*, *scattered trees* y *sparsely built*, donde la primera hace

referencia a un sitio con un paisaje muy arbolado, de especímenes de hoja caduca y/o de hoja perenne, cuya cobertura del suelo es mayoritariamente permeable (plantas bajas) y donde su función principal es de zona de bosque natural, cultivo de árboles o parque urbano. Mientras que la segunda categoría representa las zonas poco arboladas que existen en los alrededores y la tercer categoría hace referencia a la disposición dispersa de construcciones de pequeño y mediano tamaño.

9. Campo Escuela: El Campo Escuela de la Facultad de Ciencias Agropecuarias de la Universidad Nacional de Córdoba, se encuentra ubicado camino a Capilla de los Remedios, al sureste de la Ciudad, la población más cercana es Villa Corazón de María y se caracteriza por tener un uso del suelo agrícola. En este sitio se lleva a cabo la formación práctica de los alumnos de la Facultad de Ciencias Agropecuarias por lo cual se desarrollan actividades vinculadas a la agricultura y la ganadería. La cobertura de vegetación es abundante y sus calles no se encuentran pavimentadas. Las principales fuentes emisión de PM detectadas en la zona son la agricultura y la ganadería (Puliafito, 2009). Dentro de la clasificación LCZ, este sitio se clasificaría como LCZD₉, una combinación de las categorías *low plants* y *sparsely built*, donde la primera hace referencia a un sitio con un paisaje principalmente de hierbas, plantas o cultivos, con pocos árboles y cuya función principal es de pasto natural, agrícola o parque urbano. Por otra parte, la segunda categoría refleja disposición dispersa de construcciones de pequeño y mediano tamaño.
10. Juárez Celman: Se trata de un campo productivo ubicado en la ciudad de Juárez Celman al norte de la Ciudad Capital, caracterizado por tener un uso del suelo agrícola. En él se desarrollan actividades productivas agropecuarias, la cobertura de vegetación es alta y sus calles no se encuentran pavimentadas. Este sitio se

encuentra al margen de la Ruta Nacional 9, la cual podría generar un aporte de PM importante debido al tránsito vehicular. Las principales fuentes de PM detectadas en la zona son el tránsito vehicular, la agricultura y la ganadería (Puliafito, 2009). Dentro de la clasificación LCZ, este sitio se clasificaría como LCZD₉, una combinación de las categorías *low plants* y *sparsely built*, donde la primera categoría hace referencia a un sitio con un paisaje principalmente de hierbas, plantas o cultivos, con pocos árboles y cuya función principal es de pasto natural, agrícola o parque urbano. Por otra parte, la segunda categoría refleja disposición dispersa de construcciones de pequeño y mediano tamaño.

A partir de las características mencionadas para cada uno de los sitios de muestreo, puede llevarse a cabo un agrupamiento. De esta manera, el uso del suelo agrícola, agruparía los sitios de Campo Escuela y Juárez Celman, el uso del suelo de área verde se conformaría por los sitios de La Reserva y Río Ceballos, el uso del suelo urbano se formaría por los sitios de Casa Radical, Ciudad Universitaria, Tablada Park y Alta Córdoba y, finalmente el uso del suelo industrial por los sitios de Ferreyra y Los Boulevares.

3.2.4. Normativa

En la provincia de Córdoba, los niveles guía de partículas están regulados en la Resolución 105/17, la cual establece el límite para el PM_{2.5} de 35 $\frac{\mu\text{g}}{\text{m}^3}$ para la media de 24 horas. Es importante destacar que estos valores límite superan ampliamente aquellos sugeridos por la Organización Mundial de la Salud (OMS) de 15 $\frac{\mu\text{g}}{\text{m}^3}$ para la media de 24 horas y 5 $\frac{\mu\text{g}}{\text{m}^3}$ para la media anual (OMS, 2021). La brecha existente entre ambos límites se debe a abordajes conceptualmente diferentes. En el caso de la resolución de la provincia se trata de estándares para la inmisión, mientras que los sugeridos por la OMS son lineamientos guía, indicando aquellos valores para los cuales no se observan efectos adversos en la salud (OMS, 2021).

Tabla 3.2: Normativas y estándares de valores guía para la concentración de PM_{2.5}.

PM _{2.5} $\frac{\mu\text{g}}{\text{m}^3}$	OMS	EPA	Res.105/17
Promedio 24 horas	15	35	35
Promedio anual	5	12	-

3.2.5. Preprocesamiento de datos

Previo al análisis del patrón promedio de la serie de concentración de PM_{2.5} y la construcción de una base de datos a escala temporal horaria se llevó a cabo un proceso de depuración de datos. Este trabajo busca comprender la variabilidad macro o promedio de la serie de concentración de PM_{2.5} y no la debida a causas atípicas de la microescala. Por este motivo, se optó por implementar un estricto protocolo de depuración de datos puntuales aberrantes, excluyendo aquellos eventos puntuales atípicos que no representan la variabilidad promedio de la concentración de PM_{2.5}. La fundamentación del enfoque de preprocesamiento implementado se halla sustentado en los objetivos específicos de la tesis, donde se intenta la búsqueda de la relación existente entre la variabilidad promedio de la concentración de PM_{2.5} a lo largo del año con distintas variables meteorológicas. Las influencias atípicas de la microescala no aportan a describir el patrón general o promedio de los datos, a menos que estas se repitan de forma reiterada, entonces dejarían de ser atípicas y formarían parte del patrón general o promedio. Por otra parte, el objetivo final que se persigue es el modelado predictivo de la concentración de PM_{2.5} ante condiciones normales o típicas y no para eventos anómalos. El preprocesamiento de los datos fue dividido en dos etapas, una primera en la que se examinaron valores atípicos globales y otra en la que el estudio se centró en el análisis de datos aberrantes en distintas escalas temporales.

Una vez obtenidos los datos de PM_{2.5} espaciados por un intervalo de tiempo de 10

segundos, se procedió a filtrar aquellos valores que se alejaban en más de 5 desvíos estándares de la media, y que cuyo patrón de incremento de la concentración era en forma de un salto repentino, es decir que el crecimiento de la concentración no fue paulatino y continuo en el tiempo. Siendo la media de toda la base de datos de 14.39 $\frac{\mu g}{m^3}$ y la desviación estándar de 102.77 $\frac{\mu g}{m^3}$, se procedió a eliminar aquellos valores que superaban el valor de 528.26 $\frac{\mu g}{m^3}$ describiendo un salto repentino de concentración. El salto repentino se define como un crecimiento de la concentración que no es paulatino ni continuo en el tiempo con respecto a la tendencia descripta por los vecinos del dato en cuestión. Este análisis se llevó a cabo por medio de una evaluación visual. Cada uno de los datos detectado como atípicos fue comparado con sus vecinos cercanos, considerando la tendencia representada por estos puntos. La decisión de eliminar o mantener un dato en la base de datos se basó en esta evaluación. De esta forma se logró eliminar aquellos valores atípicos cuya concentración se veía fuertemente influenciada por eventos particulares, propios de la microescala que no representaban la concentración para el sitio en un momento determinado. Posteriormente, se procedió a calcular las medias horarias para los distintos sitios.

A partir de la base de datos horaria, se procedió a ajustar distintas distribuciones a la variable respuesta con el objetivo de determinar cuál era la que mejor la caracterizaba. La distribución que menores valores del criterio de información de Akaike (AIC) y del criterio de información bayesiano (BIC) entregó fue la *gamma*, hallazgo que se presenta en la sección de resultados del presente capítulo. A partir de este resultado fue que se ajustó un Modelo Lineal Generalizado Mixto que vincula la variable respuesta y las distintas variables indicadoras de la escala temporal (hora, día, mes, semana, estación) como variables de efectos aleatorios (Ecuación 3.1). Los resultados indicaron que la mayor variabilidad de la base de datos era explicada por las variables mes y hora. De esta forma, para cada uno de los sitios de muestreo se procedió a agrupar los datos en cada uno de los meses y las distintas horas del día con el objetivo de calcular

medidas de resumen estadísticas (media y desvío estándar) para cada una de las combinaciones. Posteriormente, se filtraron aquellos valores que, para un determinado sitio de muestreo, en un mes particular y a una hora determinada se alejaba en más de 5 desvíos estándares de la media de cada caso particular.

$$\log y_i = \beta_1 \text{Hora}_i + \beta_2 \text{Dia}_i + \beta_3 \text{Mes}_i + \beta_4 \text{Semana}_i + \beta_5 \text{Estacion}_i \quad (3.1)$$

Ecuación 3.1: Modelo lineal generalizado mixto de efectos aleatorios ajustado para determinar los porcentajes de varianza estimados.

Componentes:

- **Hora:** Variable numérica que toma valores en un rango de 1 a 24, indicando la variación de las horas a lo largo del día. Esta variable representa la hora del día en la que cada dato fue recopilado.
- **Dia:** Variable numérica que toma valores en un rango de 1 a 7, indicando la variación de los días a lo largo de la semana. Esta variable representa el día de la semana en el que cada dato fue recopilado.
- **Mes:** Variable numérica que toma valores en un rango de 1 a 12, indicando la variación de los meses a lo largo del año. Esta variable representa el mes del año en el que cada dato fue recopilado.
- **Semana:** Variable numérica que toma valores en un rango de 1 a 52, indicando la variación de las semanas a lo largo del año. Esta variable representa la semana del año en la que cada dato fue recopilado.
- **Estacion:** Variable numérica que toma valores en un rango de 1 a 4, indicando la variación de la estación a lo largo del año. Esta variable representa la estación climática del año en la que cada dato fue recopilado.

De esta forma, se realizó una limpieza de valores atípicos a nivel global seguida por una limpieza temporal que aseguró que los datos no estuvieran influenciados por efectos aislados de la microescala que pudieran atentar contra el patrón general de los mismos.

3.2.6. Protocolo estadístico

Formación de clústeres de series temporales

El objetivo que persigue la formación de clústeres o creación de aglomerados a partir de las series temporales de concentración de PM_{2.5} es generar un agrupamiento en función del patrón de emisión a lo largo de las horas del día en el periodo de

muestreo. De esta forma, los conglomerados además de permitir describir y encontrar similitudes entre los sitios, facilitarán el modelado predictivo de esta variable.

Para llevar a cabo el análisis de conglomerados entre las series temporales de PM pertenecientes a los distintos sitios de muestreo, fue necesario transformarlas con el fin de obtener datos que cumpliesen con los requerimientos de entrada al algoritmo. En primer lugar, se llevó a cabo su normalización, la cual constó de un centrado por medio de la sustracción de la media y un escalamiento a través de la división por la desviación estándar. A partir de las series de tiempo normalizadas se procedió a calcular la matriz de distancias implementando la *Dinamic Time Wrapping* (DTW), medida de distancia conocida por ser la más utilizada para medir similitud entre series temporales, que permite obtener buenos ajustes incluso frente a desfasajes en el tiempo (Sarda-Espinosa et al., 2018). Para la obtención de los clústeres se empleó la función *hclust()* del paquete *stats* (R Core Team, 2021), la cual recibe como entrada la matriz de distancias previamente calculada. Para la conformación de los conglomerados se empleó el método de Ward, el cual permite definir las nuevas distancias entre clústeres una vez que se ha generado un nuevo agrupamiento. La función implementada utiliza el criterio introducido por Murtagh y Legendre (Murtagh & Legendre, 2014), el cual busca minimizar la varianza dentro del conglomerado, reflejada en la suma de errores cuadrados. De esta forma, los clústeres se combinan de acuerdo con la distancia más pequeña entre los clústeres, logrando resultados más compactos y fáciles de diferenciar.

En cada etapa del proceso de formación de conglomerados, los dos clústeres más cercanos se fusionan dando origen a un nuevo grupo, una vez formado se recalculan las distancias, dando origen a una nueva matriz de distancias a partir de la cual el ajuste continúa. El proceso se repite iterativamente hasta que todo el conjunto de datos se aglomera en un único clúster. De esta forma se obtiene el histograma de conglomeración que refleja el historial de la formación de los distintos agrupamientos. Cabe destacar que es un algoritmo del tipo aglomerativo, es decir que inicialmente cada

sitio se considera como un clúster individual, de esta forma se parte de 10 clústeres distintos hasta que iterativamente se agrupan en uno único.

Para corroborar los resultados obtenidos, se implementó el algoritmo no jerárquico *K-shapes* a partir de la función *tsclust()* de la librería *dTwClust* (Sarda-Espinosa et al., 2018), que definiendo los parámetros requeridos lleva a cabo la aplicación propuesta por Paparrizos y Gravano (Paparrizos & Gravano, 2015). Las principales características de este algoritmo son que la medida de distancia se encuentra basada en la correlación cruzada entre dos series temporales y que el proceso de agrupación emplea un enfoque iterativo.

3.3. Resultados y discusión

Filtrado de *outliers* globales

En esta etapa de preprocesamiento se excluyeron las observaciones de PM_{2.5} que se encontraban alejados de la media por más de 5 desvíos estándares. Como resultado de este proceso, se excluyeron un total de 218 datos, cuyos valores de PM_{2.5} se correspondían con eventos y/o mediciones anómalas sucedidos en los sitios durante el muestreo, como por ejemplo, incendios cercanos o caídas del sensor. Estos valores resultaban notablemente superiores al resto de los datos, interpretándose como una fuente de ruido para el análisis.

Filtrado de *outliers* temporales

En primer lugar, se determinó cuál era la distribución que mejor ajustaba a la variable respuesta. Los resultados evidenciaron que la distribución *gamma* logró minimizar los criterios de AIC y BIC. Una vez establecida la distribución se procedió a ajustar un Modelo Lineal Generalizado Mixto, con el objetivo de determinar cuáles variables temporales explicaban mayor porcentaje de varianza (Tabla 3.3).

Tabla 3.3: Variabilidad explicada por las componentes temporales presentes en la base de datos.

Componente	Varianza	Desviación estándar	Porcentaje de varianza
Hora	0.070	0.265	4.975
Mes	0.172	0.414	12.184
Día	0.011	0.106	0.797
Semana	0.001	0.026	0.049
Estación	0.033	0.173	2.083
Residual	1.155	1.075	79.912

Componente temporal, varianza estimada, desviación estándar estimada y porcentaje de varianza explicado relativo al total.

A partir de la Tabla 3.3, se desprende que los meses del año son la variable temporal que más influye en la explicación de la variabilidad de la variable respuesta, seguida de las horas del día.

Una vez obtenidos estos resultados se procedió a filtrar aquellas observaciones que, para un mismo sitio, mes del año y hora del día, se encontraban a más de cinco desvíos estándares de su respectiva media. El filtro aplicado removió un total de 1184 valores, lo que representan el 1.55 % de la base de datos.

3.3.1. Características temporales de la serie de PM_{2.5}

Este apartado busca demostrar la variabilidad temporal existente en la concentración media horaria de PM_{2.5}, en las distintas escalas temporales, con el objetivo de identificar las tendencias, ciclos y patrones temporales que influyen en esta serie de tiempo. Es fundamental describir la variabilidad existente a lo largo de las horas del día en las distintas escalas temporales, ya que esto permitirá comprender la dinámica de variación de la variable respuesta. Cabe destacar que las características temporales

abordadas en el análisis son promedios generales obtenidos a partir de todos los datos.

Análisis horario

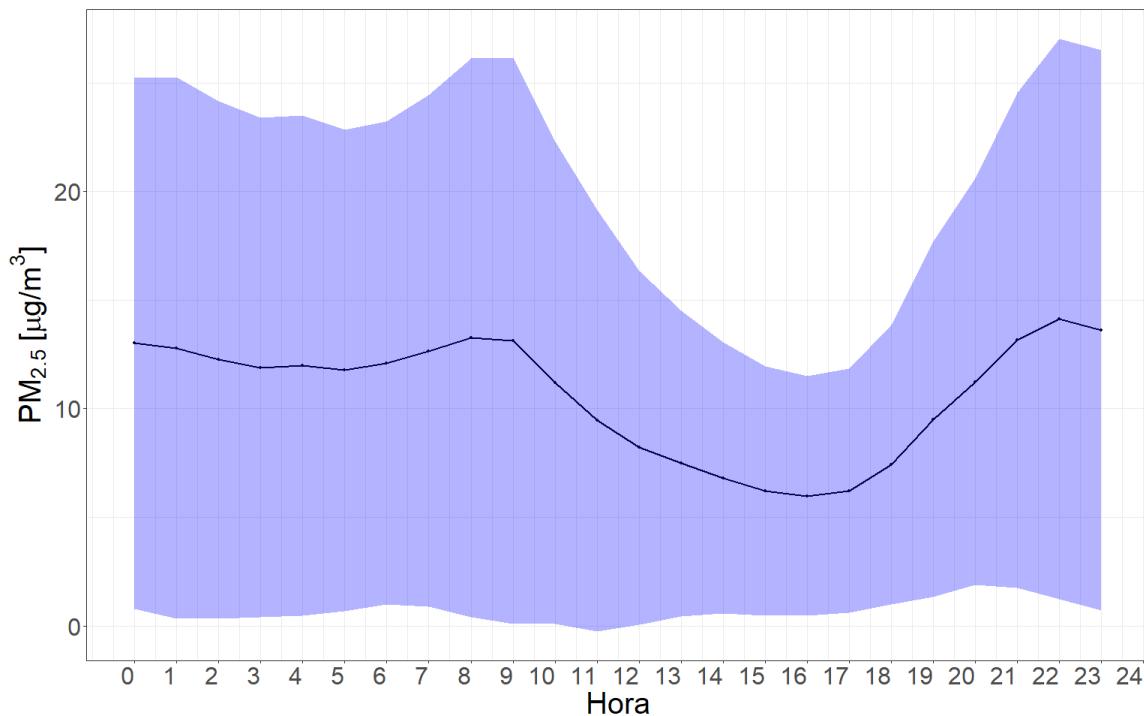


Figura 3.3: Perfil de concentración promedio horario de PM_{2.5} a lo largo de las horas en los sitios en estudio a lo largo de todo el periodo de muestreo, en azul se graficó la desviación estándar.

La Figura 3.3 ilustra el patrón de variabilidad de la concentración promedio horaria de PM_{2.5} a lo largo de las horas del día. Este se caracteriza por mostrar un mínimo de concentración entre las 15 y 17 horas, delimitado por dos picos que ocurren por la mañana entre las 7 y 9 horas y por la noche entre las 21 y 23 horas. Además, puede observarse que el menor desvió estándar promedio se encuentra localizado en la zona de los baches, mientras que los picos que encierran este bache destacan el mayor. Este patrón de concentración horaria puede estar influenciado por diversos factores, como los patrones de emisión de las fuentes de contaminación, la estabilidad de la atmósfera, los patrones de tráfico, entre otros.

Análisis diario

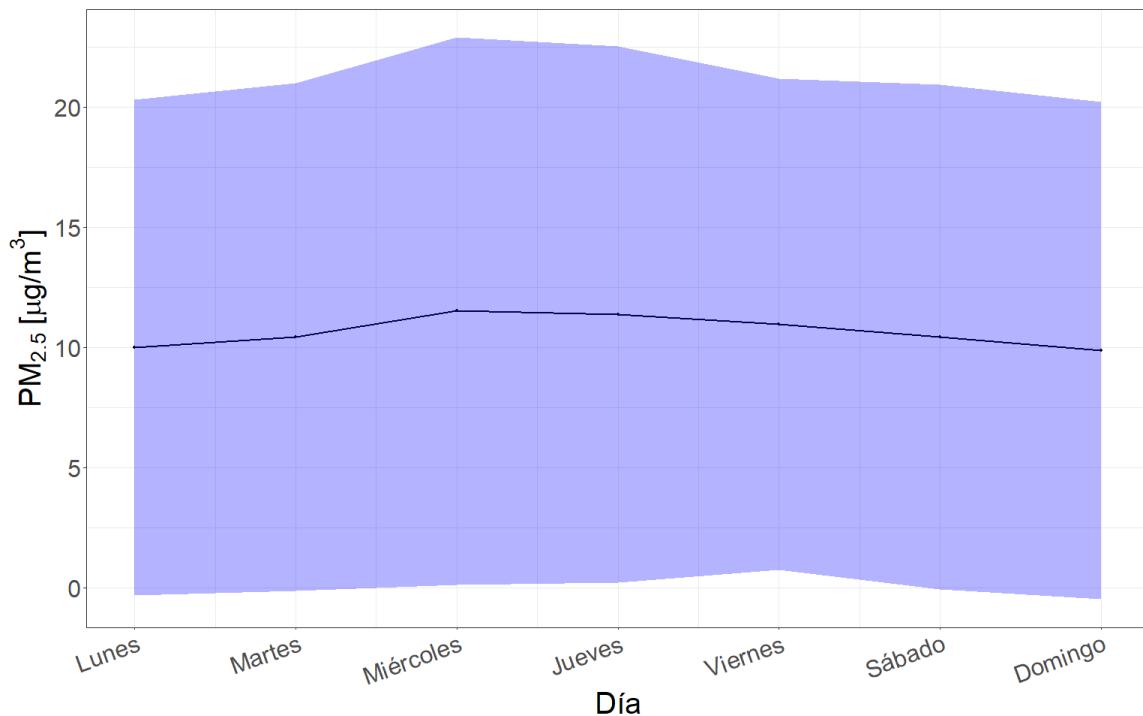


Figura 3.4: Perfil de concentración promedio diario de PM_{2.5} a lo largo de los días de la semana en los sitios en estudio a lo largo de todo el periodo de muestreo, en azul se graficó la desviación estándar.

En la Figura 3.4 puede apreciarse el patrón de variabilidad de la concentración promedio diaria de PM_{2.5} a lo largo de los días de la semana. La tendencia observada es un aumento en las concentraciones de PM_{2.5} a medida que nos adentramos en la semana, con un pico máximo de concentración los días miércoles y jueves. En contraste, los días domingo se registran las concentraciones más bajas de PM_{2.5} y a partir de ese día, éstas aumentan gradualmente hasta alcanzar su punto máximo durante la mitad de la semana. A partir de ese punto, se observa un descenso gradual en las concentraciones hasta llegar al fin de semana. Puede apreciarse que el desvió estándar promedio para los distintos días de la semana es similar a lo largo de los distintos días y demuestra una magnitud apreciable con respecto a la media general. Los valores de promedios diarios son coincidentes con los datos observados en la Figura 3.3, y

muestran una elevada variabilidad asociada a la notable variación de la concentración a lo largo de las horas del día. Esta tendencia semanal podría encontrarse principalmente influenciada por los patrones de movilidad de la población, estrechamente vinculados con su comportamiento a lo largo de los días de la semana.

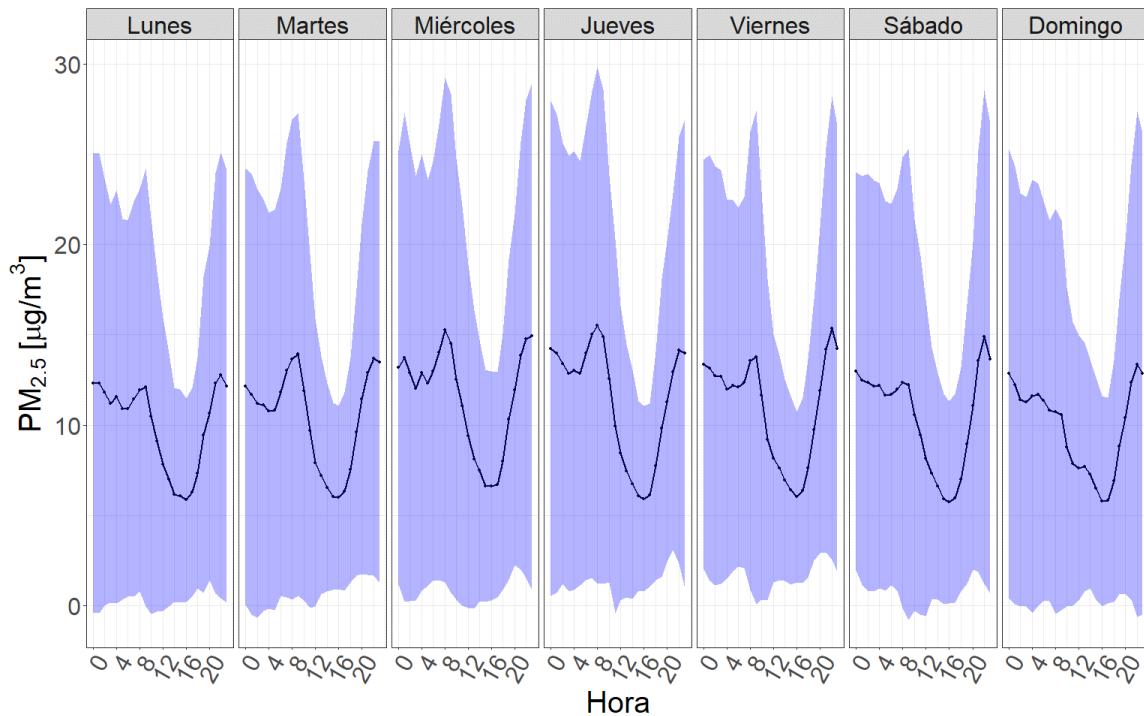


Figura 3.5: Perfil de concentración promedio horario de $\text{PM}_{2.5}$ en cada día de la semana a lo largo de la semana en los sitios en estudio a lo largo de todo el periodo de muestreo, en azul se graficó la desviación estándar.

En la Figura 3.5 puede apreciarse el patrón de variabilidad de la concentración de $\text{PM}_{2.5}$ a lo largo de las horas del día en los distintos días de la semana. Puede observarse una clara estacionalidad, demostrando que las concentraciones de $\text{PM}_{2.5}$ en una hora determinada son similares a lo largo de los distintos días de la semana. Además, las concentraciones guardan un mayor parentesco en su magnitud mientras más próximos entre sí se encuentren los días de la semana. Por otra parte, puede notarse un aumento de la variabilidad de los datos los días miércoles y jueves, variabilidad que genera que sus promedios diarios sean superiores a los de los demás días de la semana.

Análisis mensual

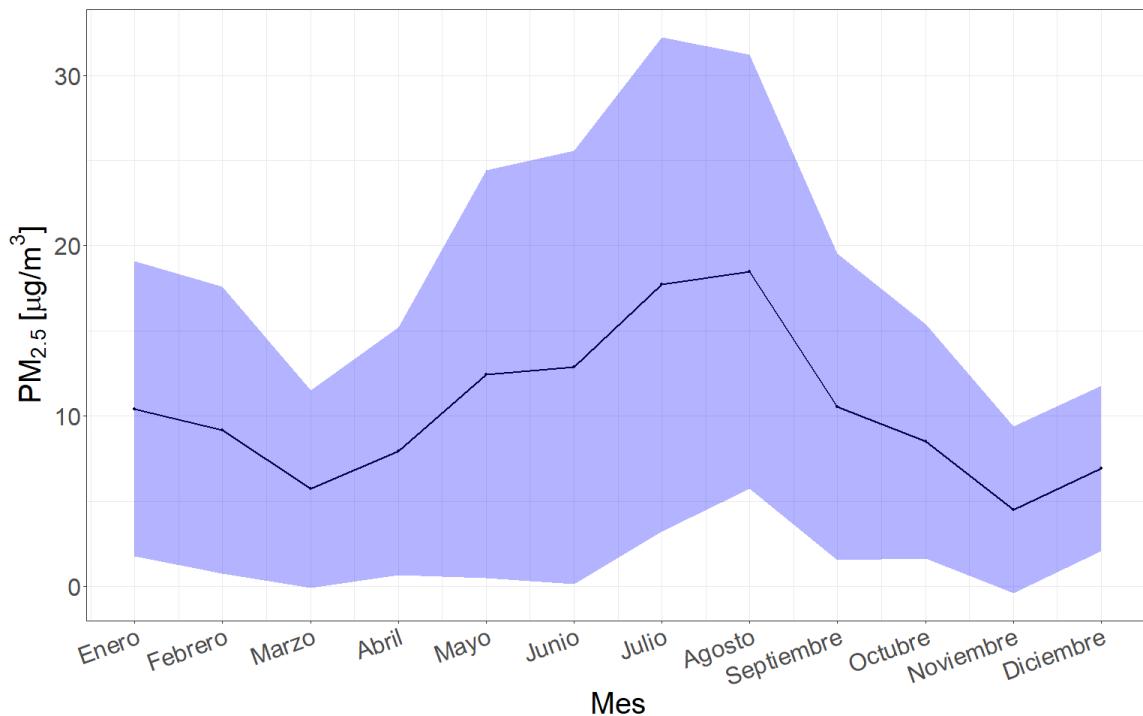


Figura 3.6: Perfil de concentración promedio mensual de PM_{2.5} a lo largo del año en los sitios en estudio a lo largo de todo el periodo de muestreo, en azul se graficó la desviación estándar.

En la Figura 3.6, se aprecian las concentraciones de PM_{2.5} promedio mensuales registradas a lo largo de los distintos meses del año. Puede notarse que las mayores concentraciones fueron registradas durante los meses de mayo, junio, julio y agosto, mientras que en los meses de noviembre, marzo y diciembre se registraron las menores concentraciones. Como era de esperarse la variabilidad aumenta notablemente en aquellos meses que demuestran mayor media de concentración de PM_{2.5}.

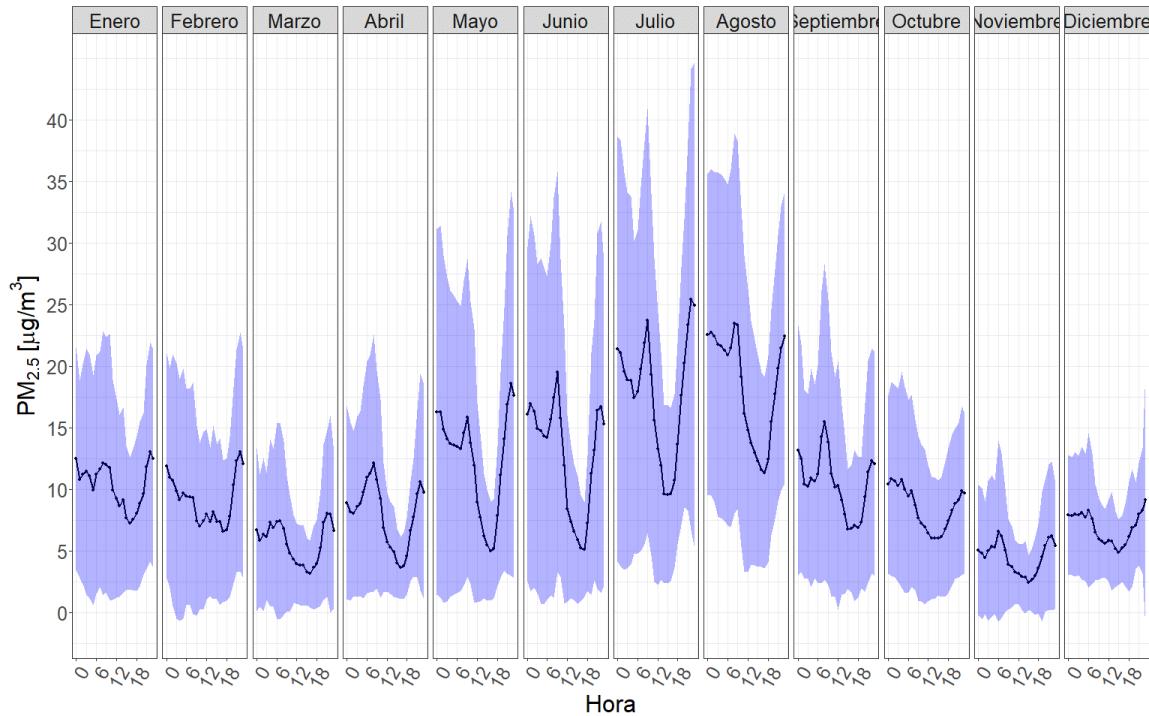


Figura 3.7: Perfil de concentración promedio horario de PM_{2.5} para cada mes a lo largo del año en los sitios en estudio a lo largo de todo el periodo de muestreo, en azul se graficó la desviación estándar.

En la Figura 3.7 se destaca el patrón de variabilidad en la concentración del PM_{2.5} a lo largo de las horas del día en los distintos meses del año, el cual refleja que el ciclo de concentración de PM_{2.5} a lo largo de las horas del día muestra una forma similar a lo largo de los distintos meses. En otras palabras, las caídas de concentración se dan de entre las 12 y 18 horas, mientras que los picos máximos pueden apreciarse entre las 6 y 12 horas y luego de las 18 horas. Además, el parentesco entre los ciclos guarda una mayor correlación mientras más cercanos se encuentren los meses a lo largo del año.

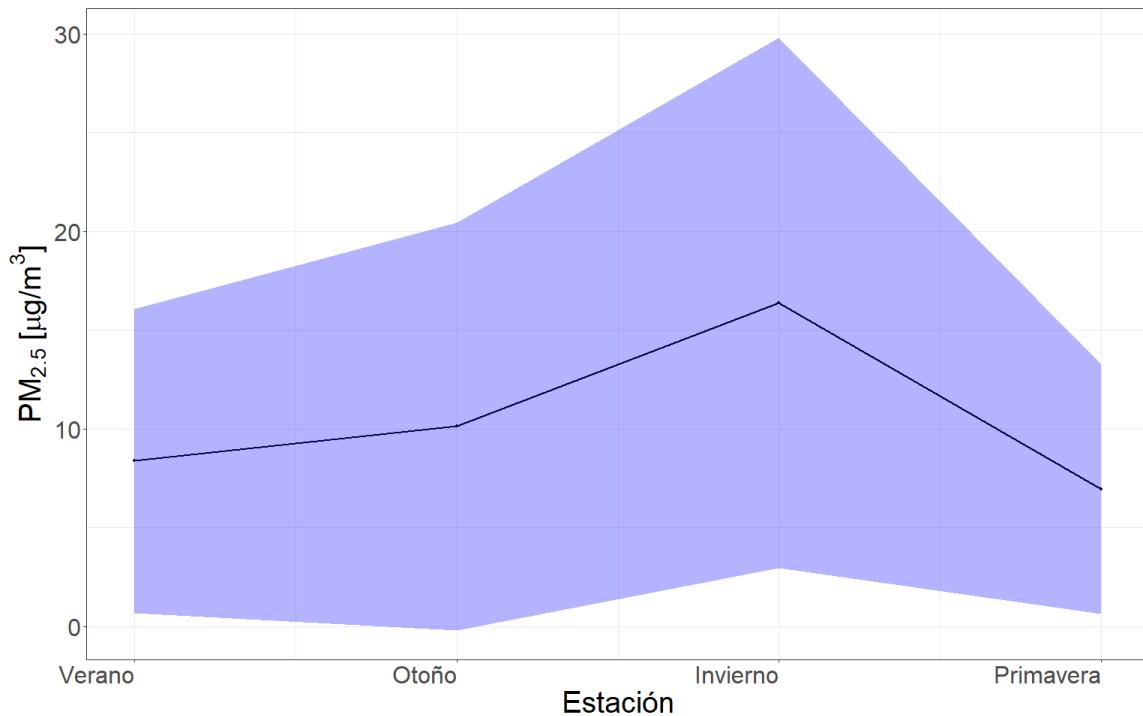
Análisis por estación del año

Figura 3.8: Perfil de concentración promedio por estación de PM_{2.5} a lo largo del año en los sitios en estudio a lo largo de todo el periodo de muestreo, en azul se graficó la desviación estándar.

La Figura 3.8 muestra el patrón de variabilidad en la concentración de PM_{2.5} promedio a lo largo de las diferentes estaciones del año. Puede notarse que las mayores concentraciones se registraron en las estaciones de otoño e invierno. Por otro lado, las menores concentraciones se registraron durante la primavera y el verano. Estos resultados indican una clara dependencia de la concentración de PM_{2.5} con la estación del año.

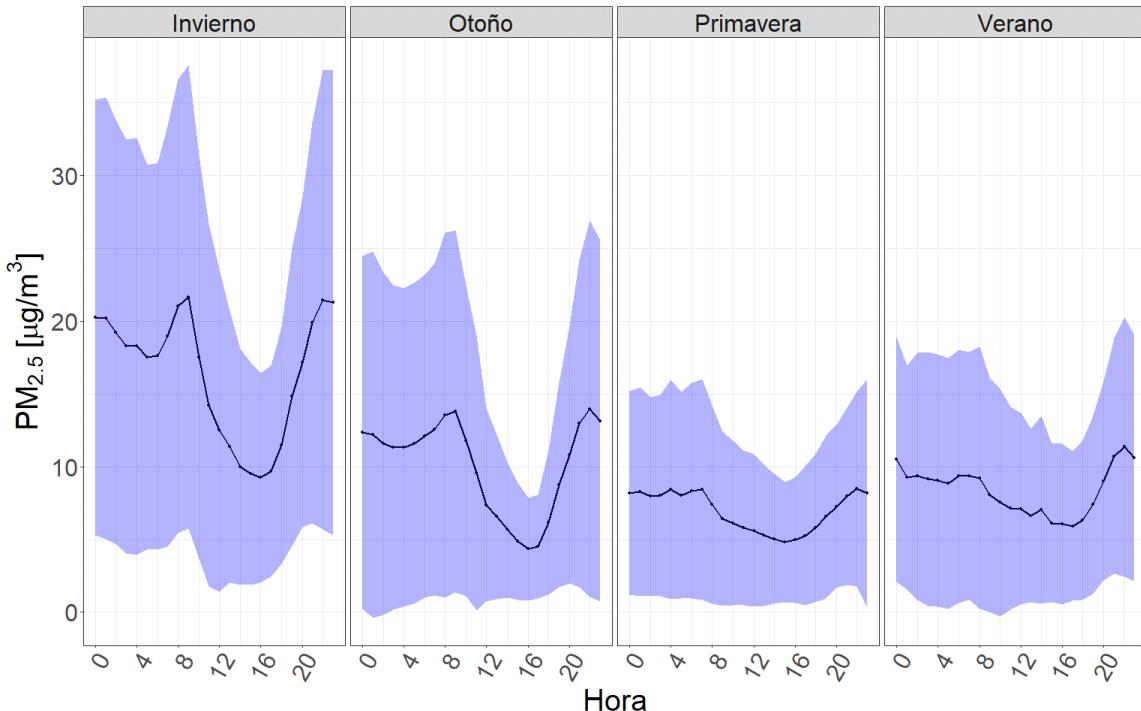


Figura 3.9: Perfil de concentración promedio horario de PM_{2.5} para cada estación a lo largo del año en los sitios en estudio a lo largo de todo el periodo de muestreo, en azul se graficó la desviación estándar.

La Figura 3.9 destaca el patrón de variabilidad de la concentración de PM_{2.5} a lo largo de las horas del día en las distintas estaciones del año. Puede notarse que la concentración de PM_{2.5} varía en magnitud según la estación del año, siendo más elevadas en invierno y otoño, mientras que en primavera y verano las concentraciones son menores. Además, se evidencia que, dentro de cada estación, las concentraciones de PM_{2.5} presentan un patrón similar a lo largo de las distintas horas del día. Este patrón evidencia que a lo largo de las horas del día ocurren dos picos de concentraciones, con valores de menor magnitud entre las 12 y 20 horas. Es pertinente destacar en la Figura 3.9, el aumento de la variabilidad que sufre la concentración de PM_{2.5} promedio en las estaciones donde se observan las mayores concentraciones a lo largo de las horas del día.

Las características temporales descritas en la serie de tiempo, demuestran la existencia de ciclos y patrones repetitivos en su perfil de variación. La estadística

descriptiva sugiere que la variación de las condiciones meteorológicas a lo largo de los meses y estaciones del año influye notablemente en la concentración de PM_{2.5}, este análisis es objeto de estudio en el capítulo 4.

Numerosos estudios han demostrando que durante el invierno, la concentración de PM_{2.5} es significativamente mayor en comparación con la estación cálida (A. Kumar et al., 2020; Lalchandani et al., 2022; Rasheed et al., 2015; Saxena et al., 2017). En algunas ciudades, la concentración de PM_{2.5} en invierno puede ser varias veces mayor que en verano. Por ejemplo, en Santiago de Chile, la concentración de PM_{2.5} en la estación fría fue tres veces mayor que en la estación cálida (Sax et al., 2007), mientras que en Cracovia, Polonia, la concentración en la estación fría fue 2.6 veces superior a la estación cálida (Samek et al., 2020). Los resultados relevados en la presente tesis muestran que para la ciudad de Córdoba, Argentina, la concentración de PM_{2.5} registrada durante la estación fría fue 1.75 veces mayor que la de la estación cálida durante todo el período analizado.

Esta variación estacional en los niveles de PM_{2.5} podría atribuirse a factores como la meteorología, la quema de biomasa agrícola y la calefacción doméstica, entre otros (Jain et al., 2020; Tiwari et al., 2013). Sin embargo, en la ciudad de Córdoba, la principal fuente de variabilidad estacional son las condiciones meteorológicas durante el invierno, que favorecen la ocurrencia de inversiones térmicas provocando que los contaminantes y aerosoles queden atrapados en una capa inferior a los 200 m (López et al., 2011).

3.3.2. Análisis espacial de la concentración de PM_{2.5}

Descripción de la concentración anual de PM_{2.5}

Tabla 3.4: Concentración anual promedio de PM_{2.5} en los sitios de muestreo dentro del área metropolitana de la provincia de Córdoba. Los valores se encuentran en $\frac{\mu\text{g}}{\text{m}^3}$.

Medida descriptiva	Concentración
Media	10.65
Desviación estándar	10.67
Máximo	150.28
Mínimo	0.53
Muestras	69587

La concentración promedio de PM_{2.5} para todo el período de muestreo fue de 10.65 $\frac{\mu\text{g}}{\text{m}^3}$ considerando todos los sitios de muestreo, con un rango de valores desde 0 a 150 $\frac{\mu\text{g}}{\text{m}^3}$ (Tabla 3.4). Estos datos indican que la concentración de PM_{2.5} supera el límite anual de PM_{2.5} sugerido por la OMS. Sin embargo, se encuentra por debajo del límite propuesto por la Agencia de Protección Ambiental de los Estados Unidos (EPA) (Tabla 3.2).

La magnitud de la concentración promedio anual de PM_{2.5} encontrada para la provincia de Córdoba (30/03/2021-30/03/2022) fue inferior a la observada en otras ciudades del mundo. La población informada para cada una de las ciudades se encuentra referida al año de la cita. Por ejemplo, en Riohacha (188.014 habitantes), Colombia, se registró una concentración de 14.57 $\frac{\mu\text{g}}{\text{m}^3}$ (04/11/2010-02/02/2011) (Rojano et al., 2013). En Brasil, ciudades como Río de Janeiro (15.5 $\frac{\mu\text{g}}{\text{m}^3}$) (6.320.446 habitantes) (01/06/2007-01/08/2008) y Porto Alegre (13.9 $\frac{\mu\text{g}}{\text{m}^3}$) (1.420.667 habitantes) (01/10/2001-01/12/2002) también presentaron niveles superiores a los observados en la provincia de Córdoba (Miranda et al., 2012). En Serbia, se reportaron concentraciones de promedios anuales de entre 13.93 $\frac{\mu\text{g}}{\text{m}^3}$ y 28.91 $\frac{\mu\text{g}}{\text{m}^3}$ para el periodo comprendido entre los años 2001 y 2016 (Stanojević et al., 2019) (6.834.000 habitantes). Por último, en Estambul, Turquía, se

registraron concentraciones de promedios anuales de entre $47.8 \frac{\mu\text{g}}{\text{m}^3}$ y $93.0 \frac{\mu\text{g}}{\text{m}^3}$ para el periodo comprendido entre el 2007 y 2008, en distintas regiones de la ciudad (Rasheed et al., 2015) (15.019.000 habitantes).

Descripción de la concentración promedio por sitio de muestreo

Tabla 3.5: Concentración promedio anual de PM_{2.5} en cada uno de los sitios de muestreo (30/03/2021-30/03/2022).

Sitio	Concentración	Desviación estándar
Ferreyra	15.10	± 12.70
Los Boulevares	13.89	± 12.80
Tablada Park	13.60	± 12.70
Casa Radical	12.10	± 10.20
Ciudad Universitaria	11.50	± 10.70
La Reserva	10.30	± 10.20
Alta Córdoba	9.08	± 8.60
Campo Escuela	8.88	± 9.22
Río Ceballos	6.40	± 6.36
Juárez Celman	6.00	± 6.88

Sitio de muestreo, concentración promedio anual de PM_{2.5} y desviación estándar estimada.

La concentración media anual registrada para todos los sitios de muestreo supera los límites sugeridos por la OMS, lo que indica una situación preocupante en términos de calidad del aire en la provincia de Córdoba. Por otro lado, la mitad de los sitios supera el estándar anual propuesto por la EPA (Tabla 3.5).

Es importante destacar que la variación en las concentraciones anuales de PM_{2.5} promedio en los distintos sitios de muestreo ponen de manifiesto tanto las diferencias como las similitudes entre ellos respecto a la concentración de PM. Esta tendencia

podría atribuirse a características que comparten estos sitios, lo que resulta fundamental para comprender la influencia de los mismos factores en la calidad del aire en estas ubicaciones (Tabla 3.5). Este hecho podría estar sugiriendo que la clasificación a priori establecida por uso del suelo en función de las características de cada uno de los sitios, podría no ser correcta.

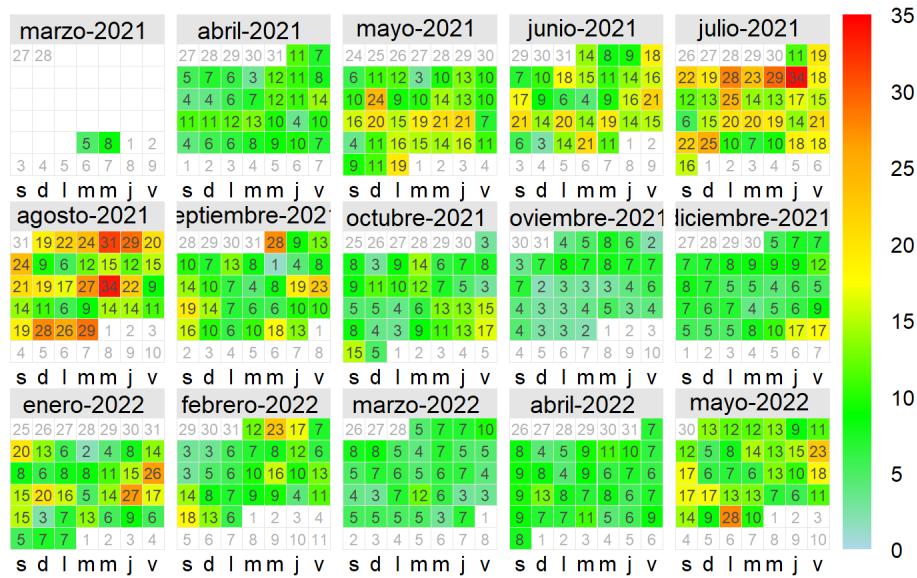


Figura 3.10: Diagrama de concentración de PM_{2.5} promedio diario para los sitios del área metropolitana de la provincia de Córdoba a lo largo del año (*Calendar Plot*).

En la Figura 3.10, pueden identificarse meses con condiciones adversas de calidad de aire, ya sea porque las condiciones atmosféricas favorecen el aumento de la concentración del contaminante o bien impiden su dispersión. Estos son los meses de mayo, junio, julio y agosto, durante los cuales se describen concentraciones sostenidas y elevadas de PM_{2.5} respecto al resto del año. Por otra parte, puede notarse la numerosa cantidad de veces en el año en que la concentración media diaria de PM_{2.5} se encuentra por encima de las normativas provinciales de calidad de aire (Tabla 3.2). Este hecho indica que existe un riesgo potencial para la población en relación con la calidad del aire en el área metropolitana de la ciudad de Córdoba.

Tabla 3.6: Concentración promedio por estación de PM_{2.5} en cada sitio de muestreo.

Sitio	Invierno	Otoño	Primavera	Verano
Ferreyra	20.5±14.7	15.2±12.8	10.3± 8.24	13.5±11.0
Los Boulevares	18.5±14.5	14.1±12.6	7.06±6.19	9.02±7.46
Tablada Park	17.6±14.3	14.4±13.7	6.80±5.91	6.30±4.25
Casa Radical	19.8±12.9	11.7± 9.52	7.85±5.85	9.53±7.40
Ciudad Universitaria	17.6±13.6	10.9±10.3	7.64±6.69	9.92±8.34
La Reserva	20.0±14.3	10.2± 9.42	5.79±4.87	7.69±6.35
Alta Córdoba	15.1±12.7	10.7± 9.79	6.00±3.73	6.72±5.83
Campo Escuela	12.4±11.4	6.54±6.99	6.36±5.29	6.21±5.12
Río Ceballos	9.65±7.96	4.78±4.69	5.20±5.28	6.59±6.34
Juárez Celman	11.3± 9.79	4.71±4.43	4.30±6.38	4.30±4.54

Sitio de muestreo, concentración promedio estacional de PM_{2.5} y desviación estándar.

A partir de la Tabla 3.6, puede observarse que las diferencias en la concentración de PM_{2.5} entre sitios de muestreo son mayores en invierno y otoño que en primavera y verano. Específicamente, se observan concentraciones máximas durante la estación invernal y mínimas durante la estación estival. Además, se destaca que los mayores desvíos estándares ocurren durante el invierno, demostrando ser la estación del año con características más variables de concentración de PM_{2.5}.

Formación de clústeres a partir del patrón de emisión

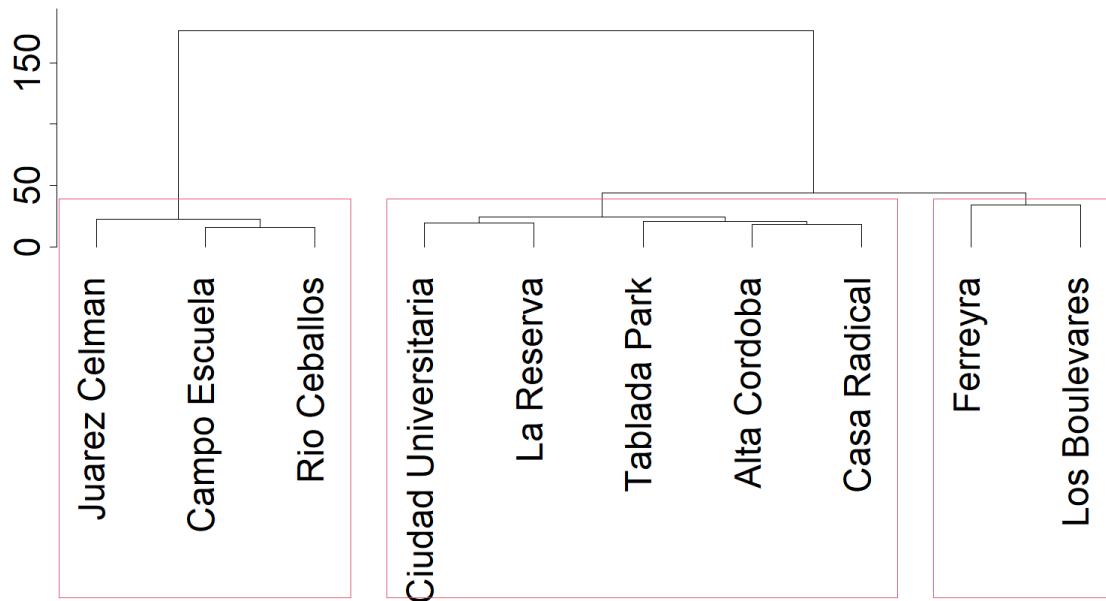


Figura 3.11: Dendrograma resultante del análisis de clúster de los sitios de muestreo en función de su patrón de variación temporal de emisión de PM_{2.5}.

A fin de corroborar la exactitud de la clasificación de los sitios de monitoreo de acuerdo a sus usos de suelo, se realizó un análisis de clúster. La Figura 3.11 muestra el agrupamiento generado por el método jerárquico, a partir de las series temporales de concentración de PM_{2.5} en los distintos sitios de muestreo. De acuerdo con las características propias de cada uno de los sitios de muestreo, su ubicación espacial, el uso del suelo declarado y sus principales fuentes de emisión, se determinó la conformación de tres grupos bien diferenciados, que se clasificaron en tres categorías en función de su uso del suelo: agrícola/área verde, urbano e industrial. Para validar los resultados obtenidos, se implementó el algoritmo aglomerativo *K-shapes*, obteniendo como mejor agrupamiento la conformación de tres grupos con los mismos constituyentes que derivó el método jerárquico.

Los sitios que se agrupan en primer lugar son los de Río Ceballos y Campo Escuela, que mostraron mayor similitud entre todos los sitios de muestreo. Luego, se agrupan los sitios de Alta Córdoba y Casa Radical evidenciando ser los segundos más similares. Posteriormente se agrupan los sitios de Ciudad Universitaria y La Reserva, y luego se añade el sitio Tablada Park al segundo clúster conformado. Posteriormente, se conforman los grupos agrícola/área verde y urbano, a través de la inclusión del sitio Juárez Celman al primer grupo formado dando origen al uso del suelo agrícola/área verde. La unión de los grupos restantes previamente conformados da origen al grupo con uso del suelo urbano. Finalmente, se conforma el grupo de uso del suelo industrial, formado por los sitios Ferreyra y Los Boulevares. En el tramo final del dendrograma hacia la formación de un único grupo, puede apreciarse que el clúster urbano e industrial se combinan, mostrando una mayor similitud entre estos usos de suelo que con respecto al uso del suelo agrícola/área verde, el cual es el último en agruparse para formar el conglomerado final.

La variable de agrupamiento derivada permite describir de mejor manera el parentesco existente en el patrón de emisión en los distintos sitios y será de gran importancia a la hora de generar un modelo predictivo para la concentración horaria de PM_{2.5}. A continuación, se analizan cada uno de los usos de suelo con el objetivo de visualizar si las características previamente observadas para los patrones de variación temporal generales en el área metropolitana de la provincia de Córdoba se mantienen constantes en cada uno de ellos.

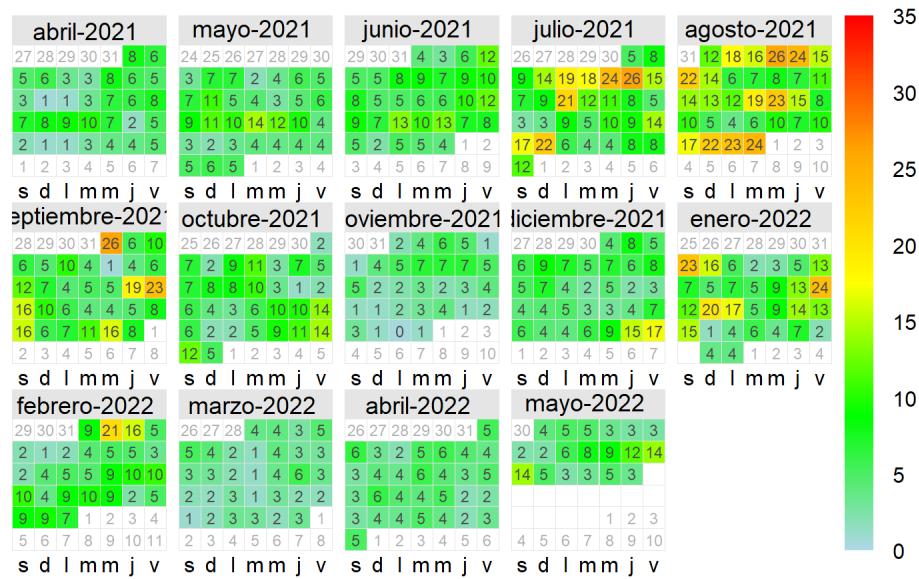


Figura 3.12: Diagrama de concentración de PM_{2.5} promedio diario para los sitios cuyo uso del suelo quedó agrupado como agrícola/área Verde, a lo largo del año (*Calendar Plot*).

Las concentraciones registradas a lo largo del año en las áreas agrícolas/verdes son menores que las encontradas en el área metropolitana de la ciudad de Córdoba (Figura 3.10), lo cual podría deberse tanto a la baja cantidad de fuentes de emisión característica de estos sitios, como a la densa vegetación existente. Esto es un hallazgo prometedor, indicando que estas áreas pueden funcionar como “pulmones verdes” en la ciudad, proporcionando un ambiente más saludable para las personas que viven en los alrededores o que frecuentan estas áreas cotidianamente (M. Chen et al., 2019; M. Wang & Wang, 2021).

A pesar de ello, las condiciones meteorológicas características de la época invernal, también afectan a estas zonas verdes (Figura 3.6), las concentraciones más altas de partículas fueron registradas durante los meses de julio y agosto (Figura 3.12). Sin embargo, en ninguno de estos días se llega a superar el estándar de 24 horas que establece la legislación provincial.

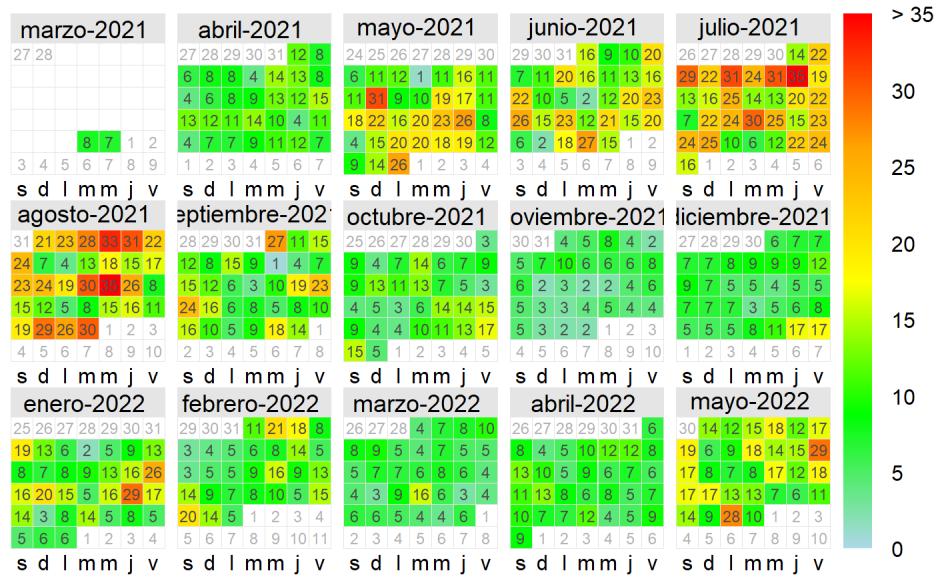


Figura 3.13: Diagrama de concentración de PM_{2.5} promedio diario para los sitios cuyo uso del suelo quedo agrupado como urbano, a lo largo del año (*Calendar Plot*).

En los puntos de muestreo urbanos, también los niveles máximos de PM_{2.5} fueron encontrados en los meses de mayo a agosto. Este hecho sugiere que, si bien el uso del suelo determina el rango de variación de la concentración del contaminante, su variabilidad se encuentra principalmente vinculada a la fluctuación de las condiciones meteorológicas a lo largo del año (Figura 3.13).

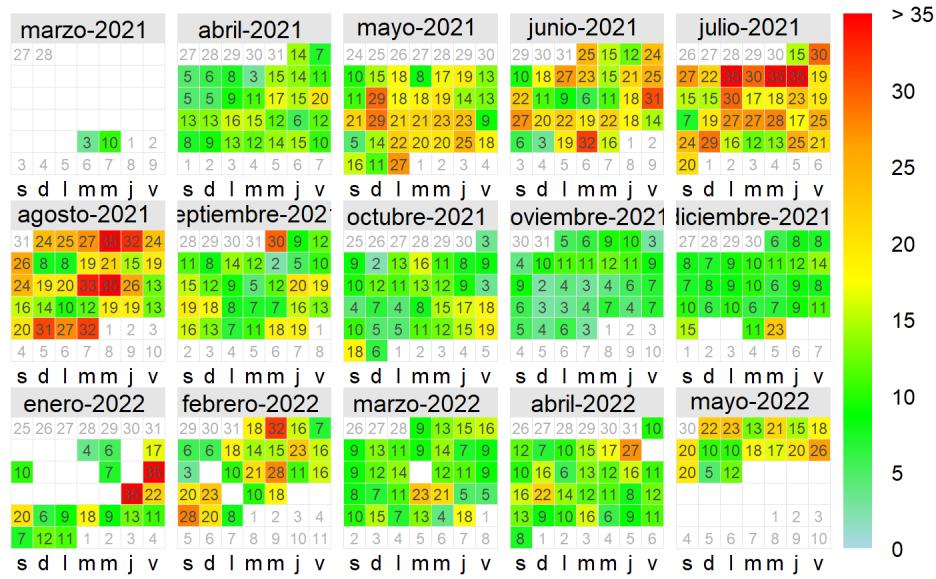


Figura 3.14: Diagrama de concentración de PM_{2.5} promedio diario para los sitios cuyo uso del suelo quedo agrupado como industrial, a lo largo del año (*Calendar Plot*).

Las concentraciones promedio diarias para el uso del suelo industrial fueron muy superiores a las encontradas en los demás usos de suelo, aunque se evidencia el mismo patrón de variación mensual. Puede notarse un notable aumento de las concentraciones desde mayo hasta agosto, mientras que durante el resto del año se observan concentraciones de menor magnitud. Además, también se destacan eventos aislados de elevadas concentraciones durante los meses de enero y febrero (Figura 3.14).

Esto demuestra que, aunque el uso del suelo influye en la concentración de PM_{2.5}, la variabilidad mensual de las condiciones meteorológicas es uno de los principales factores que contribuyen a explicar la variabilidad del contaminante. Los resultados obtenidos sugieren que la temporada invernal es especialmente crítica en términos de calidad del aire en la ciudad de Córdoba, por lo que se debería prestar especial atención a la planificación y gestión de políticas ambientales en este período.

A partir de los resultados obtenidos se demuestra que los patrones de variación de la concentración de PM_{2.5} en las distintas escalas temporales se mantienen constantes

en los distintos usos de suelo. A pesar de ello, la magnitud de su influencia varía en función del uso del suelo, lo cual podría encontrarse vinculado a los efectos de la microescala propios de cada uno de los sitios como fuentes de emisión y cobertura de vegetación.

De esta forma se evidencia que aunque el uso del suelo determina el rango de variación de la concentración del contaminante, la variabilidad del mismo se encuentra estrechamente vinculada a la fluctuación de las condiciones meteorológicas a lo largo del año.

Es importante destacar que la magnitud de la concentración varía notablemente en los distintos usos de suelo, por lo que es necesario considerar de forma diferencial estos usos al explicar la variabilidad del contaminante en cuestión.

Tabla 3.7: Cantidad de veces en que la concentración promedio diaria de PM_{2.5} supera las normativas guía en cada uno de los usos de suelos derivados del análisis de conglomerados.

Uso del suelo	OMS	CBA
Industrial	152 / 38.97 %	9 / 1.75 %
Urbano	111 / 26.93 %	4 / 0.55 %
Agrícola/Área Verde	35 / 0.85 %	0 / 0 %

Uso del suelo, cantidad de veces que supera los límites propuestos por la Organización Mundial de la Salud (OMS) / porcentaje de tiempo en que se vio superada dicha normativa y cantidad de veces que supera la resolución 105/17 de la Provincia de Córdoba (CBA) / porcentaje de tiempo en que se vio superada dicha normativa.

En la Tabla 3.7 se presenta la frecuencia en que se supera la concentración establecida por la legislación provincial, así como las concentraciones sugeridas por la OMS en cada uno de los usos de suelo considerados. Puede notarse que esta frecuencia es mayor en el uso del suelo industrial, seguido por el uso del suelo urbano y finalmente el agrícola/área verde. Es relevante destacar que en este último caso, las concentraciones de PM_{2.5} nunca superaron los valores límites mencionados por la normativa de la

provincia de Córdoba. Estos resultados permiten afirmar que el uso del suelo es un factor determinante en la variabilidad de la concentración de PM_{2.5} y que a su vez es un indicador de la cantidad de veces en las que las normativas se superan. Por lo tanto, puede establecerse que el uso del suelo que propicia concentraciones de mayor adversidad para la salud humana es en primer lugar el industrial, seguido por el urbano y en tercer lugar el agrícola/área verde.

3.3.3. Análisis espacio-temporal por uso del suelo de PM_{2.5}

Análisis descriptivo de las series temporales para cada uso del suelo

Tabla 3.8: Medidas descriptivas para las series temporales promedio para los distintos usos de suelo derivados a partir del análisis de conglomerados.

Medida	Urbano	Agrícola/Área Verde	Industrial
Media	11.54	6.78	14.22
Desviación estándar	9.90	6.10	11.59
Máximo	81.80	47.02	93.17
Mínimo	0.01	0.00	0.07
Percentil 50	8.41	4.92	10.81
Percentil 65	12.03	7.07	15.51
Percentil 75	15.72	9.30	19.96
Percentil 90	25.49	15.16	30.85
Percentil 95	32.11	19.57	37.71
RI	7.31	4.38	9.15

Rango intercuartílico (RI).

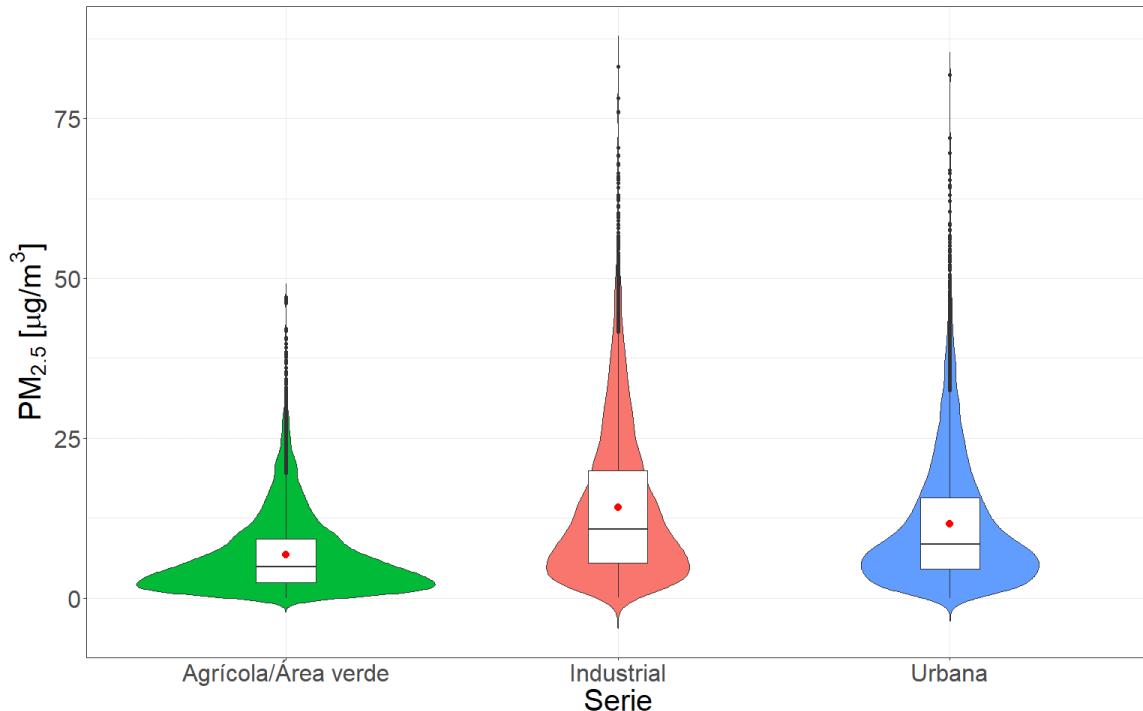


Figura 3.15: Diagrama de caja que compara la concentración promedio horaria de PM_{2.5} en los distintos usos de suelo derivados del análisis de conglomerados. La media de la serie se encuentra indicada por el punto rojo dentro de la caja.

A partir de los datos presentados en la Tabla 3.8 puede apreciarse que la media para las series de datos industrial y urbana son superiores a la de la serie agrícola y área verde. Este hecho podría indicar que en entornos urbanos e industriales los niveles de partículas son mayores que los de áreas verdes. Sin embargo, esta medida de posición puede verse fuertemente influenciada por valores extremos, lo que reduce su utilidad para representar los distintos ambientes. Para realizar tal comparación se ha empleado la mediana o percentil 50. Puede apreciarse que en entornos urbanos el percentil 50 es igual a 8.41 $\frac{\mu g}{m^3}$, indicando que la mitad de los datos toman un valor igual o menor a este. Por otra parte, la mediana para los ambientes agrícolas y área verdes es de 4.92 $\frac{\mu g}{m^3}$ y para la industrial es de 10.81 $\frac{\mu g}{m^3}$. Estos datos sugieren que los entornos urbanos e industriales muestran concentraciones superiores que las provenientes de áreas agrícolas y verdes. Además, la concentración en entornos industriales demuestra

ser superior a la registrada en entornos urbanos.

También puede apreciarse que el valor del percentil 90 para la serie de datos agrícola y de área verde es prácticamente similar al percentil 75 de la serie de datos urbana y este es a su vez similar al percentil 65 para la serie de datos industrial. Este hecho evidencia en términos generales que un mayor porcentaje de datos se encuentra por debajo del mismo límite numérico, de esta forma indica que la serie de datos industrial toma valores de mayor magnitud que la serie de datos urbana y estas a su vez demuestran mayores magnitudes que la serie de datos agrícola/área verde.

Estos datos permiten concluir que la serie agrícola y área verde posee menor dispersión de datos respecto a su media o lo que es equivalente, mayor concentración de datos en torno a su media, respecto de las series urbana e industrial. A su vez, los registros de la serie urbana se encuentran más concentrados respecto a su media en comparación con la de entornos industriales. Esto indica que hay una mayor dispersión de datos respecto de su media en la serie industrial que en la serie urbana. Asimismo, es importante destacar que estos resultados son específicos para la zona y el período de estudio y que podrían variar en otras regiones geográficas y períodos de tiempo.

La dispersión de datos en las series también se ven reflejadas en los gráficos de cajas (Figura 3.15). Existe una diferencia en el tamaño de la caja para los distintos usos de suelo, reflejando una mayor dispersión en la serie industrial, seguida por la urbana, mientras que la serie agrícola y de área verde presenta una mayor concentración de datos en torno a su media. Además, puede apreciarse la influencia de valores extremos en todas las series generando un corrimiento de la media respecto de la mediana. Este corrimiento es mayor para la serie de datos industrial y urbana, indicando un mayor porcentaje de valores extremos.

Es importante destacar que la forma de la caja y la posición de la media y la mediana indican una distribución del tipo asimétrica positiva o bien sesgada derecha para todas las series, demostrando una mayor concentración de datos en valores de

baja magnitud de la variable respuesta. Además, puede apreciarse que para la serie de datos agrícola y de área verde, la densidad de datos en la parte inferior es superior que para el resto de los usos de suelo. Por otra parte, la serie de datos urbana presenta una mayor densidad de datos en su parte inferior en comparación con la industrial, confirmando que en entornos industriales la concentración de PM_{2.5} toma valores de mayores magnitudes, además de tener una mayor dispersión respecto a su media. De esta forma, la serie de datos industrial resulta ser la más dispersa en comparación con la urbana y la agrícola/área verde, cuyos datos se encuentran notablemente concentrados en torno a su media con una baja dispersión.

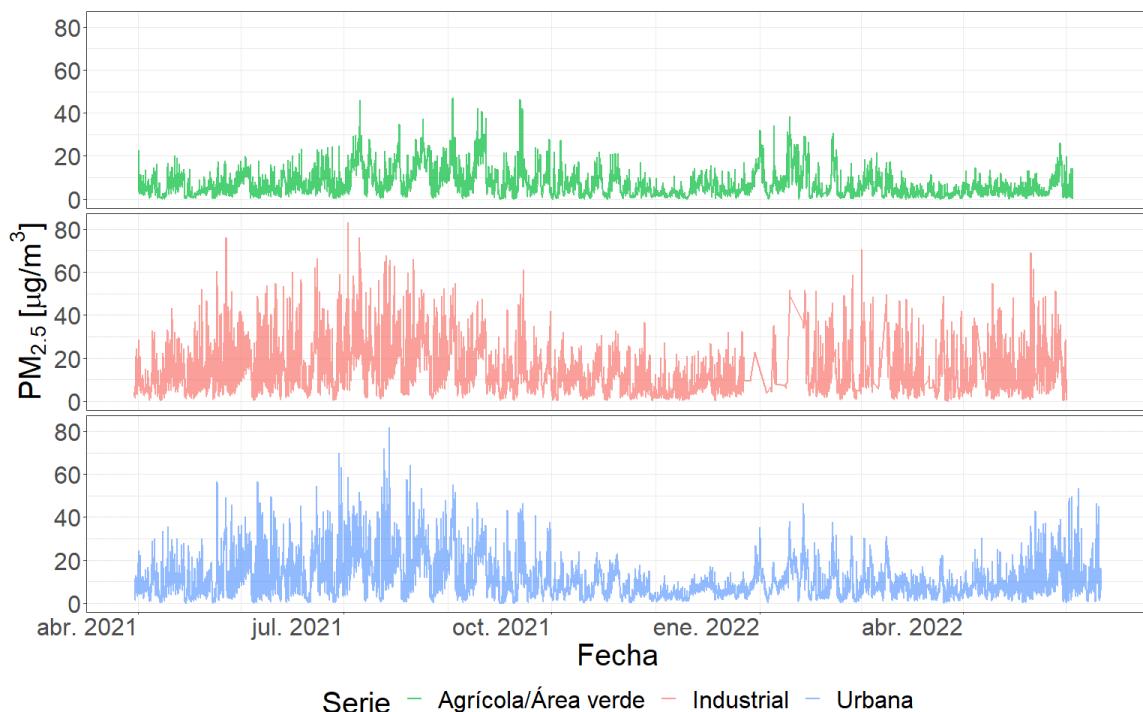


Figura 3.16: Serie de tiempo de la concentración promedio horaria de PM_{2.5} en los distintos usos de suelo derivados del análisis de conglomerados.

Las series temporales pueden diferenciarse en función de su comportamiento a lo largo del tiempo (Figura 3.16). La serie agrícola y de área verde muestra un rango de variación menor que las otras series, lo cual refuerza la idea de que los datos se encuentran concentrados alrededor de su media. Además, los valores de las series

industrial y urbana oscilan en un rango de magnitud mayor en comparación con la serie agrícola y de área verde.

La serie industrial demuestra una mayor inestabilidad, con una sucesión repetitiva de picos de elevada magnitud y baches mínimos a lo largo del tiempo, algo que también puede observarse en la serie urbana, aunque en menor medida. Estos resultados confirman que la serie industrial es la que presenta valores de mayor magnitud, dispersión respecto a su media e inestabilidad. La serie urbana presenta características similares, aunque en menor medida.

En contraste, la serie de datos agrícola/área verde posee picos menos marcados, en términos de la magnitud de los valores que toma, con una menor fluctuación entre ellos, lo cual sugiere que los datos se encuentran más concentrados respecto de su media y que poseen una menor dispersión. Estos resultados son coherentes con lo encontrado en el análisis de clúster (Figura 3.11), que indicaba mayor similitud entre las series urbanas e industriales, diferenciándose de la serie agrícola/área verde.

Análisis por uso del suelo en el tiempo

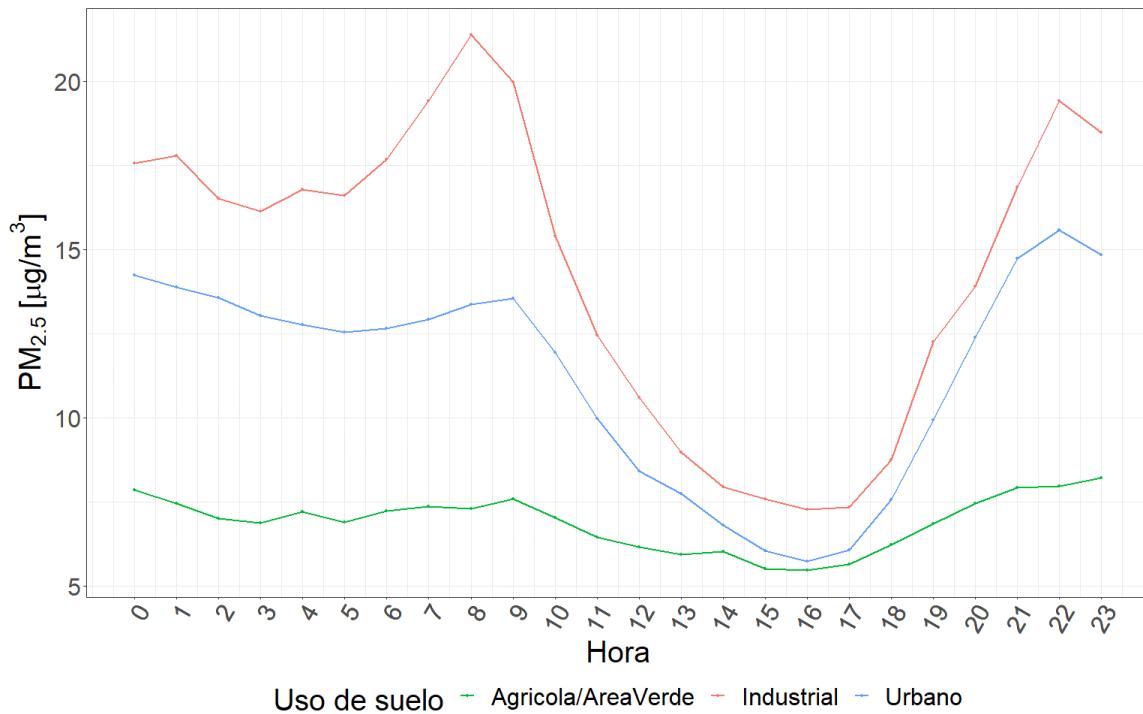


Figura 3.17: Perfil de concentración promedio horario de PM_{2.5} a lo largo de los días, discriminado por los usos de suelo derivados del análisis de conglomerados a lo largo de todo el periodo de muestreo.

El patrón de variabilidad en la concentración promedio horaria de PM_{2.5} a lo largo del día es evidente en los sitios con uso del suelo urbano e industrial, aunque se presenta en menor medida en el uso del suelo agrícola/área verde (Figura 3.17). En el caso de los sitios con uso del suelo urbano e industrial, la concentración promedio horaria de PM_{2.5} muestra un aumento en la mañana, disminuye por la tarde, para aumentar nuevamente hacia la noche. Mientras que en los sitios con uso del suelo agrícola/área verde, el patrón de variabilidad a lo largo del día se encuentra presente, aunque es menos evidente (Figura 3.17).

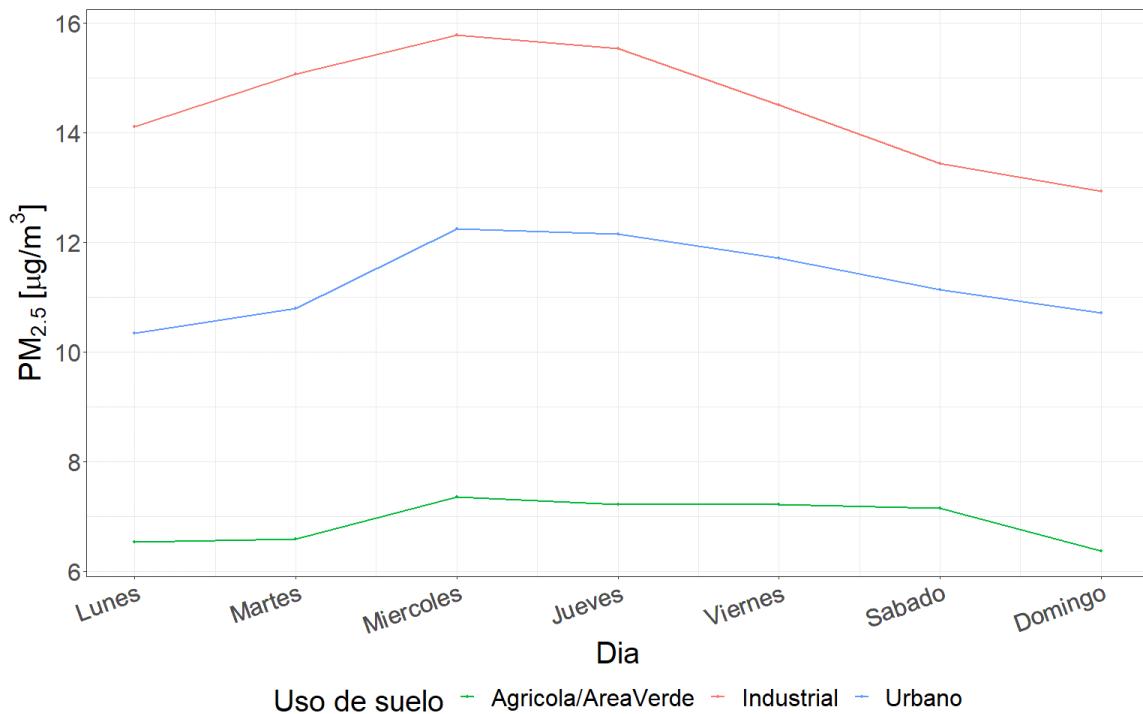


Figura 3.18: Perfil de concentración promedio diario de PM_{2.5} a lo largo de la semana, discriminado por los usos de suelo derivados del análisis de conglomerados a lo largo de todo el periodo de muestreo.

Para los usos de suelo industrial y urbano, las concentraciones de PM_{2.5} son mayores durante los días hábiles de la semana, mientras que disminuyen en los días de fin de semana, sugiriendo una relación con las actividades laborales y de tráfico vehicular asociadas a estos usos de suelo (Figura 3.18). Sin embargo, para el uso del suelo agrícola/área verde no logra observarse esta variación entre días hábiles y de fin de semana, lo que podría deberse a que en esta zona las fuentes de emisión de PM_{2.5} no están asociadas a actividades laborales, o bien a que las jornadas laborales son distintas a las que se dan en las zonas urbanas/industriales.

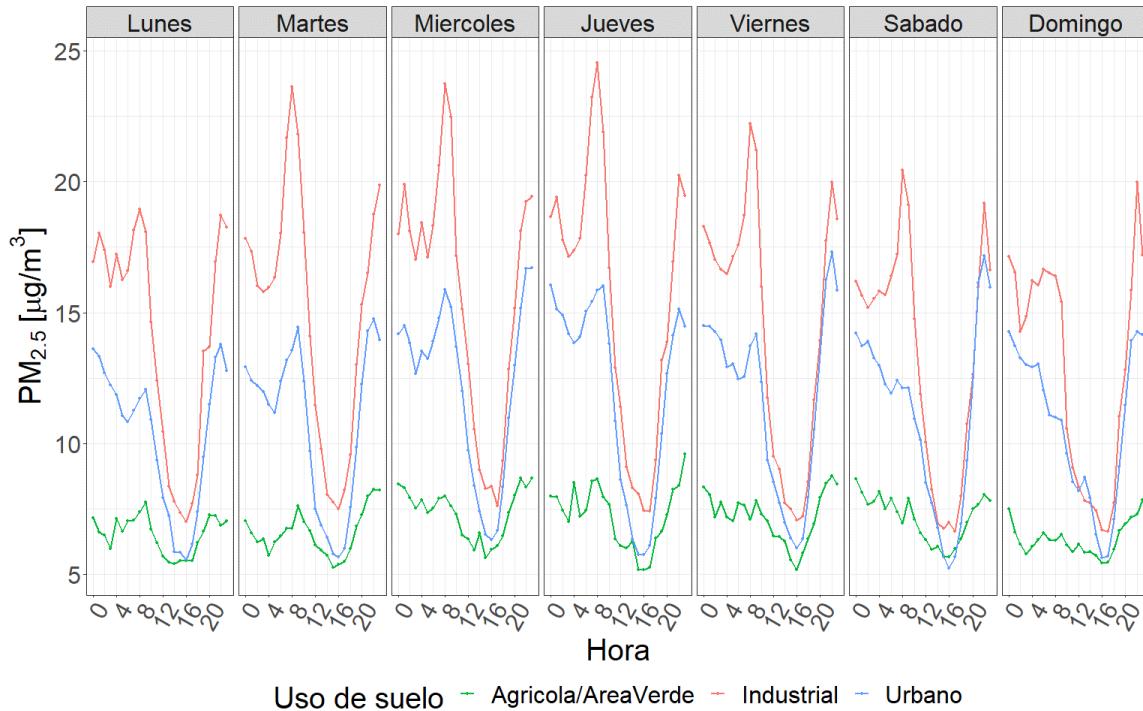


Figura 3.19: Perfil de concentración promedio horario de PM_{2.5} en el día a lo largo de la semana, discriminado por los usos de suelo derivados del análisis de conglomerados a lo largo de todo el periodo de muestreo.

La concentración de PM_{2.5} sigue un patrón de variabilidad horaria similar a lo largo de los días de la semana, independientemente del uso del suelo, a pesar de que existan diferencias en magnitud (Figura 3.19). Específicamente, se observa que los patrones de variación en zonas urbanas e industriales presentan mayor similitud entre sí que con zonas agrícola/área verde. Esto podría deberse a la dinámica de movilidad de la población, dado que el tráfico de vehículos sería una de las fuentes principales que contribuyen a su variabilidad, de manera similar a lo observado por (Pant & Harrison, 2013; Weerakkody et al., 2018). De esta forma en los sitios urbanos e industriales, las variaciones en la concentración de PM_{2.5} podrían estar relacionadas con la densidad de vehículos que circulan durante los días laborables, la cual es notablemente superior a la de los fines de semana. En estos sitios la concentración de PM_{2.5} aumenta hasta alcanzar un pico los días miércoles o jueves, mientras que la mínima se observa durante los fines de semana, generalmente los días domingo o lunes (Figura 3.18).

En contraste, en los sitios agrícola/área verde, el patrón de variabilidad no es tan evidente debido a que las principales fuentes de emisión no se relacionan directamente ni con la densidad poblacional ni el tránsito vehicular. Por lo tanto, la concentración de PM_{2.5} en estos sitios alcanza su máximo entre los días miércoles a sábados y llega a su mínimo los días domingo o lunes, sin observarse un pico pronunciado durante los días laborables.

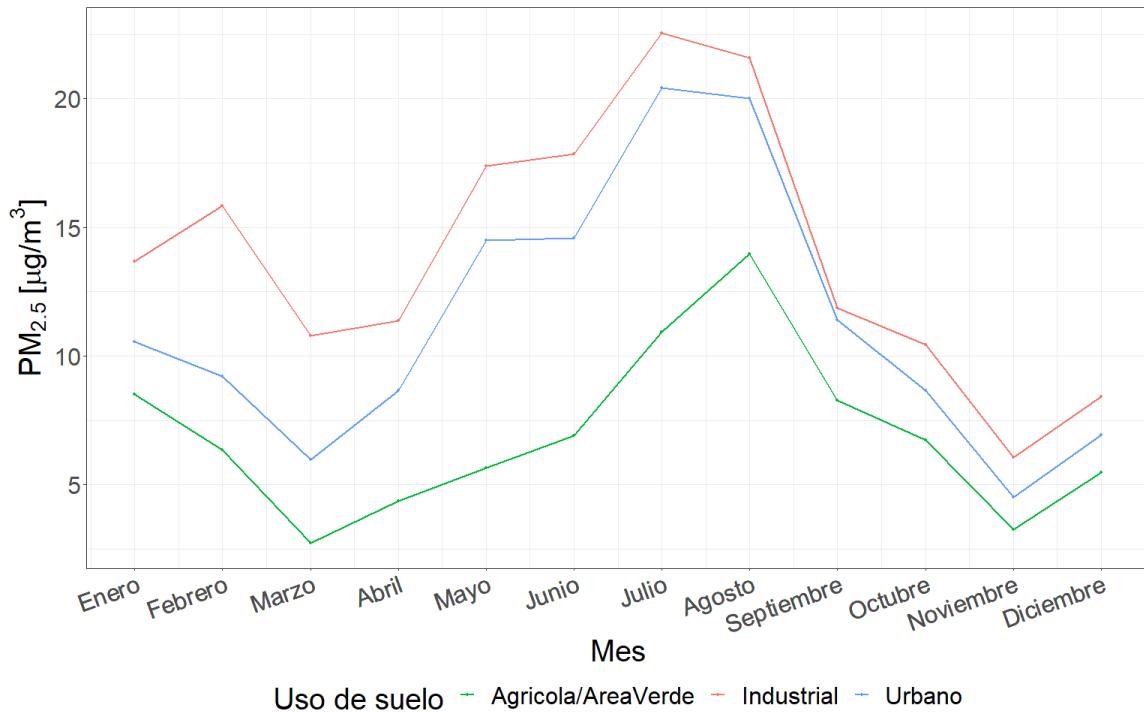


Figura 3.20: Perfil de concentración promedio mensual de PM_{2.5} a lo largo del año, discriminado por los usos de suelo derivados del análisis de conglomerados a lo largo de todo el periodo de muestreo.

En la Figura 3.20, puede destacarse la presencia de un patrón común de variación de la concentración de PM_{2.5} en ambos usos de suelo a lo largo de los meses del año estudiados durante el periodo de muestreo. Es evidente que los registros indican concentraciones máximas durante los meses de julio y agosto, mientras que mínimas durante los meses de marzo, abril y noviembre. A pesar de que los ciclos varían en magnitud en los distintos usos de suelo, estos se encuentran presentes en todos ellos. De esta forma escalas de mayor granularidad temporal como lo son los meses y

estaciones logran explicar gran parte de la variabilidad de la concentración del PM_{2.5}, independientemente del uso del suelo considerado. Este hecho podría encontrarse vinculado a la fluctuación de las condiciones meteorológicas que ocurre en escalas temporales de elevada granularidad.

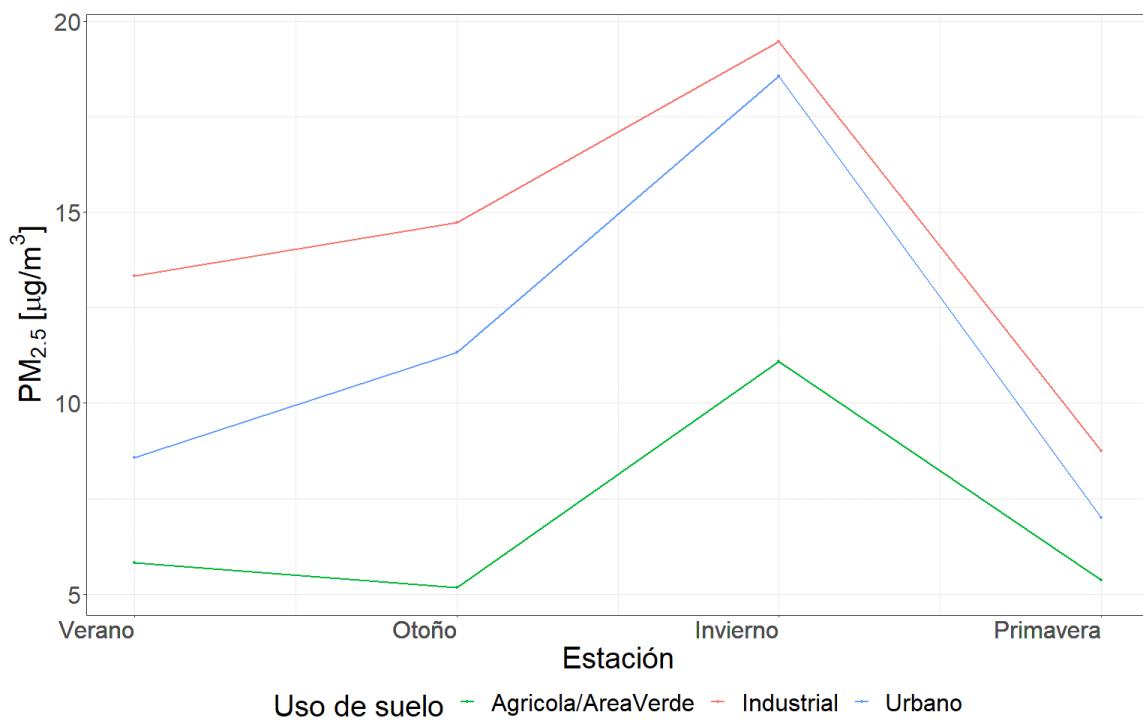


Figura 3.21: Perfil de concentración promedio estacional de PM_{2.5} a lo largo del año, discriminado por los usos de suelo derivados del análisis de conglomerados a lo largo de todo el periodo de muestreo.

El patrón estacional de concentración de PM_{2.5}, previamente observado en la Figura 3.8, se mantiene constante en los distintos usos de suelo, a pesar de variar magnitud entre ellos (Figura 3.21). Al igual que en el análisis general, se registraron menores concentraciones en el sitio con uso del suelo agrícola/área verde, lo cual refuerza la hipótesis que la estacionalidad en la concentración del PM_{2.5} se encuentra influenciada por factores climáticos y meteorológicos que actúan de manera similar en los distintos tipos de uso del suelo. Además, los resultados resaltan la importancia de considerar las características específicas del sitio de estudio tales como la topografía local, las fuentes de emisiones, la vegetación circundante y las condiciones climáticas específicas

de cada ubicación, al momento de evaluar los niveles de contaminación atmosférica y cómo éstos pueden variar en función de factores ambientales y de uso del suelo.

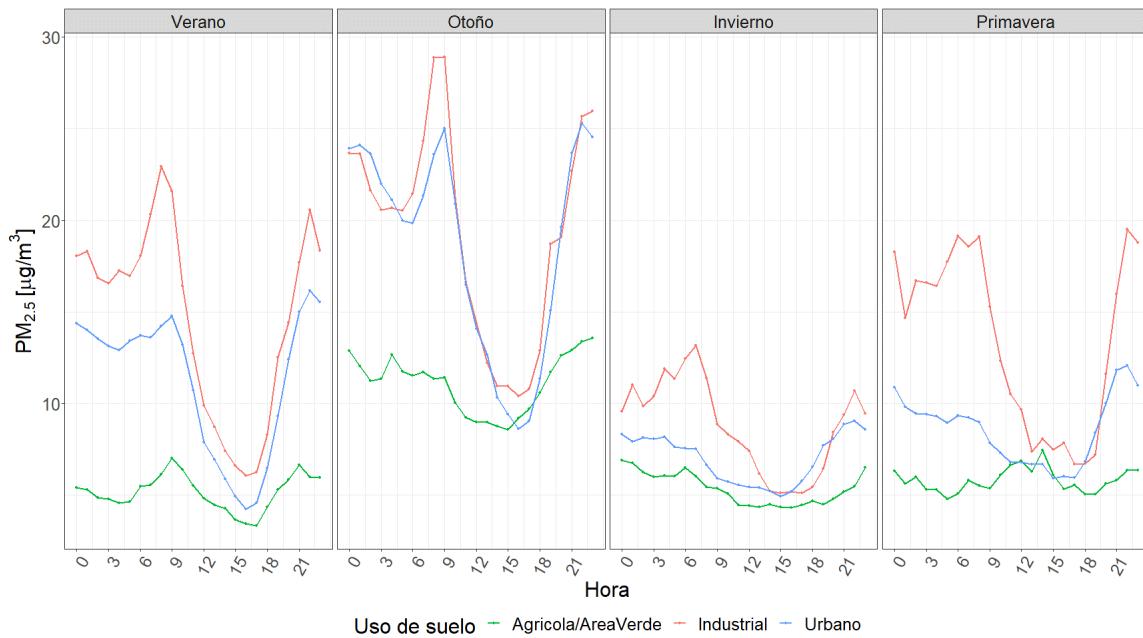


Figura 3.22: Perfil de concentración promedio horario de PM_{2.5} a lo largo del día en las distintas estaciones del año, discriminado por los usos de suelo derivados del análisis de conglomerados a lo largo de todo el periodo de muestreo.

Es interesante destacar que el patrón de variabilidad de la concentración de PM_{2.5} que se observó para el análisis general (Figura 3.8) se mantiene constante en los distintos usos de suelo, aunque disminuye en magnitud para el uso del suelo agrícola/área verde.

Los resultados sugieren que las condiciones meteorológicas durante la época fría generan condiciones adversas de calidad del aire, independientemente del uso del suelo. Este hecho podría encontrarse vinculado a las frecuentes inversiones térmicas que ocurren durante la época invernal, fenómeno que se analizara con mayor detalle en el siguiente capítulo. Además, la concentración promedio horaria de PM_{2.5} en los distintos usos de suelo presenta una tendencia a aumentar durante las horas de la mañana, disminuir en la tarde y aumentar nuevamente durante las horas de la noche.

3.4. Conclusiones

La concentración de PM_{2.5} en el área metropolitana de la ciudad de Córdoba fue menor que la informada en otras ciudades del mundo, como: Riohacha (Colombia), Río de Janeiro y Porto Alegre (Brasil) y Estambul (Turquía). Sin embargo, los estándares de calidad de aire establecidos por las normativas locales (Tabla 3.2) fueron superados múltiples veces tanto en la escala anual como de promedio de 24 horas (Tabla 3.5 y Figura 3.10).

La distribución teórica que minimizó los valores del criterio de información de Akaike (AIC) y del criterio de información bayesiano (BIC), demostrando el mejor ajuste a los datos de concentración PM_{2.5} fue la distribución *gamma*.

Las características temporales de la serie de concentración de PM_{2.5} permiten explicar parte de su variabilidad. A su vez las características espaciales, referidas a la ubicación y la influencia de la microescala en cada sitio, permiten explicar la discrepancia en la magnitud de concentraciones de PM_{2.5} existentes.

Las series temporales de concentración de PM_{2.5} muestran el patrón de emisión característico de cada uso del suelo, existiendo coincidencia entre la cantidad de veces que se superan las normativas propuestas y las características de los sitios. Puede apreciarse que en los sitios cuyo uso del suelo es industrial, la cantidad de veces en que se superan los estándares mencionados es superior a las que sucede en el uso del suelo urbano y agrícola/área verde, siendo estos últimos sitios en los que hubo menor cantidad de eventos. Las características temporales de la serie de concentración de PM_{2.5} se mantienen constantes independientemente del uso del suelo, reflejando el mismo patrón de variación a lo largo de las distintas escalas temporales. A pesar de ello el patrón de variación se ve atenuado en la magnitud de las concentraciones en la serie agrícola y área verde en comparación con la serie industrial. Este hecho indica que la concentración de PM_{2.5} responde a una dinámica temporal similar, que se encuentra fuertemente influenciada por las condiciones espaciales de los sitios que se

estén analizando. Así, la serie de datos proveniente de entornos agrícolas y área verde mostraron concentraciones de menor magnitud y menor rango de variación, que la caracterizan como una serie temporal de baja variabilidad. La variación de las escalas temporales de mayor magnitud, como los meses y estaciones, podría vincularse a la variación de las condiciones meteorológicas, la cual podría ser independiente del uso del suelo que se esté considerando.

El análisis exploratorio demostró ser fundamental para revelar las estacionalidades, ciclos y patrones presentes en la serie de concentración de PM_{2.5}, así como para identificar la influencia diferencial de las estaciones climáticas en los patrones exhibidos. Estos hallazgos son de gran utilidad para desarrollar modelos predictivos de calidad del aire, ya que evidencian una fuente de variabilidad previamente no observada, que permitirá la creación de modelos predictivos más precisos y efectivos.

Capítulo 4

Influencia de variables meteorológicas obtenidas de satélites

4.1. Introducción

La teledetección o sensado remoto es una técnica que permite adquirir información sobre el sistema terrestre mediante sensores ubicados en aviones o satélites, sin necesidad de contacto físico directo con los objetos de estudio (Chuvieco & Hantson, 2010). Estos sensores capturan y registran la radiación electromagnética reflejada o emitida por la superficie terrestre. Su aplicabilidad se extiende al estudio de aspectos clave como el clima, la meteorología, la interacción Tierra-atmósfera y la dinámica de la superficie terrestre. A partir de la información recopilada, se obtienen estimaciones horarias de diversas variables climáticas, atmosféricas, terrestres y oceánicas, que abarcan la totalidad de la superficie, con una elevada resolución tanto espacial como temporal. Estas estimaciones se generan por medio de la técnica de reanálisis, la cual combina datos de modelos teóricos, con observaciones *in situ* y leyes físicas.

Esta metodología, denominada asimilación de datos, es ampliamente utilizada por los centros de predicción meteorológica numérica. Cada un lapso de tiempo determinado se combina una previsión anterior con nuevas observaciones disponibles para producir una nueva estimación del estado de la atmósfera (Bell et al., 2021; Hersbach et al., 2020; B. Wu et al., 2020). La información generada a partir de la teledetección la convierte en una herramienta fundamental a la hora de predecir escenarios climáticos futuros (Asaly et al., 2022; Huot et al., 2022; Paul et al., 2020; Rocchini et al., 2015; Watmough et al., 2019).

La fluctuación de las condiciones meteorológicas que determinan el estado de la atmósfera en un momento determinado generan una notable influencia en la variación de la concentración de PM (Jianjun He et al., 2017; Tiwari et al., 2017). Numerosos estudios han demostrado asociaciones de vínculo inverso entre el PM con la velocidad del viento (Cichowicz et al., 2020; Usman et al., 2022; XiaoYang et al., 2019), las precipitaciones (Usman et al., 2022; XiaoYang et al., 2019), la temperatura (Nogarotto & Pozza, 2020; XiaoYang et al., 2019), el punto de rocío (Yujie Zhao et al., 2022), las precipitaciones (W. Hu et al., 2021; Kleine Deters et al., 2017) y la altura de la capa límite atmosférica (PBLH) (D. Li et al., 2020; C. Wang et al., 2019). Por otra parte, existen variables que responden a una relación directa positiva con los niveles de PM, tales como la profundidad óptica de aerosoles (AOD) (A. Gupta et al., 2021; Soni et al., 2018), el tráfico de vehículos (Pant & Harrison, 2013; Weerakkody et al., 2018), entre otras múltiples fuentes de emisión. Tanto para la presión atmosférica, como para la humedad relativa y la radiación solar, existen resultados que vinculan estas variables con el PM tanto de forma directa (H. Li et al., 2015; Nogarotto & Pozza, 2020; X. Yang et al., 2016) como inversa (Tao Chen et al., 2016; Z. Liu et al., 2015; Tian et al., 2014). Es pertinente destacar la existencia de estudios cuyas conclusiones reflejan que la concentración de PM depende más de la meteorología que de factores antropogénicos (Hooyberghs et al., 2005). Esta relación está sujeta a otras

características propias del sitio, como por ejemplo las fuentes de emisión, el tamaño de las partículas (Dimitriou et al., 2015; Khan et al., 2010; Tai et al., 2010) y la topografía (Alvarado et al., 2010; Vinoj & Pandey, 2022). Por este motivo, cobra sentido modelar la concentración de PM_{2.5} en función de variables meteorológicas para sitios con características específicas. La disponibilidad de variables meteorológicas provenientes de la teledetección o sensoramiento remoto viabiliza un ajuste de modelos predictivos más adecuado para la concentración de PM_{2.5}. Sin embargo, previo a cualquier ajuste de modelo predictivo es necesario contar con una descripción exhaustiva de la estructura de la serie temporal de PM.

Enmarcado dentro del Servicio de Cambio Climático Copernicus (C3S - *Copernicus Climate Change Service*) de la Agencia Espacial Europea (ESA), el Centro Europeo para Previsiones Meteorológicas a Medio Plazo (ECMWF - *European Centre for Medium-Range Weather Forecasts*) genera grandes conjuntos de datos meteorológicos a escala global. Este producto es la quinta generación de Reanálisis Europeo (ERA5 - *European ReAnalysis*), sucesor de *ERA-Interim* (Muñoz-Sabater et al., 2021), y se caracteriza por combinar modelos meteorológicos y datos observacionales a nivel mundial. Los productos generados ponen a disposición información para todo el globo a escala temporal horaria, diaria y mensual desde el año 1979. Así, cada vez es más factible mapear y modelar las variaciones o cambios espacio-temporales de las condiciones meteorológicas imperantes antes, durante y después de eventos de interés. El avance de la tecnología ha permitido un desarrollo notable en la capacidad computacional disponible, sumado a la disponibilidad de datos meteorológicos masivos en diversas plataformas de Internet, es posible obtener información meteorológica variada para un sitio georreferenciado sobre un territorio de interés. Actualmente, se cuenta con variadas herramientas informáticas para la obtención de productos satelitales almacenados en la nube. Tales productos pueden ser usados como covariables con potencialidad predictiva de la concentración de PM.

La búsqueda de factores que logren explicar la variabilidad de la serie temporal de concentración de PM_{2.5} se realiza tanto para la comprensión de la dinámica de variación del contaminante como para su modelado predictivo. El estudio de la estructura o patrones en la serie temporal del contaminante y su covariabilidad con las series de variables meteorológicas es fundamental para construir modelos predictivos. Estos pueden ser mecanicísticos, *i.e.* modelos que pretenden explicar y relacionar, usualmente basados en el conocimiento existente, la covariabilidad del PM_{2.5} con la de otros factores del medio. Sin embargo, los modelos predictivos también pueden ser de base computacional. En este caso, el análisis de los datos del contaminante se relaciona con la variabilidad de múltiples factores con la intención de predecir su valor aun cuando el impacto de cada uno de los factores o sus combinaciones permanezca oculto o sin justificar desde el conocimiento existente. Los modelos de inteligencia artificial, tanto de aprendizaje automático como profundo, se basan en explotar la covariabilidad entre un gran número de series de variables en una base de datos de entrenamiento con la finalidad de predecir nuevos valores de PM_{2.5} en contextos similares al del aprendizaje. Por lo tanto, para predecir valores futuros de una serie puede implementarse un marco teórico basado en información contenida en la misma y/u otro que se nutra, además de la variabilidad de factores relacionados con la serie. En particular, los métodos de aprendizaje automático permiten obtener predicciones de valores futuros en función del comportamiento de un conjunto grande de potenciales predictoras y de la covariabilidad entre ellas. El presente capítulo 4 y su posterior 5, se hallan enmarcados dentro de un enfoque de trabajo basado en la inteligencia artificial. Se incluyen métodos clásicos como la regresión lineal múltiple, que demandan un mayor control del investigador sobre las variables predictoras y supuestos sobre las relaciones subyacentes, además de otros como los algoritmos de aprendizaje automático basados en árboles y de aprendizaje profundo basados en redes neuronales, que habilitan el trabajo con múltiples series de datos cuya variabilidad podría tener potencial capacidad

predictiva de la variabilidad de PM_{2.5} con un menor nivel de control y de supuestos. Uno de los principales desafíos en el modelado predictivo a través de técnicas de aprendizaje automático es la selección de características o variables predictoras que alimentarán al modelo, dado que mientras mayor sea el número de predictoras también lo será la complejidad del resultado (Cai et al., 2018; Y. Li et al., 2019; Rumao, 2019). Es factible que las variables relevantes para el desarrollo de un modelo predictivo de elevada exactitud puedan no solo ser las recomendadas en la literatura para la construcción de modelos mecanicistas explicativos de la dinámica observada en la serie del contaminante, ya que el objetivo dentro del enfoque de aprendizaje automático está centrado en potenciar la exactitud de las predicciones y no en la explicación de cada una de las relaciones subyacentes.

En esta tesis la variable respuesta a modelar es la concentración de PM_{2.5} y las variables predictoras derivarán principalmente de mediciones satelitales, en especial meteorológicas, además de las que refieren a la estructura temporal de variabilidad de la serie de tiempo para las características de cada uso del suelo. Para abordar el problema enmarcado en determinar la potencialidad de estas variables para predecir la concentración de PM_{2.5}, se creó una base de datos satelitales. Respetando las premisas de creación de un sistema de alerta de bajo costo ante condiciones adversas de calidad del aire, las variables incluidas como predictoras en el modelado predictivo de la concentración de PM_{2.5}, son de dominio público y la variable respuesta obtenida por medio de sensores de monitoreo basados en tecnologías de bajo costo.

Las técnicas de selección de características asociadas a la predicción constituyen una herramienta útil para identificar las variables de mayor relevancia a la hora de desarrollar un modelo predictivo (Khalid et al., 2014; Venkatesh & Anuradha, 2019). Es importante destacar que, en algunos casos, las variables seleccionadas pueden no tener un respaldo teórico desde el punto de vista de los mecanismos que subyacen la contaminación ambiental. La implementación de técnicas de aprendizaje automático,

permite que vinculaciones indirectas entre las variables que alimentan el modelo puedan ser desenmascaradas. Algunas de ellas podrían tener una interpretación teórica, pero esto no necesariamente ocurrirá con todas (Lima et al., 2016; Suparwito et al., 2021).

4.2. Materiales y métodos

4.2.1. Resumen gráfico del protocolo de análisis de correlaciones entre series

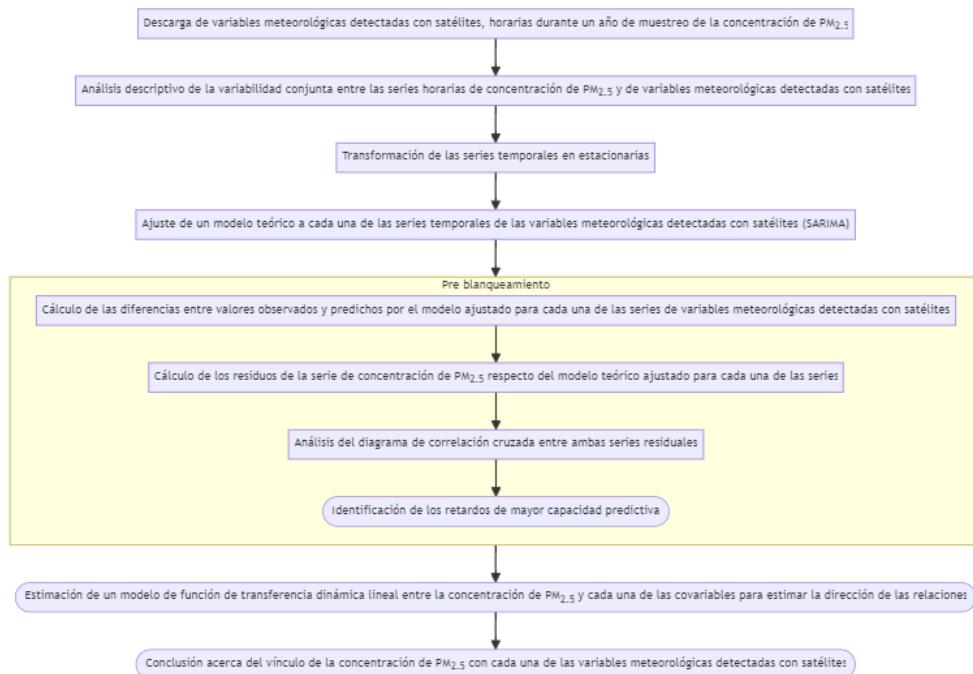


Figura 4.1: Resumen gráfico del protocolo analítico desarrollado en el capítulo 4.

La Figura 4.1 presenta el protocolo analítico desarrollado en este capítulo con la finalidad de identificar asociaciones entre la serie de concentración de PM_{2.5} y las de variables derivadas a partir de información satelital. En primer lugar, se creó la base de datos empleada durante el presente capítulo, la cual surgió a partir de la combinación de la base de datos de concentración de PM_{2.5} provista por los sensores

de monitoreo basados en tecnologías de bajo costo y las variables derivadas a partir de información satelital de acceso público descargadas desde la nube. Posteriormente se hace mención a cada uno de los pasos desarrollados durante el análisis de series temporales ejecutado en el presente capítulo. Es interesante destacar el procedimiento de pre blanqueamiento, el cual permite eliminar el ruido presente en una serie temporal. Para más detalles sobre cada uno de los procedimientos mencionados dirigirse a la sección de protocolo estadístico, más específicamente al apartado análisis de series temporales. Al finalizar el procedimiento se concluyó respecto de las correlaciones desenmascaradas para cada una de las variables derivadas a partir de información satelital analizadas.

4.2.2. Base de datos

La base de datos para el desarrollo de esta sección surge de la combinación de los datos recopilados durante el muestreo de PM_{2.5} (ver capítulo 2 para más detalles) con distintas bases de datos satelitales (Tabla 4.1). De esta forma el período a lo largo del cual se descargaron los datos derivados de satélites se corresponde con el muestreo previamente realizado en los distintos sitios de muestreo en la provincia de Córdoba, el cual abarcó la ventana temporal con inicio el 30/03/2021 y final el 01/06/2022. Cabe destacar que en el análisis exploratorio realizado en el presente capítulo se utilizaron promedios calculados a lo largo de todos los sitios de muestreo en las distintas escalas temporales consideradas de análisis, tanto para la concentración de PM_{2.5} como para las variables derivadas de información satelital. Esto quiere decir que los datos informados a partir de los cuales se realiza el análisis se corresponden con promedios (horarios, diarios, mensuales y por estación) a lo largo de todos los sitios de muestreo localizados en la provincia de Córdoba, Argentina.

Las covariables utilizadas son derivadas de información satelital a distintos niveles de resolución temporal y espacial, lo que implica una combinación de técnicas avanzadas

para combinar y analizar estos datos de diversas fuentes, asegurando una integración efectiva y precisa de la información. Algunas de estas variables se obtienen directamente a partir de la interpretación física de la reflectancia capturada por los satélites, la cual es procesada y transformada en datos utilizables, como es el caso de las mediciones del sensor espectrorradiómetro de imágenes de resolución moderada (MODIS) a bordo de los satélites Terra y Aqua de la NASA (Hulley et al., 2014; Alexei Lyapustin et al., 2018, 2019; Savtchenko et al., 2004). Otras variables son generadas a través de modelos que utilizan datos de entrada satelitales, como los producidos por el modelo meteorológico ERA5, que surge del reanálisis realizado por el Centro Europeo de Previsión Meteorológica a Plazo Medio (ECMWF) para el clima global (Bell et al., 2021; Hersbach et al., 2020), permitiendo así estimaciones y proyecciones más precisas en base a algoritmos específicos y condiciones iniciales derivadas de observaciones satelitales. Además, se emplea información surgida a partir de técnicas de asimilación de datos, como las utilizadas en el reanálisis retrospectivo para investigación y aplicaciones de era moderna en su versión 2 (MERRA-2) (Bosilovich et al., 2015), las cuales integran observaciones satelitales con modelos numéricos para mejorar la exactitud y consistencia de los datos finales. También se incluyó información proveniente del conjunto de radiómetros de imágenes infrarrojas visibles (VIIRS) (J. Wu et al., 2016). Es importante destacar que, tras un exhaustivo esfuerzo de búsqueda y revisión de la literatura, se ha confirmado que ninguna de las variables utilizadas ha sido validada con datos de campo en la provincia de Córdoba, excepto el AOD, que fue validado por De La Secca et al. en 2018. Esta precisión es esencial para contextualizar la fiabilidad y las limitaciones de los datos empleados en este estudio, subrayando la necesidad de futuras investigaciones que validen estos productos satelitales en el contexto local de Córdoba.

Para aquellas variables cuya fuente de datos se indique explícitamente entre paréntesis, se ha utilizado información derivada de métodos o sistemas distintos

a ERA5. En otras palabras, si no se indica explícitamente la fuente, entonces la información proviene del sistema ERA5. La base de datos original o completa contó con un total de 193 variables estas son: hora del día (fecha de toma de dato), día de la semana (fecha de toma de dato), mes del año (fecha de toma de dato), día del mes (fecha de toma de dato), estación del año (fecha de toma de dato), mes del año codificación seno (fecha de toma de dato), mes del año codificación coseno (fecha de toma de dato), día de la semana codificación seno (fecha de toma de dato), día de la semana codificación coseno (fecha de toma de dato), día del mes codificación seno (fecha de toma de dato), día del mes codificación coseno (fecha de toma de dato), hora del día codificación seno (fecha de toma de dato), hora del día codificación coseno (fecha de toma de dato), estación del año codificación seno (fecha de toma de dato), estación del año codificación coseno (fecha de toma de dato), componente u del viento a 100 m, componente v del viento a 100 m, componente u del viento neutral a 10 m, componente u del viento a 10 m, componente v del viento neutral a 10 m, componente v del viento a 10 m, ráfaga de viento a 10 m desde el posprocesamiento anterior, temperatura del punto de rocío a 2 m, temperatura a 2 m, ángulo de orografía a escala inferior a la cuadrícula, anisotropía de la orografía a escala inferior a la cuadrícula, disipación de la capa límite, altura de la capa límite, cielo despejado, radiación solar directa en la superficie, altura de la base de las nubes, energía potencial convectiva disponible, inhibición convectiva, precipitación convectiva, tasa de lluvia convectiva, radiación ultravioleta descendente en la superficie, tensión superficial de ondas de gravedad hacia el este, tensión superficial turbulenta hacia el este, evaporación, evaporación del suelo desnudo, evaporación desde la parte superior del dosel, evaporación de la transpiración de la vegetación, albedo previsto, logaritmo previsto de la rugosidad de la superficie para el calor, superficie prevista rugosidad, velocidad de fricción, geopotencial, disipación de ondas de gravedad, nubosidad alta, cubierta de vegetación alta, ráfaga de viento instantánea de 10 m, tensión superficial turbulenta instantánea

hacia el este, fracción de precipitación superficial instantánea a gran escala, flujo de humedad instantáneo, tensión superficial turbulenta instantánea hacia el norte, superficie sensible instantánea flujo de calor, índice k, precipitación a gran escala, fracción de precipitación a gran escala, tasa de lluvia a gran escala, índice de área foliar vegetación alta, índice de área foliar vegetación baja, cobertura de nubes baja, cobertura de vegetación baja, temperatura máxima de 2 m desde el procesamiento posterior anterior, total máximo tasa de precipitación desde el procesamiento posterior anterior, disipación media de la capa límite, tasa media de precipitación convectiva, tensión superficial turbulenta media hacia el este, tasa media de evaporación, disipación media de ondas de gravedad, fracción de precipitación media a gran escala, tasa media de precipitaciones a gran escala, tensión superficial media de ondas de gravedad hacia el norte, tensión superficial turbulenta media hacia el norte, tasa de evaporación potencial media, tasa de escorrentía media, presión media a nivel del mar, tasa de escorrentía subsuperficial media, flujo de radiación de onda corta directa en superficie media, flujo de radiación de onda larga hacia abajo en superficie media, flujo de radiación de onda corta en superficie media hacia abajo, flujo de radiación ultravioleta descendente medio en la superficie, flujo de calor latente superficial medio, flujo de radiación de onda larga neta superficial media, tasa de escorrentía superficial media, flujo de calor sensible superficial medio, flujo de radiación de onda corta descendente media superior, flujo de radiación de onda larga neta superior media, media flujo de radiación de onda corta neto superior, tasa de precipitación total media, gradiente vertical medio de refractividad dentro de la capa de captura, divergencia de humedad integrada verticalmente media, cobertura de nubes media, temperatura mínima de 2 m desde el posprocesamiento anterior, tasa de precipitación total mínima desde el posprocesamiento anterior, gradiente vertical mínimo de refractividad dentro de la capa de captura, albedo de infrarrojo cercano para radiación difusa, albedo de infrarrojo cercano para radiación directa, tensión superficial de ondas de gravedad

hacia el norte, tensión superficial turbulenta hacia el norte, evaporación potencial, tipo de precipitación, escorrentía, pendiente de la orografía a escala subcuadrícula, tipo de suelo, desviación estándar de orografía de subred filtrada, desviación estándar de la orografía, escorrentía subsuperficial, flujo de calor latente superficial, radiación solar neta superficial, radiación térmica neta superficial, presión superficial, escorrentía superficial, flujo de calor sensible superficial, radiación solar superficial hacia abajo, radiación térmica superficial hacia abajo, toa radiación solar incidente, radiación solar neta superior, radiación térmica neta superior, nubosidad total, columna total de agua helada de la nube, columna total de agua líquida de la nube, columna total de ozono, columna total de agua de lluvia, columna total de agua, columna total de vapor de agua, evaporación total, precipitación total, cielo total, radiación solar directa en la superficie, índice de totales totales, altura base de la capa de captura, altura superior de la capa de captura, tipo de vegetación alta, tipo de vegetación baja, componente u deriva de Stokes, albedo uv visible para radiación difusa, uv visible albedo para radiación directa, componente v deriva de Stokes, integral vertical de divergencia del flujo de agua congelada en la nube, integral vertical de divergencia del flujo de agua líquida en la nube, integral vertical de divergencia del flujo geopotencial, integral vertical de divergencia del flujo de energía cinética, integral vertical de divergencia de flujo de masa, integral vertical de divergencia de flujo de humedad, integral vertical de divergencia de flujo de ozono, integral vertical de divergencia de flujo de energía térmica, integral vertical de divergencia de flujo de energía total, integral vertical de nube hacia el este flujo de agua congelada, integral vertical del flujo de agua líquida en la nube hacia el este, integral vertical del flujo geopotencial hacia el este, integral vertical del flujo de calor hacia el este, integral vertical del flujo de energía cinética hacia el este, integral vertical del flujo de masa hacia el este, integral vertical del flujo de ozono hacia el este, integral vertical del flujo de energía total hacia el este, integral vertical del flujo de vapor de agua hacia el este, integral vertical de conversión

de energía, integral vertical de la energía cinética, integral vertical de la masa de la atmósfera, integral vertical de la tendencia de la masa, integral vertical del flujo de agua congelada de las nubes hacia el norte, integral vertical del flujo de agua líquida de las nubes hacia el norte, integral vertical del flujo geopotencial hacia el norte, integral vertical del calor hacia el norte integral vertical del flujo de energía cinética hacia el norte, integral vertical del flujo de masa hacia el norte, integral vertical del flujo de ozono hacia el norte, integral vertical del flujo de energía total hacia el norte, integral vertical del flujo de vapor de agua hacia el norte, integral vertical de energía potencial e interna, integral vertical de temperatura, integral vertical de energía térmica, integral vertical de energía total, divergencia de humedad integrada verticalmente, nivel de cero grados, viento a 100 m, viento a 10 m, viento neutro a 10 m, humedad relativa a 2 m, profundidad óptica de aerosoles (MERRA), incremento de la profundidad óptica de aerosoles (MERRA), incremento de la profundidad óptica de aerosoles (MAIAC), profundidad óptica de aerosoles (MAIAC), altura de la capa límite atmosférica (MERRA-2/ERA5), máscara de fuego VNP14A1 001, temperatura de la superficie terrestre (MODIS) y tráfico de vehículos (GOOGLE). Se evita brindar la definición de cada una de las variables mencionadas para no hacer engoroso el cuerpo de la tesis; sin embargo, todas las variables mencionadas se encuentran descriptas en el Anexo 3. Cabe destacar que los análisis realizados en este capítulo únicamente incluyen las variables de mayor importancia según la bibliografía, para describir la variación de la concentración de PM_{2.5}, estas son: temperatura del aire a 2 metros por encima de la superficie de la Tierra, humedad relativa a 2 metros por encima de la superficie terrestre, temperatura del punto de rocío, precipitación convectiva, velocidad del viento en su componente “neutra” a una altura de 10 metros sobre la superficie terrestre, presión atmosférica, altura de la capa límite atmosférica (PBLH), radiación solar directa sin nubosidad sobre la superficie terrestre, profundidad óptica de aerosoles (AOD), temperatura de la superficie terrestre (LST), y tráfico de automóviles.

El mayor volumen de variables lo comprenden las generadas por el modelo meteorológico ERA5, que surge del reanálisis realizado por el Centro Europeo de Previsión Meteorológica a Plazo Medio (ECMWF) para el clima global (Bell et al., 2021; Hersbach et al., 2020). Este método de estimación combina modelos predictivos con observaciones a lo largo de todo el mundo, dando por resultado un conjunto de datos completo a escala global, cuyo principio de funcionamiento se basa en la asimilación de datos. Enmarcado dentro del C3S de la Comisión Europea, el ECMWF genera el reanálisis ERA5 que representa un registro detallado de las condiciones atmosféricas globales, de la superficie terrestre y oceánicas desde 1979 en adelante. ERA5 se basa en el Sistema Integrado de Pronóstico (IFS) Cy41r2, implementado desde 2016, beneficiándose de una década de desarrollos en modelos físicos, dinámica y asimilación de datos. Esta base de datos cuenta con estimaciones horarias para 164 variables atmosféricas, oceánicas y terrestres, con una resolución espacial de 0.25 grados (Hersbach et al., 2020). Algunas de las principales variables provenientes de esta fuente de datos que están vinculadas con la concentración de PM_{2.5} son: previsión total de precipitación, evaporación, cobertura total de nubes, altura de la base de las nubes, rugosidad superficial pronosticada, velocidad de fricción, cobertura de vegetación alta, cobertura de vegetación baja, componente u del viento, componente v del viento, temperatura a 2 m, presión superficial, vapor de agua total en columna, altura de la capa límite, albedo pronosticado e índice de área foliar de vegetación alta T. Jiang et al. (2021). El listado completo de las variables extraídas a partir de esta herramienta se encuentran mencionadas en el Anexo 1, mientras que una descripción detallada de cada una de ellas se presenta en el Anexo 3 de la presente tesis.

La profundidad óptica de aerosoles (AOD) se obtuvo a partir mediciones del sensor MODIS (*Moderate Resolution Imaging Spectroradiometer*) (Hulley et al., 2014; Alexei Lyapustin et al., 2018, 2019; Savtchenko et al., 2004), el cual se encuentra a bordo de los satélites Terra y Aqua pertenecientes a la Administración Nacional de Aeronáutica

y del Espacio (NASA). Lyapustin (2011) desarrolló el algoritmo MAIAC para obtener valores de AOD sobre superficies brillantes y oscuras con una resolución de 1 km a partir de los productos de MODIS (Alexei Lyapustin et al., 2018). El algoritmo MAIAC permite así obtener características más refinadas de los aerosoles, necesarias para el monitoreo regional y a menores escalas espaciales (Tao et al., 2019). La alta resolución espacial mejora la capacidad para determinar las características de los aerosoles finos y distinguir las fuentes de emisión (Hulley et al., 2014; Mhawish et al., 2019; Savtchenko et al., 2004). Este algoritmo permite obtener el valor de AOD sobre la superficie terrestre a partir de la implementación de series temporales, logrando discriminar la contribución de aerosoles de la reflectividad de la superficie terrestre, considerando los efectos de la reflectividad bidireccional (Alexei Lyapustin et al., 2018, 2019). Además, implementa técnicas de análisis de series temporales y procesamiento basado en imágenes para llevar a cabo la recuperación de datos y la corrección en superficies brillantes y en vegetación oscura (Alexei Lyapustin et al., 2011, 2018); asimismo el algoritmo mejora la capacidad de detección de humo (polvo), nubes y cobertura de nieve (A. Lyapustin et al., 2012).

La temperatura de la superficie terrestre (LST) también se obtuvo a partir de los datos brindados por el instrumento MODIS a bordo de los satélites Terra y Aqua, proporcionando registros con una precisión objetivo de 1 K. Los sensores térmicos basados en satélites son herramientas poderosas para el monitoreo ambiental, como la detección de incendios (Justice et al., 2002) o actividad volcánica (Wright et al., 2002). Además, estos sensores ofrecen evidencia directa de la variabilidad de la superficie, la cual se utiliza para establecer su relación con la variabilidad de los aerosoles (Handschuh et al., 2023).

Los datos incluidos a partir de MERRA-2 (*Modern Era-Retrospective Reanalysis for Research and Applications Version 2*) (profundidad óptica de aerosoles, incremento de la profundidad óptica de aerosoles y altura de la capa límite atmosférica) utilizan

la nueva versión del Sistema de Asimilación de Datos del Sistema de Observación Terrestre de Goddard, Versión 5 (GEOS-5), que es un sistema de vanguardia que combina un modelo de circulación general atmosférica global (GEOS-5 AGCM) con la Interpolación Estadística de Puntos en Rejilla (GSI) del NCEP Bosilovich et al. (2015). Los datos de altura de la capa límite planetaria (PBLH) se obtuvieron del conjunto de datos de reanálisis MERRA-2 con una resolución espacial de $0.5^\circ \times 0.625^\circ$ (latitud-longitud) (Gelaro et al., 2017). Dado que los datos de reanálisis tienen la ventaja de contar con una alta resolución temporal, se pudo generar una base precisa de PBLH para la región en estudio utilizando los datos de MERRA-2 (M2T1NXFLX - Diagnóstico del Flujo de Superficie). MERRA-2 también proporciona productos de reanálisis de aerosoles, provenientes de la asimilación de numerosas observaciones provenientes de satélites y estaciones de monitoreo en superficie. Además, se incorporó el AOD obtenido a partir de MERRA-2, Zhang et al. evaluaron de manera sistemática el rendimiento de AOD de Himawari-8 y dos conjuntos de datos de AOD de reanálisis ofrecidos por MERRA-2 y el Servicio de Monitoreo de la Atmósfera de Copernicus (CAMS) (T. Zhang et al., 2020). Sus resultados mostraron que el AOD de Himawari-8 y MERRA-2 presentaron una precisión similar, mostrando variaciones diurnas significativas. Estudios previos han confirmado que el uso del AOD de MERRA-2 combinado con modelos de aprendizaje automático puede estimar las concentraciones de PM_{2.5} con una precisión razonable (P. Gupta et al., 2021). El conjunto de datos se complementó con los datos de MERRA-2 recopilados a través del servicio web de datos de radiación solar (SoDa - *Solar radiation Data*), que ofrece series temporales de temperatura (a 2 m sobre el nivel del suelo), humedad relativa (a 2 m sobre el nivel del suelo), presión (a 2 m sobre el nivel del suelo), velocidad y dirección del viento (a 10 m sobre el nivel del suelo), precipitación, nieve, espesor de nieve e irradiación horizontal global (GHI) con una resolución espacial de $0.5^\circ \times 0.5^\circ$ (aproximadamente 50 km) (Valappil et al., 2023).

El VIIRS es uno de los cinco instrumentos a bordo del satélite Suomi-NPP que opera en una órbita sincrónica con el sol a una altitud de 824 km (inclinación = 98.7°), con un nodo ascendente de tiempo solar local a la 1:30 pm (J. Wu et al., 2016). Logra una cobertura global diaria y tiene un ciclo de repetición de aproximadamente 16 días. El producto de datos de la Versión 1 de Anomalías Térmicas/Incendios (VNP14A1) proporciona información diaria sobre incendios activos y otras anomalías térmicas. El producto de datos VNP14A1 es un compuesto global en malla de 1 km de píxeles de incendios detectados a partir de las bandas del VIIRS con una resolución espacial de 375 m, lo que proporciona un rendimiento mejorado en comparación con los datos de imagen de 750 m, durante un período diario (24 horas). Los datos VNP14 están diseñados siguiendo la suite de productos de Anomalías Térmicas/Incendios del MODIS, así el producto VNP14A1 proporciona un total de cuatro capas de Conjunto de Datos Científicos (SDS) para la confianza del incendio, la potencia radiativa máxima del incendio (FRP), la evaluación de calidad (QA) y la posición del incendio dentro del escaneo (J. Wu et al., 2016).

Tabla 4.1: Fuentes de datos de variables derivadas a partir de información satelital.

Fuente	Dependencia	Espacial	Temporal	Variables
ERA5 (ECMWF reanálisis)	ECMWF	0.25°x0.25°	horaria	164
MODIS-MAIAC (MCD19A2)	NASA	0.1°×0.1°	3 a 5 (hora)	1
MERRA-2 (M2T1NXFLX)	NASA	0.625°×0.5°	horaria	1
MERRA-2 (M2I3NXGAS)	NASA	0.625°×0.5°	1 (3 horas)	1
MERRA-2 (SODA)	NASA	0.625°×0.5°	horaria	9
VIIRS (VNP14A1)	NASA	0.1°×0.1°	diaria	1
MODIS (MOD11A1/MYD11A1)	NASA	0.1°×0.1°	horaria	1

Fuente de datos, dependencia, resolución espacial, resolución temporal y número total de variables descargadas. Variables específicas descargadas de cada uno de los productos satelitales:

MODIS-MAIAC (MCD19A2) *Optical Depth 047 (Blue band (0.47 μ m) aerosol optical depth over land)*, MERRA-2 (M2T1NXFLX) *PBLH (planetary boundary layer height)*, MERRA-2 (M2I3NXGAS) *AODANA* y *AODINC* (*Aerosol Optical Depth Analysis* y *Aerosol Optical Depth Analysis Increment*), VIIRS (VNP14A1) *FireMask (Confidence of fire)* y MODIS (MOD11A1/MYD11A1) *LST Day 1km* y *LST Night 1km (Daytime Land Surface Temperature y Nighttime Land Surface Temperature)*.

Detalles Adicionales

ERA5 (ECMWF Reanalysis)

- Producto: ERA5
- Características:
 - Resolución espacial: 0.25°x0.25°
 - Resolución temporal: Horaria
- Fuente de datos: ECMWF
- Referencias: Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... & Thépaut, J. N. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999-2049.
- Sitio de descarga: Copernicus Climate Data Store

MODIS-MAIAC (MCD19A2)

- Producto: MODIS/MAIAC
- Características:
 - Resolución espacial: 0.1°x0.1°
 - Resolución temporal: 3 a 5 horas
- Fuente de datos: NASA
- Referencias: Lyapustin, A., Wang, Y., Laszlo, I., Kahn, R., Korkin, S., Remer, L., ... & Levy, R. (2018). Multi-Angle Implementation of Atmospheric Correction

(MAIAC) algorithm. Journal of Geophysical Research: Atmospheres, 123(20), 11-180.

- Sitio de descarga: NASA LAADS DAAC

MERRA-2 (M2T1NXFLX, M2I3NXGAS, SODA)

- Producto: MERRA-2
- Características:
 - Resolución espacial: 0.625°x0.5°
 - Resolución temporal: Horaria y 3 horas
- Fuente de datos: NASA
- Referencias: Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., ... & Zhao, B. (2017). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). Journal of Climate, 30(14), 5419-5454.
- Sitio de descarga: NASA GES DISC

VIIRS (VNP14A1)

- Producto: VIIRS/NPP
- Características:
 - Resolución espacial: 0.1°x0.1°
 - Resolución temporal: Diaria
- Fuente de datos: NASA
- Referencias: Csiszar, I., Schroeder, W., Giglio, L., Ellicott, E., Vadrevu, K., Justice, C. O., & Wind, B. (2014). Active fires from the Suomi NPP Visible Infrared Imaging Radiometer Suite: Product status and first evaluation results. Journal of Geophysical Research: Atmospheres, 119(2), 803-816.
- Sitio de descarga: NASA LAADS DAAC

MODIS (MOD11A1/MYD11A1)

- Producto: MODIS Terra/Aqua
- Características:
 - Resolución espacial: $0.1^\circ \times 0.1^\circ$
 - Resolución temporal: Horaria
- Fuente de datos: NASA
- Referencias: Wan, Z., & Dozier, J. (2015). A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Transactions on Geoscience and Remote Sensing*, 34(4), 892-905.
- Sitio de descarga: NASA LAADS DAAC

Variables de mayor relevancia para predecir

Aunque todas las variables utilizadas en el análisis están listadas al inicio de esta sección y se definen detalladamente en el Anexo 3, a continuación se presentan las más relevantes según la bibliografía para describir la variación en la concentración de PM_{2.5}. Estas variables son las que han sido seleccionadas para presentar su análisis exploratorio detallado en el presente capítulo.

Temperatura del aire a 2 metros por encima de la superficie de la Tierra (ERA5): calculada utilizando el modelo meteorológico ERA5 del ECMWF, se obtiene interpolando el nivel más bajo proporcionado por el modelo y el valor de temperatura en la superficie terrestre, teniendo en cuenta las condiciones atmosféricas. ERA5 es un conjunto de datos de reanálisis atmosférico que ofrece información climática de alta resolución, con una resolución espacial de aproximadamente 31 km y una resolución temporal horaria. Este parámetro se expresa en kelvin (K).

Humedad relativa a 2 metros por encima de la superficie terrestre (ERA5): establece la relación entre la presión parcial del vapor de agua y la presión de vapor de equilibrio del agua a la temperatura dada a esa altura. Esta variable es calculada a partir de

la combinación de los datos de temperatura y presión estimados por el satélite y se expresa en porcentaje (%).

Temperatura del punto de rocío (ERA5): es la temperatura a la que tendría que enfriarse el aire a 2 metros por encima de la superficie terrestre para que se produzca la saturación, es decir, para que el vapor de agua se condense en rocío o escarcha. Esta temperatura se calcula utilizando calculada utilizando el modelo meteorológico ERA5 del ECMWF, interpolando el nivel más bajo obtenido por el modelo y la medición obtenida para la superficie terrestre, teniendo en cuenta las condiciones atmosféricas. Este parámetro se expresa en kelvin (K).

Precipitación convectiva (ERA5): es la precipitación acumulada que cae sobre la superficie terrestre, obtenida a partir del esquema de convección en el Sistema Integrado de Previsión del ECMWF (IFS). Surge a partir de la combinación de información acerca de la formación y disipación de las nubes y la precipitación, derivada a partir de cambios en características atmosféricas, como presión, temperatura y humedad. Esta precipitación tiene en cuenta tanto la lluvia, como la nieve, entre otros estados intermedios de precipitación, que son acumulados horariamente. Este parámetro se expresa en metro de agua equivalente sobre la superficie terrestre, que representa la profundidad que tendría el agua si se repartiese uniformemente en la superficie de la cuadrícula de resolución en cada hora del día.

Velocidad del viento en su componente “neutra” a una altura de 10 metros sobre la superficie terrestre (ERA5): esta neutralidad se calcula a partir de la tensión superficial y de la rugosidad de la superficie terrestre, considerando una estratificación neutra de la atmósfera derivada a partir del gradiente de temperaturas. Esta componente es de menor velocidad que el viento real bajo condiciones estables, y de mayor velocidad bajo condiciones inestables. Las unidades en las que se expresa este parámetro son metros por segundo ($\frac{m}{s}$).

Presión atmosférica (ERA5): fuerza por unidad de superficie que ejerce la atmósfera

sobre la superficie terrestre. Esta es una medida del peso del aire en una columna situada verticalmente sobre un punto de la superficie terrestre. Las unidades en que se mide este parámetro son Pascales (Pa).

Altura de la capa límite atmosférica (PBLH) (ERA5): refleja la altura del estrato de aire por encima de la superficie terrestre que se ve afectado por la resistencia a la transferencia de momento, calor o humedad, que genera la superficie terrestre. En otras palabras, es la altura del estrato de aire sujeto a la influencia de la superficie terrestre. Esta puede ser tan baja como unas pocas decenas de metros, como sucede con el aire de refrigeración por la noche, o tan alta como varios kilómetros como ocurre sobre el desierto en medio de un día caluroso y soleado. Cuando es baja, pueden desarrollarse mayores concentraciones de contaminantes (emitidos en superficie). Su cálculo se basa en el número de Richardson y la unidad en la que se expresa este parámetro son los metros (m).

Radiación solar directa sin nubosidad sobre la superficie terrestre (ERA5): es la cantidad de radiación directa del sol (onda corta) que llega a la superficie terrestre, suponiendo condiciones de cielo despejado (sin nubes). Esta radiación solar puede ser dispersada en distintas direcciones por partículas presentes en la atmósfera, la parte que llega a superficie se denomina radiación solar difusa. Mientras que, la parte de la radiación solar que llega a la superficie sin ser dispersada se denomina radiación solar directa. Las cantidades de radiación a cielo despejado se calculan para las mismas condiciones atmosféricas de temperatura, humedad, ozono, gases traza y aerosoles, sin considerar la nubosidad. Las unidades en las que se expresa este parámetro son julios por metro cuadrado ($\frac{J}{m^2}$).

Profundidad óptica de aerosoles (AOD) (MODIS y MERRA-2): es un parámetro adimensional que mide la extinción de la luz debida a su dispersión y absorción por la presencia de partículas distribuidas en la columna de aire desde la superficie terrestre a la parte superior de la atmósfera. Esta variable se obtuvo a partir de dos fuentes,

en primer lugar, del algoritmo MAIAC que ha sido recientemente desarrollado para los datos obtenidos con el sensor MODIS (Alexei Lyapustin et al., 2018, 2019); y en segundo lugar, del MERRA-2 (Bosilovich et al., 2015).

Temperatura de la superficie terrestre (LST) (MODIS): se obtuvo a partir del producto MODIS LST y emisividad, (MOD21), disponible en la colección 6, el cual utiliza un algoritmo basado en la física para recuperar dinámicamente tanto esta temperatura como la emisividad simultánea para las tres bandas infrarrojas térmicas MODIS (29, 31 y 32) con una resolución espacial de 1 km en el nadir (Hulley et al., 2014; Savtchenko et al., 2004). Las unidades en las que se expresa este parámetro son kelvin (K).

Tráfico de automóviles (Google): es un parámetro adimensional obtenido a partir de los datos que se encuentran publicados en *Google Maps*; En esa plataforma puede visualizarse el tráfico promedio usual para las distintas horas en los distintos días de la semana (Y.-J. Wu et al., 2007).

4.2.3. Protocolo estadístico

A pesar de que conocemos que la variabilidad en la concentración de PM_{2.5} depende de múltiples factores actuando de forma conjunta, incluyendo las emisiones antropogénicas, la topografía y las condiciones meteorológicas, que bajo su actuar conjunto afectan los patrones de dispersión del contaminante en cuestión, el análisis exploratorio llevado a cabo en esta tesis se centra en un análisis individual (1 a 1) de cada variable contra la concentración de PM_{2.5}. Este enfoque permite identificar y verificar cómo cada variable refleja su relación con la concentración de PM_{2.5} de manera aislada. Sin embargo, no abarca la complejidad de un análisis multidimensional, que considera las interacciones simultáneas entre múltiples variables y que no es objeto de la presente tesis. Si bien reconocemos que la multiplicidad de variables contribuye a la variabilidad en la dispersión de contaminantes, el enfoque del análisis 1 a 1 es

crucial para establecer una base de datos sólida y comprensible del vínculo que reflejan las variables. Este enfoque facilita la identificación de relaciones directas entre cada variable y la concentración de PM_{2.5}. Posteriormente, durante el modelado predictivo llevado a cabo durante el capítulo 5, se considera el actuar conjunto de múltiples covariables, dejando al algoritmo predictivo de la interpretación de las complejas relaciones subyacentes entre las variables, lo cual permitirá explicar la variabilidad en la concentración de PM_{2.5}. Por este motivo en el capítulo 5, se emplean modelos predictivos avanzados que capturan estas interacciones complejas y proporcionan una interpretación más completa de cómo las distintas variables se relacionan y afectan las concentraciones de PM_{2.5}.

Análisis de series temporales

En primer lugar, se llevó a cabo un análisis descriptivo de la variabilidad conjunta de series de las variables derivadas de información satelital que ya se conoce que están asociadas a la concentración de PM_{2.5} en distintas escalas temporales (horas a lo largo del día y meses a lo largo del año). Posteriormente, se chequeó la estacionalidad de las series a través de las pruebas de Dickey-Fuller Aumentada (Dickey & Fuller, 1979) y de Kwiatkowski–Phillips–Schmidt–Shin (Kwiatkowski et al., 1992).

En una serie de tiempo estacionaria la media, varianza y covarianza no varían a lo largo del tiempo; Sin embargo tanto las series de concentración de PM_{2.5} como las de variables derivadas a partir de información satelital presentan tendencias estacionales y heterogeneidades de varianzas a lo largo del tiempo. Por este motivo, las series bajo estudio (PM_{2.5} y variables derivadas a partir de información satelital) debieron ser transformadas para lograr su estacionariedad.

La estabilización de la varianza se llevó a cabo por medio de la transformación de Box-Cox, la cual sugirió transformaciones logarítmicas y en la escala de la raíz cuadrada (Guerrero & Perera, 2004; Thompson & Fransson, 2016). Por otra parte,

la estabilización de la media dentro de cada día, en todos los casos se realizó por medio de la diferenciación de la serie considerando un paso (*lag*) de 1 hora. La diferenciación es una operación matemática que consiste en generar una nueva serie de tiempo calculando la diferencia entre los valores de la serie original espaciados en un intervalo de tiempo *k* conocido como *lag*. La diferenciación es una técnica utilizada habitualmente para eliminar la tendencia en una serie temporal (Box et al., 2015).

Posteriormente, se procedió a ajustar un modelo teórico para cada una de las series temporales de las variables derivadas a partir de información satelital. La estimación de un modelo teórico consiste en ajustar un modelo matemático a los datos observados de una serie temporal con el fin de identificar sus características. Seguidamente, para cada serie de variable satelital se obtuvieron los residuos (diferencia entre valores observados y predichos por el modelo ajustado), además luego se obtuvieron los residuos de la serie de concentración de PM_{2.5} calculados respecto del modelo teórico ajustado para cada una de las series de variables predictoras. El proceso completo se denomina pre blanqueamiento (Fuenzalida & Rosenblüth, 1990; Razavi & Vogel, 2018) y busca reducir la presencia de variación sistemática que no es relevante para la predicción.

Finalmente, se analizó la correlación cruzada entre cada una de las series filtradas de PM_{2.5} y la respectiva variable satelital. El procedimiento ejecutado permite independizarnos de la estructura de autocorrelación presente en las series temporales, cuya interpretación podría conducir a la determinación de relaciones espurias entre las variables (Razavi & Vogel, 2018). Una relación se considera espuria o aleatoria cuando las variables aparentan tener una conexión numérica, forzada por el azar o factores externos y no por una verdadera relación de vinculación entre las variables. La correlación cruzada entre series temporales indica la relación lineal entre las series a lo largo del tiempo. Este análisis es comúnmente implementado en los casos en se quiere determinar si una serie de tiempo puede predecir otra. En otras palabras, la correlación cruzada mide la covariabilidad entre dos series y cómo se relacionan

cuando están distanciadas en un determinado paso (*lag*) de tiempo. A partir de estos resultados se busca reflejar la potencial correlación de las series de variables derivadas a partir de información satelital con la variable concentración de PM_{2.5} (Probst et al., 2012). El ajuste de los modelos teóricos con los cuales se filtró la serie de concentración de PM_{2.5} fue realizado a través de modelos de media móvil integrada autorregresiva estacional o conocido por sus siglas en inglés como SARIMA (*Seasonal Autoregressive Integrated Moving Average*) (Box et al., 2015; Carmen Bas et al., 2017).

Los modelos SARIMA son una técnica estadística que permiten modelar y predecir series de tiempo con patrones estacionales. Estos se ajustaron luego de estabilizar la media intra-diaria y la varianza de las series temporales. Un modelo SARIMA se escribe como SARIMA (p, d, q) x (P, D, Q) s , donde los parámetros p, d y q denotan el orden de los términos autorregresivos (*AR*), de diferenciación y de media móvil (*MA*), respectivamente; mientras que los órdenes estacionales autorregresivos de medias móviles y de diferenciación vienen dados por los parámetros P, D y Q , respectivamente. El periodo de estacionalidad s , en este caso, es de 24 horas dado al periodo cíclico de los datos a lo largo de las horas del día. Para identificar el modelo teórico de mejor ajuste a cada variable, se inspeccionó la función de autocorrelación simple (*FAS*) y la función de autocorrelación parcial (*FAP*), junto con el criterio de información de Akaike (AIC) para cada modelo (metodología de Box-Jenkins, Box et al., 2015).

Por otra parte, el ajuste de los modelos teóricos se llevó a cabo por medio de la metodología de Box-Jenkins, en donde el modelo teórico se deduce del análisis iterativo gráfico de la *FAS* y *FAP*, y el cálculo del criterio AIC. La *FAS* es una medida de la correlación entre una serie de tiempo y sus retrasos, es decir, la *FAS* de un retraso k mide la correlación entre los valores de la serie de tiempo en el tiempo t y los valores de la serie de tiempo en el tiempo $t-k$. Por otra parte, la *FAP* mide la correlación directa entre los valores de la serie de tiempo en diferentes momentos, pero eliminando la influencia de los retrasos intermedios. Es decir, la *FAP* de un retraso k mide la

correlación entre los valores de la serie de tiempo en el tiempo t y los valores de la serie de tiempo en el tiempo $t-k$, eliminando la influencia de los valores intermedios entre t y $t-k$ (Box et al., 2015).

Una vez obtenidos los modelos teóricos para cada una de las series temporales de las variables derivadas a partir de información satelital, se procedió a filtrar la serie de concentración de PM_{2.5} y se calculó el diagrama de correlación cruzada. Este diagrama permite identificar cuáles *lags* de la variable independiente poseen mayor capacidad predictiva para estimar la variable dependiente. A la hora de analizar el gráfico de correlación cruzada entre las series filtradas, es importante destacar que la variable satelital es la independiente, mientras que la concentración de PM_{2.5} es la variable dependiente. De esta forma, debe prestarse atención a los coeficientes de correlación posicionados en los desfasajes negativos, los cuales indican la capacidad predictiva de la variable independiente desfasada un determinado intervalo de tiempo para predecir la concentración de PM_{2.5} en el tiempo t .

Una vez identificados estos retardos, se procedió a ajustar un modelo de función de transferencia dinámica lineal para estimar la dirección de las relaciones entre las series en cuestión (S. Bisgaard & Kulahci, 2006; Box et al., 1994). Los resultados obtenidos fueron corroborados a partir de la estimación de la misma relación funcional en diferencias. Si la relación se mantiene con la misma direccionalidad tanto en la estimación de la función de transferencia dinámica lineal con y sin diferenciación, entonces puede concluirse que esta ocurre en la realidad y que no se debe a efectos espurios (S. ren Bisgaard & Kulahci, 2006). Los resultados completos del proceso desarrollado pueden encontrarse en el Anexo 2. El símbolo virgulilla (~) colocado posterior al nombre de una variable (serie temporal), indica que esta ha sido previamente normalizada, es decir, centrada y estandarizada (Ecuación 4.1).

$$\tilde{x}_i = \frac{x_i - \mu}{\sigma} \quad (4.1)$$

Ecuación 4.1: Normalización de una variable centrada y estandarizada (\tilde{x}_i), donde x_i es el valor original, μ es la media de la variable, y σ es la desviación estándar de la variable.

En la práctica, el análisis de la correlación cruzada se lleva a cabo mediante la construcción de un gráfico de la función de correlación cruzada luego de haber realizado el proceso de pre blanqueamiento (*prewhitening*) de las series temporales. Este paso es crucial ya que permite eliminar la autocorrelación presente en las series, garantizando una evaluación aún más precisa de la correlación entre las dos series a distintos desfasajes (*lags*). La significancia de los coeficientes de correlación se determina a partir de los gráficos presentes en el Anexo 2. Estos gráficos muestran los valores de los coeficientes de correlación cruzada en el eje vertical, mientras que los desfasajes en el eje horizontal. Las líneas azules horizontales en el gráfico indican los intervalos de confianza al 95 %. Si los coeficientes de correlación cruzada superan estas líneas, se consideran estadísticamente significativos, independientemente de que puedan tener valores de baja magnitud, comparando con coeficientes de correlación convencionales (lo cual no es lo mismo). En estos gráficos pueden observarse coeficientes de correlación cruzada que superan los intervalos de confianza, indicando que son significativos y deben ser tenidos en cuenta en el análisis, a pesar de que algunos de ellos presenten valores relativamente pequeños.

4.3. Resultados

En esta sección se reportan los resultados obtenidos en el análisis descriptivo de la variabilidad conjunta de las variables meteorológicas satelitales más relevantes en relación con la concentración de PM_{2.5}.

Temperatura

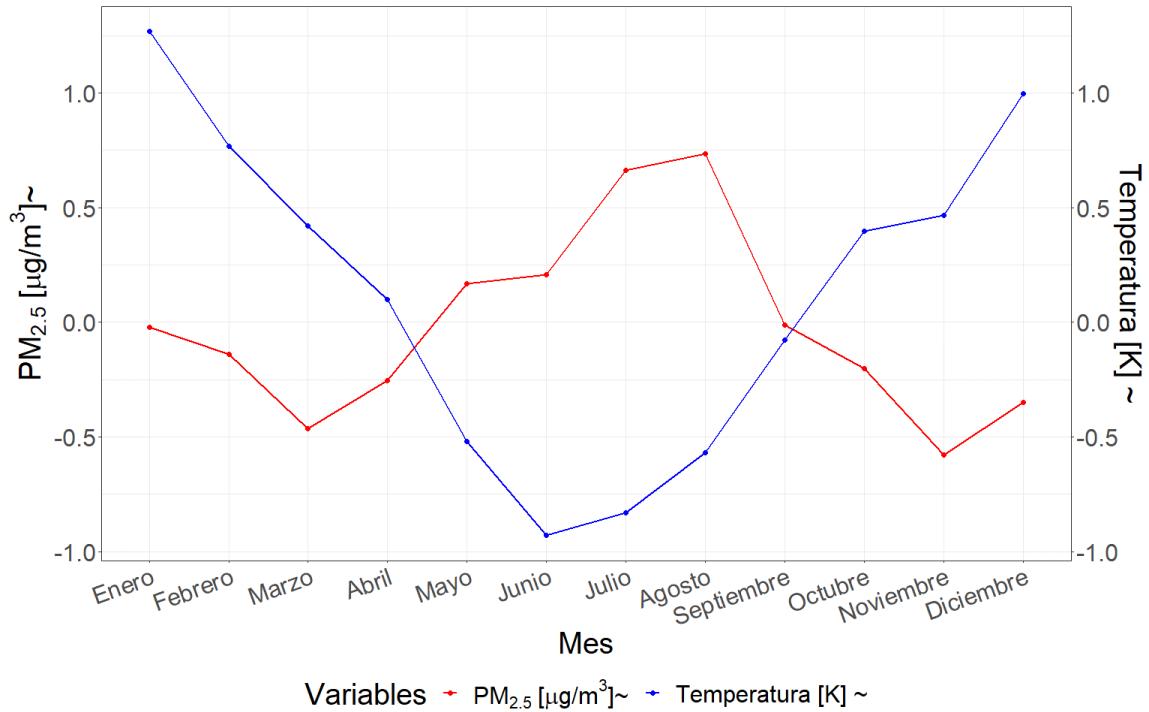


Figura 4.2: Perfil de variación conjunto de los promedios mensuales del $\text{PM}_{2.5}$ y la temperatura a lo largo del año (2022-2023), los promedios calculados consideran todos los sitios de muestreo ubicados en la provincia de Córdoba, Argentina.

En la Figura 4.2 se presenta la relación existente entre la temperatura y la concentración de $\text{PM}_{2.5}$ a lo largo del año. Puede observarse que durante los meses en que las temperaturas promedio son más bajas, la concentración promedio de $\text{PM}_{2.5}$ aumenta, mientras que durante los meses en que la temperatura se eleva por encima de su media, las concentraciones disminuyen.

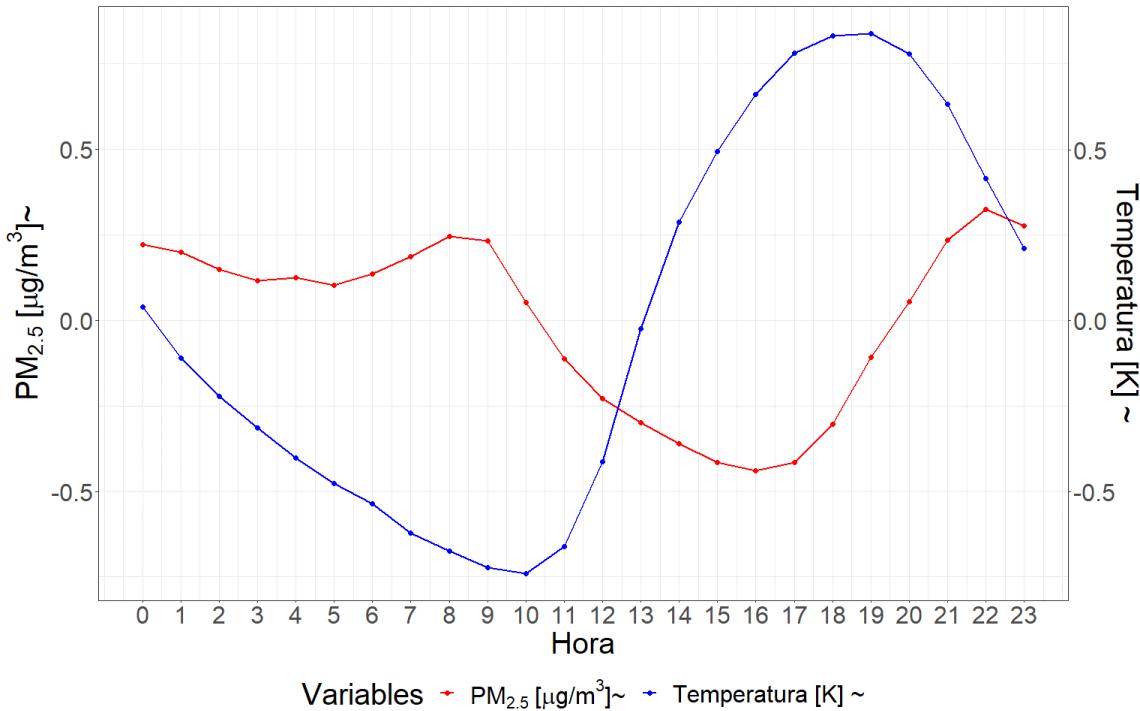


Figura 4.3: Perfil de variación conjunto de los promedios horarios del PM_{2.5} y la temperatura a lo largo de un día promedio en la provincia de Córdoba, Argentina.

La Figura 4.3 muestra la variación cíclica que describe la temperatura promedio a lo largo de las horas del día, caracterizada por dos puntos de inflexión para la curva de la temperatura: uno en la mañana a las 10 horas y otro en la tarde a las 19 horas. Durante las horas de la mañana, la temperatura desciende continuamente hasta llegar a su mínimo a las 10 horas, mientras que en las horas de la tarde, la temperatura aumenta constantemente hasta alcanzar su máximo a las 19 horas.

Los resultados indican que temperaturas bajas favorecen mayores concentraciones de PM_{2.5}, mientras que temperaturas altas se asocian con una reducción en la concentración sugiriendo una relación inversa entre la temperatura y la concentración de PM_{2.5} promedio en las escalas de variación temporal tanto mensual.

Tabla 4.2: Coeficientes de la función de transferencia dinámica lineal estimados para la covariable analizada en sus promedios horarios a lo largo de todos los sitios de muestreo localizados en la provincia de Córdoba, Argentina, considerando todo el periodo de muestreo analizado que abarca desde el 30/03/2021 hasta el 01/06/2022.

Desfasaje	Con diferenciación	Sin diferenciación
t	-0.1573001	-1.04115
t-1	0.2441853	1.44786
t-12	-0.0197213	-0.06497
t-24	-0.2628158	-0.68048

Las correlaciones significativas que perduran luego del procedimiento de filtrado de las series se observaron en los tiempos t, t-1, t-12 y t-24, que representan al tiempo a 0, 1, 12 y 24 horas, indicando que son los desfasajes temporales de mayor capacidad predictiva para predecir la concentración de PM_{2.5} en el presente a partir de la información pasada de temperatura.

Luego de la estimación de la relación funcional se determinó que tres de los cuatro coeficientes significativos demostraron una relación inversa estadísticamente significativa entre temperatura (t, t-12 y t-24) y la concentración de PM_{2.5}. Sin embargo, para el *lag* t-1 se detectó una relación directa positiva (Tabla 5.5).

Humedad relativa

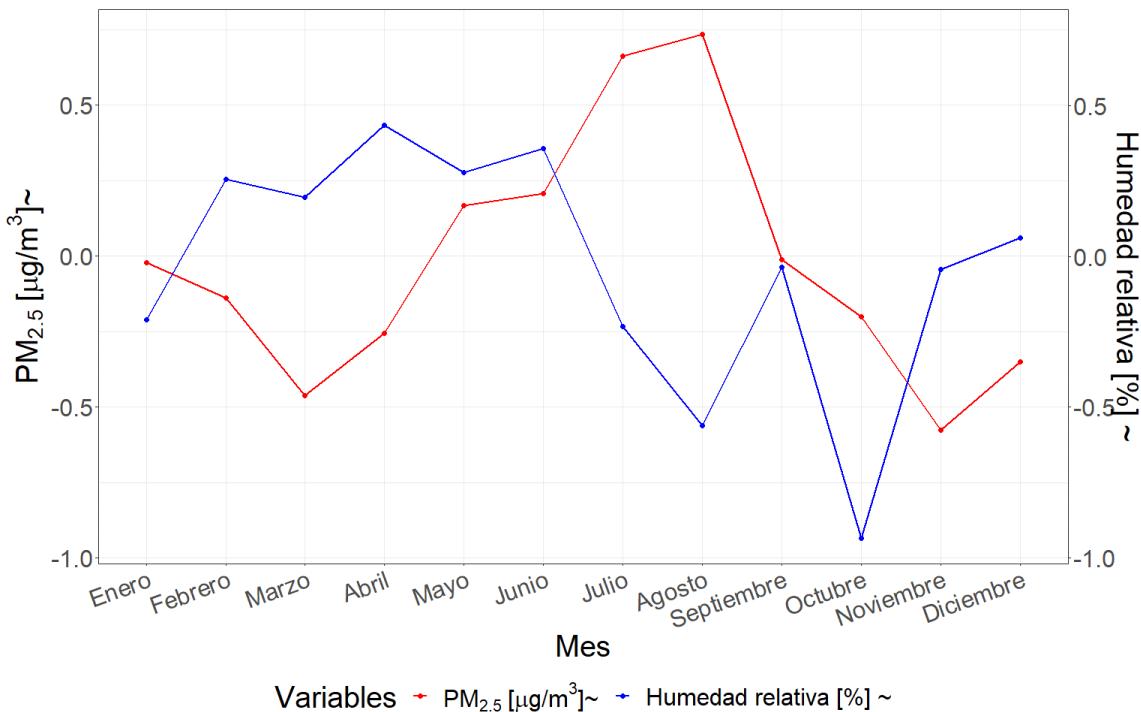


Figura 4.4: Perfil de variación conjunto de los promedios mensuales del PM_{2.5} y la humedad relativa a lo largo del año (2022-2023), los promedios calculados consideran todos los sitios de muestreo ubicados en la provincia de Córdoba, Argentina.

En la Figura 4.4 puede visualizarse la relación que describen los promedios mensuales de la humedad relativa y la concentración de PM_{2.5} a lo largo del año. Esta escala temporal permite apreciar que los meses en que la humedad relativa toma valores por debajo de su media, la concentración de PM_{2.5} promedio es alta, mientras que para los meses en que la humedad relativa es alta la concentración de PM_{2.5} es baja, con respecto a su media. Sin embargo, esta relación no se mantiene constante a lo largo del año, como puede observarse por ejemplo en el mes de octubre donde la relación se modifica mostrando una relación de proporcionalidad directa.

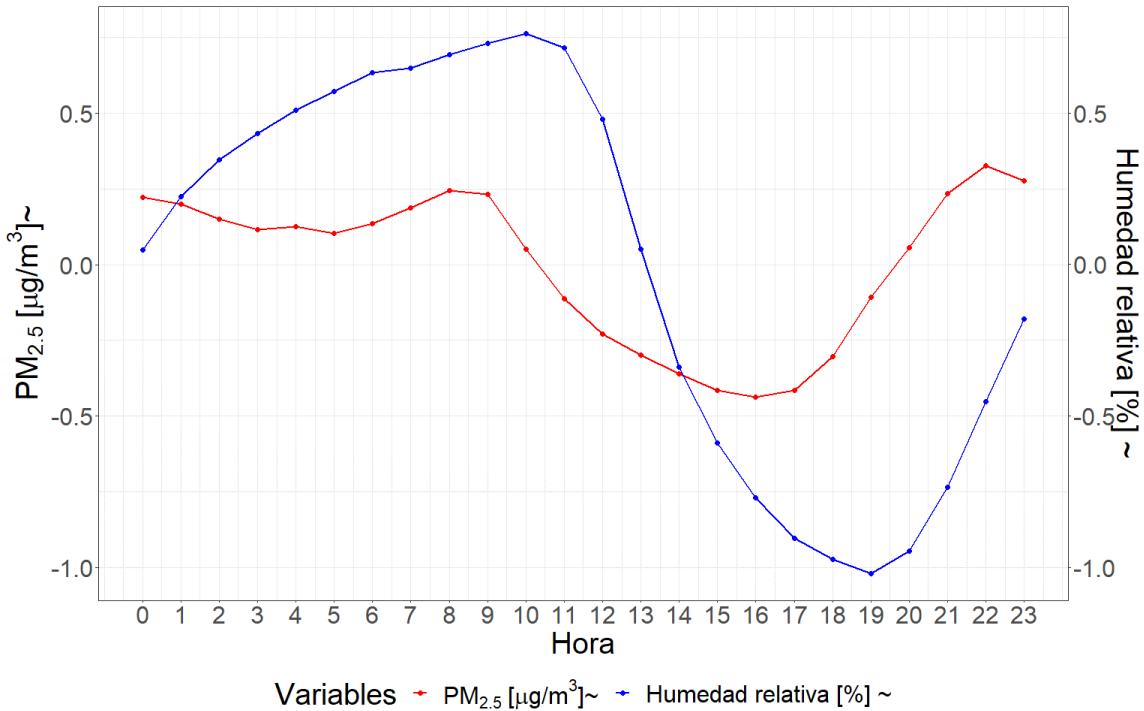


Figura 4.5: Perfil de variación conjunto de los promedios horarios del PM_{2.5} y la humedad relativa a lo largo de un día promedio, en la provincia de Córdoba, Argentina.

La Figura 4.5 exhibe la variación promedio de la humedad relativa a lo largo del día. Este patrón cíclico presenta dos puntos de inflexión para la curva de la humedad relativa, uno a las 10 horas y otro a las 19 horas, lo que sugiere la existencia de dos momentos antagónicos durante el día. Entre las 10 y las 19 horas, la humedad relativa disminuye, mientras que en el resto del día aumenta constantemente para el período de muestreo analizado (30/03/2021-01/06/2022). Puede inferirse que una disminución de la humedad relativa con respecto a su media favorece un incremento de la concentración de PM_{2.5}, mientras que un aumento favorece su disminución.

Tabla 4.3: Coeficientes de la función de transferencia dinámica lineal estimados para la covariable analizada en sus promedios horarios a lo largo de todos los sitios de muestreo localizados en la provincia de Córdoba, Argentina, considerando todo el periodo de muestreo analizado que abarca desde el 30/03/2021 hasta el 01/06/2022.

Desfasaje	Con diferenciación	Sin diferenciación
t-5	-0.2681536	-0.2548
t-24	0.1367988	0.1309

Los desfasajes temporales de mayor capacidad predictiva para predecir la concentración de PM_{2.5} en el presente a partir de la información pasada de humedad relativa, fueron los tiempos t-5 y t-24, que perduraron luego del procedimiento de filtrado de las series. Luego de la estimación de la relación funcional se determinó que cada uno de los coeficientes significativos demostró una relación opuesta de vinculación. El *lag* temporal de 5 horas resultó en una relación inversa estadísticamente significativa, mientras que el *lag* de 24 horas se asoció de manera directa y estadísticamente significativa, entre la humedad relativa y la concentración de PM_{2.5}.

Temperatura del punto de rocío

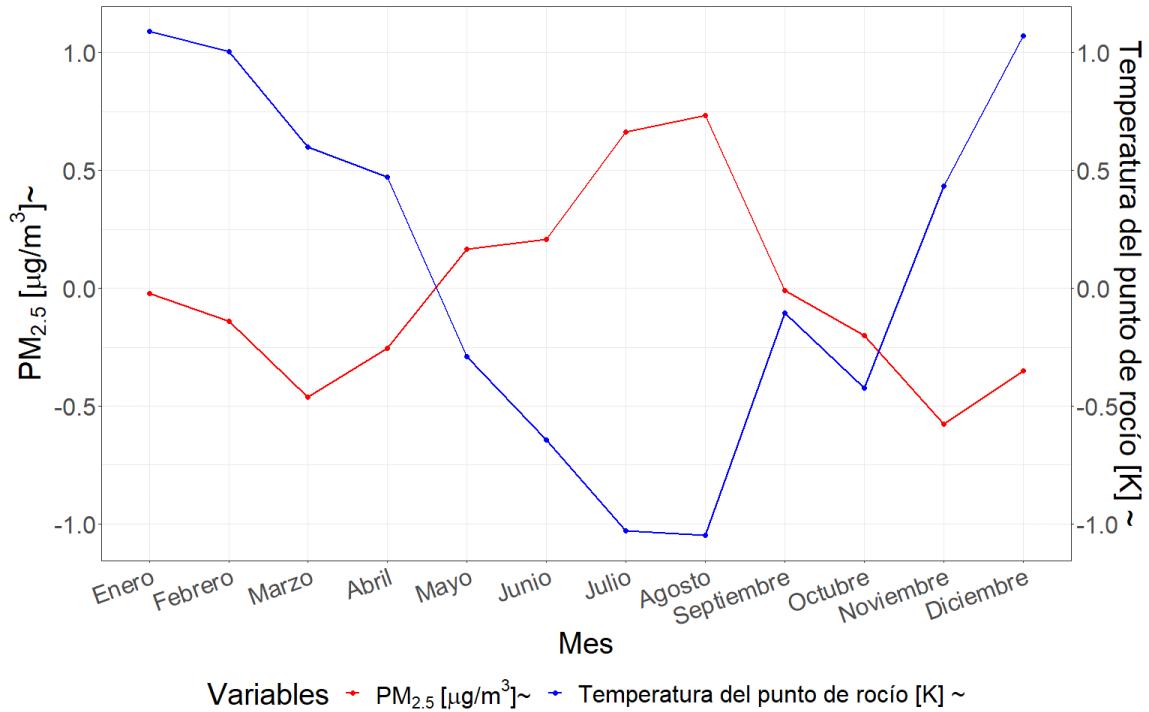


Figura 4.6: Perfil de variación conjunto de los promedios mensuales del PM_{2.5} y la temperatura del punto de rocío a lo largo del año (2022-2023), los promedios calculados consideran todos los sitios de muestreo ubicados en la provincia de Córdoba, Argentina.

Los meses en los que la temperatura del punto de rocío está por debajo de su media, la concentración de PM_{2.5} promedio es alta, mientras que en los meses en los que la temperatura del punto de rocío es alta, la concentración de PM_{2.5} también lo es (Figura 4.6).

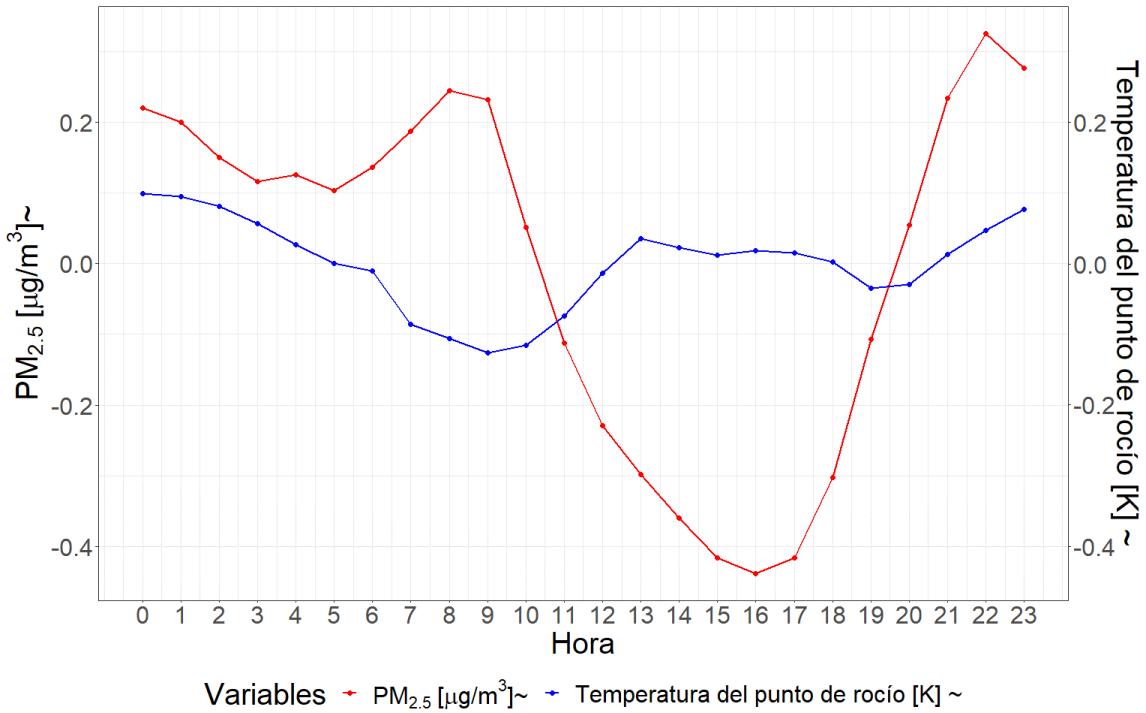


Figura 4.7: Perfil de variación conjunto de los promedios horarios del PM_{2.5} y la temperatura del punto de rocío a lo largo de un día promedio en la provincia de Córdoba, Argentina.

La Figura 4.7 muestra el ciclo de variación que describe la temperatura del punto de rocío promedio a lo largo de las horas del día. A medida que transcurren las horas, la temperatura del punto de rocío experimenta un ciclo de variación inverso al patrón de concentración media de PM_{2.5}. En otras palabras, cuando la temperatura del punto de rocío aumenta en sus promedio calculados para las distintas horas del día, se observan menores concentraciones del contaminante, como sucede entre las 10 y 20 horas, el periodo del día en que las concentraciones son más bajas. En contraposición, el pico de concentraciones entre las 6 y 10 horas se produce al mismo tiempo en que se registra un descenso en la temperatura del punto de rocío. A partir de esta información puede interpretarse que la relación que mejor describe el vínculo entre las variables en cuestión es una relación inversa.

Tabla 4.4: Coeficientes de la función de transferencia dinámica lineal estimados para la covariable analizada en sus promedios horarios a lo largo de todos los sitios de muestreo localizados en la provincia de Córdoba, Argentina, considerando todo el periodo de muestreo analizado que abarca desde el 30/03/2021 hasta el 01/06/2022.

Desfasaje	Con diferenciación	Sin diferenciación
t-11	-0.1597575	-0.2921
t-14	-0.3212396	-0.3222
t-20	0.3752728	-0.3377

Los desfasajes temporales de mayor capacidad predictiva para predecir la concentración de PM_{2.5} en el presente a partir de la información pasada de temperatura del punto de rocío fueron los tiempos t-11, t-14 y t-20. Luego de la estimación de la relación funcional se determinó que dos de los tres coeficientes significativos demostraron una relación inversa estadísticamente significativa entre la temperatura del punto de rocío (t-11 y t-14) y la concentración de PM_{2.5}. Sin embargo, para el *lag* t-20 se detectó una relación espuria o aleatoria, invirtiendo su signo en ambas estimaciones de la relación funcional.

Precipitación

Es importante destacar que el análisis exploratorio de la precipitación en la escala temporal horaria no es representativo dada la frecuencia de ocurrencia de este fenómeno a lo largo del año, resultando en un gran porcentaje de datos nulos que pueden interferir con las interpretaciones. Por lo tanto, se decidió no incluir el análisis exploratorio en esta escala temporal para evitar posibles conclusiones erróneas. Sin embargo, para analizar la correlación entre ambas series temporales, se obtuvieron resultados significativos de correlación para los días en que este fenómeno ocurría.

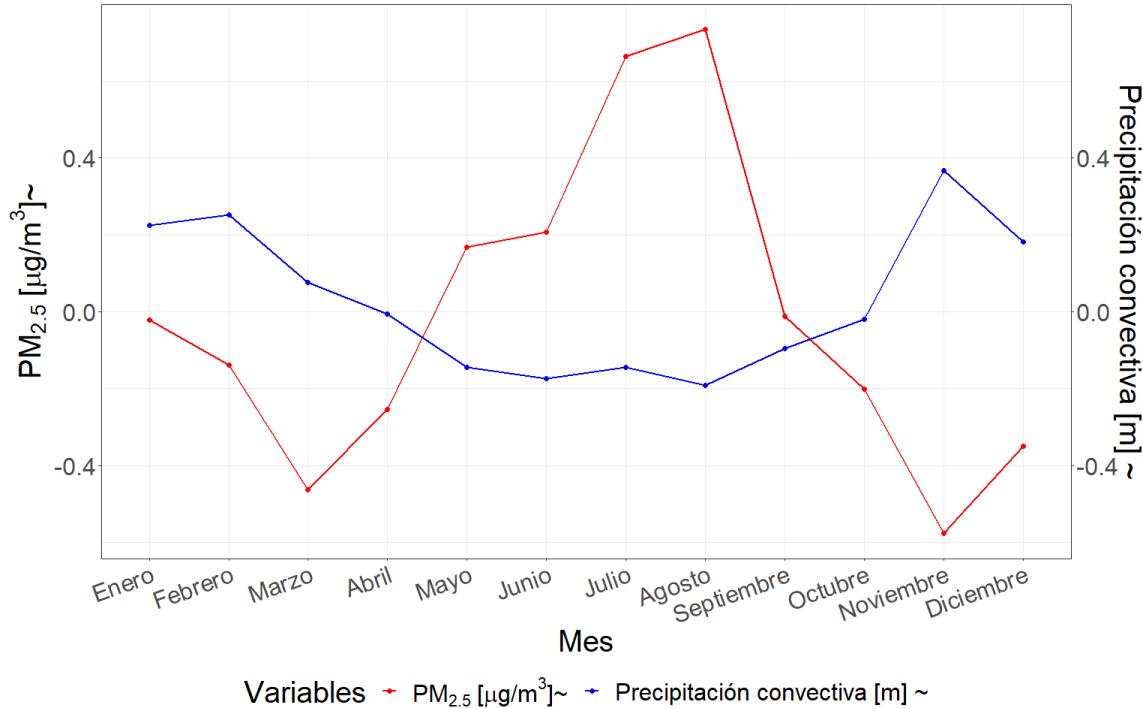


Figura 4.8: Perfil de variación conjunto de los promedios mensuales del PM_{2.5} y la precipitación a lo largo del año (2022-2023), los promedios calculados consideran todos los sitios de muestreo ubicados en la provincia de Córdoba, Argentina.

Durante la primavera y el verano, las precipitaciones aumentan notablemente en comparación con el otoño e invierno (Figura 4.8). Además, los mayores valores de precipitación se registraron en los meses de noviembre y febrero, mientras que los meses de mayo a octubre presentaron las menores precipitaciones en todo el periodo de muestreo analizado que abarca desde el 30/03/2021 hasta el 01/06/2022..

Tabla 4.5: Coeficientes de la función de transferencia dinámica lineal estimados para la covariable analizada en sus promedios horarios a lo largo de todos los sitios de muestreo localizados en la provincia de Córdoba, Argentina, considerando todo el periodo de muestreo analizado que abarca desde el 30/03/2021 hasta el 01/06/2022.

Desfasaje	Con diferenciación	Sin diferenciación
t-1	-6.162e+01	-3466.73

La única correlación significativa que perdura luego del procedimiento de filtrado de las series se observa en el tiempo t-1. Este coeficiente demostró una relación inversa estadísticamente significativa entre la precipitación y la concentración de PM_{2.5}.

Velocidad del viento

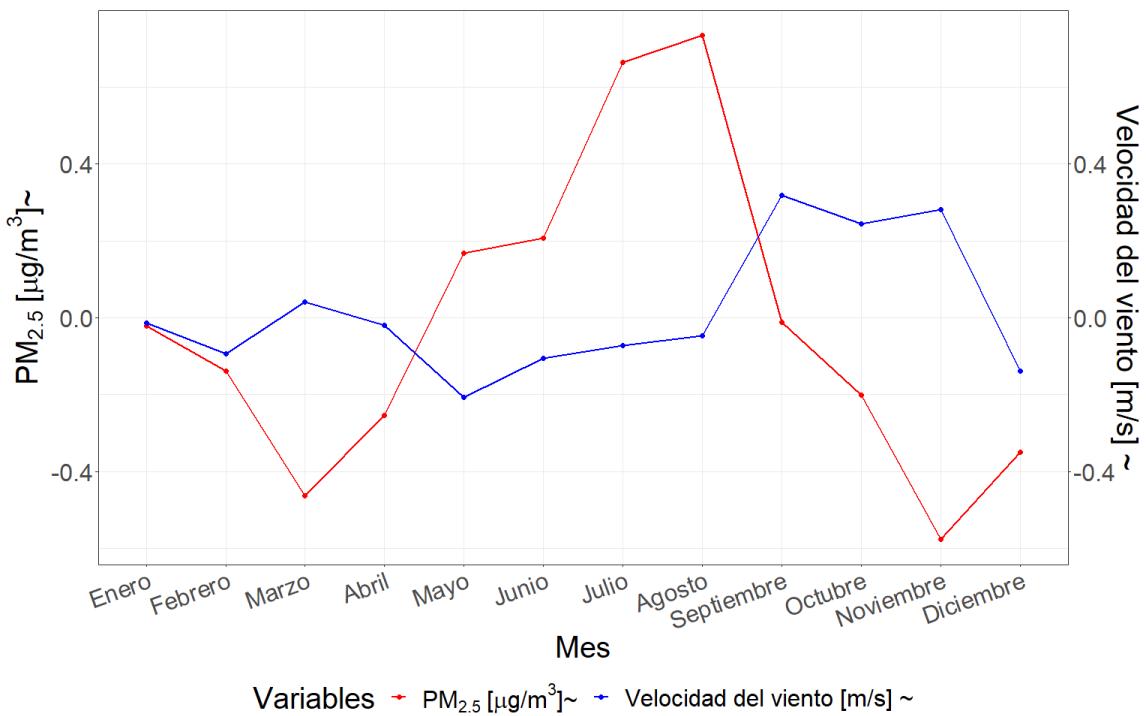


Figura 4.9: Perfil de variación conjunto de los promedios mensuales del PM_{2.5} y la velocidad del viento a lo largo del año (2022-2023), los promedios calculados consideran todos los sitios de muestreo ubicados en la provincia de Córdoba, Argentina.

La Figura 4.9 presenta el patrón de variación de los promedios mensuales de la velocidad del viento a lo largo del año con respecto a la concentración media de PM_{2.5}. Los mayores valores de velocidad del viento ocurren durante parte del invierno y la primavera, alcanzando sus máximos registros en los meses de septiembre y octubre. Por otra parte, los valores de los registros disminuyen durante el verano y el otoño, describiendo el mínimo durante el mes de febrero.

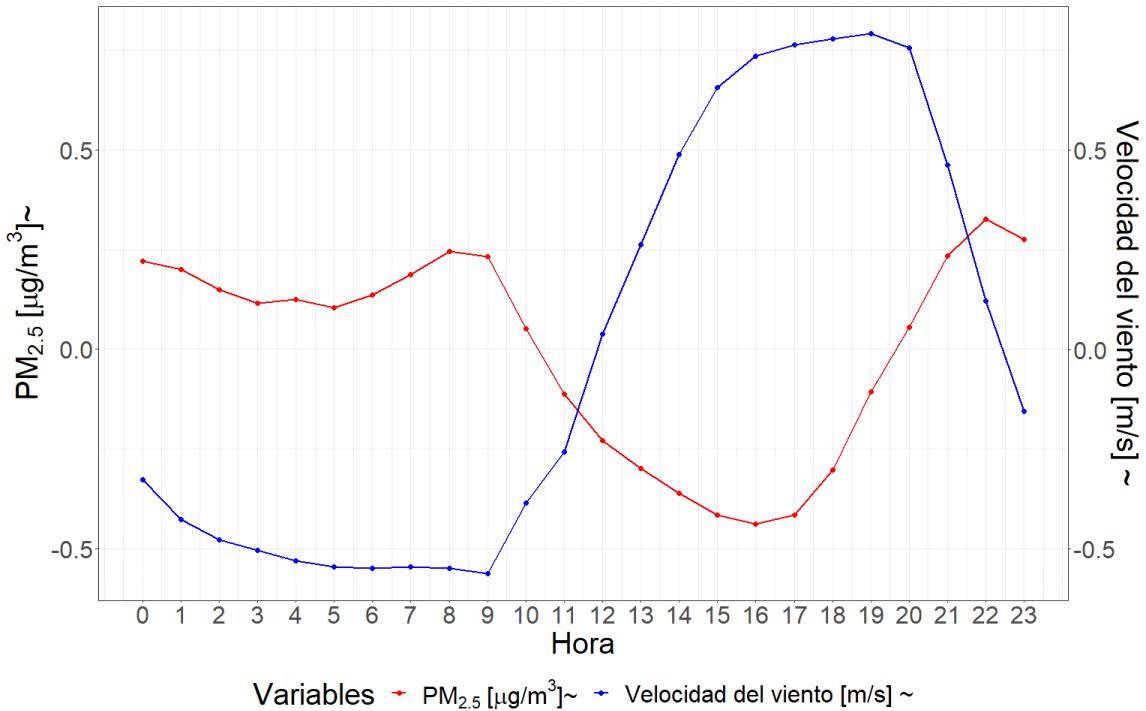


Figura 4.10: Perfil de variación conjunto de los promedios horarios del $\text{PM}_{2.5}$ y la velocidad del viento a lo largo de un día promedio en la provincia de Córdoba, Argentina.

La Figura 4.10 muestra el ciclo de variación que describe la velocidad del viento promedio a lo largo de las horas del día. Se observa un incremento entre las 10 y 20 horas, y las máximas magnitudes promedio entre las 15 y 20 horas, mientras que el resto del día la intensidad del viento es baja, mostrando su mínimo a las 9 horas. La variación conjunta con respecto a la concentración promedio horaria de $\text{PM}_{2.5}$ permite establecer que cuando la velocidad decrece, la concentración de $\text{PM}_{2.5}$ aumenta, por lo que puede deducirse una relación inversa entre ambas variables.

Tabla 4.6: Coeficientes de la función de transferencia dinámica lineal estimados para la covariable analizada en sus promedios horarios a lo largo de todos los sitios de muestreo localizados en la provincia de Córdoba, Argentina, considerando todo el periodo de muestreo analizado que abarca desde el 30/03/2021 hasta el 01/06/2022.

Desfasaje	Con diferenciación	Sin diferenciación
t-5	0.1016645	-0.1288
t-7	0.1317611	-0.03429
t-8	0.1445327	-0.006416
t-24	-0.1998146	-0.3433

Las correlaciones significativas que perduran luego del procedimiento de filtrado de las series se observaron en los tiempos t-5, t-7, t-8 y t-24, tres de estos cuatro coeficientes significativos demostraron estar influenciados por una relación aleatoria (t-5, t-7 y t-8). Mientras que, el coeficiente con un *lag* de t-24 horas, resultó en una relación inversa estadísticamente significativa entre velocidad del viento y la concentración de PM_{2.5}.

Presión atmosférica

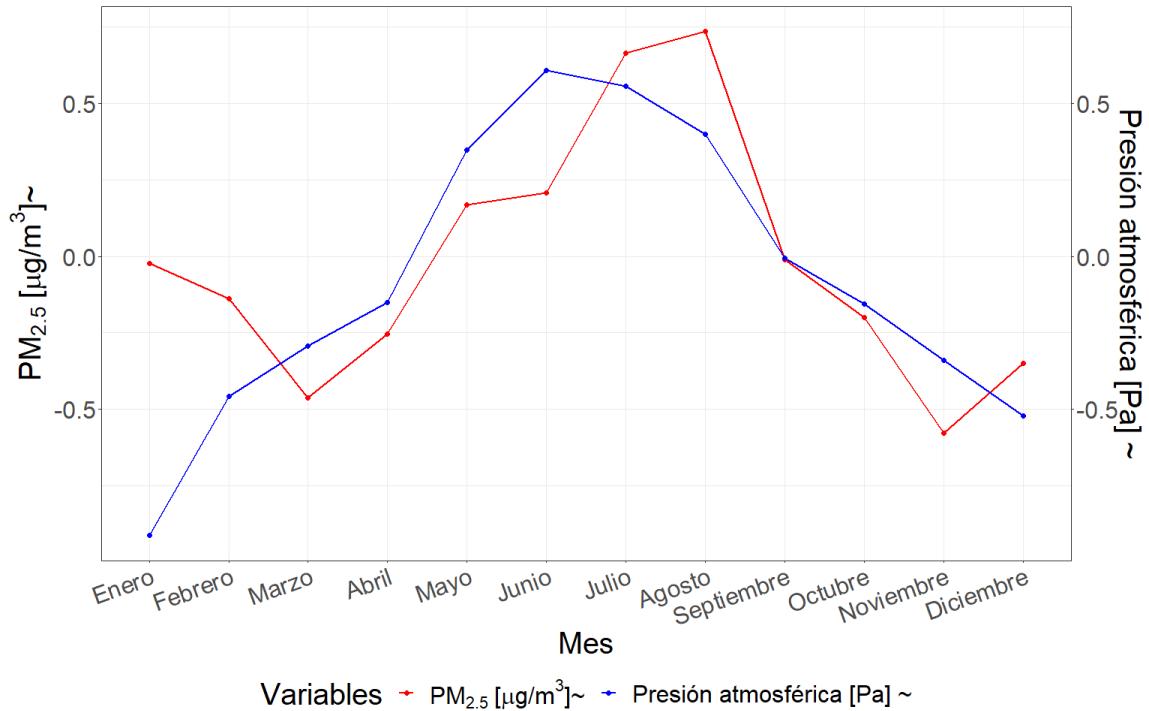


Figura 4.11: Perfil de variación conjunto de los promedios mensuales del PM_{2.5} y la presión a lo largo del año (2022-2023), los promedios calculados consideran todos los sitios de muestreo ubicados en la provincia de Córdoba, Argentina.

En la Figura 4.11 se observa la variación conjunta de los promedios mensuales de la presión atmosférica con respecto a la concentración promedio de PM_{2.5} a lo largo del año. Esta variable alcanza sus valores máximos durante el otoño y el invierno, en los meses de junio y julio, mientras que en la primavera y el verano, la magnitud de esta variable decrece. Por otra parte, la concentración promedio de PM_{2.5} se relaciona de manera inversa con la presión atmosférica, alcanzando sus valores máximos durante los meses de primavera y verano y disminuyendo durante el otoño e invierno. Esto indica una relación directa entre ambas variables.

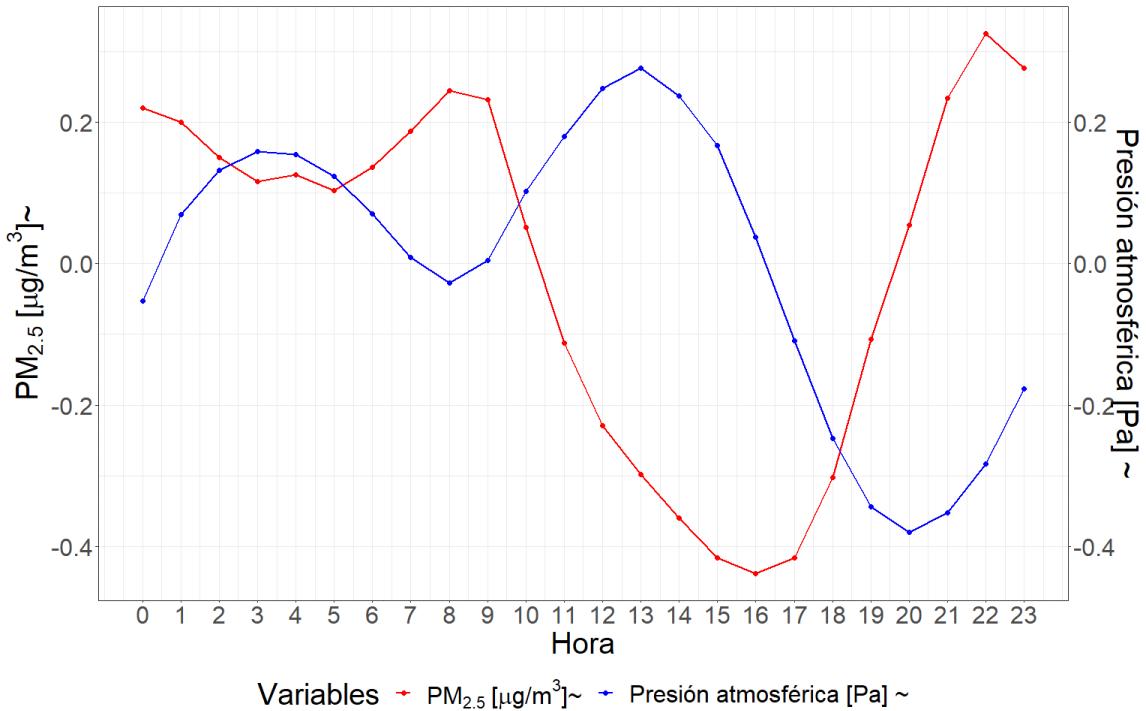


Figura 4.12: Perfil de variación conjunto de los promedios horarios del PM_{2.5} y la presión a lo largo de un día promedio en la provincia de Córdoba, Argentina.

En la Figura 4.12, se presenta el patrón de variación promedio de la presión atmosférica a lo largo de las horas del día. Puede notarse que la presión alcanza sus valores mínimos en dos momentos del día: entre las 7 y 9 horas y entre las 19 y 21 horas. De la misma forma, alcanza sus valores máximos en dos momentos del día: entre las 2 y 4 horas y entre las 12 y 14 horas. Al analizar la variación horaria conjunta entre ambas variables, puede apreciarse una relación inversa. En otras palabras, los picos de presión generan una notable reducción en la concentración de PM_{2.5}, mientras que reducciones de presión generan picos en la concentración del contaminante.

Tabla 4.7: Coeficientes de la función de transferencia dinámica lineal estimados para la covariable analizada en sus promedios horarios a lo largo de todos los sitios de muestreo localizados en la provincia de Córdoba, Argentina, considerando todo el periodo de muestreo analizado que abarca desde el 30/03/2021 hasta el 01/06/2022.

Desfasaje	Con diferenciación	Sin diferenciación
t	-0.003024	-3.434e-03
t-9	0.002529	-7.623e-04
t-19	0.008081	6.478e-03

Las correlaciones significativas que perduran luego del procedimiento de filtrado de las series se observaron en los tiempos t, t-9 y t-19, uno de los tres coeficientes significativos demostró una relación del tipo aleatoria (t-9), mientras que los dos restantes exhibieron una relación estadísticamente significativa, tanto inversa en el presente (t) como directa teniendo en cuenta un *lag* de t-19, entre la presión atmosférica y la concentración de PM_{2.5}.

Radiación solar

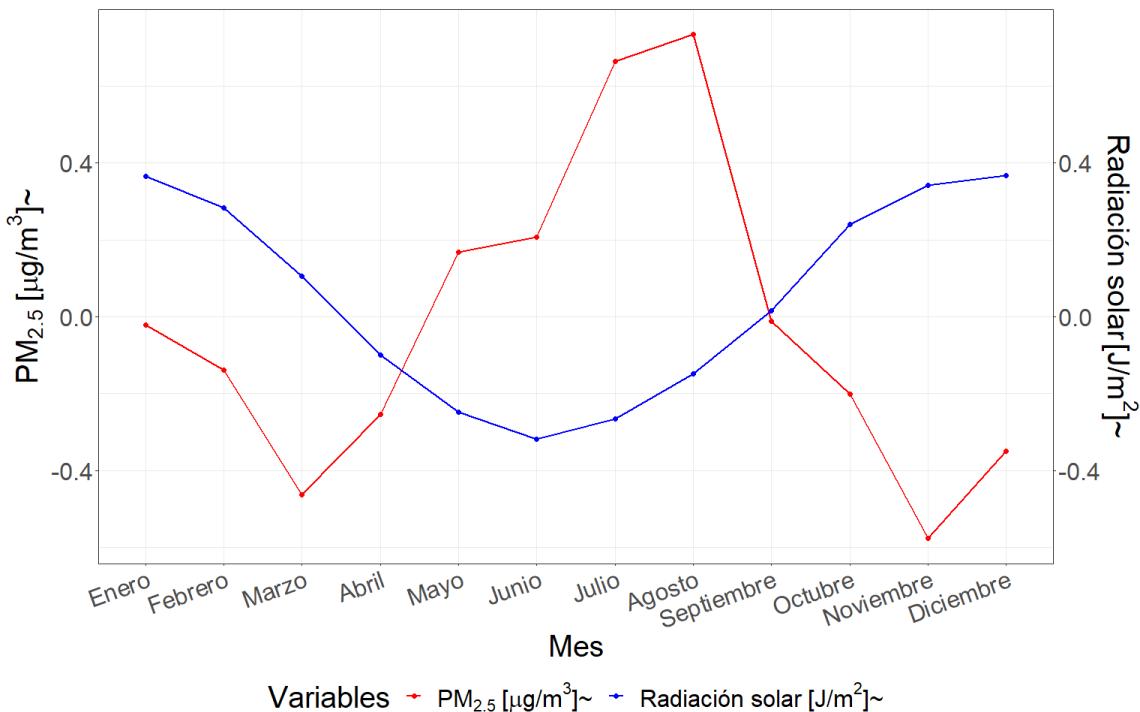


Figura 4.13: Perfil de variación conjunto de los promedios mensuales del $\text{PM}_{2.5}$ y la radiación de corta longitud de onda a lo largo del año (2022-2023), los promedios calculados consideran todos los sitios de muestreo ubicados en la provincia de Córdoba, Argentina.

La Figura 4.13 muestra el patrón de variación de la radiación solar promedio a lo largo de los meses del año. Puede observarse que esta disminuye durante el otoño e invierno, alcanzando sus mayores valores durante la primavera y el verano, específicamente durante los meses de noviembre y diciembre, mientras que los mínimos se alcanzan durante mayo y junio. La relación de vinculación aparenta ser del tipo inversa, ya que radiaciones de baja magnitud coinciden con el aumento de la concentración de $\text{PM}_{2.5}$, mientras que aumentos de radiación están asociados con disminuciones de los niveles de $\text{PM}_{2.5}$.

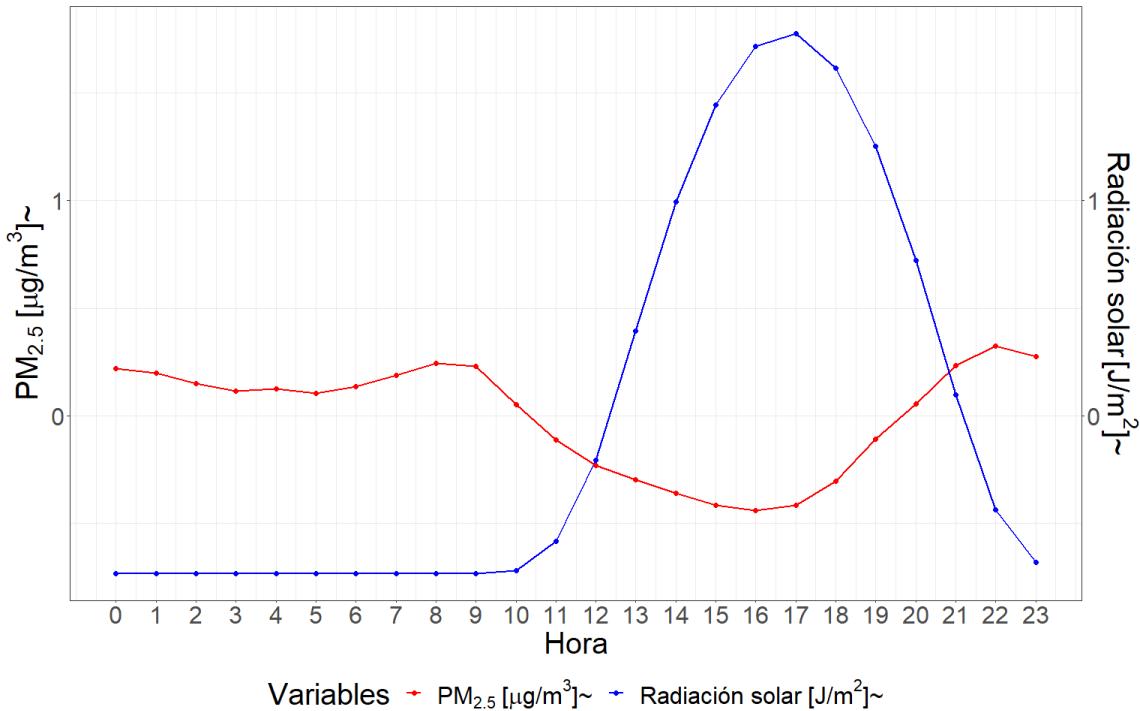


Figura 4.14: Perfil de variación conjunto de los promedios horarios del PM_{2.5} y la radiación de corta longitud de onda a lo largo de un día promedio en la provincia de Córdoba, Argentina.

La Figura 4.14 presenta el perfil de variación promedio de la radiación solar a lo largo de las horas del día, el cual exhibe dos puntos de inflexión. El primero ocurre a las 10 horas en la mañana, a partir del cual se describe una tendencia incremental en las magnitudes registradas. El segundo se alcanza a las 17 horas, momento a partir del cual se describe una tendencia decreciente hasta alcanzar los valores mínimos de radiación, que se dan durante la noche y la madrugada. Por otra parte, la relación que mejor describe la vinculación entre la radiación solar y la concentración de PM_{2.5} es inversa (Figura 4.14).

Desfasaje	Con diferenciación	Sin diferenciación
t-13	4.023e-07	2.753e-07
t-23	-1.884e-06	-2.413e-06

Las correlaciones significativas que perduran luego del procedimiento de filtrado de las series se observaron en los tiempos t-13 y t-23. El desfasaje temporal con un *lag* de 13 horas resultó en una relación directa estadísticamente significativa, mientras que el *lag* de 23 horas resultó en una inversa estadísticamente significativa entre la radiación solar y la concentración de PM_{2.5}.

Profundidad óptica de los aerosoles (AOD)

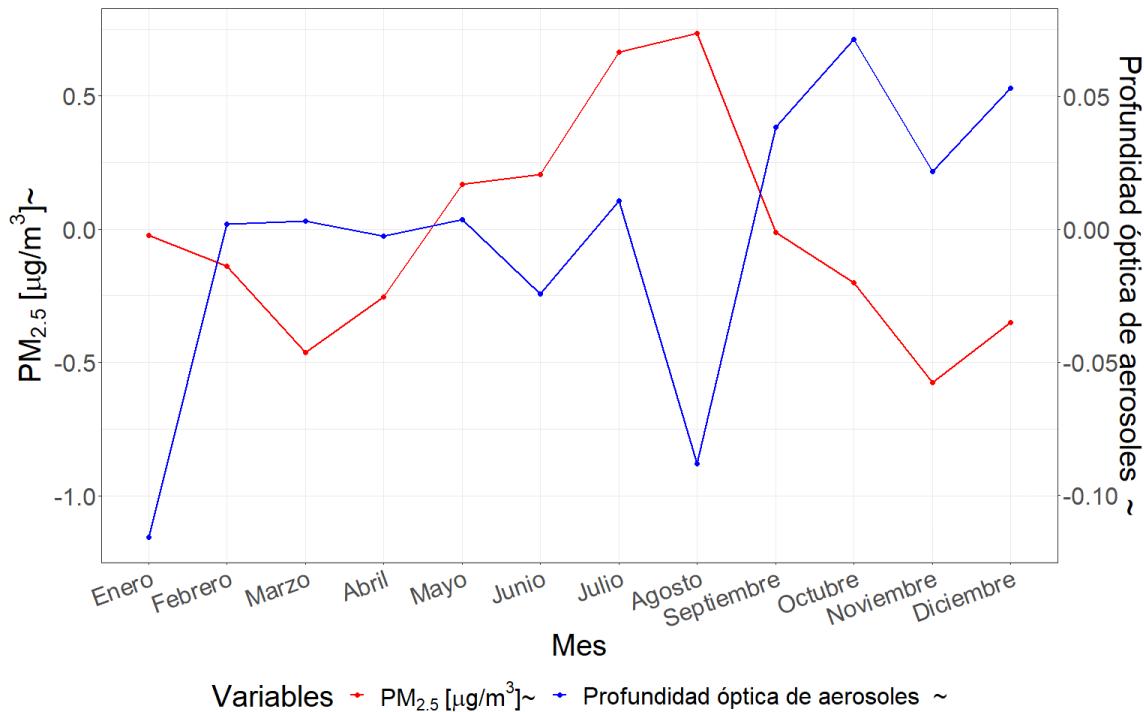


Figura 4.15: Perfil de variación conjunto de los promedios mensuales del PM_{2.5} y la profundidad óptica de los aerosoles a lo largo del año (2022-2023), los promedios calculados consideran todos los sitios de muestreo ubicados en la provincia de Córdoba, Argentina.

En la Figura 4.15 se observa el patrón de variación conjunto entre los promedios mensuales de AOD y la concentración de PM_{2.5} a lo largo del año. Durante los meses de baja AOD la concentración de PM_{2.5} aumenta, mientras que durante los meses en que el AOD es alto los niveles de PM_{2.5} son bajos. Estos resultados indican una relación inversa entre ambas variables en la escala mensual.

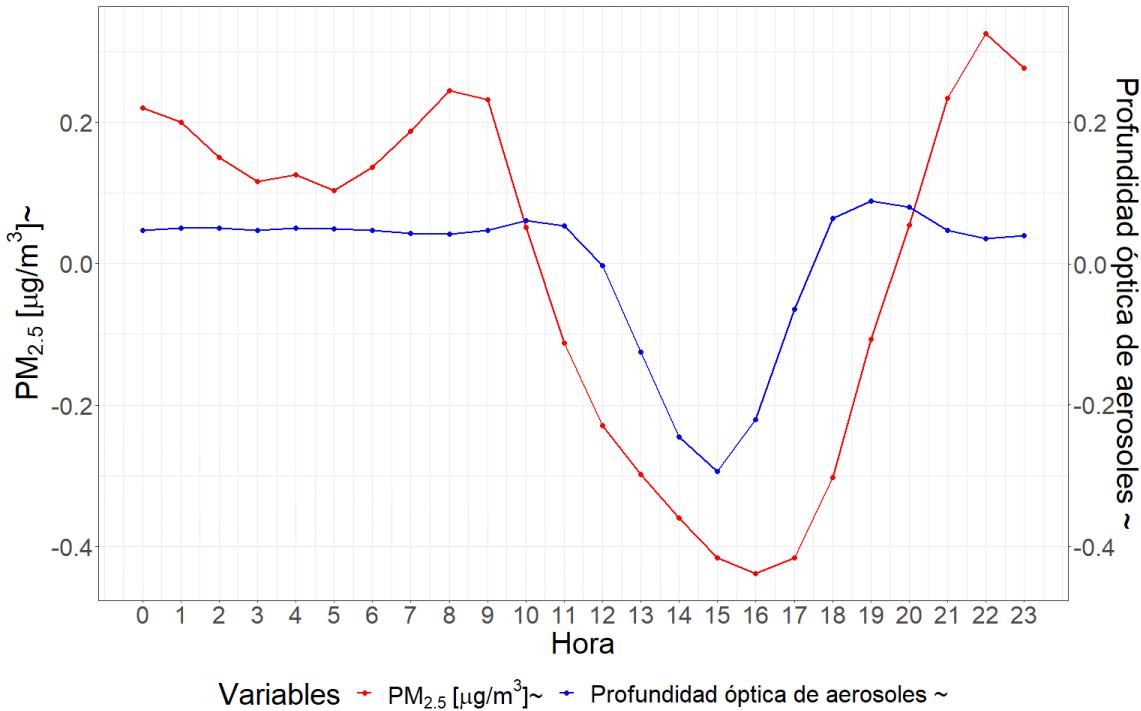


Figura 4.16: Perfil de variación conjunto de los promedios horarios del PM_{2.5} y la profundidad óptica de los aerosoles a lo largo de un día promedio en la provincia de Córdoba, Argentina.

En la Figura 4.16, se presenta el patrón de variación conjunto del AOD y la concentración de PM_{2.5} a lo largo de las horas del día, lo cual indica una relación directa entre ambas variables. Los menores registros tanto de concentración de PM_{2.5}, como de AOD ocurren de forma simultánea, mientras que los incrementos en el valor medio de ambas variables también ocurren de forma conjunta y simultanea.

Tabla 4.9: Coeficientes de la función de transferencia dinámica lineal estimados para la covariable analizada en sus promedios horarios a lo largo de todos los sitios de muestreo localizados en la provincia de Córdoba, Argentina, considerando todo el periodo de muestreo analizado que abarca desde el 30/03/2021 hasta el 01/06/2022.

Desfasaje	Con diferenciación	Sin diferenciación
t	6.777e+00	0.4331
t-7	7.795e+00	-0.9688

Desfasaje	Con diferenciación	Sin diferenciación
t-8	1.384e+01	-11.15574

Las correlaciones significativas que perduran luego del procedimiento de filtrado de las series se observaron en los tiempos t, t-7 y t-8. Dos de los tres coeficientes se vieron influenciados por una relación espuria o aleatoria (t-7 y t-8), demostrando que no existe una relación real entre estas variables en esos *lags* temporales. Sin embargo, el coeficiente restante mostró una relación de vinculación directa estadísticamente significativa entre el AOD y la concentración de PM_{2.5}.

Altura de la capa límite atmosférica (PBLH)

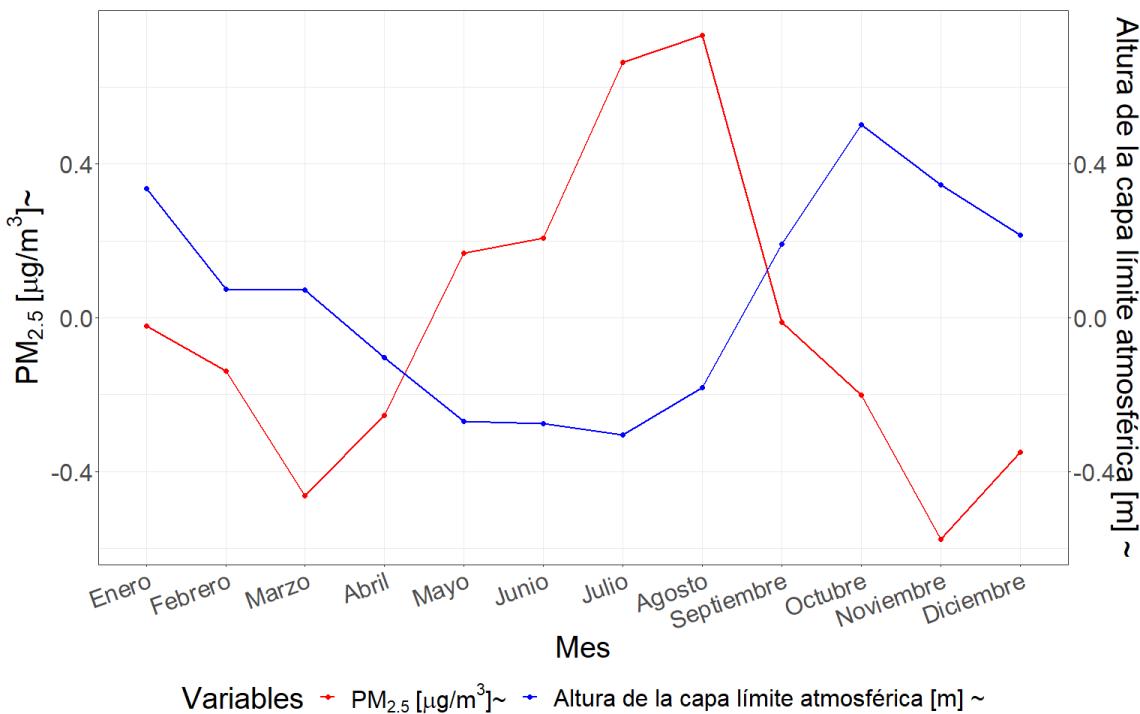


Figura 4.17: Perfil de variación conjunto de los promedios mensuales del PM_{2.5} y el PBLH a lo largo del año (2022-2023), los promedios calculados consideran todos los sitios de muestreo ubicados en la provincia de Córdoba, Argentina.

La Figura 4.17 ilustra el patrón de variación de los promedios mensuales del PBLH y la concentración de PM_{2.5} a lo largo del año. El PBLH muestra sus valores mínimos durante el otoño e invierno, mientras que los máximos durante la primavera y el verano. A medida que el PBLH disminuye, las concentraciones de PM_{2.5} aumentan, y a medida que la altura aumenta, las concentraciones disminuyen.

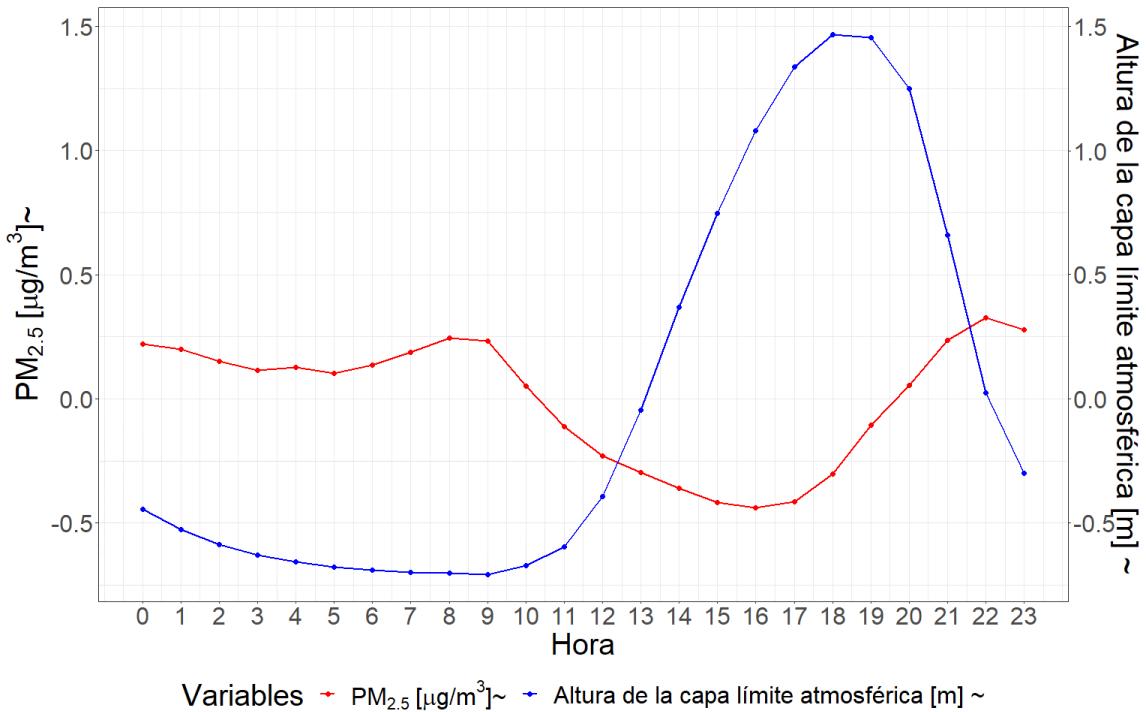


Figura 4.18: Perfil de variación conjunto de los promedios horarios del PM_{2.5} y el PBLH a lo largo de un día promedio en la provincia de Córdoba, Argentina.

A partir de la Figura 4.18, puede apreciarse el patrón de variación del PBLH a lo largo de las horas del día. Puede notarse que el PBLH alcanza su punto máximo al anochecer, alrededor de las 18 horas, a partir del cual, disminuye gradualmente hasta alcanzar su mínimo a las 9 horas de la mañana. Esta variación horaria sugiere que el PBLH está inversamente relacionada con la concentración de PM_{2.5}. Estos resultados confirman la relación inversa entre ambas variables observada en la Figura 4.17 y sugiere que el PBLH puede ser un factor determinante en la concentración de PM_{2.5} a nivel local.

Tabla 4.10: Coeficientes de la función de transferencia dinámica lineal estimados para la covariable analizada en sus promedios horarios a lo largo de todos los sitios de muestreo localizados en la provincia de Córdoba, Argentina, considerando todo el periodo de muestreo analizado que abarca desde el 30/03/2021 hasta el 01/06/2022.

Desfasaje	Con diferenciación	Sin diferenciación
t	-0.0019158	-0.00436
t-21	-0.003367	-0.004821
t-24	-1.928e-03	-0.003588

Las correlaciones significativas que perduran luego del procedimiento de filtrado de las series se observaron en los tiempos t-0, t-21 y t-24, presentando todos una relación de vinculación inversa estadísticamente significativa entre el PBLH y la concentración de PM_{2.5}.

Tráfico (Google Traffic)

Es relevante señalar que el análisis del tráfico a escala mensual no resulta significativo debido a que los datos disponibles representan promedios horarios a lo largo del año, manteniéndose constantes en los distintos meses. Esto quiere decir que los datos a lo largo de los distintos meses no variaban, si no que solo variaban a lo largo de las horas del día, es decir se tenía como dato el promedio del tráfico para las distintas horas del día pero no para los distintos meses, por este motivo fue que se excluyó el análisis a lo largo de los meses. Cabe destacar que los datos analizados se corresponden a promedios horarios a lo largo de todos los sitios de muestreo localizados en la provincia de Córdoba, Argentina, considerando todo el periodo de muestreo analizado que abarca desde el 30/03/2021 hasta el 01/06/2022. Por esta razón, se optó por no incluir el análisis exploratorio en esta escala temporal, con el fin de evitar posibles conclusiones erróneas que pudieran surgir de una interpretación equivocada

de los datos.

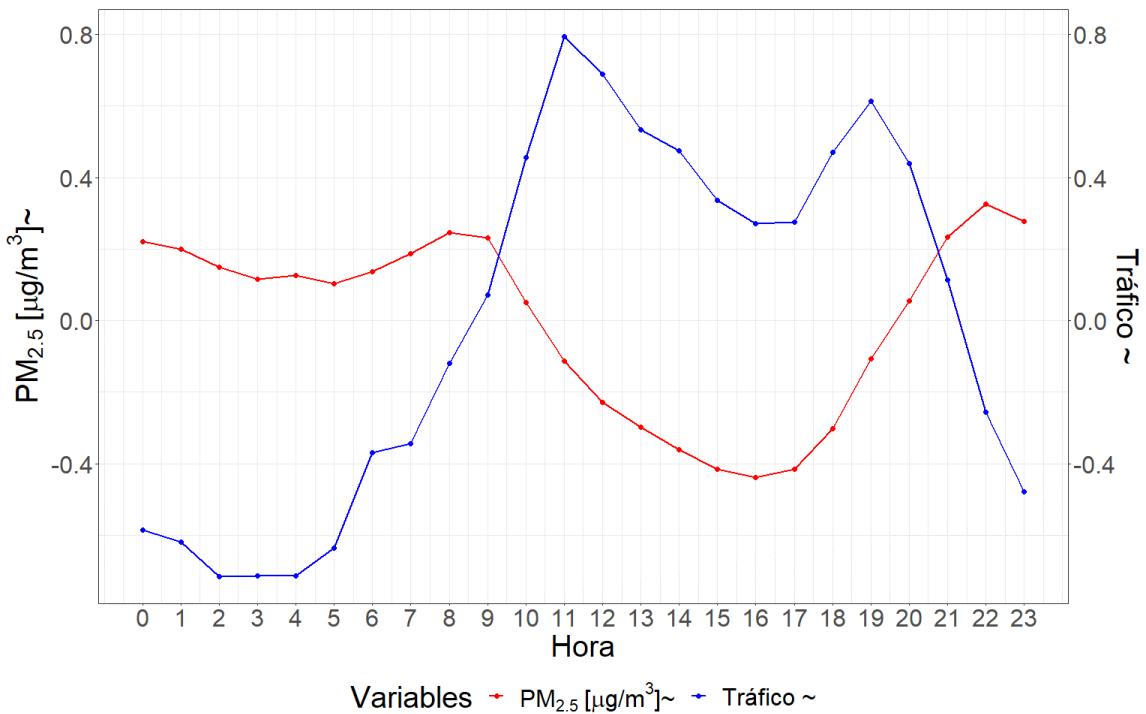


Figura 4.19: Perfil de variación conjunto de los promedios horarios del PM_{2.5} y el tráfico a lo largo de un día promedio, los promedios calculados consideran todos los sitios de muestreo ubicados en la provincia de Córdoba, Argentina.

A partir del patrón de variación del tráfico promedio a lo largo de las horas del día (Figura 4.19), pueden identificarse dos momentos en que la variable alcanza valores máximos, a las 11 y 19 horas. Durante el resto del día el tráfico de automóviles disminuye gradualmente hasta alcanzar sus valores mínimos entre las 0 y las 5 horas. A pesar de no existir una clara asociación entre tráfico y concentración de PM_{2.5}, la relación aparenta ser del tipo directa. Esto se debe a que el pico de concentraciones entre las 8 y las 9 de la mañana coincide con un aumento en el tráfico. Además, entre las 12 y las 18 horas, el tráfico disminuye de manera constante, lo que se relaciona con una reducción en la concentración de PM_{2.5} en la misma franja horaria. Por último, entre las 18 y las 20 horas, se observa un aumento simultáneo en las magnitudes de ambas variables, a pesar de que luego se descoordinen, dado a que el tráfico disminuye

hacia el anochecer y la concentración de PM_{2.5} aumenta.

Tabla 4.11: Coeficientes de la función de transferencia dinámica lineal estimados para la covariable analizada en sus promedios horarios a lo largo de todos los sitios de muestreo localizados en la provincia de Córdoba, Argentina, considerando todo el periodo de muestreo analizado que abarca desde el 30/03/2021 hasta el 01/06/2022.

Desfasaje	Con diferenciación	Sin diferenciación
t-8	1.085e-01	2.442

La única correlación significativa que perdura luego del procedimiento de filtrado de las series se observa en el tiempo t-8, que mostró una relación directa estadísticamente significativa entre el tráfico y la concentración de PM_{2.5}.

Temperatura de la superficie terrestre (LST).

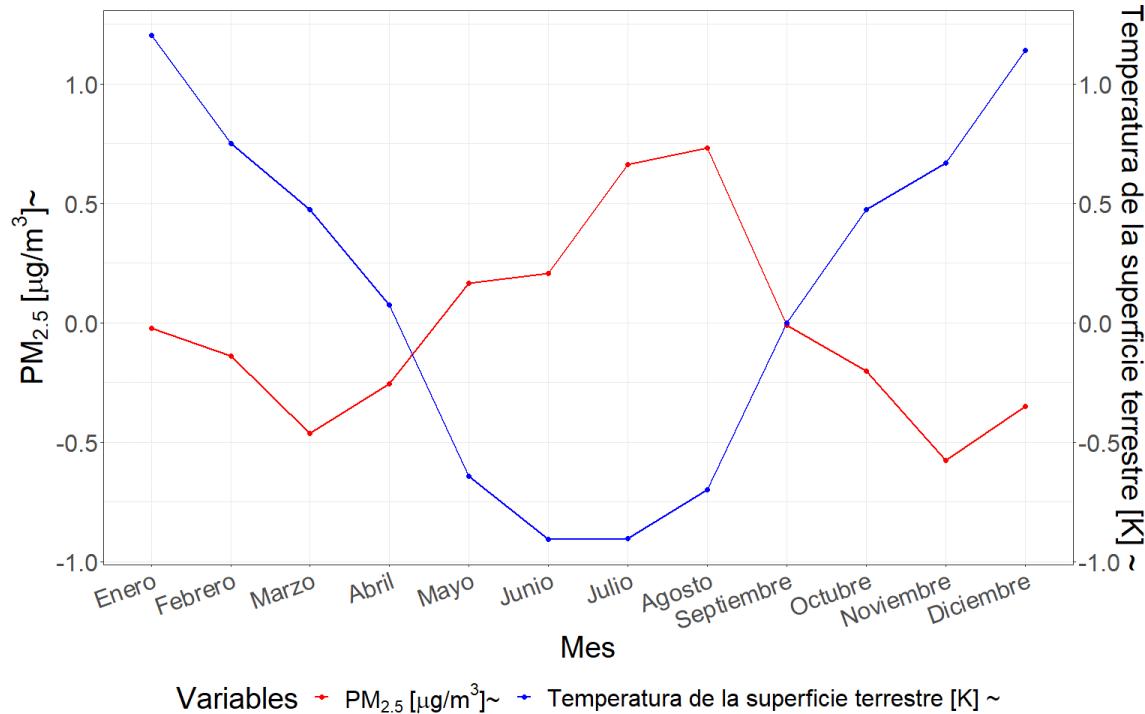


Figura 4.20: Perfil de variación conjunto de los promedios mensuales del PM_{2.5} y la LST a lo largo del año (2022-2023), los promedios calculados consideran todos los sitios de muestreo ubicados en la provincia de Córdoba, Argentina.

La Figura 4.20 presenta el patrón de variación conjunto que describen los promedios mensuales de la LST y la concentración de PM_{2.5} a lo largo del año. Puede observarse que bajas temperaturas de la superficie terrestre favorecen elevadas concentraciones de PM_{2.5}, mientras que altas temperaturas se asocian con una reducción en las concentraciones, describiendo una relación inversa.

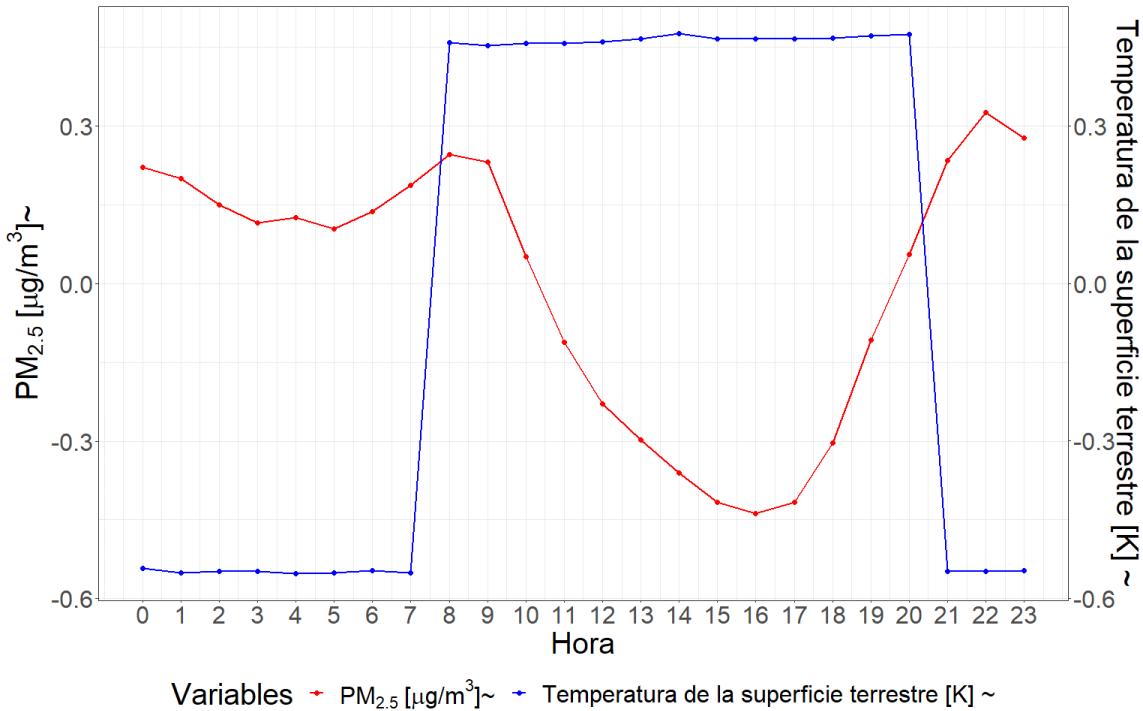


Figura 4.21: Perfil de variación conjunto de los promedios horarios del PM_{2.5} y la LST a lo largo de un día promedio en la provincia de Córdoba, Argentina.

La Figura 4.21 muestra el patrón de variación de la LST promedio a lo largo de las horas del día. Como puede apreciarse, las magnitudes de la temperatura aumentan durante las horas de sol y disminuyen al anochecer, por lo que puede inferirse que existe una relación inversa entre la LST y la concentración de PM_{2.5} horaria.

Tabla 4.12: Coeficientes de la función de transferencia dinámica lineal estimados para la covariable analizada en sus promedios horarios a lo largo de todos los sitios de muestreo localizados en la provincia de Córdoba, Argentina, considerando todo el periodo de muestreo analizado que abarca desde el 30/03/2021 hasta el 01/06/2022.

Desfasaje	Con diferenciación	Sin diferenciación
t-5	-0.0139078	-0.3025

La única correlación significativa que perdura luego del procedimiento de filtrado de las series se observa en el tiempo t-5, que demostró una relación inversa estadísticamente

significativa entre LST y la concentración de PM_{2.5}.

4.4. Discusión

Esta discusión contempla los resultados revelados por el presente capítulo donde los datos analizados refieren a la combinación de los datos recopilados durante el muestreo de PM_{2.5} (ver capítulo 2 para más detalles) con distintas bases de datos obtenidas a partir de información satelitales (Tabla 4.1). El período analizado se corresponde con el muestreo previamente realizado en los distintos sitios de muestreo en la provincia de Córdoba (30/03/2021 - 01/06/2022). Cabe destacar que en el análisis del presente capítulo se calcularon promedios a lo largo de todos los sitios de muestreo en las distintas escalas temporales consideradas de análisis, tanto para la concentración de PM_{2.5} como para las variables derivadas de información satelital.

Los resultados obtenidos evidencian la existencia de relaciones entre la concentración de PM_{2.5} y las variables meteorológicas estudiadas. Los desfases temporales significativos identificados en este trabajo amplían la comprensión de la relación existente entre las variables meteorológicas estudiadas y el PM. Por otro lado, es importante destacar que el análisis de la correlación entre PM_{2.5} con respecto a las variables meteorológicas en diferentes *lags* no ha sido reportado aun en la literatura, según nuestro conocimiento.

La asociación inversa entre la temperatura y la concentración de PM_{2.5} indica que a temperaturas elevadas hay mayor dispersión del PM_{2.5} en la atmósfera, mientras que las bajas temperaturas características del invierno podrían favorecer la acumulación del contaminante. Este hecho podría encontrarse vinculado a las frecuentes inversiones térmicas durante la época invernal que impiden la dispersión de contaminantes. Además, el aumento del uso de combustibles fósiles para calefacción en invierno también podría contribuir al aumento de la concentración de PM_{2.5}. Esta misma asociación inversa

también fue reportada previamente por otras investigaciones, Xiaoyang et al. (2019) obtuvo un resultado similar para un estudio que involucró múltiples ciudades chinas, mientras que Nogarotto et al. (2020) obtuvo un resultado similar al considerar datos recopilados a lo largo de todo el mundo. Los resultados obtenidos sugieren la existencia de mecanismos físicos y químicos que influyen en el transporte y la dispersión de partículas en la atmósfera. Durante los días más cálidos, el aire cercano a la superficie se calienta, volviéndose menos denso que el aire circundante, generando un gradiente de presión vertical que impulsa el aire caliente hacia arriba, dando lugar a corrientes ascendentes. Estas corrientes podrían arrastrar consigo las partículas finas presentes en la atmósfera, reduciendo su concentración a nivel del suelo. Asimismo, la temperatura también influye en la cantidad y la composición de precursores químicos, como los óxidos de nitrógeno y los compuestos orgánicos volátiles, los cuales pueden reaccionar para formar partículas finas. La temperatura también afecta la estabilidad de la capa límite atmosférica, la cual en función de su altura puede modificar la concentración de contaminantes en las capas bajas (Deng et al., 2023). La correlación significativa entre la temperatura y la concentración de PM_{2.5} simultáneas demuestra el efecto instantáneo que tienen las fluctuaciones térmicas en los niveles de partículas. Dado al patrón cíclico de la temperatura se interpreta de igual forma la correlación inversa significativa entre una medición y la observada a la misma hora el día anterior (*lag t-24*), la cual informa un efecto inmediato entre ambas variables. Por otra parte, el desfasaje temporal de 12 horas podría estar sugiriendo el tiempo necesario para que las reacciones químicas impulsadas por variaciones de la temperatura resulten en aportes a la concentración de PM_{2.5}. Estas reacciones fotoquímicas, que involucran compuestos como NOx y VOCs, pueden tardar varias horas en completarse dada la complejidad de los procesos y las condiciones atmosféricas, como la intensidad de la radiación solar y la temperatura. Estudios han mostrado que la formación de PM_{2.5} secundario puede continuar hasta 12 horas después de las emisiones iniciales de los precursores (Cao et

al., 2024; Hodan & Barnard, 2004). Además, la variación diurna en la concentración de PM_{2.5}, que a menudo presenta picos en la tarde y noche, sugiere que las reacciones químicas y la estabilidad de la capa límite atmosférica influyen significativamente en la producción y dispersión de estos contaminantes (Cao et al., 2024).

La humedad relativa desempeña un papel fundamental en la formación y dispersión de partículas, y sus variaciones pueden tener impacto significativo en la concentración de PM_{2.5} (Sun et al., 2020). La relación de vinculación entre la concentración de PM_{2.5} y la humedad relativa demostró ser del tipo inversa, a excepción de meses como por ejemplo el de octubre donde esta forma de relación se invierte, lo que podría estar indicando que existen factores adicionales que influyen la concentración de PM_{2.5} en la atmósfera; Entre ellos podemos mencionar la ocurrencia de eventos meteorológicos extremos y la posible interacción con otros contaminantes atmosféricos. La asociación horaria entre ambas variables podría ser interpretada como un relación directa a pesar de que esto no sea consistente con la bibliografía (Nogarotto & Pozza, 2020 ; Tao Chen et al., 2016; Jianhua He et al., 2019; Q. Liu et al., 2020). A pesar de no coincidir con la relación de vinculación reportada por la bibliografía citada, cabe destacar que ninguno de los estudios mencionados estudio la variación de la concentración de PM_{2.5} en términos de promedios horarios, los estudios previamente citados en su granularidad mínima consideran promedios mensuales. En el presente estudio se observó una relación inversa estadísticamente significativa entre la concentración de PM_{2.5} y la humedad relativa al considerar un desfasaje temporal de 5 horas, por otra parte, el desfasaje temporal de 24 horas significativo reportó una asociación directa entre las variables. La relación inversa podría explicarse mediante el efecto de lavado causado por el proceso de coalescencia o coagulación entre partículas finas y las gotas de agua presentes en la atmósfera, lo que resulta en una disminución de los niveles de partículas en el aire. Por otra parte, al reducir la granularidad del análisis y considerar promedios horarios puede que la relación varíe con respecto a mayores granularidades. De esta forma

la relación directa para la granularidad horaria podría estar asociada a los ciclos de fluctuación conjunta que describen ambas variables en esta escala temporal horaria. Al variar la escala de temporalidad de análisis, los promedios observados difieren para ambas variables analizadas, generando que la relación observada entre ellas cambie con respecto a la escala mensual. Además, podría pensarse que durante aumentos de humedad relativa, los sensores de bajo costo podrían estar detectando micro gotas de agua como un aumento de la concentración de PM_{2.5}, dicha afirmación no es correcta ya que la influencia de la humedad relativa en las concentraciones registradas para los sensores implementados se da para valores mayores al 80 % de esta variable, dato que nunca fue registrado en el muestreo realizado (Badura et al., 2018). Es relevante mencionar que el ciclo promedio a lo largo de las horas del día descripto por ambas variables en este trabajo es similar al reportado por otros estudios de investigación (Anjos & Lopes, 2019; Dhaka et al., 2022; Gama et al., 2018; Grossi et al., 2016; Jena et al., 2021; McFarlane et al., 2021; Omar & Syed-Fadzil, 2011; Saaroni et al., 2015). Además, dadas las características cíclicas de ambas series a lo largo del periodo de muestreo analizado, el desfasaje temporal de 24 horas refleja la relación existente entre las variables en un mismo momento de tiempo o bien sin considerar desfasaje. Esto es así como consecuencia de que al cabo de 24 horas el ciclo de promedios horarios a lo largo del día se vuelve a iniciar, y como consecuencia de las caracterizadas cíclicas las promedios calculados son similares. De esta forma, el vínculo directo encontrado, podría deberse solo a la variabilidad conjunta descripta por ambas variables en la escala horaria a lo largo del día. Se entiende por covarianza al grado de variación conjunta de dos variables aleatorias respecto a sus medias.

Así como en el estudio realizado por Mishra et al., (2015), en este trabajo también se ha encontrado una relación inversa entre la temperatura del punto de rocío y la concentración de PM_{2.5}. Este fenómeno podría atribuirse al hecho de que bajas temperaturas de rocío indican una mayor probabilidad de condensación del vapor agua

contenido en el aire, resultando en un incremento de micro gotas de agua en forma de rocío o neblina, que resulta en una condensación y precipitación de partículas. Las correlaciones negativas halladas con *lags* de 11 y 14 horas anteriores a la lectura del PM_{2.5} sugieren que el proceso físico de influencia de la temperatura del punto de rocío necesita aproximadamente esa cantidad de horas para resultar en efectos significativos sobre la concentración de PM_{2.5}.

La relación inversa encontrada entre la precipitación y la concentración de PM_{2.5} también ha sido previamente documentada en otros estudios (Usman et al., 2022; XiaoYang et al., 2019). En este trabajo se observó que los días en los que se produce precipitación hay una reducción del 40 % en la concentración de PM_{2.5}, en comparación con los días en que esta no sucede. Esta disminución podría atribuirse al efecto de lavado causado por las gotas de lluvia, las cuales capturan las partículas presentes en el aire, reduciendo así su concentración en la atmósfera. Además, las precipitaciones limpian la superficie terrestre; de esta forma se evita la resuspención de partículas del suelo y se reduce la concentración de partículas en el aire. La correlación estadísticamente significativa se observó teniendo en cuenta un *lag* de una hora entre la precipitación y la concentración de PM_{2.5}, y fue negativa, reflejando el efecto de limpieza prácticamente instantánea de la atmósfera al iniciarse las lluvias.

La relación inversa entre la velocidad del viento y la concentración de PM_{2.5}, también ha sido documentada en otros trabajos de investigación (Cichowicz et al., 2020; Usman et al., 2022; XiaoYang et al., 2019). Esta relación podría explicarse a través del proceso de dispersión de partículas debido al movimiento de las masas de aire. Cuando la velocidad del viento es alta, las corrientes de aire arrastran y dispersan las partículas en la atmósfera, conduciendo a una reducción de la concentración de PM_{2.5}. Por el contrario, cuando la velocidad del viento es baja, no existe tal dilución, resultando en mayores concentraciones de PM_{2.5}. Cabe destacar que, dadas las características cíclicas a lo largo de las horas del día, el desfase temporal de 24

horas a su vez refleja la asociación entre estas variables sin considerar desfasaje o *lag*.

A pesar de los resultados contradictorios hallados para la relación entre la presión atmosférica y la concentración de PM_{2.5} que indicaron un vínculo inverso estadísticamente significativo al no considerar *lag*, mientras que uno directo al considerar un *lag* de 19 horas. Es importante destacar que múltiples estudios científicos resaltan una correlación positiva entre las variables, sustentada en que altas presiones atmosféricas suelen estar asociadas con una baja altura de la capa límite atmosférica, lo que dificulta la dispersión de contaminantes (Gao et al., 2022; Q. Yang et al., 2017).

La relación inversa entre la radiación solar directa y la concentración de PM_{2.5} se basa en estudios previos, como el realizado por Liu et al. en 2015. Se postula que el aumento en la radiación solar directa puede elevar la temperatura de la superficie terrestre, lo que a su vez incrementaría la temperatura del aire cercano a la superficie. Este calentamiento induciría corrientes ascendentes impulsadas por variaciones en la densidad del aire, lo que resultaría en la dispersión de las partículas finas presentes en la atmósfera y, por ende, en la reducción de su concentración. Durante períodos de baja radiación solar, este proceso puede no ocurrir, lo que podría contribuir a concentraciones más altas de PM_{2.5} debido a la falta de fenómenos de dispersión vertical. Además, la relación inversa encontrada entre la radiación solar y el PM_{2.5} se ve respaldada por datos que muestran una reducción prácticamente instantánea en las magnitudes de radiación solar seguida de un aumento en la concentración de PM_{2.5}. Esto se refleja en el análisis de *lags*, donde se encontró un coeficiente significativo para la radiación solar con un *lag* de 23 horas, lo que sugiere una relación inversa con la concentración de PM_{2.5}.

Por otra parte, la relación entre la temperatura de la superficie terrestre (LST) y la concentración de PM_{2.5} también es significativa. El aumento de la LST puede desencadenar corrientes ascendentes de aire, que requieren un tiempo considerable para tener un efecto significativo en la limpieza de la atmósfera. Esto sugiere que

el aumento de la temperatura de la superficie terrestre podría contribuir a la reducción de la concentración de PM_{2.5} a través de procesos de dispersión atmosférica. Además, se observa un coeficiente directo significativo entre la LST y la radiación solar, indicando una ventana temporal específica en la que ocurren los fenómenos de limpieza atmosférica como resultado de la interacción entre estos dos factores. Sin embargo, es importante considerar que la LST puede ser influenciada por una variedad de factores, como fuentes de calor cercanas, lo que podría afectar la interpretación de su relación con la concentración de PM_{2.5}. Esto se evidencia en el análisis de *lags*, donde se encontró un coeficiente significativo para la LST con un *lag* de 5 horas, lo que sugiere una relación inversa con la concentración de PM_{2.5}.

La relación entre el AOD y la concentración de PM_{2.5} generalmente es positiva en la literatura científica. Esto implica que un aumento en el AOD, está correlacionado con un aumento en la concentración de PM_{2.5} (A. Gupta et al., 2021; Q. He et al., 2021; Soni et al., 2018). Este vínculo se basa en el principio de que una mayor cantidad de partículas finas en el aire resulta en una mayor dispersión y absorción de la radiación solar, reflejándose como un mayor AOD. No obstante, algunos estudios, como el de Della Ceca et al., 2018, han encontrado tendencias inversas en ciertas regiones rurales, lo que sugiere que esta relación puede variar dependiendo del contexto específico, como las características locales de las fuentes de emisión y las condiciones meteorológicas prevalentes. El análisis gráfico realizado en el presente capítulo para los promedios mensuales podría llevarnos a interpretar una relación entre el AOD y la concentración de PM_{2.5} del tipo inversa. Además, Della Ceca et al., 2018, quien describió la variación espacio temporal del AOD en escala mensual analizando una base de datos que contempló una década. La abundancia de datos le permitió observar una tendencia positiva de diciembre a marzo en las zonas urbanas. En este estudio alcanza a observarse elevados valores en diciembre alcanzando el pico máximo que termina decayendo notablemente en enero, lo cual podría encontrarse vinculado a

una tendencia atípica en el año de análisis. Por otra parte, en la investigación llevada a cabo por Della Ceca et al., 2018 se observó una tendencia negativa en las zonas rurales, resultados que se asemejan más a lo revelado en los gráficos exploratorios de la presente tesis, lo que podría revelar una predominancia de los sitios rurales por sobre los urbanos involucrados en esta tesis. Cabe destacar que es importante examinar en escalas temporales de menor granularidad como la horaria, que permite detectar la relación directa previamente documentado por otras investigaciones. Los resultados obtenidos en este estudio indican una relación significativa directa entre el AOD y la concentración de PM_{2.5}, sugiriendo que los aerosoles presentes en la atmósfera se encuentran directamente relacionados con la concentración de PM_{2.5}. Esta relación podría encontrarse sustentada en que al aumentar la cantidad de partículas finas en la atmósfera, una mayor proporción de radiación solar es atenuada por los aerosoles, reflejándose en un aumento del AOD. Por otro lado, valores bajos de AOD se traducen en bajas concentraciones de PM_{2.5}, lo cual sugiere que el AOD es una medida indirecta que permite estimar la concentración de PM_{2.5} en la atmósfera. Estos resultados son consistentes con otras investigaciones actuales que reportan una relación directa entre el AOD y la concentración de PM_{2.5} (A. Gupta et al., 2021; Soni et al., 2018). El coeficiente de correlación significativo que conserva su signo antes y después del procedimiento de diferenciación se caracteriza por no involucrar ningún desfase temporal, sugiriendo que las variaciones del AOD reflejan de forma instantánea las fluctuaciones de la concentración de PM_{2.5}.

Los resultados obtenidos en la presente investigación demuestran la existencia de una relación inversa entre la altura de la capa límite atmosférica (PBLH) y la concentración de PM_{2.5}. Esta relación se debe a que la PBLH define el volumen de la capa de la atmósfera más cercana a la superficie, donde se diluyen los contaminantes emitidos. Investigaciones recientes, como las de Li et al., en 2020 y Wand et al., en 2019, han reportado resultados similares. Cabe destacar, que todos los coeficientes de

correlación que resultaron significativos indican la existencia de una relación inversa. Además, es pertinente resaltar que los coeficientes tanto para un *lag* nulo como para uno de 24 horas, ambos significativos, demuestran que dadas las características cíclicas de las series a lo largo de las horas del día, reflejan una vinculación en el mismo momento temporal. La interpretación nace como consecuencia de que al cabo de 24 horas el ciclo de promedios horarios a lo largo del día se vuelve a iniciar, y como consecuencia de las caracterizadas cíclicas las promedios calculados son similares. De esta forma, sugieren que la influencia de la variación del PBLH en la concentración del PM_{2.5} se manifiesta de manera instantánea.

Los resultados obtenidos sobre la relación entre el tránsito vehicular y la concentración de PM_{2.5} se encuentran en línea con los obtenidos por otros estudios en los que se ha demostrado una relación directa (Pant & Harrison, 2013; Weerakkody et al., 2018). Las emisiones vehiculares son una fuente importante de partículas finas en la atmósfera, debido a la combustión incompleta de los combustibles fósiles. Por lo tanto, se espera que aumentos en el tráfico vehicular conduzcan a un incremento en las emisiones generadas, lo que a su vez resulta en mayor concentración de PM_{2.5}. El coeficiente de correlación que perdura luego del procedimiento de diferenciación resultando significativo tiene lugar con un desfase temporal de 8 horas. Este desfase podría deberse al tiempo necesario desde la emisión hasta su registro por parte de los sistemas de monitoreo, los cuales se hallan dispuestos en edificaciones y no a la salida inmediata del caño de escape de los vehículos. Finalmente hay que destacar que la calidad de los datos de tráfico no es óptima, ya que solo disponemos de promedios horarios para un sitio en particular, lo que limita considerablemente la variabilidad conjunta de los datos con respecto a la concentración de PM_{2.5}. Esta limitación ha dificultado la obtención de una correlación fuerte con la concentración de PM_{2.5}. No obstante, los datos proporcionados por *Google*, aunque fiables, no poseen la resolución temporal deseada para este estudio. Sin embargo, estos datos han sido ampliamente utilizados

en otros estudios que abordan diversos problemas, lo que respalda su utilización en este trabajo (Ganji et al., 2020; Mostafi & Elgazzar, 2021).

Es importante destacar que el impacto de las condiciones meteorológicas en la concentración de PM_{2.5} es un proceso intrínsecamente ligado a la emisión de contaminantes atmosféricos. Este fenómeno, complejo y multifacético, involucra una serie de interacciones dinámicas entre los factores meteorológicos y las emisiones contaminantes, que influyen significativamente en la concentración de partículas en suspensión en la atmósfera. Sin fuentes de emisión no habría contaminantes, en la práctica, estas fuentes son omnipresentes en áreas urbanas y peri urbanas. Por tanto, es crucial entender cómo las condiciones meteorológicas modulan la concentración y distribución de PM_{2.5}, previamente tomando en consideración lo mencionado respecto de las fuentes de emisión. Las fuentes de PM_{2.5} incluyen tanto emisiones primarias como las derivadas de procesos de combustión (vehículos, industria, calefacción) y emisiones secundarias resultantes de reacciones químicas en la atmósfera (N. Jiang et al., 2018). Además, es importante considerar las variaciones espaciales de las fuentes de PM_{2.5}. Las zonas industriales y las áreas cercanas a carreteras principales tienden a mostrar concentraciones más altas debido a la proximidad a fuentes de emisión intensivas (Guo et al., 2020).

4.5. Conclusiones

Las condiciones meteorológicas tienen un impacto sustancial en la presencia y concentración de PM_{2.5}, pero este impacto no puede ser comprendido sin considerar las fuentes de emisión. Los resultados del estudio confirman que las variables derivadas a partir de información satelital y meteorológica guardan una estrecha relación con la variación temporal de la concentración de PM_{2.5} en el Área Metropolitana de Córdoba. Estas variables poseen potencialidad para explicar la variabilidad espacio temporal de

la concentración de PM_{2.5} en el área de estudio.

Es importante destacar que, aunque en este estudio se ha analizado la covariación entre las variables derivadas de información satelital de forma individual con respecto a la concentración de PM_{2.5}, la variabilidad conjunta de estas variables es un hecho y refleja un sistema dinámico e interdependiente. Las variables meteorológicas no actúan de manera aislada, sino que forman un sistema integrado cuya dinámica conjunta determina los patrones de dispersión de contaminantes atmosféricos, regulando así la concentración de PM_{2.5}. El comportamiento integral de este sistema es esencial para comprender las fluctuaciones en la concentración de PM_{2.5} en un momento dado.

La complejidad de estas interacciones justificó un análisis exploratorio que vinculó cada variable de manera aislada, análisis que justificó su inclusión en una base de datos para el desarrollo de modelos predictivos de la concentración de PM_{2.5}. El modelo es el encargado de interpretar las relaciones subyacentes entre las diversas variables para obtener predicciones, las cuales serán precisas dada a la influencia de cada una de estas variables sobre la concentración de PM_{2.5}.

Capítulo 5

Modelado predictivo de la concentración de PM_{2.5}

5.1. Introducción

Preservar la calidad del aire no solo reviste una importancia fundamental para salvaguardar la salud pública, sino también para asegurar el equilibrio ecosistémico y climático. Sin embargo, esta calidad se ve constantemente amenazada por el avance del desarrollo urbano, la creciente industrialización, la alteración activa de la superficie terrestre y los efectos del cambio climático. En este contexto, el monitoreo de la calidad del aire emerge como una valiosa práctica capaz de brindar información crucial sobre las emisiones atmosféricas. Esto, a su vez, posibilita la evaluación de los niveles de contaminación y la formulación de estrategias efectivas de mitigación. En este marco, el desarrollo de modelos predictivos se ha consolidado como una herramienta fundamental dentro de los sistemas de gestión de calidad del aire, permitiendo obtener predicciones con el propósito de alertar a la población ante condiciones de contaminación que podrían ser adversas para la salud.

Los modelos predictivos constituyen herramientas estadísticas diseñadas para

recrear con la máxima fidelidad posible los procesos que se observan en la naturaleza. En el contexto de los modelos que abordan las emisiones atmosféricas, su utilidad no solo se limita a aplicaciones científicas, sino que también desempeñan un papel crucial en la gestión de la calidad del aire. Su principal aporte radica en la capacidad de controlar y prever, tanto en tiempo real como en el futuro, la contaminación atmosférica en zonas urbanas o regionales específicas. La generación de un modelo predictivo eficiente es una tarea compleja que exige la integración sistemática de un gran volumen de información generada a partir de diferentes fuentes de datos.

El avance de la tecnología ha posibilitado la recopilación de información relativa a diversos fenómenos atmosféricos relacionados a la variabilidad de la contaminación del aire a lo largo del tiempo. Una análisis común de los datos en forma de series temporales radica en el análisis de su estructura temporal para desentrañar sus patrones de variabilidad. Este análisis promueve la extracción de información representativa en términos de sus orígenes y relaciones subyacentes, además de evaluar la posibilidad de extrapolar y anticipar comportamientos de la serie de datos en el futuro. Por otra parte, la correcta selección de la metodología de predicción depende intrínsecamente del fenómeno específico o situación que se pretenda predecir.

Son numerosas y variadas las aplicaciones de series temporales en monitores ambientales. En esta tesis, dado al creciente número de estudios científicos que establecen vínculos entre la contaminación del aire con enfermedades respiratorias (Hahad et al., 2024), cardiovasculares, cáncer (N. Chen et al., 2022), trastornos del sistema nervioso (Costa et al., 2022), así como enfermedades transmitidas por el aire e inducidas por el calor, la aplicación se centra en la predicción de la concentración de la calidad del aire, específicamente de la concentración de PM_{2.5} (Sacks et al., 2011; Yan Zhao et al., 2020). A medida que se profundiza en el estudio de este fenómeno, se evidencia un vínculo cada vez más estrecho entre la calidad del aire y el sistema meteorológico-climático (Elminir, 2005; Pearce et al., 2011; Sfetsos & Vlachogiannis, 2010). De allí,

los numerosos intentos de predecir la distribución de partículas, a partir de información satelital meteorológica.

El desarrollo tecnológico ha permitido un notable progreso en la capacidad de cómputo disponible. Esta evolución, sumada a la disponibilidad de datos meteorológicos masivos en diversas plataformas de Internet, posibilitan el entrenamiento de modelos predictivos de calidad de aire en un sitio específico en el contexto de la inteligencia artificial. Actualmente, se cuenta con variadas herramientas informáticas para la obtención de productos satelitales almacenados en la nube. Tales productos pueden ser empleados como covariables con potencialidad predictiva de la concentración de PM_{2.5}. Sin embargo, aun cuando existen modelos predictivos mecanicistas basados en la estimación de efectos para variables reportadas en la literatura, el desarrollo de modelos de base empírica sustentados en la abundancia de información hoy disponible es más incipiente (Al-Janabi et al., 2020; Gu et al., 2021; Kang et al., 2018; Muthukumar et al., 2022). Un paso importante en la creación de un sistema de alerta frente a condiciones adversas de calidad de aire de bajo costo recae en el desarrollo y la selección del algoritmo de mayor desempeño predictivo para la concentración de PM_{2.5}. El objetivo de este capítulo es comparar distintos enfoques de modelado predictivo basados en el aprendizaje automático y aprendizaje profundo para predecir la concentración de PM_{2.5} horarias. Los procesos de modelado se basan en el uso de variables predictoras meteorológicas, ambientales y temporales, obtenidas principalmente de satélites, para cada uno de los usos de suelo evaluados. La finalidad que persigue este capítulo es completar el desarrollo de un sistema de alerta basado en tecnologías de bajo costo que detecte condiciones adversas de calidad de aire.

5.2. Materiales y métodos

5.2.1. Resumen gráfico del protocolo analítico de modelado predictivo

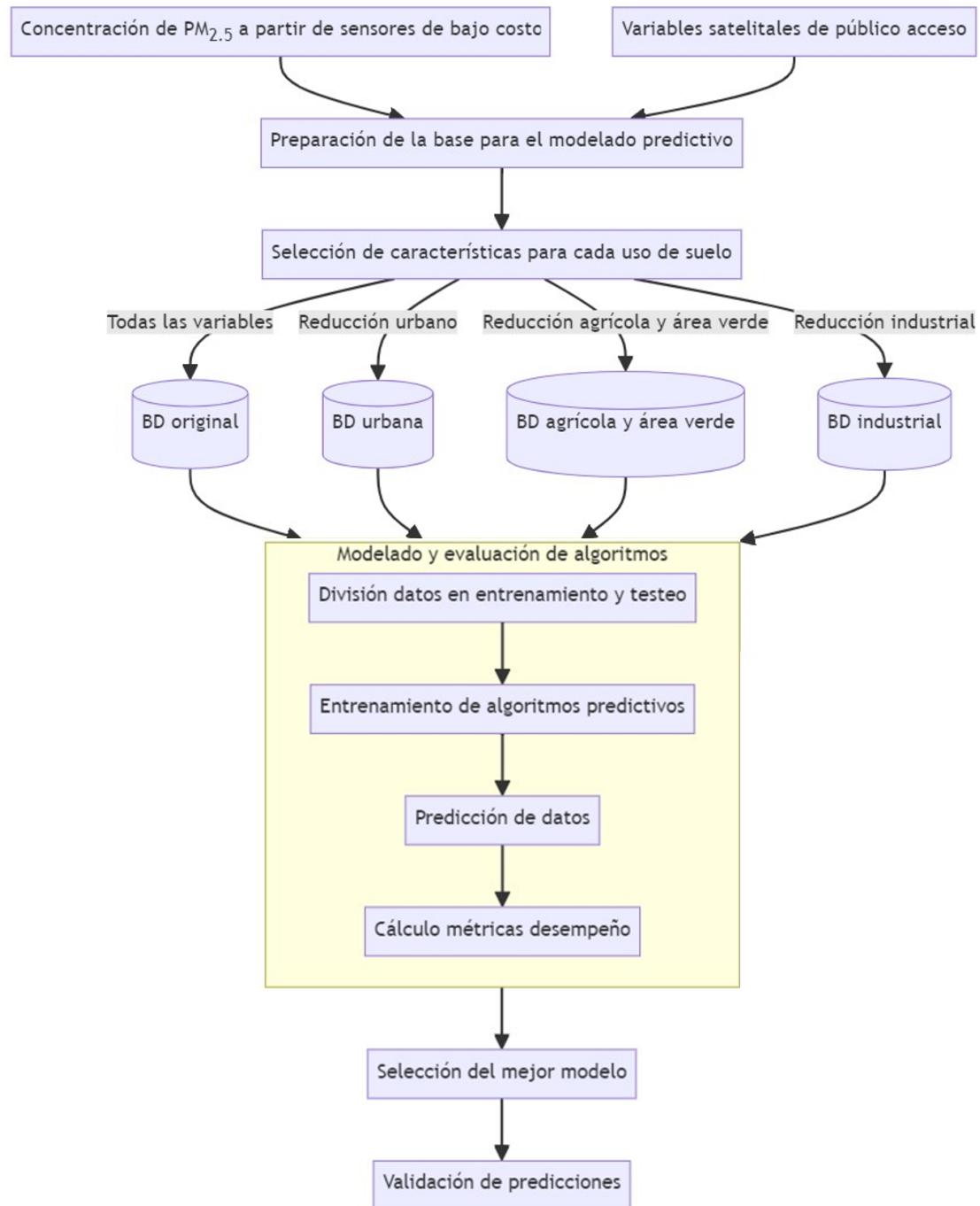


Figura 5.1: Resumen gráfico del protocolo analítico para selección del mejor modelo predictivo.

La Figura 5.1 ilustra el protocolo analítico desarrollado en este capítulo, cuyo objetivo central es la construcción de un modelo predictivo para la concentración de PM_{2.5} en el Área Metropolitana de Córdoba, Argentina. Este modelo se destaca por prever la concentración de PM_{2.5} horaria, con cinco días de anticipación, motivo por el cual la variable objetivo con la cual se entrena el modelo es la concentración de PM_{2.5} horaria recopilada por las estaciones de monitoreo, y emplea como entrada variables meteorológicas satelitales y temporales relacionadas a la variabilidad periódica de este contaminante. Se compararon modelos alternativos basados en aprendizaje automático y profundo, llevando a cabo el ajuste de cada uno de ellos, evaluación y selección del modelo de mayor capacidad predictiva para distintos usos de suelo. Es importante destacar que la presente sección se haya enmarcada dentro de un enfoque computacional de inteligencia artificial.

En primer lugar, se llevó a cabo una comparación exhaustiva de distintos algoritmos de aprendizaje automático: regresión lineal múltiple (Bottou & Bousquet, 2007; Saunders et al., 1998; Tibshirani, 1996; Zou & Hastie, 2005), máquinas de vectores de soporte (Cortes & Vapnik, 1995), k vecinos más cercanos (Fix, 1985), bosques aleatorios (Breiman, 2001a) y máquina de aumento de gradiente (Friedman, 2001); y profundo (redes neuronales recurrentes (Bishop et al., 1995), redes de memoria a largo plazo (Hochreiter & Schmidhuber, 1997) y redes recurrentes con compuerta (Cho et al., 2014), con el propósito de identificar cuál de ellos demostraba el mejor rendimiento predictivo. La metodología de modelado abarcó el desarrollo de un modelo predictivo que considera un desfasaje temporal de 5 días, es decir 120 horas, para generar una predicción de la concentración de PM_{2.5} horaria con este desfasaje de antelación. Esto quiere decir a partir del modelo predictivo, ingresando las variables de entrada con las que haya sido entrenado el mismo, se logran obtener predicciones de la concentración de PM_{2.5} horaria 120 horas en el futuro o bien dentro de 5 días. Además, se aplicó una estrategia de modelado que discriminó según el uso del suelo, permitiendo ajustar

distintos modelos predictivos para cada uno de los usos considerados, de esta forma se permitió la obtención de predicciones discriminadas para cada uno de los usos de suelo involucrados en el presente estudio. Los datos fueron recopilados en cada uno de los sitios de muestreo descritos en el capítulo 3 con una frecuencia de variación horaria, tanto la concentración de PM_{2.5}, como las variables regresoras que en su mayoría son variables derivadas de información satelital. Posteriormente, estos resultados fueron promediados entre los sitios correspondientes a un determinado uso de suelo, es decir que de las variables regresoras, así como la respuesta, se obtuvo la media a lo largo de los sitios de muestreo involucrados en cada uno de los usos de suelo. De esta forma se redujo la información a promedios horarios para cada uno de los usos de suelo, la cual permitió desarrollar los modelos predictivos para cada uno de ellos. Además se evaluaron distintas alternativas de bases de datos como soporte para el modelo predictivo, una incluyó todas las variables predictoras y otra involucró una selección de variables para trabajar con una base de datos reducida. Esta última se creó mediante técnicas de reducción de dimensionalidad y selección de características (featurewiz (Bakheet et al., 2023), Boruta (Kursa et al., 2010) y Boruta alternativo (Jamei et al., 2023)). Las variables incluidas en cada una de las bases de datos se encuentran mencionadas en el Anexo 1 para cada uso del suelo. Los algoritmos de aprendizaje evaluados fueron optimizados por medio de la búsqueda de hiperparámetros e interpretados a través de la implementación del método de Shapley (valores SHAP) (Lubo-Robles et al., 2020). La elección del horizonte de predicción se encuentra fundada en que la disponibilidad de las variables derivadas a partir de información satelital es gratuita solo cinco días previos al momento presente, de esta forma para poder obtener la predicción de la hora siguiente, es necesario emplear como predictoras los datos de la hora anterior con cinco días de antelación. Por este motivo, el entrenamiento de los algoritmos predictivos se realizó con un desfasaje de la variable respuesta respecto a las predictoras de 120 horas, es decir, cinco días, dadas las características de las bases

de datos satelitales. La evaluación de los modelos se realizó mediante la comparación de la métrica RMSE obtenida a partir de una validación cruzada *K-fold* con un K igual a 10. En el contexto de esta tesis de ingeniería, se ha decidido utilizar la raíz del error cuadrático medio (RMSE) como la métrica principal para la evaluación del modelo predictivo. A pesar de que se recomienda el uso de múltiples métricas para una evaluación más completa, la elección de RMSE se fundamenta en varias razones sólidas y respaldadas por la literatura científica. La RMSE es ampliamente reconocida y utilizada en la evaluación de modelos de regresión debido a su capacidad para penalizar los errores de predicción más grandes, proporcionando una medida clara y directa de la precisión del modelo. Esta característica es crucial cuando se necesita una métrica que resalte las predicciones extremadamente inexactas, lo cual es fundamental en aplicaciones donde los grandes errores pueden tener consecuencias significativas (Hodson, 2022). Además, la RMSE está en las mismas unidades que la variable de interés, lo que facilita su interpretación y comparación de los resultados. Esta propiedad es particularmente relevante en estudios de ingeniería, donde es esencial comprender el desempeño del modelo en términos prácticos y aplicables (Hodson, 2022). El uso exclusivo de la RMSE en este estudio se alinea con prácticas comunes en la literatura científica, donde se utiliza como la métrica principal debido a su capacidad para proporcionar una visión clara y concisa del desempeño del modelo. Además, estudios recientes han destacado que la RMSE, al ser sensible a los grandes errores, puede ser más informativa que otras métricas en ciertos contextos de modelado predictivo (Hodson, 2022).

5.2.2. Base de datos

En esta sección, se utilizó la misma base de datos que fue implementada en el capítulo 4, generada a partir de la combinación de los datos de concentración de PM_{2.5} recopilados durante el muestreo con sensores de bajo costo, analizados en el capítulo 3,

y los datos satelitales y meteorológicos analizados en el capítulo 4, descargados desde la nube.

Además de las variables mencionadas, se incorporaron variables temporales como la estación del año, el mes, el día del mes, el día de la semana y la hora de la lectura del contaminante. Éstas se incluyeron tanto en su forma categórica original como luego de ser codificadas de manera cíclica mediante la introducción de las componentes de seno y coseno correspondientes. Es importante destacar que todas las variables numéricas se estandarizaron o normalizaron con el objetivo de homogeneizar sus rangos de variabilidad y ajustar su desviación estándar a 1. El centrado y estandarización fue llevado a cabo a través de la Ecuación 4.1. Este proceso permitió evitar problemas de escala y facilitó la comparación entre las variables.

Para lograr una comprensión integral de la relación de vinculación entre las variables que resultaron relevantes en el proceso de modelado predictivo y la concentración de PM_{2.5} se incluyó el análisis exploratorio realizado durante el capítulo 4 de forma resumida haciendo especial énfasis en la conexión de las variables. Por otra parte, para aquellas variables que no fueron abordadas en el análisis del capítulo 4 dado a que no son habitualmente citadas en sustentos bibliográficos para describir la variación de la concentración de PM_{2.5}, se incluyó un análisis exploratorio equivalente al desarrollado en dicho capítulo. Las variables mencionadas son: la rugosidad de la superficie terrestre, la disipación en la capa límite atmosférica, la integración vertical flujo ozono norte, tasa media de escorrentía subterránea y la integración vertical de temperatura. En primer lugar, a pesar de existir una descripción detallada de cada una de las variables en el Anexo 3, se incluyó en el presente capítulo la definición de cada una de las variables exploradas para cada uno de los usos de suelo, proseguida de su análisis exploratorio. La inclusión de este análisis tiene por objetivo garantizar una interpretación e inferencia adecuada acerca de la manera en que se manifiesta la variabilidad conjunta desenmascarada entre cada una de las variables analizadas y la

concentración de PM_{2.5}, permitiendo generar independencia entre los capítulos de la presente tesis.

Integración de datos

Resolución Espacial Para integrar datos provenientes de diversas fuentes con diferentes resoluciones espaciales, se adoptó una metodología de extracción puntual. Este enfoque se centra en la utilización de datos puntuales, independientemente de la resolución del píxel de origen. De este modo, aunque los datos sean obtenidos de fuentes con distintas resoluciones espaciales, la información se homogeneiza a través de la extracción puntual. Esta estrategia es especialmente útil para evitar las complicaciones derivadas de la heterogeneidad en las resoluciones espaciales de las fuentes de datos. La elección de este método se justifica en que la resolución original de la fuente no impacta significativamente en la precisión del modelo predictivo, siempre que se utilicen las mejores fuentes de datos disponibles, las cuales proporcionan la calidad necesaria para los objetivos del estudio.

Resolución Temporal La integración de datos con diferentes resoluciones temporales se manejó ajustando la frecuencia de los datos a la frecuencia de análisis requerida por el modelo. En casos donde las variables presentaban variabilidad diaria y el análisis requería una resolución horaria, se procedió a replicar el dato diario a lo largo de todas las horas del día. Esta técnica asegura que, aunque la variabilidad horaria no esté presente en los datos originales, la consistencia temporal se mantenga al incorporar el dato diario como constante en cada hora. Esta metodología permite armonizar datos de distintas resoluciones temporales, asegurando que todas las variables sean representadas adecuadamente en la frecuencia de análisis del modelo. La repetición del dato diario a nivel horario garantiza una serie temporal completa y continua, aunque a expensas de la variabilidad intra-diaria.

Cantidad de datos

En este estudio, se recopilaron datos horarios durante un periodo de un año y tres meses, totalizando aproximadamente 70,000 datos horarios teniendo en cuenta todos los sitios de muestreo. Este volumen se redujo al calcular los promedios para los distintos usos de suelo, dando por resultado 10,152 datos horarios para uso urbano, 8,648 para uso industrial y 9,695 para uso agrícola/área verde. Aunque estos volúmenes de datos pueden parecer limitados en comparación con otros estudios, investigaciones recientes han demostrado que es posible desarrollar modelos predictivos efectivos con conjuntos de datos de tamaño similar. Huang et al., (1998) demostraron que mediante técnicas adecuadas de modelado y preprocesamiento, es posible obtener predicciones precisas incluso con conjuntos de datos pequeños, alrededor de 10,000 observaciones (Maior & Silva, 2024). Asimismo, el estudio de Musah Lampo et al., (2024) sobre el uso de herramientas de *AutoML* en series temporales mostró que volúmenes de datos que oscilan entre 8,000 y 12,000 observaciones horarias son suficientes para generar modelos predictivos robustos y precisos. Evaluaron diversas herramientas y modelos en múltiples conjuntos de datos, concluyendo que la cantidad de datos utilizada, cuando se combina con técnicas avanzadas de manejo de datos, puede ser altamente efectiva (Westergaard et al., 2024).

5.2.3. Protocolo estadístico

Modelos predictivos comparados

Regresión lineal: La regresión lineal múltiple (*multiple linear regression*) es una técnica estadística que se utiliza para modelar una variable dependiente como una combinación lineal de múltiples variables independientes. Este metodología es aplicable tanto a datos de sección transversal como a series temporales contemplando la autocorrelación temporal (Aldas Manzano & Uriel Jimenez, 2017). El supuesto subyacente

en este modelo es que existe una relación lineal entre la variable dependiente y las variables independientes. En el contexto del aprendizaje automático, el aprendizaje de la interrelación entre las variables se logra mediante el ajuste iterativo de los coeficientes de regresión. El algoritmo de regresión busca determinar el hiperplano que mejor ajusta a los datos, lo cual implica encontrar coeficientes que minimicen la diferencia entre los valores observados y predichos por el modelo.

En esta investigación, se emplearon diversos algoritmos para ajustar la regresión lineal múltiple. El primero fue el método de mínimos cuadrados ordinarios, el cual deriva en un modelo lineal cuyos coeficientes resultantes minimizan la suma de los residuos cuadrados entre los valores observados y los valores predichos. Este método representa el enfoque clásico para ajustar modelos de regresión lineal múltiple, y proporciona estimaciones insesgadas y eficientes de los coeficientes.

Otro algoritmo implementado fue la regresión de Ridge, que incorpora un término de penalización al tamaño de los coeficientes estimados (Saunders et al., 1998). Esto implica que la función que se busca minimizar se encuentra además penalizada por un parámetro de complejidad que controla el nivel de contracción de los coeficientes. Este algoritmo es útil en los casos en que existe colinealidad entre las variables predictoras, proporcionando un modelo más robusto en comparación con el enfoque de mínimos cuadrados ordinarios bajo la situación de múltiples variables predictoras correlacionadas.

También se evaluó la alternativa de regresión LASSO, que incluye un término de penalización distinto en la función a minimizar. El objetivo de este término es reducir la cantidad de variables que integran el modelo final al mínimo, lo que conlleva a una disminución fuerte en la cantidad de variables involucradas en la predicción Santosa & Symes (1986). Este algoritmo es especialmente útil cuando se desea seleccionar un subconjunto de variables predictoras, proporcionando un modelo de mayor interpretabilidad.

Además, se implementó el algoritmo de red elástica, el cual fusiona las regularizaciones y penalizaciones introducidas por los métodos de regresión Ridge y LASSO (Zou & Hastie, 2005). Este enfoque posibilita la estimación de un modelo que combina las ventajas de ambos métodos, ofreciendo la estabilidad característica de Ridge y la habilidad de selección de características de LASSO.

También, se empleó el algoritmo de descenso por el gradiente estocástico, que es una alternativa comúnmente usada en el contexto del aprendizaje automático donde se busca ajustar iterativamente los parámetros del modelo de regresión lineal múltiple de manera eficiente (regresión SGD). Este método implica calcular el error entre las predicciones del modelo y los valores reales, y luego ajustar los parámetros en la dirección opuesta del gradiente de la función de costo, permitiendo una adaptación precisa. Al ser de naturaleza “estocástica”, en cada iteración se utiliza un subconjunto aleatorio de los datos disponibles para actualizar los parámetros, lo que agiliza la convergencia y se torna especialmente útil en situaciones donde los conjuntos de datos son de dimensiones considerables y presentan colinealidad (Bottou & Bousquet, 2007).

Máquinas de vectores de soporte: Las máquinas de vectores de soporte (*support vector machine (SVM)*) (Cortes & Vapnik, 1995) es otro algoritmo de aprendizaje automático que puede ser empleado tanto para la clasificación de variables discretas como para la predicción de variables continuas. A diferencia de los modelos de regresión lineal, las SVM buscan encontrar un hiperplano en un espacio de características de mayor dimensión, logrado mediante una transformación no lineal de los datos (mediante el uso de funciones kernel) (Figura 5.2). El enfoque se centra en la identificación de los vectores de soporte, puntos de datos clave que influyen en la ubicación del hiperplano y los márgenes. La efectividad de las SVM radica en su capacidad para descubrir el hiperplano que maximiza los márgenes alrededor de estos vectores de soporte, permitiendo el ajuste. Asimismo, esta técnica es especialmente idónea para tratar

conjuntos de datos que no sean separables linealmente en el espacio original, gracias a la transformación llevada a cabo en el espacio de características de mayor dimensión.

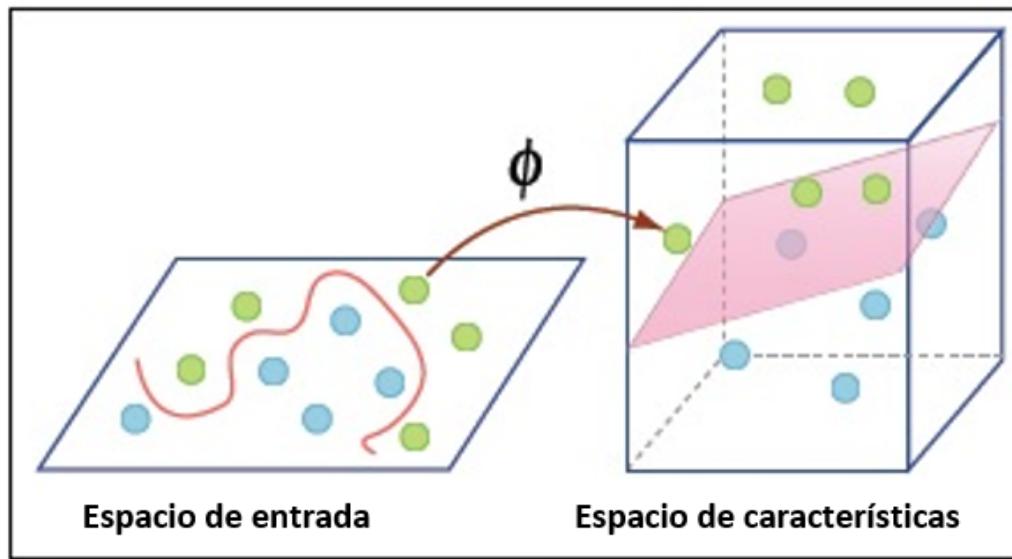


Figura 5.2: Diagrama representativo del algoritmo de máquinas de vectores de soporte.

K vecinos más cercanos: El algoritmo de k vecinos más cercanos (*k nearest neighbours (KNN)*) es un método basado en distancias. En el contexto del aprendizaje automático es capaz de aprender funciones objetivo complejas sin hacer suposiciones sobre la distribución de los datos, lo que lo hace útil en problemas de relaciones no lineales (Fix, 1985). Su funcionamiento se basa en el cálculo de las distancias entre las observaciones respecto a las variables predictoras y el promedio de las observaciones más cercanas. Así, selecciona las K observaciones de mayor similitud a la observación que se desea predecir y, asigna como predicción el promedio de la respuesta en el grupo de vecinos. En lugar de generar una función que describa la relación entre las variables de entrada y salida, KNN guarda en memoria las muestras de entrenamiento y busca en esta base la respuesta más cercana para cada consulta de entrada. Este enfoque lo hace especialmente útil para problemas de predicción donde la estructura subyacente del problema es compleja y no se conoce a priori.

Bosques aleatorios: En los últimos años ha aumentado el interés en los métodos de aprendizaje automático que usan ensambles de modelos como los bosques aleatorios (*random forest*) (Breiman, 1996). Este método combina los resultados de múltiples árboles de regresión (Breiman, 1996) generados independientemente por remuestreo del conjunto de datos original para mejorar el rendimiento predictivo mediante ensamblado (Figura 5.3). Uno de los métodos de ensamblado más reconocido es el empaquetado (*bagging* o *bootstrap aggregating*) (Breiman, 1996). En el aprendizaje automático, se crean múltiples conjuntos de datos mediante remuestreo con reemplazo de los datos de entrenamiento y luego se ensamblan los resultados obtenidos en cada muestra mediante el promedio.

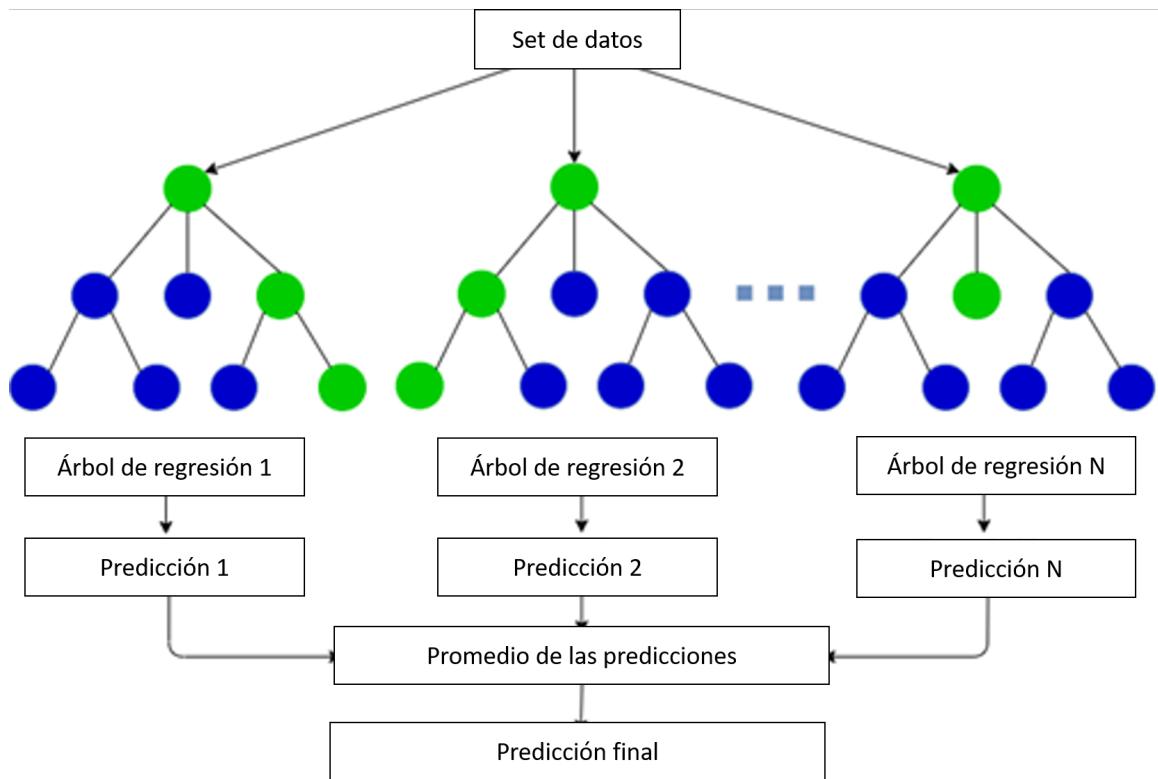


Figura 5.3: Diagrama de predicción por ensamble, en verde se muestra el camino de predicción resultante para cada árbol de regresión.

Breiman (Breiman, 2001a) propuso el modelo de bosques aleatorios, que añade una capa de aleatoriedad en las variables de entrada a cada uno de los empaquetados,

proporcionando mayor robustez contra el sobreajuste. Al ajustar el modelo, se considera un número limitado de variables predictoras elegidas aleatoriamente. Esta aleatoriedad hace que cada uno de los subconjuntos de datos considere distintos subconjuntos de variables predictoras garantizando mayor independencia entre los modelos estimados a partir de las distintos subconjuntos generados por el remuestreo.

Máquina de aumento de gradiente: Otra técnica de ensamblaje utilizada es el *boosting*, un método que busca mejorar el rendimiento del modelo a través de la combinación de múltiples estimadores individuales. En el contexto del aprendizaje automático, el *boosting* se emplea en conjunción con árboles de regresión, dando origen al algoritmo conocido como máquina de aumento de gradiente (*gradient boosting machine (GBM)*), propuesto por Friedman en 2001. La esencia de este enfoque reside en la incorporación secuencial de nuevos modelos de árbol al conjunto (*ensemble*), lo que contribuye a disminuir el sesgo de los estimadores previamente ajustados. Esta reducción del sesgo se refleja a través de una función de error (Figura 5.4). Cada árbol adicional es entrenado con respecto al error acumulado hasta esa iteración del modelo (Friedman, 2001).

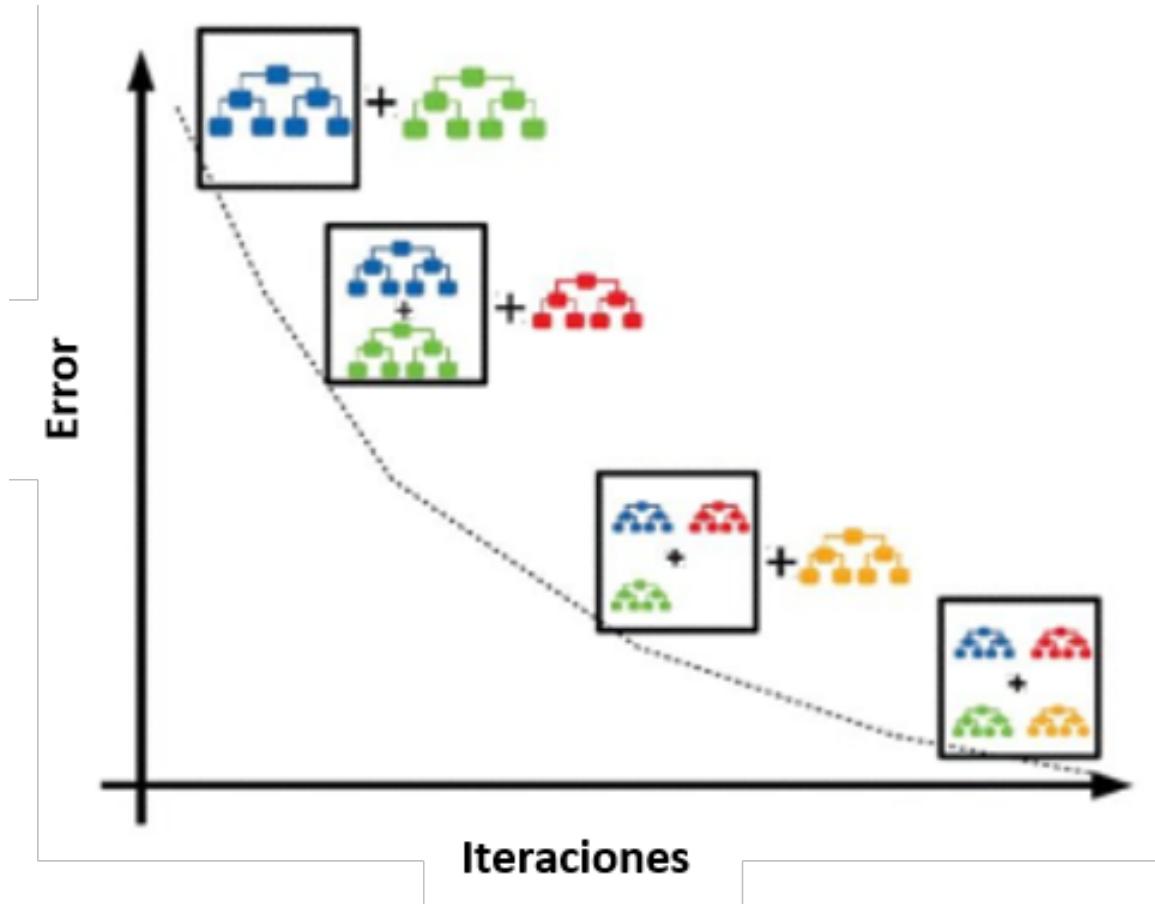


Figura 5.4: Diagrama representativo del algoritmo de máquina de aumento de gradiente.

Uno de los mayores desafíos en el modelado por aprendizaje automático es la capacidad de generalizar, ya que los modelos pueden sesgarse durante el aprendizaje, dando como resultado modelos inadecuados que podrían sobreajustar los datos. Para mitigar los efectos del sobreajuste, Friedman propuso la técnica de contracción (Friedman, 2001). Esta técnica comprende un enfoque de regularización que actúa penalizando los coeficientes de regresión. En el contexto del GBM, la contracción penaliza la importancia de los árboles individuales en cada paso consecutivo. La función objetivo final del modelo consta de dos términos: una función de pérdida de entrenamiento representada por error cuadrático medio y la regularización que mide la complejidad del modelo.

Redes neuronales artificiales: Las redes neuronales artificiales (*artificial neural networks (ANN)*) (Bishop et al., 1995; Haykin, 1999) son algoritmos de aprendizaje profundo que permiten modelar procesos no lineales de manera eficiente. La información de entrada es recolectada a través de un vector denominado capa de entrada. Una vez ingresada se propaga y fluye capa a capa de la red, estableciendo las relaciones entre la capa de entrada y la capa final denominada de salida. El aprendizaje de las interrelaciones entre las variables predictoras se desarrolla a través del ajuste de parámetros característicos de este tipo de modelo denominados pesos sinápticos. Las capas intermedias u ocultas consisten en múltiples unidades denominadas neuronas que se encuentran interconectadas con las neuronas de las capas anterior y posterior. El número de capas ocultas y el número de neuronas en cada una de ellas define la topología de la red (Figura 5.5). Cada neurona genera una respuesta excitatoria a las señales recibidas por medio de una función de activación. Existen distintas funciones de activación, de las cuales algunas pueden resultar más recomendadas para un objetivo y problema en particular (Bishop et al., 1995; Haykin, 1999).

El aprendizaje de la red se basa en capturar las relaciones entre la capa de entrada y de salida a partir de la optimización de una métrica de error. En los problemas de predicción de variables continuas la métrica más implementada es la raíz del error cuadrático medio (*RMSE*), la cual se calcula comparando las salidas predichas por la red con los valores observados en un conjunto de entrenamiento. A partir de la pérdida registrada es que se genera una actualización de los pesos sinápticos mediante el algoritmo de retropropagación (*backpropagation*) (Rumelhart et al., 1986). Una de las mayores dificultades de este tipo de enfoque es encontrar la topología de red que mejor ajuste a los datos. Este proceso normalmente se aborda a partir del ajuste iterativo de múltiples arquitecturas. Generalmente, se parte de una red simple de una capa oculta y pocas neuronas, y se procede aumentando la complejidad de la red (incrementando secuencialmente el número de neuronas, así como el número de capas

ocultas) hasta obtener un modelo satisfactorio que optimice los resultados.

Las redes neuronales recurrentes (*recurrent neural network (RNN)*), a diferencia de las ANN, incluyen bucles o conexiones entre unidades que permiten que la información persista en mayor medida a lo largo de las sucesivas capas de la red (Figura 5.5). La capacidad de incorporar información sobre el contexto pasado en sus sucesivas salidas las hace especialmente prometedores para tareas en las que la información pasada es relevante para la predicción, como en el caso de series temporales. Las *RNN* pueden ser consideradas como múltiples copias de la misma red neuronal actuando conjuntamente, cada una de las cuales transfiere información a su sucesora y forma una arquitectura en cadena capaz de preservar la información de la secuencia. Las *RNN* son eficientes implementando información pasada en un intervalo de tiempo corto, para realizar una tarea de predicción. Sin embargo, la información relevante no siempre está cercana al pasado y, a medida que aumenta el intervalo de tiempo, las *RNN* no logran conectar la información para realizar las predicciones, de esta forma disminuye su capacidad predictiva cuando la información importante se encuentra contenida en mayores desfases de tiempo.

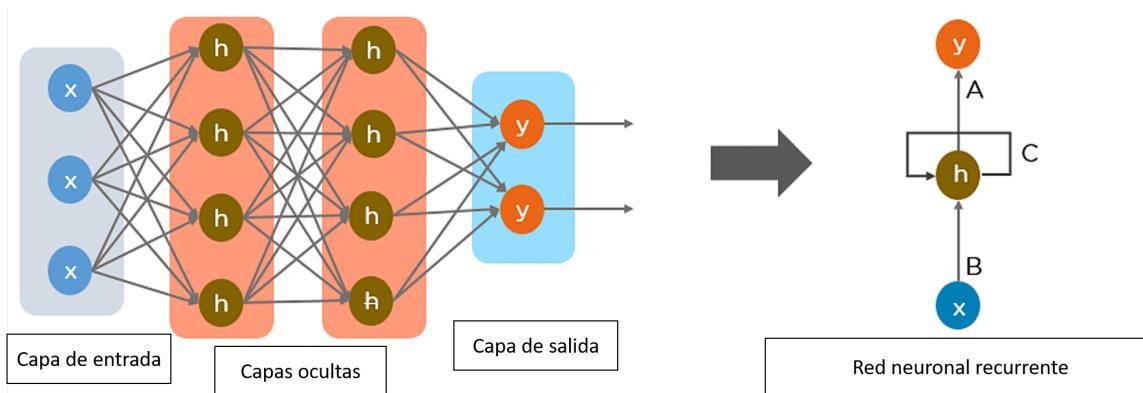


Figura 5.5: Diagrama representativo de algoritmos de redes neuronales.

Para solventar la falencia de las RNN, en 1997 fueron introducidas las redes de memoria a largo plazo (*Long Short-Term Memory (LSTM)*) por Hochreiter y Schmidhuber (Hochreiter & Schmidhuber, 1997) y mejoradas en el 2000 por Gers

et al. (Gers et al., 2000). Estas son una variación de las *RNN* capaz de aprender dependencias de largo plazo, gracias a la inclusión de unidades especiales llamadas bloques de memoria. Además, de otras unidades multiplicativas llamadas puertas que controlan el flujo de información de una unidad *LSTM* a otra. Una variante de este tipo de red es la red recurrente con compuerta (*Gated Recurrent Unit (GRU)*), introducidas por Cho et al. (Cho et al., 2014), la cual mantiene las características propias de las *LSTM* con una estructura más simple, reduciendo la cantidad de parámetros a ajustar.

Selección de variables

Fueron analizadas dos alternativas respecto de la cantidad de variables predictoras a incluir en el modelado predictivo (base original y base reducida). Para la reducción de la cantidad de variables se implementó una combinación de múltiples algoritmos que se describen a continuación.

El algoritmo *featurewiz* (Bakheet et al., 2023; A. Hu et al., 2023; Novaes Neto, 2022), surge de la combinación de dos algoritmos de selección de características. En primer lugar, emplea el algoritmo de *SULOV* (*Searching for Uncorrelated List of Variables*), el cual se encuentra basado en el algoritmo de *MRMR* (*Minimum-Redundancy-Maximum-Relevance*) (Radovic et al., 2017). Este permite reducir la base de datos a un número óptimo de variables no correlacionadas que contienen la máxima información mutua con respecto a la variable respuesta. Una vez que *SULOV* ha seleccionado las variables más relevantes, se emplea un método de eliminación recursiva de variables (*RFE* de *Recursive Feature Elimination*). La *RFE* es un método de selección de características que ajusta de forma iterativa un algoritmo predictivo a una base de datos, eliminando aquellas características de menor importancia, hasta alcanzar el número óptimo de variables que permiten maximizar el desempeño predictivo. El algoritmo de *featurewiz* emplea una eliminación recursiva basada en el algoritmo predictivo de *XGBoost* (Tianqi Chen et al., 2015; Tianqi Chen & Guestrin, 2016). De esta forma, el algoritmo de

featurewiz posteriormente a la reducción por *SULOV* aplica un algoritmo de *XGBoost Recursivo*, que logra encontrar el conjunto óptimo de características, que permite entrenar un modelo de aprendizaje automático. A modo de resumen se desea destacar que el algoritmo de *featurewiz* surge de la combinación de dos algoritmos de selección de características: *SULOV* basado en el algoritmo de *MRMR* y *RFE* basado en el algoritmo de *XGBoost*.

El algoritmo de *Boruta* (Kursa et al., 2010), se encuentra basado en el método de bosques aleatorios y tiene por objetivo identificar la importancia de las variables en la explicación de la variable respuesta. Su algoritmo de búsqueda se basa en el ajuste del modelo de bosques aleatorios en la base original y en una base de datos creada a partir de la mezcla aleatoria de las realizaciones de las variables. A partir de la comparación del desempeño de los distintos modelos ajustados determina cuáles son las variables de mayor relevancia en la base de datos.

Finalmente, se implementó otro algoritmo denominado como *Boruta alternativo* (Jamei et al., 2023), cuya variación reside en su paso inicial, el cual busca una reducción de las variables en función del coeficiente de correlación de Pearson. Este algoritmo busca eliminar las variables en la base de datos que se encuentren fuertemente correlacionadas, priorizando conservar aquellas de mayor correlación con la variable respuesta. Posteriormente, implementa el algoritmo de eliminación recursiva basado en bosques aleatorios implementado por *Boruta* para determinar cuáles son las variables más importantes en la base de datos.

Para crear la base de datos reducida final, se optó por combinar los resultados obtenidos por los tres métodos de reducción de dimensionalidad (*featurewiz*, *Boruta* y *Boruta alternativo*). El procedimiento acatado fue otorgar un punto a cada una de las variables seleccionadas por cada uno de los métodos. Para obtener un conjunto de datos reducido se seleccionaron todas las variables predictoras cuya sumatoria de puntos era igual a tres, es decir que fueron seleccionadas de forma independiente

por los tres algoritmos de reducción. La decisión de combinar múltiples métodos de selección de variables para la reducción de dimensionalidad se basa en la evidencia de que esta estrategia puede mejorar significativamente la precisión del modelo y reducir su complejidad. Cada método de selección de variables tiene sus propias fortalezas y debilidades, y su combinación permite captar las ventajas de cada uno, resultando en un conjunto de características más robusto y relevante (R.-C. Chen et al., 2020; Htun et al., 2023).

Procedimiento de ajuste y evaluación de modelos

Para comparar el desempeño predictivo de los distintos algoritmos ajustados tanto en el conjunto de datos de entrenamiento como en el de testeo, se implementó la métrica de rendimiento raíz del error cuadrático medio de predicción (*RMSE*), la cual se deriva a partir del error cuadrático medio (*MSE*).

Error cuadrático medio (*mean squared error (MSE)*): es un estimador que mide el promedio de los errores al cuadrado, es decir, la diferencia entre el valor observado y su predicción por el modelo (Ecuación 5.1). Es una función de riesgo, correspondiente al valor esperado de la pérdida del error al cuadrado o pérdida cuadrática. Mientras menor es su valor, más exactos serán los pronósticos del modelo predictivo (Lehmann & Casella, 2006).

$$\sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (5.1)$$

Ecuación 5.1: Error cuadrático medio (*mean squared error*).

Componentes:

- y_i : valor real u observado en el i-ésimo punto de datos en un conjunto de datos. Representa el resultado o la variable de respuesta que se está tratando de predecir o modelar en el contexto de un problema de regresión.
- \hat{y}_i : predicción del modelo en el i-ésimo punto de datos. Es el valor estimado por el modelo para y_i .
- n : número total de puntos en el conjunto de datos.

Raíz del error cuadrático medio (*root mean squared error (RMSE)*): expresa el promedio de la diferencia cuadrática entre los valores predichos por el modelo y

los valores observados (Ecuación 5.2). Esta es siempre no negativa, y un valor de 0 (poco común en la práctica) indicaría un ajuste perfecto a los datos. Generalmente, mientras menor $RMSE$ mejor ajuste o mayor exactitud en las predicciones. Esta métrica es sensible a valores atípicos y pone mayor peso en los grandes errores, ya que su coeficiente elevado al cuadrado influirá notablemente en el resultado (Lehmann & Casella, 2006; Pontius et al., 2008; Willmott & Matsuura, 2006). Esta métrica se encuentra definida hace muchos años y es ampliamente recomendada para evaluar el desempeño de modelos de calidad del aire (Thunis et al., 2011; H. Zhang et al., 2014).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (5.2)$$

Ecuación 5.2: Raíz del error cuadrático medio (*root mean squared error (RMSE)*).

Componentes:

- y_i : valor real u observado en el i-ésimo punto de datos en un conjunto de datos. Representa el resultado o la variable de respuesta que se está tratando de predecir o modelar en el contexto de un problema de regresión.
- \hat{y}_i : predicción del modelo en el i-ésimo punto de datos. Es el valor estimado por el modelo para y_i .
- n : número total de puntos en el conjunto de datos.

La metodología implementada para llevar a cabo el ajuste y evaluación de los modelos predictivos, se basó en un enfoque iterativo que involucró la división de datos, entrenamiento del modelo y evaluación por medio de las métricas de desempeño. La base de datos se dividió en subconjuntos de entrenamiento, validación y testeо a partir de un muestreo aleatorio estratificado por meses, días y horas. Esta estratificación buscó maximizar la representatividad de las distintas condiciones temporales existentes en los distintos subconjuntos de datos para el ajuste del modelo predictivo. Las proporciones correspondientes para cada uno de los subconjuntos (entrenamiento, validación y testeо), fue de 70 %, 15 % y 15 %, respectivamente. Posteriormente, se procedió a ajustar cada uno de los algoritmos de modelado predictivo en los subconjuntos de entrenamiento y validación. Una vez que los modelos fueron entrenados, se realizaron

predicciones tanto en la base de datos de entrenamiento como en la de testeo. A partir de estas predicciones, se llevó a cabo el cálculo de las métricas de desempeño en la base de datos de testeo. Este procedimiento de entrenamiento y evaluación se repitió de forma iterativa 10 veces, variando las semillas aleatorias que determinaron las particiones de los datos en cada iteración. Finalizadas las iteraciones, se calculó el promedio de las métricas de desempeño para las 10 corridas del procedimiento en los de testeo. Finalmente, se compararon los resultados promedio de las métricas de desempeño para determinar cuál de los algoritmos predictivos fue el que generó el ajuste de mayor capacidad predictiva. En el contexto de entrenamiento de modelos de aprendizaje automático, los datos de validación se utilizan para evaluar el rendimiento del modelo de manera iterativa durante el proceso de entrenamiento. Esto permite ajustar los hiperparámetros del modelo y realizar mejoras antes de hacer una evaluación final con los datos de testeo.

La inclusión de la variable categórica máscara de fuego (VNP14A1) en la base de datos para desarrollar el modelo predictivo se llevó a cabo mediante el empleo de la técnica conocida de *One-Hot Encoding* (Draper & Smith, 1998). Este proceso implica la transformación de las variables categóricas en una representación binaria, donde cada categoría se convierte en una nueva variable binaria independiente. La razón detrás del uso de *One-Hot Encoding* radica en la necesidad de representar las variables categóricas de manera que puedan ser incorporadas efectivamente en modelos matemáticos y algoritmos de aprendizaje automático.

Todos los algoritmos predictivos ajustados fueron optimizados durante el procedimiento iterativo de entrenamiento y evaluación a partir de las particiones de entrenamiento y validación. Esta optimización se llevó a cabo variando los valores de los hiperparámetros, hasta encontrar las configuraciones que aseguraron el mayor rendimiento predictivo. Los resultados de los hiperparámetros que demostraron el mayor rendimiento predictivo en la base de datos seleccionada se encuentran repor-

tados en el Anexo 4 para cada uno de los usos de suelo. Además, en el caso de las redes neuronales, se profundizó en el estudio de diferentes arquitecturas. Las redes neuronales recurrentes entrenadas en este capítulo se encuentran conformadas por un modelo apilado (múltiples capas superpuestas). Cada una de las capas del modelo apilado consta de una capa recurrente y una de dilución (*dropout*). La capa recurrente permite comprender dependencias temporales y secuenciales, mientras que la capa de dilución contribuye a la regularización al prevenir el sobreajuste, al mismo tiempo que fomenta la generalización del modelo. Además, se evaluaron distintas metodologías de recurrencia implementando diferentes variantes de apilamiento para determinar si mejoraba el rendimiento del modelo.

Además, cabe destacar que para abordar la comparación de la capacidad predictiva de los modelos ajustados se implementó una línea de base contra la cual contrastar los algoritmos ajustados. Esta refiere a un modelo de referencia (*base line*), que predice la concentración de PM_{2.5} para las horas sin dato (testeo) como la media de la concentración en la hora anterior y posterior. Es relevante resaltar que esta predicción no es posible en la práctica, ya que no se dispone de datos de concentración de PM_{2.5} en tiempo real. Un desempeño del modelo ajustado superior o similar al de este modelo de referencia indica una elevada capacidad predictiva.

Las contribuciones de cada una de las variables a las predicciones se cuantificaron con el método de Shapley (SHAP) (Lubo-Robles et al., 2020). Cada valor SHAP representa la contribución de una variable independiente a una predicción individual específica, para el caso de la concentración de PM_{2.5}, en unidades de $\frac{\mu\text{g}}{\text{m}^3}$. En concreto, el valor SHAP para un predictor y una observación dada es la diferencia existente en los resultados de predicción entre un modelo que se ajusta incluyendo y otro excluyendo a dicho predictor. Para cada observación, la suma de todos los valores SHAP, más la media general (concentración media global de PM_{2.5} en los datos de entrenamiento), es igual a la predicción del modelo predictivo. La matriz resultante de los valores SHAP

puede resumirse para comprender cómo contribuye cada predictor a las predicciones resultantes. El valor SHAP absoluto medio de todas las observaciones resume la contribución global de cada variable predictora, y es posible una interpretación local del modelo mediante la visualización a través de un gráfico de dispersión que compare los valores de dicha variable y la contribución SHAP que generan dichos valores.

5.3. Resultados

5.3.1. Modelado de PM_{2.5} en entornos urbanos

Comparación de algoritmos para el uso del suelo urbano

Esta sección desarrolla el modelado predictivo de la concentración de PM_{2.5} a partir de las base de datos original y reducida. La base de datos original involucró todas las variables independientes (193), mientras que la reducida contó con un total de 77 variables. Una descripción detallada de cada una de ellas puede hallarse en el Anexo 1. En el contexto del análisis y modelado de datos, el término “variable independiente” se refiere a cualquier variable que se utiliza como entrada para predecir la variable objetivo (también conocida como variable dependiente). No se refiere a la independencia estadística entre las variables. Los resultados del modelo predictivo desarrollado en esta sección se presentan en la tabla 5.1.

Tabla 5.1: Error cuadrático medio de predicción (*RMSE*) de modelos ajustados para predecir la concentración de PM_{2.5} en el uso del suelo urbano a partir de 193 variables derivadas a partir de información satelital (base de datos original) y 77 variables seleccionadas (base de datos reducida).

Modelo / Base de datos	Original	Reducida
GRU	3.44	3.46
LSTM	3.50	3.50

Modelo / Base de datos	Original	Reducida
RNN	4.44	4.25
Bosques aleatorios	4.10	4.31
KNN	3.55	3.60
SVM	5.77	5.69
GBM	7.12	7.28
Regresión lineal	8.02	8.30
Regresión SGD	8.03	8.26
Regresión Ridge	8.02	8.30
Regresión LASSO	9.01	9.02
Regresión de red elástica	9.85	9.85
<i>Modelo de referencia</i>	3.67	3.67

Los valores de *RMSE* se encuentran en las mismas unidades que la concentración de PM_{2,5}, es decir [$\frac{\mu g}{m^3}$]. Un menor valor de *RMSE* indica mejor desempeño del modelo predictivo. El PM_{2,5} promedio para el uso del suelo urbano fue de 11.54 $\frac{\mu g}{m^3}$. El *RMSE* reportado se calculó sobre la base de datos de testeo.

Los resultados presentados en la Tabla 5.1 demuestran que el algoritmo predictivo de mayor desempeño predictivo para el uso del suelo urbano fueron las redes neuronales recurrentes en sus versiones bajo la base de datos original y reducida. El gran desempeño predictivo demuestra cómo las características propias de aprendizaje del algoritmo logran captar las fuentes de variabilidad y realizar estimaciones acertadas de la variable respuesta. De entre las distintas arquitecturas de redes neuronales recurrentes evaluadas, específicamente el modelo que se destaca por su sobresaliente capacidad predictiva es aquel que incorpora redes con compuerta (*GRU*).

Además, es importante destacar que el único algoritmo predictivo que también es capaz de superar al modelo de referencia propuesto es el modelo de k vecinos más

cercanos. Sin embargo, este algoritmo no garantiza el aprendizaje de la estructura de variabilidad presente en los datos, ya que su proceso de predicción se basa en identificar las k observaciones más similares en la base de datos de entrenamiento a partir de la matriz de distancias, y calcular su media para obtener la predicción final. Este algoritmo de modelado puede ser una forma novedosa y no convencional de calcular predicciones condicionadas por el parentesco entre los datos, a diferencia de los demás métodos que buscan calcular parámetros y pesos que les permiten inferir acerca de las interrelaciones entre las variables.

Para evitar caer en interpretaciones sesgadas respecto a la relación de vinculación que guardan las variables, se optó por seleccionar el modelo predictivo desarrollado sobre la base de datos reducida. Esta elección se fundamenta en el hecho de que el modelo construido sobre la base de datos original, al contar con un número superior de variables predictoras, aumenta drásticamente el espacio de posibles soluciones que pueden encontrarse, pudiendo existir soluciones en las cuales se atribuya importancia a variables de forma equivocada. Cabe destacar que la base de datos reducida cuenta con las variables más importantes seleccionadas a partir de la combinación de múltiples algoritmos de selección de características. Al trabajar con esta base de datos, el espacio de posibles soluciones se reduce considerablemente, asegurando que todas las soluciones posibles involucren las variables de mayor relevancia. Además, puede apreciarse que la diferencia en la precisión promedio entre los algoritmos predictivos en ambas bases de datos es de $0.020 \frac{\mu g}{m^3}$. Tomando en consideración que la base de datos reducida involucra un 60 % menos de variables, se asume que se introduce un menor porcentaje de ruido al modelado y que no disminuye drásticamente el desempeño predictivo. Por este motivo, además de encontrar soluciones de mayor relevancia, permite reducir la complejidad y el tiempo de ejecución en el entrenamiento del modelo predictivo. Otra ventaja inherente a la base de datos reducida es su mayor viabilidad para la implementación en producción, ya que implica una menor

carga y procesamiento para generar predicciones, además de simplificar las tareas de actualización y reentrenamiento del modelo. Durante el proceso de modelado, se evidenció que la inclusión de todas las variables presentes en la base de datos original puede introducir ruido y sobreajuste, lo que a su vez complica la identificación de las verdaderas relaciones subyacentes entre las variables. Esto se observó en las múltiples iteraciones de validación cruzada, donde la varianza en las métricas de evaluación era considerablemente alta cuando se utilizaba la base de datos original. A diferencia de esto, la base de datos reducida, al ser construida seleccionando las variables más importantes a partir de la combinación de múltiples algoritmos de selección de características, redujo notablemente el espacio de posibles soluciones y aseguró menor varianza y mayor homogeneidad entre las métricas obtenidas durante validación cruzada. A partir de estos resultados fue se fundamentó la elección de la base de datos reducida frente a la original.

La elección de utilizar la base de datos reducida está basada en una combinación de técnicas de selección de características que aseguran que solo las variables más relevantes se utilicen en el modelo. Esta selección no solo mantiene una alta precisión predictiva, sino que también reduce el ruido, la complejidad y el tiempo de entrenamiento, además de mejorar la viabilidad para la implementación en producción. Estas ventajas se traducen en un modelo más robusto, interpretable y eficiente, adecuado para aplicaciones prácticas y sostenibles a largo plazo. En primer lugar la elección de una base de datos reducida se fundamenta en que la reducción de variables de 193 a 77 implica un 60 % menos de variables, lo que disminuye la posibilidad de introducir ruido en el modelo. Menos variables irrelevantes significa menos riesgo de que el modelo encuentre patrones espurios, lo que puede llevar a interpretaciones incorrectas y a un desempeño menos robusto en datos no vistos. En segundo lugar, la diferencia en la precisión promedio entre los algoritmos predictivos en ambas bases de datos es mínima ($0.020 \frac{\mu g}{m^3}$). Esto demuestra que la reducción de variables no impacta drásticamente

el desempeño predictivo, manteniendo una alta precisión mientras se simplifica el modelo. En tercer lugar, el modelo entrenado con la base de datos reducida es menos complejo y más rápido de entrenar. Menor complejidad y tiempo de entrenamiento facilitan iteraciones más rápidas, pruebas más frecuentes y ajustes de modelo más eficientes, lo cual es crucial en entornos de producción. Finalmente, un modelo con menos variables es más fácil de mantener y actualizar. Reduce la carga computacional y el tiempo necesario para generar predicciones, además de simplificar las tareas de reentrenamiento y actualización del modelo en producción.

El modelo de mejor ajuste quedó conformado en su primer aglomerado del apilamiento por un total de 200 neuronas en la capa con compuerta, seguido por una capa de dilución con un 30 % de neuronas apagadas. Los dos aglomerados subsiguientes que dan estructura a esta red se encuentran conformados por un total de 100 neuronas en cada una de sus capas con compuerta y una capa de dilución con un 30 % de porcentaje de neuronas apagadas. Previo a la capa de salida, se encuentra la estructura clásica de capa densa totalmente conectada que permite la combinación de los coeficientes aprendidos por la red para dar como resultado la predicción final. El algoritmo de mayor desempeño permitió obtener predicciones que difieren en promedio de los valores reales medidos de concentración de PM_{2.5} en $3.461 \frac{\mu g}{m^3}$ (*RMSE*).

Interpretación modelo predictivo:

Variables más importantes

Exploración de las variables de mayor importancia dada su contribución a las predicciones.

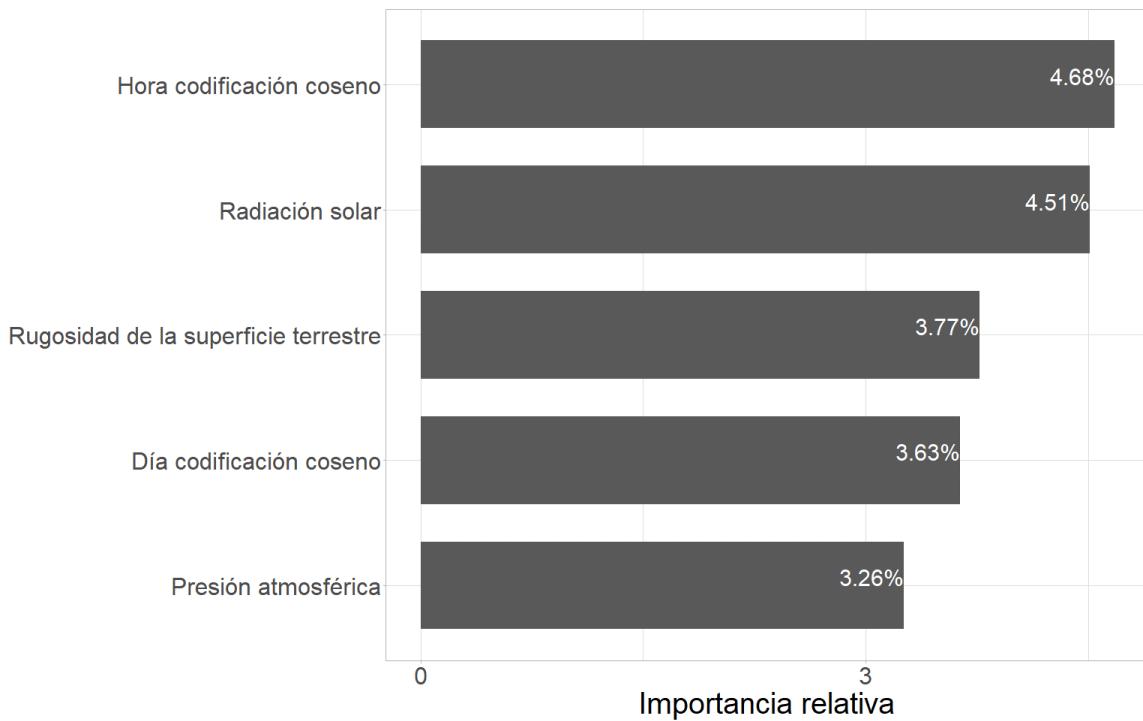


Figura 5.6: Importancia relativa de las variables en las redes neuronales recurrentes para el uso del suelo urbano.

A partir de la Figura 5.6 pueden apreciarse las cinco variables de mayor importancia relativa para el modelo predictivo en el uso del suelo urbano bajo la base de datos reducida. Dos de las variables se encuentran vinculadas a la escala temporal, lo que destaca la influencia de los patrones de variación temporal en la variable objetivo, como se exploró en el capítulo 3. Las variables restantes están asociadas con condiciones meteorológicas vinculadas a la variabilidad analizada en el capítulo 4. Los resultados muestran que las cinco variables de mayor importancia relativa abarcan menos del 25 % de la importancia total. A la hora de predecir la concentración de PM_{2.5} en entornos urbanos, es crucial considerar la hora del día, la radiación solar, la rugosidad de la superficie terrestre, el día de la semana y la presión atmosférica.

Radiación solar

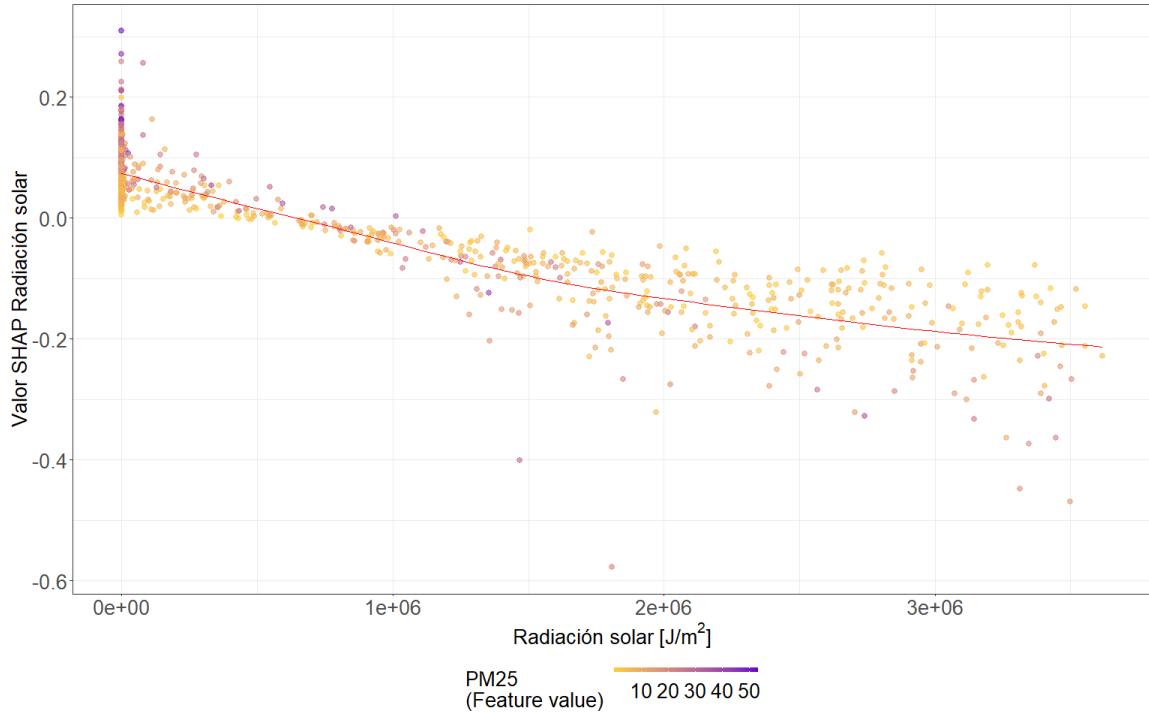


Figura 5.7: Gráfico de la variable radiación solar contra su valor SHAP coloreado por el valor de PM_{2.5} para el uso del suelo urbano.

La Figura 5.7 muestra qué valores de baja magnitud de radiación solar están asociados a incrementos en las predicciones de la variable respuesta, mientras que valores de radiación solar alrededor de $2\text{e}6 \frac{\text{J}}{\text{m}^2}$ están asociados con menores concentraciones de PM_{2.5} predichas. De esta forma se aprecia una relación del tipo inversa similar a una de las identificadas a partir del análisis de vinculación entre ambas variables. Además, puede notarse como los valores de mayor concentración de PM_{2.5}, en función de la escala de graduación de colores, se hallan concentrados en la zona de menores valores de radiación solar, lo cual se condice con la relación inversa descripta en el análisis exploratorio del capítulo 4 en la sección de resultados de la radiación solar, demostrando que el modelo logra capturar la relación entre las variables.

Rugosidad de la superficie terrestre

La rugosidad de la superficie terrestre es la altura equivalente en metros de la rugosidad aerodinámica que posee la superficie del terreno. Esta es una medida de la resistencia que ejerce la superficie al flujo de aire, particularmente este parámetro se utiliza para determinar la transferencia de momento del aire a la superficie. Una mayor rugosidad de la superficie genera una reducción de la velocidad del viento cerca de la superficie. Esta variable se deriva de la cobertura, tipo de vegetación y presencia de nieve en la superficie terrestre y se expresa en metros (m). Aunque la presencia de nieve puede influir en la rugosidad de la superficie terrestre, en el contexto del área metropolitana de Córdoba, este aspecto no es relevante debido a la escasa o nula presencia de nieve. Por lo tanto, para este caso específico, la rugosidad de la superficie se determina principalmente por la cobertura rocosa y el tipo de vegetación.

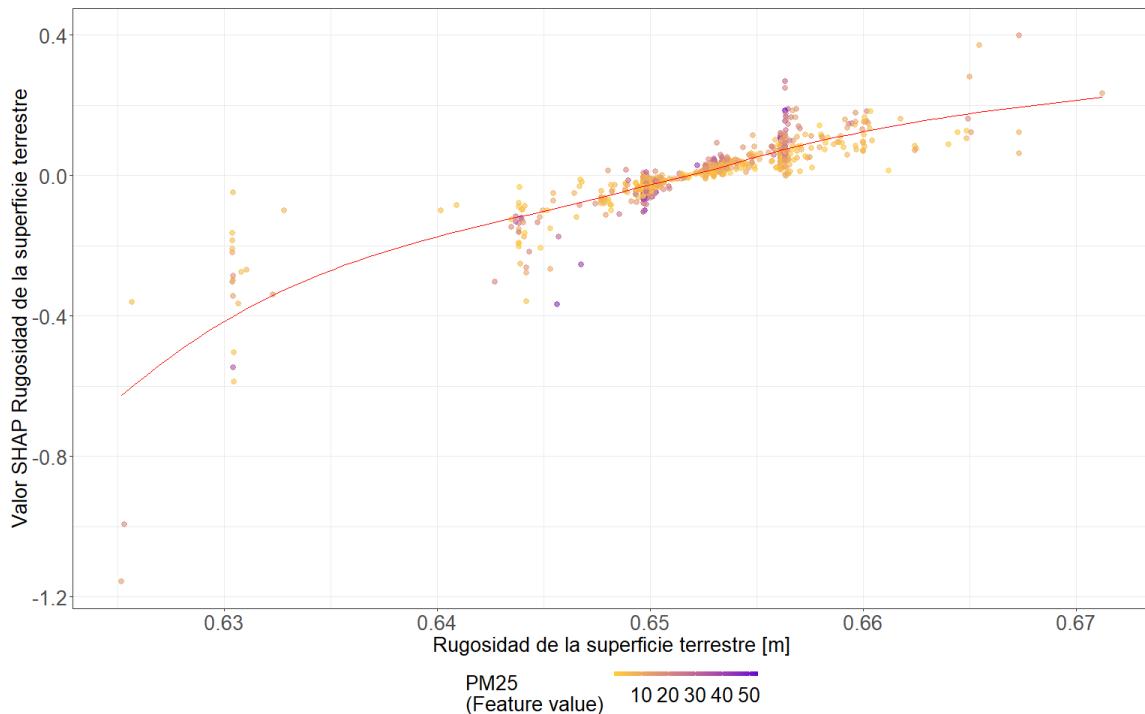


Figura 5.8: Gráfico de la variable rugosidad de la superficie terrestre contra su valor SHAP coloreado por el valor de PM_{2.5} para el uso del suelo urbano.

La Figura 5.8 muestra una relación directa que coincide con la reportada por el análisis de vinculación conjunto entre ambas variables que se encuentra en el Anexo 5. Es decir que valores alrededor de 0.63 m de la rugosidad de la superficie terrestre se asocian con disminuciones en las predicciones de la variable respuesta.

Presión atmosférica

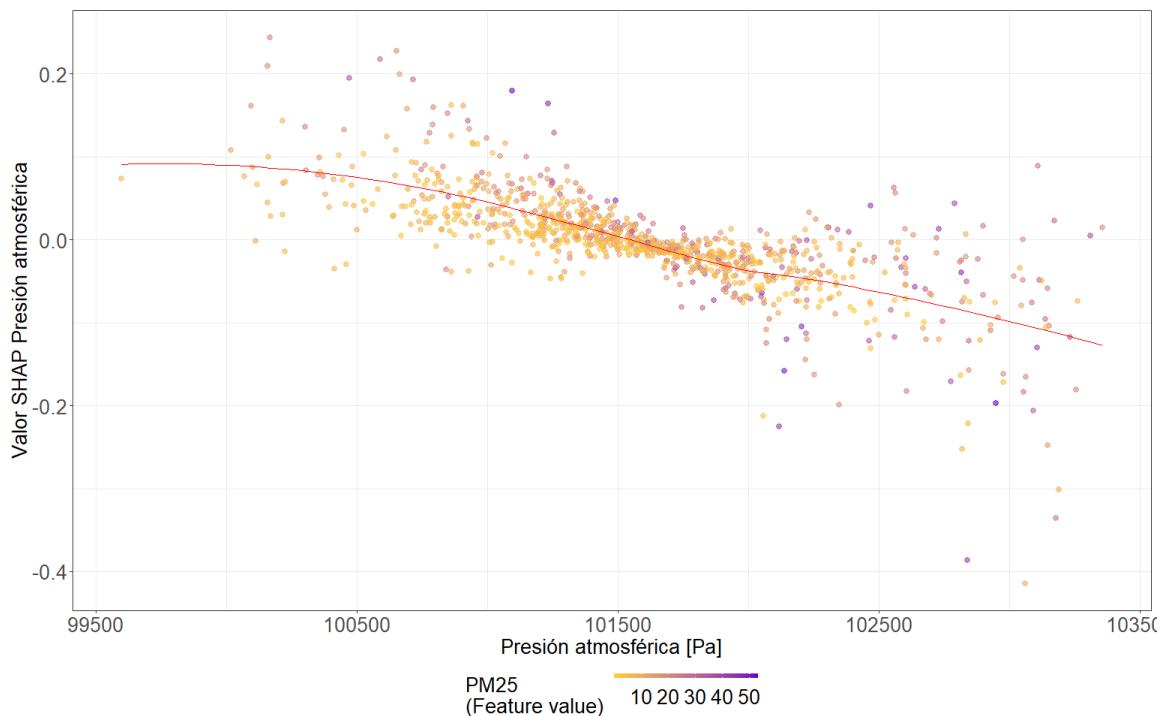


Figura 5.9: Gráfico de la variable presión atmosférica contra su valor SHAP coloreado por el valor de PM_{2.5} para el uso del suelo urbano.

La Figura 5.9 muestra una asociación entre variables de tipo inversa similar a la reportada por el análisis de vinculación conjunto entre ambas variables sin considerar desfasaje (Capítulo 4). Esto indica que valores de presión atmosférica menores a 100,500 Pa favorecen incrementos en las predicciones de la variable respuesta, mientras que valores mayores a 102,500 Pa resultan en disminuciones de las concentraciones de PM_{2.5} predichas.

5.3.2. Modelado de PM_{2.5} en entornos industriales

Comparación de algoritmos para el uso del suelo industrial

Esta sección desarrolla el modelado predictivo de la concentración de PM_{2.5} a partir de la base de datos original y reducida. La base de datos reducida contó con un total de 53 variables. Una descripción detallada de cada una de ellas puede hallarse en el Anexo 1. Los resultados del modelo predictivo desarrollado en esta sección se presentan en la tabla 5.2.

Tabla 5.2: Error cuadrático medio de predicción (*RMSE*) de modelos ajustados para predecir la concentración de PM_{2.5} en el uso del suelo industrial a partir de 193 variables derivadas a partir de información satelital (base de datos original) y 53 variables seleccionadas (base de datos reducida).

Modelo / Base de datos	Original	Reducida
GRU	6.10	5.96
LSTM	6.19	6.17
RNN	6.72	6.59
Bosques aleatorios	6.52	6.54
KNN	6.24	6.26
SVM	7.87	8.02
GBM	9.34	9.53
Regresión lineal	10.10	10.33
Regresión SGD	10.11	10.24
Regresión Ridge	10.10	10.33
Regresión LASSO	10.92	10.93
Regresión de red elástica	11.82	11.82
<i>Modelo de referencia</i>	6.47	6.47

Los valores de *RMSE* se encuentran en las mismas unidades que la concentración de PM_{2,5}, es decir [$\frac{\mu\text{g}}{\text{m}^3}$]. Un menor valor de *RMSE* indica mejor desempeño del modelo predictivo. El PM_{2,5} promedio para el uso del suelo industrial fue de 14.22 $\frac{\mu\text{g}}{\text{m}^3}$. El *RMSE* reportado se calculó sobre la base de datos de testeo.

La Tabla 5.2 presenta los resultados obtenidos en el proceso de modelado predictivo de los datos correspondientes al uso del suelo industrial. Estos resultados exhiben similitudes notables con los obtenidos para el uso del suelo urbano (Tabla 5.1). En primer lugar, se observa que las redes neuronales recurrentes logran el mejor desempeño predictivo, lo que sugiere su robustez y eficacia en la predicción de la concentración de PM_{2,5}. La distinción principal radica en que el óptimo desempeño bajo la base de datos industrial se obtiene sobre los datos reducidos. De igual forma que para el uso del suelo urbano, el algoritmo de redes neuronales recurrentes que destaca por su desempeño predictivo es el modelo que involucra redes con compuerta (*GRU*).

Adicionalmente, para este tipo de uso del suelo, se evidencia que el único algoritmo predictivo que supera al modelo de referencia propuesto es el modelo de k vecinos más cercanos. A través de la estimación de la media entre k observaciones, este modelo logra un alto poder predictivo, aunque no necesariamente captura las interacciones existentes entre las variables.

Cabe destacar que los mejores resultados predictivos fueron obtenidos al emplear la base de datos reducida, proporcionando evidencia acerca de las ventajas de este tipo de base para el modelado predictivo. Por lo tanto, esta será la base de datos elegida para realizar el modelado y, posteriormente, para analizar la relación de vinculación entre las variables.

El modelo de mejor ajuste quedó conformado en su primer aglomerado del apilamiento por un total de 200 neuronas en la capa con compuerta, seguido por una capa de dilución con un 30 % de neuronas apagadas. La subsiguiente capa del aglomerado quedó conformada por un total de 100 neuronas en su capa con compuerta y un 30 % de porcentaje de neuronas apagadas en su capa de dilución. Previo a la capa de salida,

se encuentra la estructura clásica de capa densa totalmente conectada que permite la combinación de los coeficientes aprendidos por la red para dar como resultado la predicción final. En términos de rendimiento, las predicciones del modelo de mejor ajuste difieren en promedio de los valores reales medidos de concentración de PM_{2.5} en 5.964 $\frac{\mu\text{g}}{\text{m}^3}$ (*RMSE*).

Interpretación modelo predictivo:

Variables más importantes

Exploración de las variables de mayor importancia dada su contribución a las predicciones.

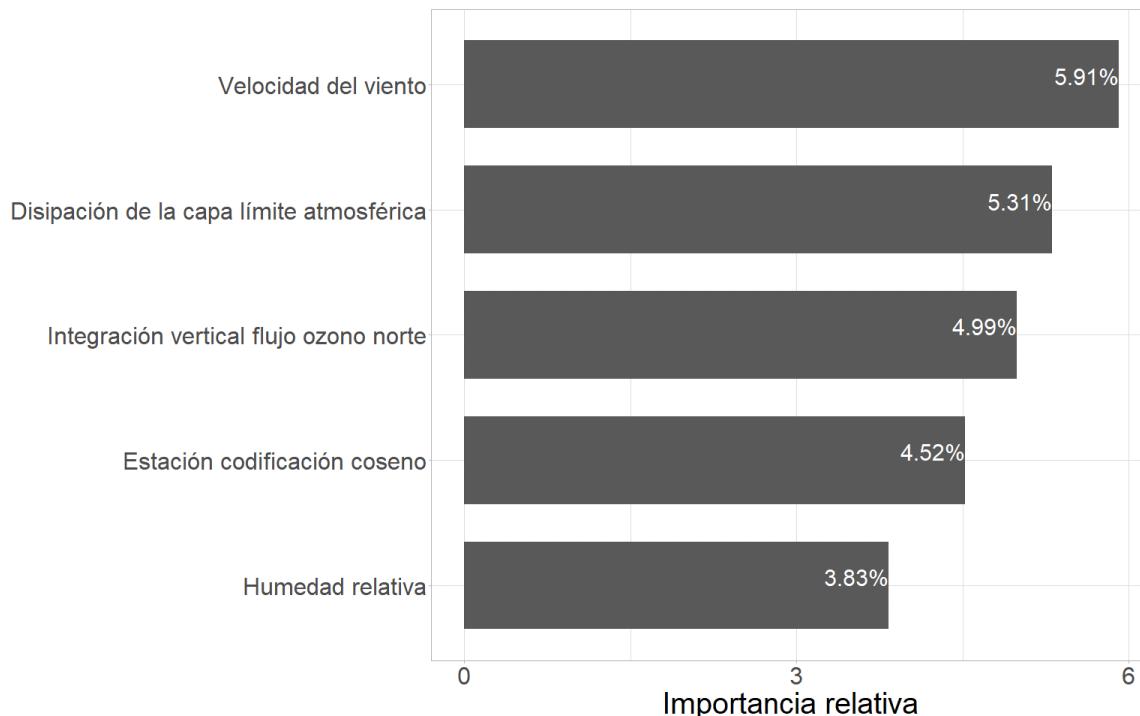


Figura 5.10: Importancia relativa de las variables en las redes neuronales recurrentes para el uso del suelo industrial.

A partir de la Figura 5.10 pueden apreciarse las cinco variables de mayor importancia relativa para el modelo predictivo en el uso del suelo industrial bajo la base de datos reducida. Solo una de las variables se encuentra vinculada a la escala temporal,

destacando la dependencia de la variable respuesta respecto a sus patrones de variación temporal explorados en el capítulo 3. Las restantes son variables meteorológicas vinculadas a la variabilidad estudiada en el capítulo 4. Los resultados muestran que las cinco variables de mayor importancia relativa reúnen menos del 25 % de la misma.

A la hora de predecir la concentración de PM_{2.5} en entornos industriales, es crucial considerar la velocidad del viento, la disipación en la capa límite atmosférica, el flujo de ozono, la estación del año y la humedad relativa.

Velocidad del viento

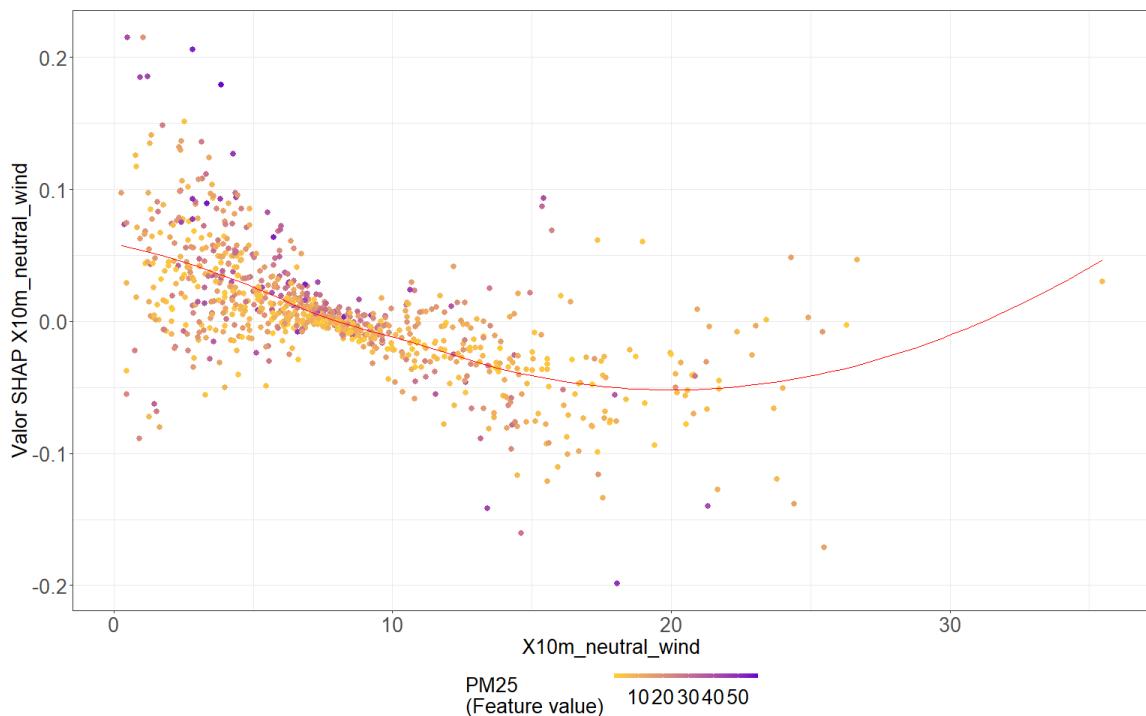


Figura 5.11: Gráfico de la variable velocidad del viento contra su valor SHAP coloreado por el valor de PM_{2.5} para el uso del suelo industrial.

La Figura 5.11 muestra que valores de baja magnitud de velocidad del viento (menores a $10 \frac{m}{s}$) resultan en incrementos de las predicciones de la variable respuesta, mientras que valores de elevada magnitud de la variable independiente (mayores a $20 \frac{m}{s}$) promueven disminuciones en las concentraciones de PM_{2.5} predichas. De esta

forma se identifica una relación del tipo inversa similar a la encontrada durante el análisis de vinculación conjunto entre ambas variables (Capítulo 4). Además, se puede apreciar cómo los valores de mayores concentraciones de PM_{2.5} se localizan en zonas de reducida velocidad del viento, demostrando la gran capacidad predictiva de la variable.

Disipación en la capa límite atmosférica

La disipación en la capa límite atmosférica, indica la conversión acumulada de la energía cinética de flujo medio en calor en toda la columna atmosférica, por unidad de superficie, como consecuencia de los efectos de la tensión asociada a remolinos turbulentos cerca de la superficie y al arrastre de forma orográfica turbulenta. Los remolinos turbulentos cerca de la superficie están relacionados con la rugosidad superficial, mientras que el arrastre de forma orográfica turbulenta es la tensión debida a los valles, colinas y montañas en escalas horizontales inferiores a 5 km. Se espera que incrementos de la disipación en la capa límite atmosférica promueva reducciones de la concentración de partículas finas, como consecuencia de la influencia en la disipación de las mismas, por otra parte, reducciones de la disipación no favorecería el fenómeno de dispersión resultando en mayores concentraciones de PM_{2.5}. Las unidades en las que se mide este parámetro son julios por metro cuadrado ($\frac{J}{m^2}$).

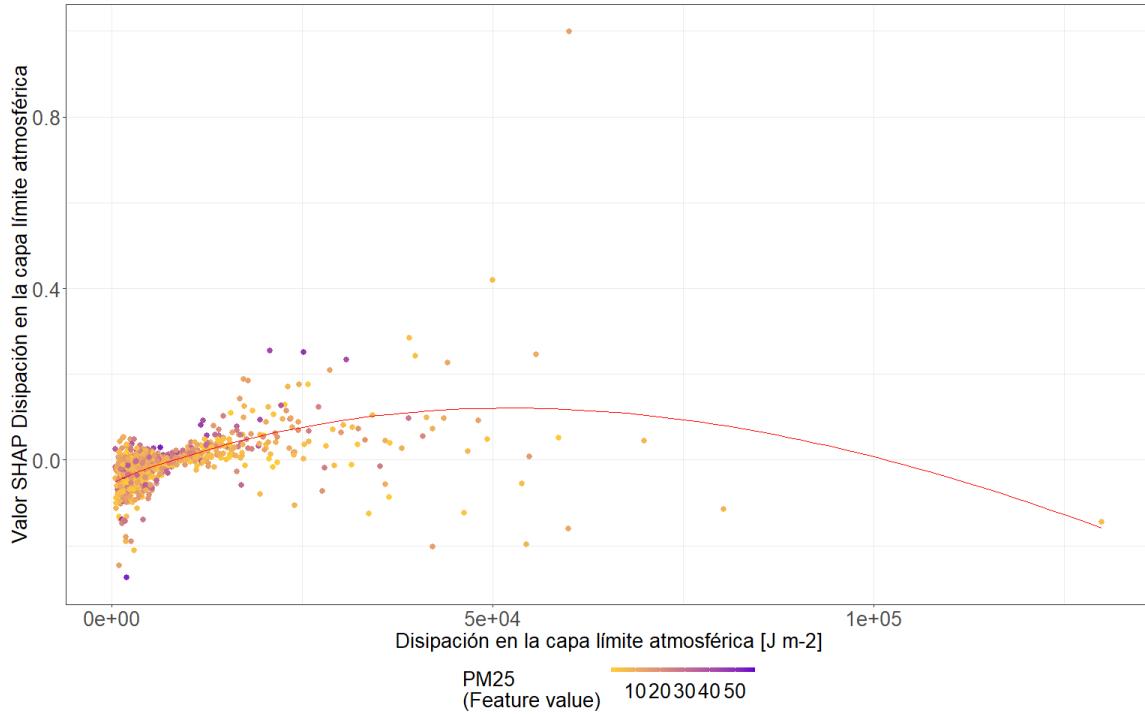


Figura 5.12: Gráfico de la variable disipación en la capa límite atmosférica contra su valor SHAP coloreado por el valor de PM_{2.5} para el uso del suelo industrial.

La Figura 5.12 permite apreciar que el mayor porcentaje de los datos se encuentra concentrado en valores menores a $5e4 J/m^2$ de disipación en la capa límite atmosférica. Por este motivo, no logra apreciarse una clara tendencia en cuanto al aporte de la variable predictora en el modelo predictivo dado a que en todo el rango de variación los aportes son tanto de incremento como de decremento de la concentración. Dada la inconsistencia existente en el rango de variación de los aportes, se concluye que para este modelo predictivo la importancia que adquiere la variable se encuentra vinculada al actuar conjunto con otras variables independientes.

Integración vertical flujo ozono norte

La variable integración vertical de flujo de ozono norte refiere a la tasa horizontal de flujo de ozono en dirección norte, por metro a través del flujo, para una columna de aire que se extiende desde la superficie hasta el tope de la atmósfera. Los valores positivos

denotan un flujo de sur a norte. Las unidades en las que se mide este parámetro son kilogramos por metro por segundo ($\frac{Kg}{m.s}$).

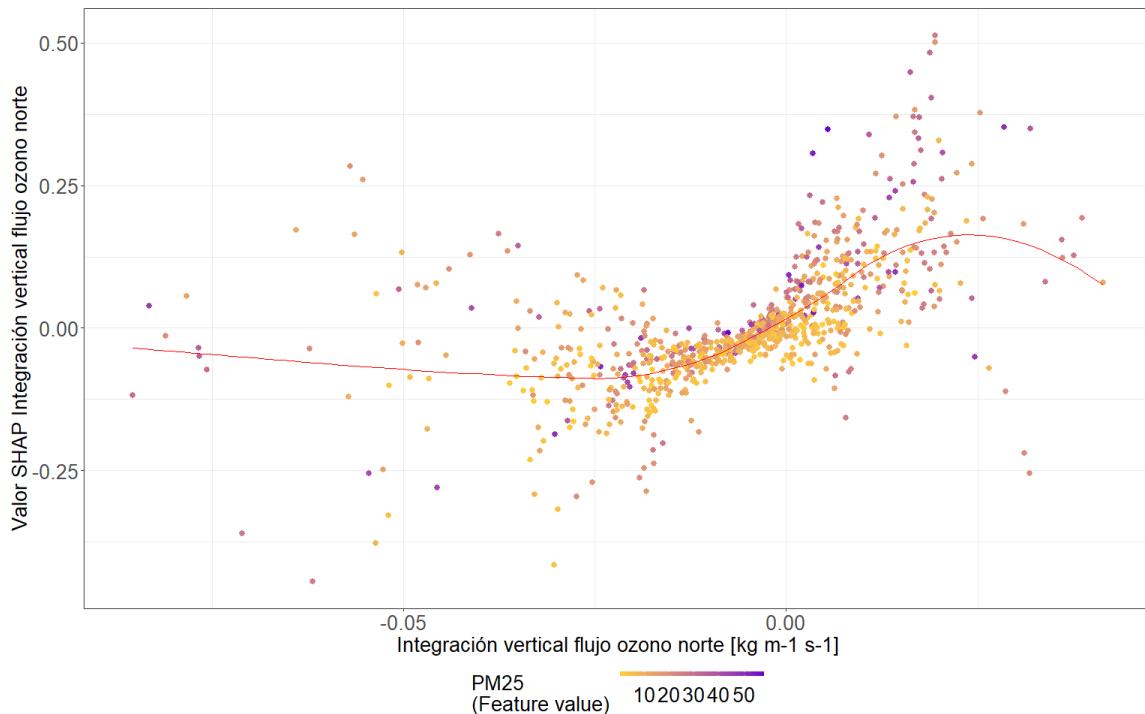


Figura 5.13: Gráfico de la variable integración vertical flujo ozono norte contra su valor SHAP coloreado por el valor de PM_{2.5} para el uso del suelo industrial.

La Figura 5.8 muestra que valores de baja magnitud de la integración vertical flujo ozono norte (menores a $0 \frac{kg}{m.s}$) promueven reducciones en las predicciones de la variable respuesta, mientras que valores de elevada magnitud de la variable independiente (mayores a $0 \frac{kg}{m.s}$) resultan en incrementos de las concentraciones de PM_{2.5} predichas. De esta forma se identifica una relación del tipo directa similar a la reportada por el análisis de variación conjunto entre ambas variables (Anexo 5).

Humedad relativa

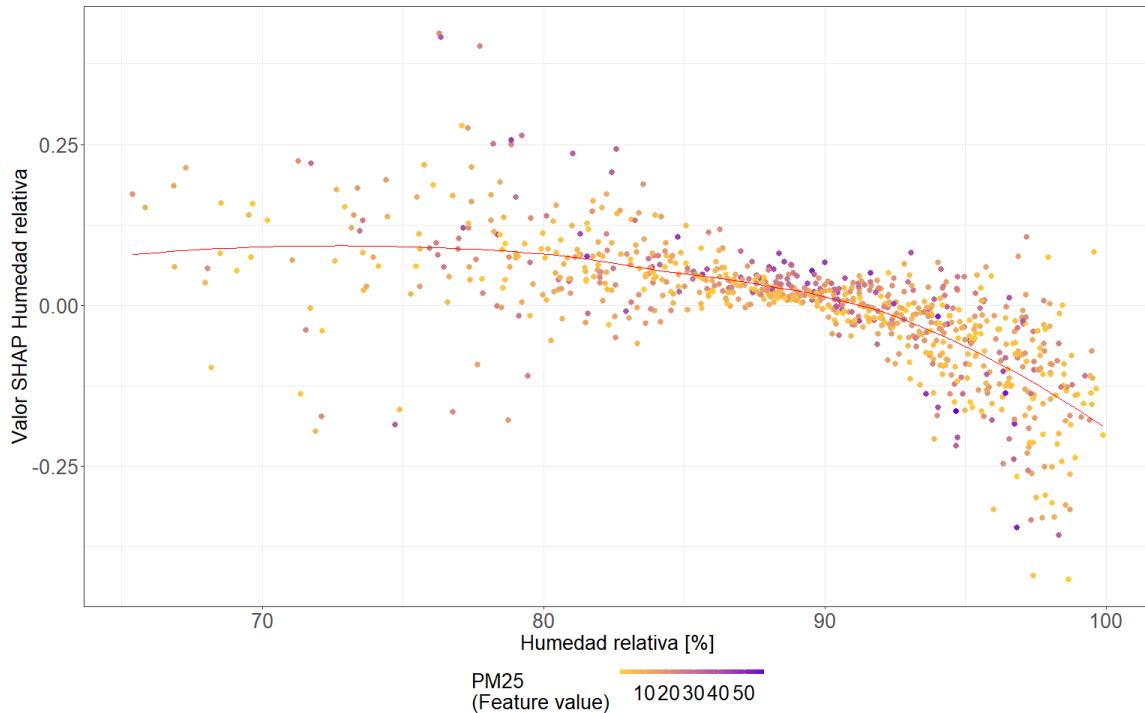


Figura 5.14: Gráfico de la variable humedad relativa contra su valor SHAP coloreado por el valor de PM_{2.5} para el uso del suelo industrial.

La Figura 5.14 muestra que valores menores al 90 % de humedad relativa favorecen incrementos de las concentraciones de PM_{2.5}, mientras que humedades relativas mayores tienden a favorecer decrementos de la concentraciones. Además, puede apreciarse una mayor densidad de valores de elevada concentración de PM_{2.5} entre el 80 % y 90 % de humedad relativa. De esta forma, puede notarse que en este modelo predictivo valores por encima del 90 % de humedad relativa favorecen la remoción de PM_{2.5}, mientras que valores por debajo de este límite favorecen aportes a la concentración de dicho contaminante. De esta forma, puede concluirse que la relación de vinculación que prevalece en el modelo predictivo es del tipo inversa, similar a una de las reportadas durante el análisis de conexión entre las mismas (Capítulo 4).

5.3.3. Modelado de PM_{2.5} en entornos agrícola y área verde

Comparación de algoritmos para el uso del suelo agrícola/ área verde

Esta sección desarrolla el modelado predictivo de la concentración de PM_{2.5} a partir de las base de datos original y reducida. La base de datos reducida contó con un total de 84 variables. Una descripción detallada de cada una de ellas puede hallarse en el Anexo 1. Los resultados del modelo predictivo desarrollado en esta sección se presentan en la tabla 5.3.

Tabla 5.3: Error cuadrático medio de predicción (*RMSE*) de modelos ajustados para predecir la concentración de PM_{2.5} en el uso del suelo agrícola/ área verde a partir de 193 variables derivadas a partir de información satelital (base de datos original) y 84 variables seleccionadas (base de datos reducida).

Modelo / Base de datos	Original	Reducida
GRU	2.83	2.76
LSTM	2.91	2.86
RNN	3.35	3.10
Bosques aleatorios	2.61	2.32
KNN	2.74	2.65
SVM	3.82	3.65
GBM	4.50	4.60
Regresión lineal	5.27	5.38
Regresión SGD	5.18	5.25
Regresión Ridge	5.38	5.38
Regresión LASSO	5.80	5.80
Regresión de red elástica	6.18	6.18
<i>Modelo de referencia</i>	2.77	2.77

Los valores de *RMSE* se encuentran en las mismas unidades que la concentración de PM_{2,5}, es decir [$\frac{\mu g}{m^3}$]. Un menor valor de *RMSE* indica mejor desempeño del modelo predictivo. El PM_{2,5} promedio para el uso del suelo agrícola y área verde fue de 2.76 $\frac{\mu g}{m^3}$. El *RMSE* reportado se calculó sobre la base de datos de testeo.

La Tabla 5.3 presenta los resultados derivados del proceso de modelado predictivo empleando los datos correspondientes al uso del suelo agrícola y área verde. En este caso, puede apreciarse que el algoritmo que demuestra el mayor rendimiento predictivo es el de bosques aleatorios. Este resultado es independiente de la versión de la base de datos utilizada, destacando su mayor precisión en la versión desarrollada a partir de la base de datos reducida. A diferencia de los resultados obtenidos con respecto a los demás usos de suelo (Tabla 5.1 y Tabla 5.2) las redes neuronales recurrentes no logran superar el modelo de referencia propuesto. Este resultado podría encontrarse vinculado a las características propias de la serie de tiempo en entornos agrícolas y áreas verdes.

El modelo óptimo en términos de capacidad predictiva dentro del enfoque de bosques aleatorios se encontró compuesto por un total de 200 árboles individuales que forman el bosque. Además, es interesante destacar que para lograr la separación en un nuevo nodo, se requiere contar con al menos dos muestras en cada uno de estos. Las predicciones generadas por este algoritmo difieren en promedio de los valores reales medidos de concentración de PM_{2,5} en 2.324 $\frac{\mu g}{m^3}$ (*RMSE*).

Interpretación modelo predictivo:

Variables más importantes

Exploración de las variables de mayor importancia dada su contribución a las predicciones.

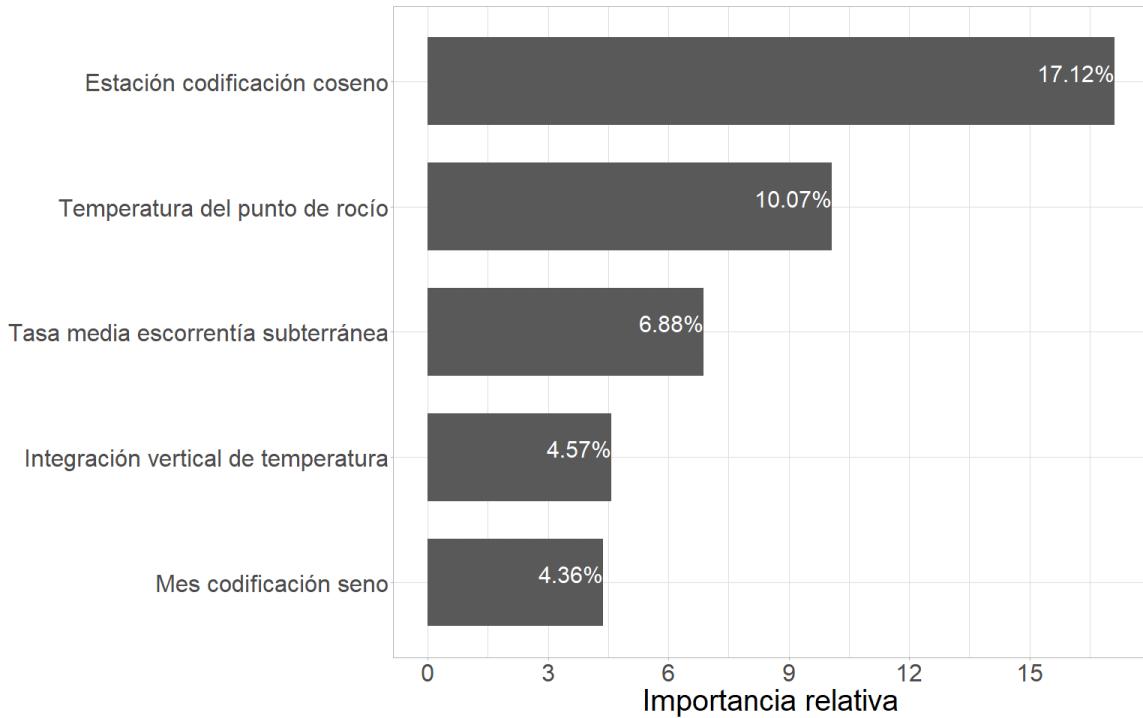


Figura 5.15: Importancia relativa de las variables en el bosque aleatorio para el uso del suelo agrícola/ área verde.

A partir de la Figura 5.15 pueden apreciarse las cinco variables de mayor importancia relativa para el modelo predictivo en el uso del suelo agrícola y área verde bajo la base de datos reducida. Dos de las mismas se encuentran vinculadas a la escala temporal, destacando la dependencia de la variable respuesta respecto a sus patrones de variación temporales explorados en el capítulo 3. Las restantes son variables meteorológicas vinculadas a la variabilidad estudiada en el capítulo 4. Los resultados muestran que las cinco variables de mayor importancia relativa reúnen más del 40 % de la misma. De esta forma puede apreciarse como en este uso del suelo se requiere una menor cantidad de variables para reunir el mismo porcentaje de importancia relativa con respecto a los demás usos de suelo, demostrando de esta forma, mayor capacidad para explicar la variabilidad en la concentración de PM_{2.5} a partir de un porcentaje menor de información.

A la hora de predecir la concentración de PM_{2.5} en entornos agrícolas y áreas

verdes, es crucial considerar la estación del año, temperatura de rocío, escorrentía, temperatura y el mes del año.

Temperatura del punto de rocío

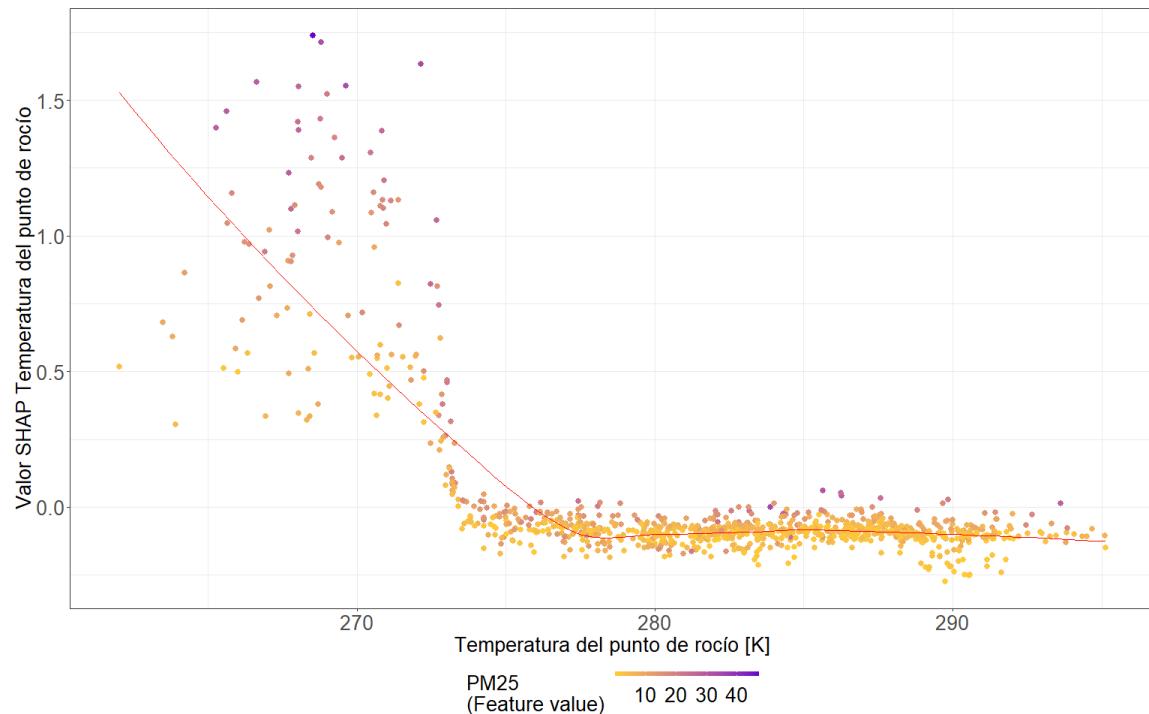


Figura 5.16: Gráfico de la variable temperatura del punto de rocío contra su valor SHAP coloreado por el valor de PM_{2.5} para el uso del suelo agrícola / área verde.

La Figura 5.16 muestra que valores de baja magnitud de la temperatura del punto de rocío (menores a 275 K) se asocian con altas concentraciones de PM_{2.5}, mientras que valores de elevada magnitud (mayores a 280 K) lo hacen con bajas concentraciones de PM_{2.5}. De esta forma se identifica una relación del tipo inversa similar a la reportada durante el estudio de la relación de vinculación entre ambas variables (Capítulo 4).

Tasa media de escorrentía subterránea

La tasa media de escorrentía subterránea, es la parte del agua procedente de las precipitaciones, del deshielo o de las napas, que escurre por debajo de la superficie

cie terrestre. El agua que no se almacena escurre, esto puede ser por la superficie (escorrentía superficial) o por debajo del suelo (escorrentía subsuperficial), la suma de ambas se denomina escorrentía. Este parámetro refiere a la tasa que tendría la escorrentía si se repartiera uniformemente por la cuadrícula de resolución espacial de 31 km x 31 km. La escorrentía es una medida de la disponibilidad de agua en el suelo y puede utilizarse, por ejemplo, como indicador de sequía o inundación. Se espera que la influencia de la variable en la concentración de PM_{2.5}, sea a través del proceso físico de refrigeración de la atmósfera, influyendo en los niveles de la capa límite atmosférica. Las unidades en las que se mide este parámetro son kilogramos por metro cuadrado por segundo ($\frac{Kg}{m^2.s}$). La resolución espacial utilizada proviene del reanálisis ERA5, que tiene una cuadrícula de resolución espacial de aproximadamente 0.25 grados (alrededor de 31 km).

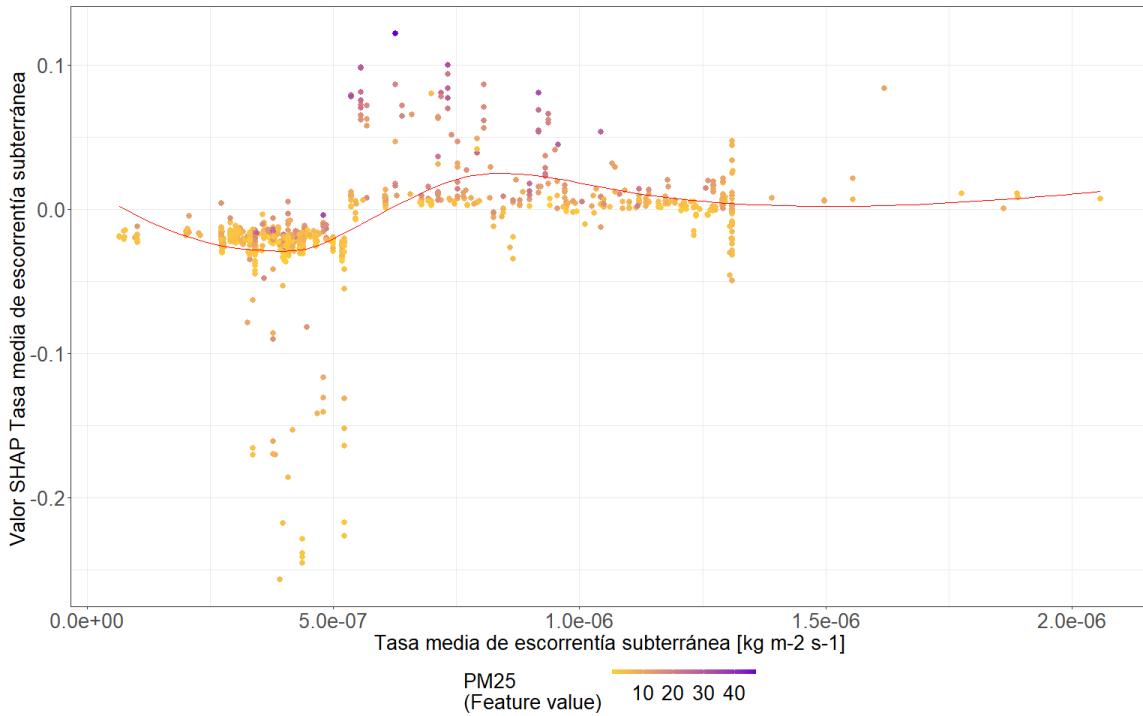


Figura 5.17: Gráfico de la variable tasa media de escorrentía subterránea contra su valor SHAP coloreado por el valor de PM_{2.5} para el uso del suelo agrícola / área verde.

La Figura 5.17 muestra que valores de baja magnitud de la tasa media de es-

correntía subterránea (menores a $5.07e-07 \frac{Kg}{m^2.s}$) favorecen bajas concentraciones de PM_{2.5}, mientras que valores superiores de la variable promueven un aumento en la concentración de PM_{2.5} predicha. Para valores superiores a $1.0e-06 \frac{Kg}{m^2.s}$, la relación entre las variables permanece sin variaciones importantes respecto del aporte a la concentración. De esta forma se identifica una relación del tipo directa similar a la reportada en el análisis de conexión entre las mismas (Anexo 5).

Integración vertical de temperatura

La integración vertical de temperatura es la integral vertical de la temperatura para una columna de aire ponderada por su masa, que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Este parámetro puede utilizarse para estudiar el balance energético atmosférico. Se espera que esta variable se vincule con la concentración de PM_{2.5}, por medio de la relación existente entre la variación de la temperatura a lo largo del espesor de la atmósfera y las estratificaciones que en esta se generan, resultando en aumentos o disminuciones de la concentración PM_{2.5}, en función de la altura a la cual ocurre la estratificación. Este fenómeno se encuentra vinculado también con la variación de la altura de la capa límite atmosférica. Las unidades en las que se mide este parámetro son en kilogramos por Kelvin por metro cuadrado ($\frac{Kg.K}{m^2}$).

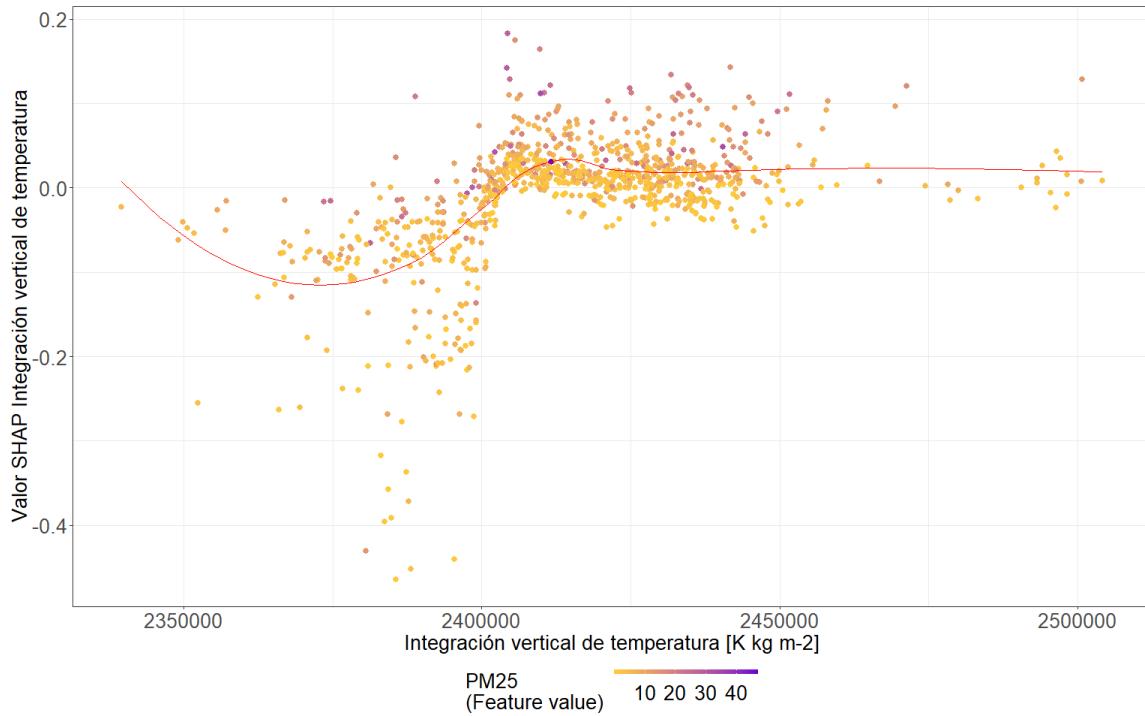


Figura 5.18: Gráfico de la variable integración vertical de temperatura contra su valor SHAP coloreado por el valor de PM_{2.5} para el uso del suelo agrícola / área verde.

La Figura 5.18 muestra que valores de baja magnitud de la integración vertical de temperatura (menores a 2400000 $\frac{K\text{kg}}{m^2}$) se asocian con bajas predicciones de PM_{2.5}, mientras que valores de elevada magnitud de la variable (mayores a 2450000 $\frac{K\text{kg}}{m^2}$) resultan en incrementos de las concentraciones de PM_{2.5} predichas. De esta forma se identifica una relación del tipo directa similar a la reportada en el análisis de conexión entre las mismas (Anexo 5).

5.3.4. Evaluación de tiempos de cómputo

En este estudio, se evaluaron los tiempos de cómputo requeridos para entrenar modelos de series temporales en diferentes usos de suelo: industrial, urbano, agrícola y áreas verdes. Cabe destacar que todos los entrenamientos se ejecutaron sobre la Unidad de Procesamiento Central (CPU), además es influyente sobre el tiempo de

computo la cantidad de variables y observaciones involucradas en cada una de las bases de datos (original y reducida). Los modelos se entrenaron utilizando Keras durante 200 épocas.

Uso de suelo industrial

Para el uso de suelo industrial, el modelo que demostró mayor desempeño fue el de GRU en sus dos versiones de conjuntos de datos: completo involucrando 193 variables y reducido con 53 entradas. Cada conjunto contenía 8,648 observaciones. El entrenamiento sobre la base de datos completa requirió aproximadamente 3 horas, mientras que el entrenamiento sobre la base de datos reducida se completó en alrededor de 1.5 horas. Esta reducción del tiempo de cómputo se debe a la menor cantidad de datos que el modelo debe procesar en cada iteración, lo que optimiza significativamente el tiempo de entrenamiento.

Uso de suelo urbano

En el caso del uso de suelo urbano, el modelo que demostró mayor desempeño fue el de GRU en sus dos versiones de conjuntos de datos. Se compararon una base de datos completa con 193 variables y una reducida con 77 entradas. Cada conjunto contenía 10,152 observaciones. El tiempo de entrenamiento sobre la base de datos completa fue de aproximadamente 3.5 horas, mientras que la base de datos reducida permitió completar el entrenamiento en 2 horas. Esta mejora refleja cómo la reducción de datos no solo ahorra tiempo sino que también puede facilitar el ajuste del modelo a datos más específicos y relevantes.

Uso de suelo agrícola y área verde

Para el uso de suelo agrícola y área verde, el modelo que demostró mayor desempeño predictivo fue el de *Bosques Aleatorios*. En este caso, se utilizaron bases de datos con

9,695 observaciones: la base de datos completa con 193 variables y una reducida con 84 entradas. Los tiempos de cómputo para el modelo de *Bosques Aleatorios* sobre la base completa fueron de aproximadamente 2 horas, mientras que para la base reducida el tiempo disminuyó a 1 hora. Este modelo, al ser menos complejo que un modelo de *GRU*, se benefició igualmente de la reducción de datos, mejorando la eficiencia del cómputo sin comprometer la precisión.

5.4. Discusión

Es interesante destacar que tanto para el uso del suelo urbano como para el industrial el algoritmo predictivo de mejor ajuste demostraron ser las redes neuronales recurrentes. Resultados similares fueron hallados por Ayturan et. al., 2020 evidenciando que las redes neuronales recurrentes, en especial las *GRU* destacan de entre múltiples algoritmos predictivos como las de mayor capacidad para predecir la concentración de PM_{2.5} (Ayturan et al., 2020). Además, otros estudios han demostrado la supremacía de las redes neuronales recurrentes para predecir la concentración de PM_{2.5} (Casallas et al., 2021; Dhakal et al., 2021; R. Feng et al., 2019). En esta tesis el algoritmo de mayor capacidad predictiva para el uso del suelo urbano alcanzó un *RMSE* de 3.46 $\frac{\mu\text{g}}{\text{m}^3}$ en el set de datos de testeo, mientras que para el uso del suelo industrial las redes neuronales recurrentes de mayor capacidad alcanzaron un *RMSE* de 5.96 $\frac{\mu\text{g}}{\text{m}^3}$ en el set de datos de testeo. Cabe destacar que el modelo predictivo sobre el uso de suelo urbano alcanzo un mejor desempeño que el desarrollado sobre el uso de suelo industrial, lo cual podría encontrarse asociado a la variabilidad propia de los datos en cada una de las series de tiempo.

Por otra parte, en el uso del suelo agrícola y área verde, el algoritmo predictivo que demostró mayor capacidad predictiva fue el de bosques aleatorios. Múltiples estudios han expuesto resultados coincidentes con los informados en este trabajo, destacando

el algoritmo de bosques aleatorios como el de mayor capacidad para predecir la concentración de PM_{2.5} (G. Chen et al., 2018; Huang et al., 2018; Suleiman et al., 2020; Xu et al., 2021). En el presente trabajo la mayor capacidad predictiva para este uso del suelo reportó un *RMSE* de 2.32 $\frac{\mu\text{g}}{\text{m}^3}$ en el set de datos de testeo.

Los resultados de ajuste obtenidos para los distintos algoritmos predictivos ajustados provenientes tanto del contexto de aprendizaje automático como profundo demuestran que el mejor ajuste varía en función del uso del suelo analizado. Este resultado podría encontrarse vinculado no solo a las características propias de cada una de las series temporales, en cada uno de los usos de suelo descriptos en el capítulo 3, si no también a las de aprendizaje de los algoritmos empleados en el modelado predictivo.

Para el caso de la serie de temporal proveniente del uso del suelo agrícola y área verde, se observó que los datos de concentración de PM_{2.5} variaban en un entorno reducido respecto de su media. Esto indica una baja dispersión de los datos, lo que se traduce en una baja frecuencia de fluctuación entre valores de elevada y baja magnitud. Esta característica permite que la variable respuesta pueda ser explicada por un menor número de variables predictoras (Figura 5.15). Cabe destacar que más del 40 % de la importancia relativa es acumulada por las cinco variables de mayor importancia para el modelo predictivo. En un set de datos que alberga un total de 70 variables, es destacable que prácticamente el 7 % de las mismas concentren la información necesaria para describir la variable respuesta. Esta concentración de la información podría encontrarse sustentada en que al ser una serie de baja variabilidad se requiere menor información para describir sus fluctuaciones, ya que pueden ser explicadas la mayor parte del tiempo por las mismas variables predictoras.

La concentración de información en un número reducido de variables podría favorecer el algoritmo de bosques aleatorios. Los bosques aleatorios resultaron eficientes detectando y utilizando la información de las variables de mayor importancia en la

base de datos. Una vez detectadas, estas podrían haber sido retenidas en múltiples árboles individuales del bosque, generando una influencia de mayor peso en las predicciones. Dado al gran porcentaje de información contenido en una cantidad reducida de variables, los bosques aleatorios podrían haber logrado internalizar de forma efectiva las interrelaciones existentes entre las distintas variables, lo que resultó en un modelo predictivo de elevada exactitud. Resultados coincidentes fueron publicados por Pengcheng et. al., y por Wang et. al., quienes mostraron que el algoritmo de bosques aleatorios aumenta su capacidad predictiva al ser entrenado en bases de datos previamente reducidas por medio de algoritmos de selección de características (Pengcheng et al., 2020; Y. Wang et al., 2021).

Por otra parte, es importante destacar que las series provenientes de entornos urbanos e industriales presentan una notoria dispersión en sus realizaciones. Esto implica que sus datos son menos homogéneos en comparación con la serie agrícola y de área verde. Las series urbana e industrial muestran continuas fluctuaciones entre valores de elevada y baja magnitud de concentración de PM_{2.5}. Por este motivo, a diferencia de la serie temporal agrícola y área verde, en estos caso se requiere un número de variables superior para reunir el mismo porcentaje de importancia relativa (Figuras 5.6 y 5.10). Es importante destacar que la importancia relativa otorgada a cada una de las variables participantes en los modelos para cada uno de estos usos de suelo varía en un rango acotado. A diferencia del uso del suelo agrícola y área verde, no hay variables en ninguno de los restantes usos de suelo que demuestren supremacía en cuanto a su importancia relativa. Los resultados muestran que las primeras cinco variables explican menos del 25 % de la importancia relativa para los usos de suelo de mayor variabilidad.

Los algoritmos de redes neuronales de mejor ajuste para los usos de suelo urbano e industrial no identificaron variables predictoras de elevada importancia relativa, a diferencia de lo que sucede con los bosques aleatorios en el uso del suelo agrícola y

área verde. En las redes neuronales, todas las variables aportaron de forma homogénea a las predicciones, siendo el algoritmo predictivo el encargado de internalizar las ponderaciones de las mismas. En las figuras 5.6 y 5.10, se presentaron los porcentaje de importancia relativa para los usos de suelo urbano e industrial respectivamente, los cuales evidenciaron ser inferiores a los del uso del suelo agrícola y área verde.

Las características de la serie temporal urbana e industrial podrían favorecer el modelado predictivo a través de redes neuronales recurrentes, cuyo algoritmo posee numerosos parámetros que logran reflejar el aporte de cada una de las variables a la predicción bajo las distintas situaciones posibles presentes en la base de datos. Al calcularse una gran cantidad de parámetros, podría notarse reflejado en cada predicción la influencia de cada una de las variables presentes en la base de datos, siendo el algoritmo predictivo el encargado de ponderar las influencias en cada caso particular. Al tratarse de series en donde la importancia relativa atribuida a las distintas predictoras es similar, el contar con un algoritmo predictivo que tenga en cuenta la influencia conjunta de todas las variables podría resultar en un mayor desempeño predictivo. De esta forma en cada predicción se vería reflejada la influencia de cada una de las variables predictoras. Numerosos estudios han reflejado de forma similar la mejoría en la capacidad predictiva de la concentración de PM_{2.5} experimentada al aplicar redes neuronales recurrentes (Ding & Zhu, 2022 ; Ayturan et al., 2020 ; Y. Kim et al., 2023 ; Casallas et al., 2021; Dhakal et al., 2021; R. Feng et al., 2019).

La selección de bases de datos reducidas se justifica no solo por la significativa reducción en los tiempos de cómputo, sino también por la capacidad de los modelos de generalizar mejor cuando se entrena con datos más específicos y manejables. La reducción del tiempo de computo hasta en la mitad comparando el entrenamiento en ambas bases de datos permite establecer una optimización de recursos computacionales. Esta reducción de variables facilita la iteración en el ajuste y la validación del modelo.

En la presente tesis se observó que el modelado en áreas industriales presentó que

tres de las variables de mayor relevancia no habían sido relacionadas previamente con la contaminación del aire por otros estudios. En primer lugar, la inclusión de estas variables se encuentra fundamentada en el enfoque de modelado computacional que se persigue en este trabajo, a diferencia de uno mecanicista. En un enfoque computacional, la selección de variables predictoras se basa en relaciones subyacentes que pueden no ser evidentes en la realidad observable. La ausencia de estas relaciones en estudios previos no descalifica su inclusión en el modelo. Es importante destacar que, aunque la correlación no implica causalidad, las variables que muestran una fuerte correlación con la variable de respuesta contienen un porcentaje significativo de información relevante para el modelo predictivo. Este enfoque permite capturar información útil que puede no estar comprendida en un marco teórico convencional. Además, las bases de datos empleadas fueron reducidas utilizando múltiples técnicas de selección de características, lo que asegura que la información contenida en una variable no es redundante en otras. Esta metodología fortalece la hipótesis de que la inclusión de estas variables en el modelo es crucial para mejorar su desempeño predictivo. La reducción de la base de datos mediante estas técnicas asegura que solo se mantengan las variables con mayor capacidad explicativa, optimizando así la eficiencia y precisión del modelo.

Para el uso de suelo urbano, las variables derivadas de información satelital más importantes para describir la variabilidad de la concentración de PM_{2.5} fueron la radiación solar, la rugosidad de la superficie terrestre, y la presión atmosférica. En áreas urbanas, la radiación solar juega un papel crucial debido a la presencia de numerosos edificios y calles pavimentadas que absorben y retienen el calor, creando “islas de calor”. Estas islas de calor pueden influir significativamente en la dispersión de contaminantes, aumentando la concentración de PM_{2.5} debido a la reducción de la dispersión vertical en condiciones de alta temperatura (H. Wu et al., 2024). La rugosidad de la superficie terrestre es determinante en la dispersión de contaminantes

en entornos urbanos. La presencia de edificios, árboles, vehículos y otros objetos aumenta la turbulencia del aire, lo que puede afectar la dispersión horizontal y vertical de PM_{2.5}. Un mayor nivel de rugosidad puede reducir la velocidad del viento a nivel del suelo, limitando la capacidad del aire para dispersar los contaminantes (Yan et al., 2021). La presión atmosférica influyó notablemente en la dispersión y concentración de contaminantes en áreas urbanas. Generalmente, una baja presión atmosférica está asociada con condiciones climáticas inestables y un aumento en la capacidad de dispersión de los contaminantes. Sin embargo, en condiciones de alta presión, la dispersión se reduce, resultando en mayores concentraciones de PM_{2.5} (Yan et al., 2021).

En el uso de suelo industrial, las variables derivadas de información satelital más importantes para describir la variabilidad de la concentración de PM_{2.5} fueron la velocidad del viento, la disipación en la capa límite atmosférica, el flujo de ozono, y la humedad relativa. En áreas industriales, la velocidad del viento es un factor crítico para la dispersión de contaminantes debido a la menor densidad poblacional y reducida presencia de estructuras de gran altura que impidan las corrientes de viento. Esto facilita la dispersión de PM_{2.5}, reduciendo las concentraciones locales (Yan et al., 2021; H. Zhao et al., 2022). En áreas industriales, una capa límite bien desarrollada y con alta disipación puede aumentar la mezcla y dispersión de PM_{2.5}, reduciendo así su concentración. El flujo de ozono puede estar relacionado con las emisiones industriales específicas y las reacciones fotoquímicas que ocurren en estas áreas. El ozono puede influir en la formación secundaria de partículas finas, afectando las concentraciones de PM_{2.5}. Además, en áreas industriales, una mayor humedad puede promover la aglomeración de partículas y su eventual eliminación a través de la deposición húmeda.

En el uso de suelo agrícola y área verde, las variables derivadas de información satelital más importantes para describir la variabilidad de la concentración de PM_{2.5} fueron la temperatura del rocío, la escorrentía, y la temperatura. La temperatura

del rocío es un indicador de la cantidad de humedad en el aire y puede influir en la formación de niebla y la deposición de partículas. En áreas agrícolas, donde no hay terrenos pavimentados, la humedad del suelo y la formación de rocío pueden jugar un papel importante en la deposición húmeda de PM_{2.5} (Yan et al., 2021; H. Zhao et al., 2022). La escorrentía es influyente porque en áreas agrícolas y verdes, una gran parte de la lluvia se infiltra en el suelo, contribuyendo a la refrigeración del aire y a la deposición húmeda de partículas. Este proceso puede reducir las concentraciones de PM_{2.5} al eliminar partículas del aire. La temperatura influye en la formación de capas de inversión térmica, que pueden atrapar contaminantes cerca del suelo, aumentando las concentraciones de PM_{2.5}. En áreas agrícolas y verdes, las variaciones de temperatura pueden afectar significativamente la dispersión y eliminación de contaminantes (H. Wu et al., 2024).

La relación directa entre la rugosidad de la superficie terrestre y la concentración de PM_{2.5} podría estar respaldada por el hecho de que la presencia de objetos y estructuras en la superficie, tales como edificios y árboles, puede obstaculizar el flujo de aire y dificultar la dispersión de los contaminantes. Como resultado, un aumento en la rugosidad de la superficie podría llevar a un aumento en las concentraciones de PM_{2.5}. En contraste, superficies terrestres con baja rugosidad permiten un flujo libre de las corrientes de aire, lo que facilita la dispersión de los contaminantes. Todos los coeficientes estadísticamente significativos exhiben una relación de vinculación directa entre la rugosidad de la superficie terrestre y la concentración de PM_{2.5}.

Investigaciones previas han demostrado una correlación positiva entre la concentración de ozono troposférico y la concentración de PM_{2.5} (Kalashnikov et al., 2022; Qu et al., 2022), la cual también ha sido constatada en este trabajo. Aunque la variable utilizada en este estudio indica la columna total de ozono, es importante destacar que esta incluye el ozono troposférico. Dado que el ozono troposférico constituye una parte significativa de la columna total de ozono, la correlación observada se mantiene válida.

Según lo señalado por Kalashnikov, esta relación podría ser atribuida a incendios forestales o eventos meteorológicos específicos que contribuyen al aumento simultáneo de ambos contaminantes (Kalashnikov et al., 2022). Durante los incendios forestales, se liberan grandes cantidades de gases y partículas al ambiente. Entre estos gases se encuentran los compuestos orgánicos volátiles (COV), que reaccionan con los óxidos de nitrógeno (NO_x) para producir ozono. Además, el ozono puede interactuar con las partículas en la atmósfera para formar compuestos químicos más complejos, como los llamados nitratos orgánicos, que pueden contribuir al aumento de la concentración de partículas y ozono en áreas distantes de los incendios forestales. El ozono troposférico es un oxidante fotoquímico que puede reaccionar con precursores del material particulado, tales como los compuestos orgánicos volátiles y los óxidos de nitrógeno, para formar partículas secundarias. Por lo tanto, el aumento en la concentración de ozono podría estar relacionado con un incremento en la concentración PM_{2.5} (Qu et al., 2022). El coeficiente estadísticamente significativo detectado se obtuvo considerando un *lag* temporal de 19 horas demostrando un vínculo directo. Este coeficiente refleja que los factores que relacionan la concentración de ozono y PM_{2.5}, como las múltiples reacciones químicas en cadena, requieren intervalos temporales amplios para que sus efectos se manifiesten y contribuyan a las concentraciones observadas.

Los hallazgos de este estudio sugieren una relación directa entre la tasa media de escorrentía subterránea y la concentración de PM_{2.5}. Esta relación podría encontrarse vinculada al hecho de que un aumento en la tasa de escorrentía podría indicar una mayor disponibilidad de agua en la superficie terrestre. Esta presencia de agua podría tener un efecto refrigerante, reduciendo la LST y dificultando la generación de corrientes verticales de dispersión de contaminantes. Por otro lado, una tasa de escorrentía baja podría indicar una situación de sequía lo que facilitaría la generación de corrientes ascendentes de masas de aire, generando un aumento en la dispersión de los contaminantes (Chauhan et al., 2022; Z. Ma et al., 2023; Zhou et al., 2024).

Por otra parte, además del efecto refrigerante, las precipitaciones son conocidas por favorecer la deposición húmeda de partículas del aire, lo que reduce temporalmente las concentraciones de PM_{2.5} al eliminar las partículas a través del agua de lluvia. Múltiples estudios han demostrado que la precipitación actúa como un mecanismo de limpieza atmosférica, disminuyendo las concentraciones de contaminantes durante y poco después de los eventos de lluvia (Lim et al., 2020; Zhou et al., 2024). El coeficiente estadísticamente significativo se encontró considerando un *lag* de 15 horas demostrando un vínculo directo, sugiriendo que la influencia del agua contenida en la superficie y el efecto refrigerante de esta requiere de un desfasaje temporal de esta magnitud para tener efectos significativos sobre la concentración de PM_{2.5}. Esto podría deberse a que, durante las precipitaciones, la radiación solar incidente tiende a ser baja, y al cabo de 15 horas, la radiación vuelve a incidir en la superficie de manera significativa desencadenando efectos de refrigeración de la atmósfera que disminuyen la concentración del PM_{2.5}. Este fenómeno se relaciona con la influencia de la radiación solar en la temperatura superficial. Un aumento en la radiación solar lleva a un calentamiento de la superficie terrestre y el aire circundante, lo que provoca corrientes ascendentes que dispersan las partículas finas en la atmósfera, reduciendo su concentración. En contraste, durante períodos de baja radiación solar, las partículas pueden acumularse debido a la falta de dispersión vertical. La nubosidad ya se ha tenido en cuenta en este proceso. Por lo tanto, la radiación solar directa influye en la concentración de PM_{2.5} en la atmósfera. Cabe mencionar que, efectivamente, las precipitaciones favorecen la deposición húmeda de partículas del aire, un proceso conocido que reduce temporalmente las concentraciones de PM_{2.5} en la atmósfera al eliminarlas a través del agua de lluvia (Seinfeld & Pandis, 1998). Sin embargo, en este estudio se ha observado que, tras un período de 15 horas post-precipitación, la radiación solar vuelve a incidir en la superficie terrestre de manera significativa, desencadenando procesos de refrigeración de la atmósfera que influyen en la concentración de PM_{2.5}.

Este fenómeno sugiere que, aunque la deposición húmeda inicialmente reduce las concentraciones de partículas, la influencia de la radiación solar y las condiciones meteorológicas posteriores juegan un papel crucial en la dinámica y distribución temporal de PM_{2.5} en la atmósfera.

La integración vertical de la temperatura es un parámetro esencial para comprender el balance energético en la atmósfera. Esta refiere a la integral vertical de la temperatura ponderada por la masa en una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera, desempeña un papel esencial en la comprensión del balance energético en la atmósfera. Esta variable nos permite analizar cómo la temperatura varía a lo largo de la altura en la atmósfera y cómo estas variaciones están relacionadas con las estratificaciones que se forman. Estas estratificaciones pueden llevar a aumentos o disminuciones en la concentración de PM_{2.5}, dependiendo de la altura a la que se produzcan. Además, la integración vertical de la temperatura está vinculada con la variación en la altura de la capa límite atmosférica. Se mide en kilogramos por Kelvin por metro cuadrado ($\frac{Kg.K}{m^2}$) y su estudio es fundamental para comprender los efectos en la concentración de PM_{2.5} en la atmósfera. Los resultados de este estudio demostraron una relación directa estadísticamente significativa con la concentración de PM_{2.5}. Aumentos de la integración vertical de la temperatura podrían favorecer la formación de inversiones térmicas, atrapando los contaminantes cerca de la superficie terrestre y, por lo tanto, favoreciendo el aumento de la concentración de PM_{2.5}. Los coeficientes significativos detectados se presentaron considerando un *lag* temporal de 16 y 18 horas, reflejando una relación directa. De esta forma, podrían sugerir que la influencia de esta variable requiere un desfasaje temporal prolongado para incidir en la concentración de PM_{2.5}. Esto podría deberse a la complejidad que vincula la relación entre la variable en cuestión y el PBLH. Por otra parte, el coeficiente detectado para el desfasaje temporal de 15 horas refleja un vínculo inverso que contradice los resultados propuestos y aún

podría encontrarse influenciado por efectos espuria.

La disipación en la capa límite atmosférica refiere a la capacidad de esta capa para disipar la energía mecánica de las corrientes de aire, lo cual afecta la dispersión de contaminantes en la atmósfera. Aumentos en la disipación podrían expresar condiciones favorables para la ocurrencia de la dispersión de contaminantes, como por ejemplo el aumento de remolinos turbulentos cerca de la superficie, reduciendo la concentración de partículas. Por el contrario, valores bajos de disipación en la capa límite atmosférica podrían vincularse a condiciones favorables para el aumento de la concentración de contaminantes y, en consecuencia, el aumento de la concentración de PM_{2.5}. Esta variable se encuentra influenciada por diversos factores como la rugosidad de la superficie, la velocidad del viento y la estabilidad atmosférica, por lo que su interacción puede estar sujeta también a la fluctuación de otras variables (H. Ma et al., 2022). Los coeficientes estadísticamente significativos encontrados para la disipación reflejan múltiples desfasajes temporales que la relacionan a través de un vínculo inverso con la concentración de PM_{2.5}. Por una parte, el coeficiente sin considerar un *lag* temporal sugiere que algunos de los procesos que involucra la disipación en la capa límite atmosférica influencian de forma inmediata la concentración y dispersión de contaminantes en la atmósfera. Mientras que, los desfasajes restantes, de mayor *lag* temporal (20 y 22 horas) demuestran que la variable también incluye procesos que efectivizan su influencia al cabo de grandes intervalos de tiempo.

5.5. Conclusiones

Los resultados obtenidos reflejan que los modelos predictivos desarrollados logran explicar la variabilidad de la concentración de PM_{2.5} a partir de variables derivadas a partir de información satelital y temporales. Estos modelos también han demostrado una interpretación acertada de las relaciones subyacentes entre la concentración de

PM_{2.5} y las variables climáticas.

La similitud existente entre las métricas obtenidas para el modelo de referencia y las de los modelos predictivos desarrollados evidencian su potencial para cumplir con el objetivo de alertar a la población ante condiciones adversas de calidad de aire con cinco días de anticipación.

Las características propias de la variabilidad de cada una de las series temporales en los distintos usos de suelo, así como las características de aprendizaje de cada uno de los modelos predictivos ajustados, fueron determinantes en el tipo de algoritmo que demostró mayor exactitud en la predicción para cada uno de los usos de suelo. En el uso del suelo de menor variabilidad, el algoritmo de bosques aleatorios se desempeñó mejor que los algoritmos de aprendizaje profundo. Mientras que los de aprendizaje profundo permitieron ajustar modelos de mayor capacidad predictiva en usos de suelo donde la serie de PM_{2.5} se caracterizaba por una mayor variabilidad.

Comentarios finales

Este apartado sintetiza los aspectos más relevantes de la tesis, destacando los procedimientos realizados, los resultados obtenidos y las discusiones que han proporcionado información novedosa. Su objetivo principal esta orientado a dar respuesta a los objetivos específicos planteados al principio de esta tesis, así como al objetivo general, que busca desarrollar un modelo predictivo de concentración de aerosoles para el Área Metropolitana de Córdoba utilizando datos de sensores de bajo costo, variables derivadas de información satelital de acceso público, variables meteorológicas, variables temporales y variables relacionadas al uso del suelo, y validar el empleo de estos sensores en áreas urbanas.

El proceso de ensamblaje y calibración de los sistemas de monitoreo de bajo costo, descrito en detalle en el capítulo 2, fue crucial para asegurar la integración efectiva y funcional de los sensores. Este meticuloso proceso incluyó la conexión de sensores a la placa Arduino mediante cables pelados y soldados con mercurio líquido, garantizando uniones firmes y conductoras al solidificarse. La carga del *firmware* se realizó directamente desde una computadora, estableciendo una configuración robusta y fiable del sistema de monitoreo. Los sistemas de bajo costo mostraron una alta consistencia en las mediciones de PM_{2.5}, comparables con los datos de la estación de referencia de la ciudad, sugiriendo que estos sistemas pueden ser empleados eficazmente para monitorear la calidad del aire en zonas urbanas.

La variación espacio-temporal de la concentración de aerosoles en el área metropo-

litana de Córdoba, explorada en el capítulo 3, reveló que, aunque las concentraciones de PM_{2.5} en Córdoba son menores que en otras ciudades del mundo, los estándares de calidad del aire locales fueron superados en múltiples ocasiones. La magnitud de la concentración promedio anual de PM_{2.5} encontrada para la provincia de Córdoba (30/03/2021-30/03/2022) fue inferior a la observada en otras ciudades del mundo. Por ejemplo, en Riohacha, Colombia (188.014 habitantes), se registró una concentración de 14.57 $\frac{\mu\text{g}}{\text{m}^3}$ (04/11/2010-02/02/2011) (Rojano et al., 2013). En Brasil, ciudades como Río de Janeiro (15.5 $\frac{\mu\text{g}}{\text{m}^3}$) (6.320.446 habitantes) (01/06/2007-01/08/2008) y Porto Alegre (13.9 $\frac{\mu\text{g}}{\text{m}^3}$) (1.420.667 habitantes) (01/10/2001-01/12/2002) también presentaron niveles superiores a los observados en la provincia de Córdoba (Miranda et al., 2012). En Serbia, se reportaron concentraciones de promedios anuales de entre 13.93 $\frac{\mu\text{g}}{\text{m}^3}$ y 28.91 $\frac{\mu\text{g}}{\text{m}^3}$ para el periodo comprendido entre los años 2001 y 2016 (Stanojević et al., 2019) (6.834.000 habitantes). Por último, en Estambul, Turquía, se registraron concentraciones de promedios anuales de entre 47.8 $\frac{\mu\text{g}}{\text{m}^3}$ y 93.0 $\frac{\mu\text{g}}{\text{m}^3}$ para el periodo comprendido entre 2007 y 2008, en distintas regiones de la ciudad (Rasheed et al., 2015) (15.019.000 habitantes). Sin embargo, los estándares de calidad de aire establecidos por las normativas de la OMS y la EPA fueron superados en múltiples ocasiones tanto en la escala anual como de promedio de 24 horas. Al segmentar los datos en función del uso del suelo característico del patrón de emisión en cada sitio de muestreo, se observó una marcada correspondencia entre la frecuencia de superación de estos límites y las particularidades de cada uno de los clústeres identificados (agricultura/área verde, urbano e industrial).

El análisis de las series temporales de PM_{2.5} mostró patrones de emisión característicos de cada uso del suelo, con concentraciones más altas en zonas industriales en comparación con áreas urbanas y agrícolas/áreas verdes. La distribución *gamma* se ajustó mejor a los datos de concentración de PM_{2.5}, minimizando los valores del criterio de información de Akaike (*AIC*) y del criterio de información Bayesiano (*BIC*),

y proporcionando un modelo teórico que explicó adecuadamente la variabilidad observada. Tanto las características temporales como espaciales de estas series demostraron potencial para explicar la variabilidad de los niveles de PM_{2.5}, convirtiéndolas en una valiosa entrada para la construcción de modelos predictivos. Además, se observaron diferencias notables en las características de las series de tiempo de concentración de PM_{2.5} en función del uso del suelo, resaltando la importancia de considerar estas variaciones al analizar y abordar la calidad del aire.

Por otra parte, el análisis exploratorio resaltó el potencial que podrían tener las variables estudiadas para explicar la variabilidad de la concentración de PM_{2.5}. En busca de cumplir con el objetivo de ajustar modelos predictivos, se recopilaron numerosas variables que pudieran reflejar las características de la atmósfera en un momento determinado. La base de datos final incorporó un total de 193 variables, a partir de la cual se exploraron distintas técnicas de modelado predictivo. Aunque el enfoque principal del modelado utilizado en este trabajo se basa en técnicas de aprendizaje automático, que buscan maximizar la capacidad predictiva a partir de patrones en los datos, sin requerir una comprensión explícita de las relaciones causales subyacentes entre las variables, como lo es en el caso de los modelos mecanicistas (Breiman, 2001b; Palmer & Hagedorn, 2006), no se busca establecer una distinción rígida entre enfoques computacionales y mecanicistas. Si bien en los enfoques mecanicistas las relaciones entre variables están predefinidas en función de principios teóricos y conocimiento previo sobre el fenómeno, en nuestro enfoque, muchas de las variables incluidas, como las relacionadas con el clima, también se basan en un entendimiento profundo de los procesos atmosféricos. Esto permite que el modelo aproveche tanto la disponibilidad de datos como las capacidades de inferencia automática que ofrecen los algoritmos, sin desvincularse por completo de los fundamentos científicos. Así, el enfoque computacional basado en aprendizaje automático se complementa con el conocimiento mecanicista, al permitir que el modelo explore relaciones adicionales

de manera más flexible, potenciando la capacidad predictiva. Existen sofisticados modelos mecanistas como el *Regional Air Quality Deterministic Prediction System (RAQDPS)* desarrollado por *Environment Canada* que ha demostrado alta eficacia en la predicción de concentraciones de PM_{2.5} en diferentes regiones (Moran et al., 2018). Este sistema integra la meteorología y la química atmosférica para proporcionar predicciones precisas de la concentración de partículas hasta 72 horas antes, permitiendo alertar a la población ante condiciones adversas de calidad de aire y detectar incendios forestales de manera temprana. Además, existen numerosos estudios que han implementado modelos predictivos utilizando diferentes enfoques y técnicas para la predicción horaria de la concentración de PM_{2.5}. Chen et al., (2023) desarrollaron un modelo híbrido *CNN-RF* (*Convolutional Neural Network - Random Forest*) que mostró un rendimiento robusto en la predicción de la concentración de PM_{2.5}, logrando un *RMSE* de 3.67 $\frac{\mu\text{g}}{\text{m}^3}$ para la hora siguiente (M.-H. Chen et al., 2023). Por otra parte, Zhong et al., (2021) comparó un modelo de red neuronal profunda con el *Community Multiscale Air Quality (CMAQ)* y encontró que el modelo de red neuronal proporcionaba mejores predicciones horarias de PM_{2.5} con un RMSE de 19.19 $\frac{\mu\text{g}}{\text{m}^3}$ para la hora siguiente en Beijin (Zhong et al., 2021). Estos estudios subrayan la capacidad de los modelos de aprendizaje profundo para captar la complejidad de los datos derivados de información satelital y ofrecer predicciones precisas a partir de una estrategia de modelado mecanicista.

Las variables derivadas a partir de información satelital demostraron potencial para explicar la variabilidad espacio temporal de la concentración de PM_{2.5}, como se detalla en el capítulo 4. Estas variables, que forman un sistema dinámico e interdependiente, no actúan de manera aislada sino que su dinámica conjunta determina los patrones de dispersión de contaminantes atmosféricos. El análisis exploratorio de la covariación entre estas variables y la concentración de PM_{2.5} justificó su inclusión en la base de datos para el desarrollo de modelos predictivos, resaltando su potencial para explicar la variabilidad espacio-temporal de la concentración de aerosoles en el área de estudio.

Su inclusión en el modelado predictivo asegura la captura de la variabilidad de la variable objetivo y la obtención de estimaciones precisas. La similitud existente entre las métricas obtenidas para el modelo de referencia y las de los modelos predictivos desarrollados indican que estos poseen potencial para cumplir el objetivo de alertar a la población ante condiciones adversas de calidad de aire con cinco días de anticipación.

Los resultados obtenidos en el capítulo 5 indican que los modelos predictivos desarrollados demuestran potencial para alertar a la población ante condiciones adversas de calidad de aire con cinco días de anticipación. Estos modelos demostraron ser capaces de interpretar adecuadamente las relaciones subyacentes entre la concentración de PM_{2.5} y las variables derivadas de información satelital. La similitud entre las métricas obtenidas para el modelo de referencia y las de los modelos predictivos desarrollados sugiere que estos poseen el potencial para cumplir con el objetivo de alerta temprana. Además, se observó que las características propias de la variabilidad de cada una de las series temporales en los diferentes usos de suelo, junto con las características de aprendizaje de cada uno de los modelos predictivos ajustados, influyeron significativamente en el rendimiento de los algoritmos. Para el uso del suelo agrícola y área verde (menor variabilidad) el algoritmo de bosques aleatorios demostró mayor capacidad predictiva. Mientras que en el uso del suelo urbano e industrial (mayor variabilidad) los de aprendizaje profundo permitieron ajustar modelos de mayor capacidad predictiva.

5.6. Repuesta a los objetivos

5.6.1. Objetivo general

El objetivo general de esta tesis fue desarrollar un modelo predictivo de concentración de aerosoles para el área metropolitana de la provincia de Córdoba a partir de datos relevados por sensores de bajo costo, en función de variables derivadas a partir de información satelital de acceso público, variables meteorológicas, variables

temporales y variables relacionadas al uso del suelo, y validar el empleo de sensores de bajo costo para determinar partículas finas en áreas urbanas.

5.6.2. Objetivos específicos

1. Describir la variación espacio temporal de la concentración de aerosoles colectados con sensores de bajo costo en zonas del Área Metropolitana de Córdoba, Argentina.
 - Conclusiones capítulo 2 y 3: La implementación de sistemas de monitoreo de bajo costo demostró ser efectiva y funcional, proporcionando datos consistentes con las mediciones de estaciones de referencia. Se observó que la concentración de PM_{2.5} en Córdoba fue menor que en otras ciudades del mundo, aunque se superaron los estándares locales en varias ocasiones. Las características temporales y espaciales de la concentración de PM_{2.5} permitieron explicar la variabilidad observada, con diferencias significativas en función del uso del suelo (industrial, urbano y agrícola/área verde).
2. Evaluar la utilidad de sensores de bajo costo para el estudio de aerosoles en zonas urbanas del Área Metropolitana de Córdoba, Argentina.
 - Conclusiones capítulo 2: Los sensores de bajo costo mostraron una alta consistencia en sus mediciones y una correlación significativa con las estaciones de referencia, validando su utilidad para el monitoreo de la calidad del aire en áreas urbanas de Córdoba. Estos sistemas representan una alternativa viable y económica para la gestión y control de la calidad del aire.
3. Analizar la influencia de distintos tipos de variables obtenidas a través de satélites, entre ellas meteorológicas, en los patrones de variación temporal de aerosoles en el Área Metropolitana de Córdoba, Argentina.

- Conclusiones capítulo 4: Las variables meteorológicas y derivadas de información satelital demostraron tener un impacto sustancial en la concentración de PM_{2.5}. Estas variables, integradas en un sistema dinámico e interdependiente, explican la variabilidad espacio-temporal de la concentración de PM_{2.5} y son fundamentales para el desarrollo de modelos predictivos precisos.
4. Validar el empleo de productos satelitales, tales como AOD (del inglés *Aerosol Optical Depth*), para estimar la concentración de material particulado en el Área Metropolitana de Córdoba, Argentina, a través de análisis estadísticos.
 - Conclusiones capítulo 4: El análisis de la covariación entre variables satelitales y la concentración de PM_{2.5} confirmó la validez de estos productos para estimar la concentración de material particulado. Este análisis justificó su inclusión en la base de datos para el desarrollo de modelos predictivos. 5. Proponer un modelo predictivo que incluya variables de uso del suelo, satelitales y temporales para la estimación de la concentración de aerosoles en el Área Metropolitana de Córdoba, Argentina.
 - Conclusiones capítulo 5: Los modelos predictivos desarrollados lograron explicar la variabilidad de la concentración de PM_{2.5} utilizando variables satelitales y temporales. Los modelos demostraron una capacidad adecuada para alertar a la población con anticipación sobre condiciones adversas de calidad del aire, con una precisión variable según el tipo de uso del suelo y el algoritmo empleado.

5.6.3. Conclusión Final

La presente tesis ha demostrado que los sensores de bajo costo son herramientas eficaces para monitorear la concentración de PM_{2.5} en áreas urbanas. La integración de variables satelitales y meteorológicas ha permitido desarrollar modelos predictivos

precisos, capaces de alertar a la población sobre condiciones adversas de calidad del aire con suficiente antelación. Los resultados obtenidos no solo validan el empleo de sensores de bajo costo, sino que también destacan la importancia de considerar la variabilidad espacio-temporal y las características del uso del suelo al desarrollar estrategias de gestión de la calidad del aire. La tesis cumple con el objetivo general y los objetivos específicos planteados, aportando soluciones prácticas y conocimiento científico relevante para el monitoreo y la gestión de la calidad del aire en la provincia de Córdoba.

Futuras líneas de investigación

Se propone como futuro lineamiento de continuación de esta tesis el estudio de la variabilidad conjunta de la concentración de PM_{2.5} con respecto a cada una de las variables derivadas a partir de información satelital con el objetivo de identificar cuáles son las que más covarían con la concentración de este contaminante. Sería interesante considerar también las correlaciones entre las potenciales variables explicativas para caracterizar el sistema de correlaciones y dependencias, como se hace en un análisis de sendero o modelo de ecuaciones estructurales.

También se promueve hacer hincapié en la comparación del desempeño de los modelos predictivos de concentración de PM_{2.5} implementados en esta tesis con otros algoritmos predictivos entrenados en el contexto de aprendizaje automático que aprovechen datos de imágenes satelitales o bien la combinación de datos satelitales y meteorológicos tabulares con imágenes.

Además, se podría implementar un procedimiento de simulación para verificar los resultados obtenidos en la presente tesis. Este tendría por objetivo simular series con distinto grado de variabilidad (alto y bajo), además de múltiples variables predictoras. Una vez llevada a cabo la simulación proceder al ajuste de los modelos predictivos evaluados en la presente tesis, para confirmar las conclusiones obtenidas a partir de los datos observados desde las obtenidas con los datos simulados. Una conclusión de la presente tesis a confirmar es la que establece que el modelado predictivo de series temporales con características de gran variabilidad y fluctuaciones entre sus valores

es mejor de ser realizado a través de redes neuronales recurrentes; Mientras que el modelado predictivo de series temporales de baja variabilidad y fluctuaciones es mejor bajo el algoritmo de bosques aleatorios.

Además, se podría implementar un procedimiento de simulación para verificar los resultados obtenidos en la presente tesis. Este tendría por objetivo simular series temporales con distintos grados de variabilidad (alto y bajo) y múltiples variables predictoras. La simulación consistiría en generar datos sintéticos que imiten las características de las series temporales observadas, permitiendo un control más riguroso sobre su variabilidad y las relaciones entre variables. Una vez llevada a cabo la simulación, se procedería al ajuste de los modelos predictivos similares a los evaluados en la presente tesis, como las redes neuronales recurrentes (*RNN*) y los bosques aleatorios (*Random Forest*), utilizando estos datos simulados. Esto permitiría confirmar si las conclusiones obtenidas a partir de los datos observados se mantienen con los datos simulados. La conclusión de la presente tesis a confirmar es la que establece que el modelado predictivo de series temporales con características de gran variabilidad y fluctuaciones considerables entre sus valores es mejor realizado a través de redes neuronales recurrentes. Estas redes son capaces de capturar dependencias temporales complejas y no lineales. Mientras que el modelado predictivo de series temporales de baja variabilidad y fluctuaciones más estables es mejor bajo el algoritmo de bosques aleatorios, que es robusto y eficaz para captar relaciones en datos con menor ruido y variabilidad. La simulación ayudaría a validar estas conclusiones al proporcionar un entorno controlado en el que se pueda evaluar el rendimiento de cada modelo bajo diferentes condiciones de variabilidad.

Finalmente, se propone la implementación del modelo predictivo en producción, este inciso persigue el objetivo de poner en funcionamiento el modelo desarrollado, permitiendo la consulta de datos por medio de la población.

Anexo 1: Bases de datos

En el presente apartado se detallan las bases de datos empleadas en esta tesis, las cuales son fundamentales para el análisis y desarrollo de los resultados obtenidos. Es importante destacar que cada una de las variables aquí mencionadas, salvo indicación en contrario dentro de paréntesis, se considera que han sido obtenidas a partir del modelo de reanálisis ERA5. ERA5 es un producto del Centro Europeo de Previsión Meteorológica a Plazo Medio (ECMWF), el cual proporciona datos climáticos globales de alta resolución mediante técnicas avanzadas de reanálisis. Este modelo integra múltiples fuentes de datos observacionales y simula diversas variables meteorológicas y climáticas con gran precisión temporal y espacial. Para aquellas variables que explícitamente indiquen su fuente de datos entre paréntesis, se ha utilizado información derivada de otros métodos o sistemas distintos al ERA5.

5.6.4. Base de datos original

La base de datos original contó con un total de 193 variables, estas son: hora del día (fecha de toma de dato), día de la semana (fecha de toma de dato), mes del año (fecha de toma de dato), día del mes (fecha de toma de dato), estación del año (fecha de toma de dato), mes del año codificación seno (fecha de toma de dato), mes del año codificación coseno (fecha de toma de dato), día de la semana codificación seno (fecha de toma de dato), día de la semana codificación coseno (fecha de toma de dato), día del mes codificación seno (fecha de toma de dato), día del mes codificación coseno

(fecha de toma de dato), hora del día codificación seno (fecha de toma de dato), hora del día codificación coseno (fecha de toma de dato), estación del año codificación seno (fecha de toma de dato), estación del año codificación coseno (fecha de toma de dato), componente u del viento a 100 m, componente v del viento a 100 m, componente u del viento neutral a 10 m, componente u del viento a 10 m, componente v del viento neutral a 10 m, componente v del viento a 10 m, ráfaga de viento a 10 m desde el posprocesamiento anterior, temperatura del punto de rocío a 2 m, temperatura a 2 m, ángulo de orografía a escala inferior a la cuadrícula, anisotropía de la orografía a escala inferior a la cuadrícula, disipación de la capa límite, altura de la capa límite, cielo despejado, radiación solar directa en la superficie, altura de la base de las nubes, energía potencial convectiva disponible, inhibición convectiva, precipitación convectiva, tasa de lluvia convectiva, radiación ultravioleta descendente en la superficie, tensión superficial de ondas de gravedad hacia el este, tensión superficial turbulenta hacia el este, evaporación, evaporación del suelo desnudo, evaporación desde la parte superior del dosel, evaporación de la transpiración de la vegetación, albedo previsto, logaritmo previsto de la rugosidad de la superficie para el calor, superficie prevista rugosidad, velocidad de fricción, geopotencial, disipación de ondas de gravedad, nubosidad alta, cubierta de vegetación alta, ráfaga de viento instantánea de 10 m, tensión superficial turbulenta instantánea hacia el este, fracción de precipitación superficial instantánea a gran escala, flujo de humedad instantáneo, tensión superficial turbulenta instantánea hacia el norte, superficie sensible instantánea flujo de calor, índice k, precipitación a gran escala, fracción de precipitación a gran escala, tasa de lluvia a gran escala, índice de área foliar vegetación alta, índice de área foliar vegetación baja, cobertura de nubes baja, cobertura de vegetación baja, temperatura máxima de 2 m desde el procesamiento posterior anterior, total máximo tasa de precipitación desde el procesamiento posterior anterior, disipación media de la capa límite, tasa media de precipitación convectiva, tensión superficial turbulenta media hacia el este, tasa media de evaporación, disipación

media de ondas de gravedad, fracción de precipitación media a gran escala, tasa media de precipitaciones a gran escala, tensión superficial media de ondas de gravedad hacia el norte, tensión superficial turbulenta media hacia el norte, tasa de evaporación potencial media, tasa de escorrentía media, presión media a nivel del mar, tasa de escorrentía subsuperficial media, flujo de radiación de onda corta directa en superficie media, flujo de radiación de onda larga hacia abajo en superficie media, flujo de radiación de onda corta en superficie media hacia abajo, flujo de radiación ultravioleta descendente medio en la superficie, flujo de calor latente superficial medio, flujo de radiación de onda larga neta superficial media, tasa de escorrentía superficial media, flujo de calor sensible superficial medio, flujo de radiación de onda corta descendente media superior, flujo de radiación de onda larga neta superior media, media flujo de radiación de onda corta neto superior, tasa de precipitación total media, gradiente vertical medio de refractividad dentro de la capa de captura, divergencia de humedad integrada verticalmente media, cobertura de nubes media, temperatura mínima de 2 m desde el posprocesamiento anterior, tasa de precipitación total mínima desde el posprocesamiento anterior, gradiente vertical mínimo de refractividad dentro de la capa de captura, albedo de infrarrojo cercano para radiación difusa, albedo de infrarrojo cercano para radiación directa, tensión superficial de ondas de gravedad hacia el norte, tensión superficial turbulenta hacia el norte, evaporación potencial, tipo de precipitación, escorrentía, pendiente de la orografía a escala subcuadrícula, tipo de suelo, desviación estándar de orografía de subred filtrada, desviación estándar de la orografía, escorrentía subsuperficial, flujo de calor latente superficial, radiación solar neta superficial, radiación térmica neta superficial, presión superficial, escorrentía superficial, flujo de calor sensible superficial, radiación solar superficial hacia abajo, radiación térmica superficial hacia abajo, toa radiación solar incidente, radiación solar neta superior, radiación térmica neta superior, nubosidad total, columna total de agua helada de la nube, columna total de agua líquida de la nube, columna total de ozono,

columna total de agua de lluvia, columna total de agua, columna total de vapor de agua, evaporación total, precipitación total, cielo total, radiación solar directa en la superficie, índice de totales totales, altura base de la capa de captura, altura superior de la capa de captura, tipo de vegetación alta, tipo de vegetación baja, componente u deriva de Stokes, albedo uv visible para radiación difusa, uv visible albedo para radiación directa, componente v deriva de Stokes, integral vertical de divergencia del flujo de agua congelada en la nube, integral vertical de divergencia del flujo de agua líquida en la nube, integral vertical de divergencia del flujo geopotencial, integral vertical de divergencia del flujo de energía cinética, integral vertical de divergencia de flujo de masa, integral vertical de divergencia de flujo de humedad, integral vertical de divergencia de flujo de ozono, integral vertical de divergencia de flujo de energía térmica, integral vertical de divergencia de flujo de energía total, integral vertical de nube hacia el este flujo de agua congelada, integral vertical del flujo de agua líquida en la nube hacia el este, integral vertical del flujo geopotencial hacia el este, integral vertical del flujo de calor hacia el este, integral vertical del flujo de energía cinética hacia el este, integral vertical del flujo de masa hacia el este, integral vertical del flujo de ozono hacia el este, integral vertical del flujo de energía total hacia el este, integral vertical del flujo de vapor de agua hacia el este, integral vertical de conversión de energía, integral vertical de la energía cinética, integral vertical de la masa de la atmósfera, integral vertical de la tendencia de la masa, integral vertical del flujo de agua congelada de las nubes hacia el norte, integral vertical del flujo de agua líquida de las nubes hacia el norte, integral vertical del flujo geopotencial hacia el norte, integral vertical del calor hacia el norte integral vertical del flujo de energía cinética hacia el norte, integral vertical del flujo de masa hacia el norte, integral vertical del flujo de ozono hacia el norte, integral vertical del flujo de energía total hacia el norte, integral vertical del flujo de vapor de agua hacia el norte, integral vertical de energía potencial e interna, integral vertical de temperatura, integral vertical de

energía térmica, integral vertical de energía total, divergencia de humedad integrada verticalmente, nivel de cero grados, viento a 100 m, viento a 10 m, viento neutro a 10 m, humedad relativa a 2 m, profundidad óptica de aerosoles (MERRA), incremento de la profundidad óptica de aerosoles (MERRA), incremento de la profundidad óptica de aerosoles (MAIAC), profundidad óptica de aerosoles (MAIAC), altura de la capa límite atmosférica, máscara de fuego VNP14A1 001, temperatura de la superficie terrestre (MODIS) y tráfico de vehículos (GOOGLE).

5.6.5. Bases de datos reducidas

uso del suelo urbano

La base de datos reducida para el uso del suelo urbano contó con un total de 77 variables, estas son: hora del día (fecha de toma de dato), día de la semana (fecha de toma de dato), mes del año (fecha de toma de dato), día del mes (fecha de toma de dato), estación del año (fecha de toma de dato), mes del año codificación seno (fecha de toma de dato), mes del año codificación coseno (fecha de toma de dato), día de la semana codificación seno (fecha de toma de dato), día de la semana codificación coseno (fecha de toma de dato), día del mes codificación seno (fecha de toma de dato), día del mes codificación coseno (fecha de toma de dato), hora del día codificación seno (fecha de toma de dato), hora del día codificación coseno (fecha de toma de dato), estación del año codificación seno (fecha de toma de dato), estación del año codificación coseno (fecha de toma de dato), componente v de viento neutral a 10 m, temperatura a 2 m, energía potencial convectiva disponible, tasa de lluvia convectiva, tensión de la superficie de la onda de gravedad hacia el este, evaporación del suelo desnudo, evaporación de la transpiración de la vegetación, pronóstico del logaritmo de la rugosidad de la superficie para el calor, pronóstico de la rugosidad de la superficie, disipación de la onda de gravedad, nubosidad alta, tensión superficial turbulenta instantánea hacia el este, flujo de humedad instantáneo, tasa media de escorrentía, presión media a

nivel del mar, tasa media de escorrentía subsuperficial, flujo de radiación de onda larga neta superficial media, divergencia de humedad integrada verticalmente media, cubierta de nubes media, gradiente vertical mínimo de refractividad dentro de la capa de captura, cerca del albedo ir para la radiación difusa, tensión superficial de la onda de gravedad hacia el norte, escorrentía subsuperficial, radiación térmica neta superior, cobertura total de nubes, columna total de ozono, columna total de agua de lluvia, columna total de agua, precipitación total, total índice de totales, altura superior de la capa de captura, integral vertical de divergencia del flujo de agua congelada en la nube, integral vertical de divergencia del flujo de agua líquida en la nube, integral vertical de divergencia del flujo geopotencial, integral vertical de divergencia del flujo de energía cinética, integral vertical de divergencia de flujo de masa, integral vertical de la divergencia del flujo de ozono, integral vertical de la nube hacia el este flujo de agua congelada, integral vertical de la nube hacia el este flujo de agua líquida, integral vertical del flujo de energía total hacia el este, integral vertical del flujo de vapor de agua hacia el este, integral vertical de la nube hacia el norte flujo de agua congelada, integral vertical del flujo de agua líquida de la nube hacia el norte, integral vertical del flujo de energía cinética hacia el norte, integral vertical del flujo de ozono hacia el norte, integral vertical del flujo de vapor de agua hacia el norte, nivel de cero grados, viento neutro a 10 m, humedad relativa a 2 m, punto de rocío temperatura a 2 m, tasa de precipitación media a gran escala, presión superficial, cielo despejado, radiación solar directa en la superficie, altura de la capa límite, disipación de la capa límite, profundidad óptica de aerosoles (MERRA), incremento de la profundidad óptica de aerosoles (MERRA), incremento de la profundidad óptica de aerosoles (MAIAC), altura de la capa límite atmosférica, VNP14A1 001 máscara de fuego, temperatura de la superficie terrestre (MODIS) y tráfico de vehículos (GOOGLE).

uso del suelo industrial

La base de datos reducida para el uso del suelo industrial contó con un total de 53 variables, estas son: hora del día (fecha de toma de dato), día de la semana (fecha de toma de dato), mes del año (fecha de toma de dato), día del mes (fecha de toma de dato), estación del año (fecha de toma de dato), mes del año codificación seno (fecha de toma de dato), mes del año codificación coseno (fecha de toma de dato), día de la semana codificación seno (fecha de toma de dato), día de la semana codificación coseno (fecha de toma de dato), día del mes codificación seno (fecha de toma de dato), día del mes codificación coseno (fecha de toma de dato), hora del día codificación seno (fecha de toma de dato), hora del día codificación coseno (fecha de toma de dato), estación del año codificación seno (fecha de toma de dato), estación del año codificación coseno (fecha de toma de dato), tensión superficial de la onda de gravedad hacia el este, evaporación desde la parte superior del dosel, evaporación de la transpiración de la vegetación, ráfaga de viento instantánea de 10 m, tensión superficial turbulenta instantánea hacia el este, flujo de humedad instantáneo, índice k, presión media a nivel del mar, tensión superficial de la onda de gravedad hacia el norte, escorrentía, sub escorrentía superficial, columna total de ozono, integral vertical de divergencia del flujo de agua líquida en la nube, integral vertical de divergencia del flujo de energía cinética, integral vertical de divergencia del flujo de humedad, integral vertical de divergencia del flujo de ozono, integral vertical de divergencia del flujo total de energía , integral vertical del flujo de agua líquida de la nube hacia el este, integral vertical del flujo de energía cinética hacia el este, integral vertical del flujo de vapor de agua hacia el este, integral vertical de tendencia de masa, integral vertical del flujo de calor hacia el norte, integral vertical del flujo de ozono hacia el norte, nivel de cero grados, temperatura a 2 m, humedad relativa a 2 m, punto de rocío temperatura a 2 m, tasa media de precipitación a gran escala, viento neutro a 10 m, presión superficial, cielo despejado, radiación solar directa en la superficie,

altura de la capa límite, disipación de la capa límite, profundidad óptica de aerosoles (MERRA), incremento de la profundidad óptica de aerosoles (MERRA), incremento de la profundidad óptica de aerosoles (MAIAC), altura de la capa límite atmosférica, VNP14A1 001 máscara de fuego, temperatura de la superficie terrestre (MODIS) y tráfico de vehículos (GOOGLE).

uso del suelo agrícola y área verde

La base de datos reducida para el uso del suelo agrícola y área verde contó con un total de 84 variables, estas son: hora del día (fecha de toma de dato), día de la semana (fecha de toma de dato), mes del año (fecha de toma de dato), día del mes (fecha de toma de dato), estación del año (fecha de toma de dato), mes del año codificación seno (fecha de toma de dato), mes del año codificación coseno (fecha de toma de dato), día de la semana codificación seno (fecha de toma de dato), día de la semana codificación coseno (fecha de toma de dato), día del mes codificación seno (fecha de toma de dato), día del mes codificación coseno (fecha de toma de dato), hora del día codificación seno (fecha de toma de dato), hora del día codificación coseno (fecha de toma de dato), estación del año codificación seno (fecha de toma de dato), estación del año codificación coseno (fecha de toma de dato), componente v del viento a 100 m, componente u del viento neutral a 10 m, energía potencial convectiva disponible, tasa de lluvia convectiva, tensión de la superficie de la onda de gravedad hacia el este, evaporación de la transpiración de la vegetación, albedo pronosticado, logaritmo pronosticado de la rugosidad de la superficie para el calor, disipación de la onda de gravedad, nubes altas cobertura, temperatura máxima de 2 m desde el posprocesamiento anterior, tasa media de evaporación, tasa media de precipitación a gran escala, tensión superficial media de la onda de gravedad hacia el norte, tasa media de escorrentía, presión media a nivel del mar, tasa media de escorrentía subsuperficial, flujo de radiación de onda larga neta superficial media, gradiente vertical medio de refractividad dentro de la capa

atrapante, nubosidad media, tipo de precipitación, escorrentía subsuperficial, flujo de calor latente superficial, flujo de calor sensible superficial, radiación térmica neta superior, nubosidad total, columna de agua total, precipitación total, índice de totales totales, altura de la base de la capa de captura, tipo de vegetación baja, albedo visible uv para radiación difusa, integral vertical de divergencia del flujo de agua congelada en la nube, integral vertical de divergencia del flujo de agua líquida en la nube, integral vertical de divergencia del flujo de energía cinética, integral vertical de divergencia del flujo de ozono, integral vertical de la divergencia del flujo de energía térmica, integral vertical de la divergencia del flujo de energía total, integral vertical del flujo de agua congelada en la nube hacia el este, integral vertical del flujo de agua líquida en la nube hacia el este, integral vertical del flujo geopotencial hacia el este, integral vertical de flujo de ozono hacia el este, integral vertical del flujo de vapor de agua hacia el este, integral vertical de la masa de la atmósfera, integral vertical de la tendencia de la masa, integral vertical del flujo de agua congelada en las nubes hacia el norte, integral vertical del flujo de agua líquida en las nubes hacia el norte, integral vertical del flujo de calor hacia el norte, integral vertical del flujo de ozono hacia el norte, integral vertical del flujo de vapor de agua hacia el norte, integral vertical de temperatura, divergencia de humedad integrada verticalmente, nivel de cero grados, viento neutro a 10 m, temperatura a 2 m, humedad relativa a 2 m, temperatura del punto de rocío a 2 m, presión superficial, cielo despejado directo radiación solar en la superficie, pbl, flujo de humedad instantáneo, altura de la capa límite, disipación de la capa límite, profundidad óptica de aerosoles (MERRA), incremento de la profundidad óptica de aerosoles (MERRA), incremento de la profundidad óptica de aerosoles (MAIAC), altura de la capa límite atmosférica, máscara de fuego VNP14A1 001, temperatura de la superficie terrestre (MODIS) y tráfico de vehículos (GOOGLE).

Anexo 2: Procedimiento de análisis de series temporales

La presente sección incluye los resultados obtenidos a partir del procedimiento completo de análisis de series temporales desarrollado durante el capítulo 4, para el análisis de las variables derivadas a partir de información satelital y meteorológicas. En la misma se incluye desde los resultados del análisis descriptivo hasta los obtenidos en el procedimiento de ajuste del modelo teórico y el análisis de correlación cruzada para cada una de las variables analizadas.

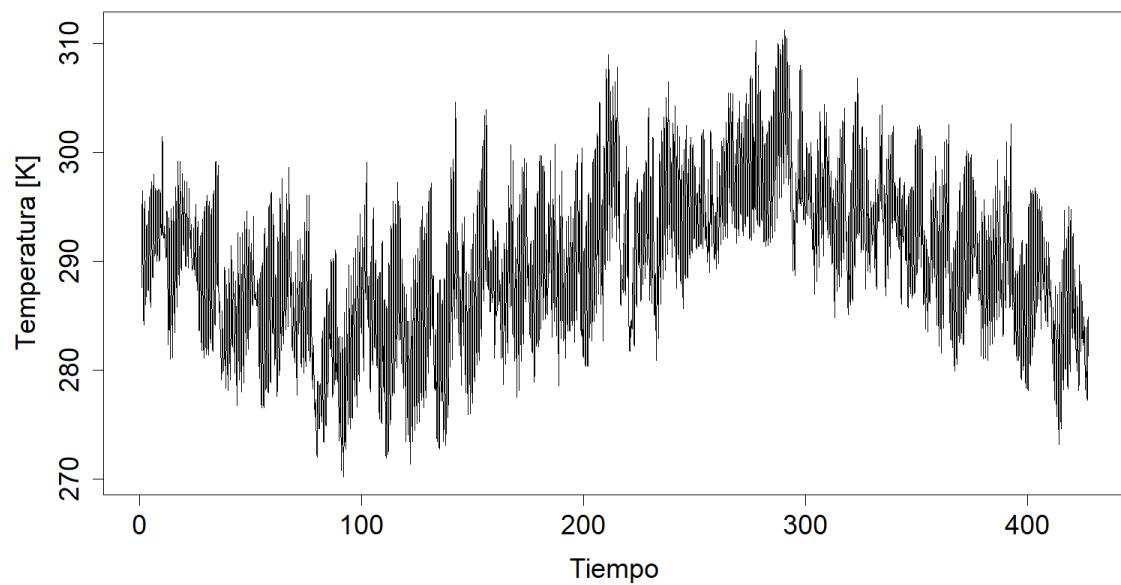
Temperatura

Figura 5.19: Gráfico de la serie horaria promedio de temperatura.

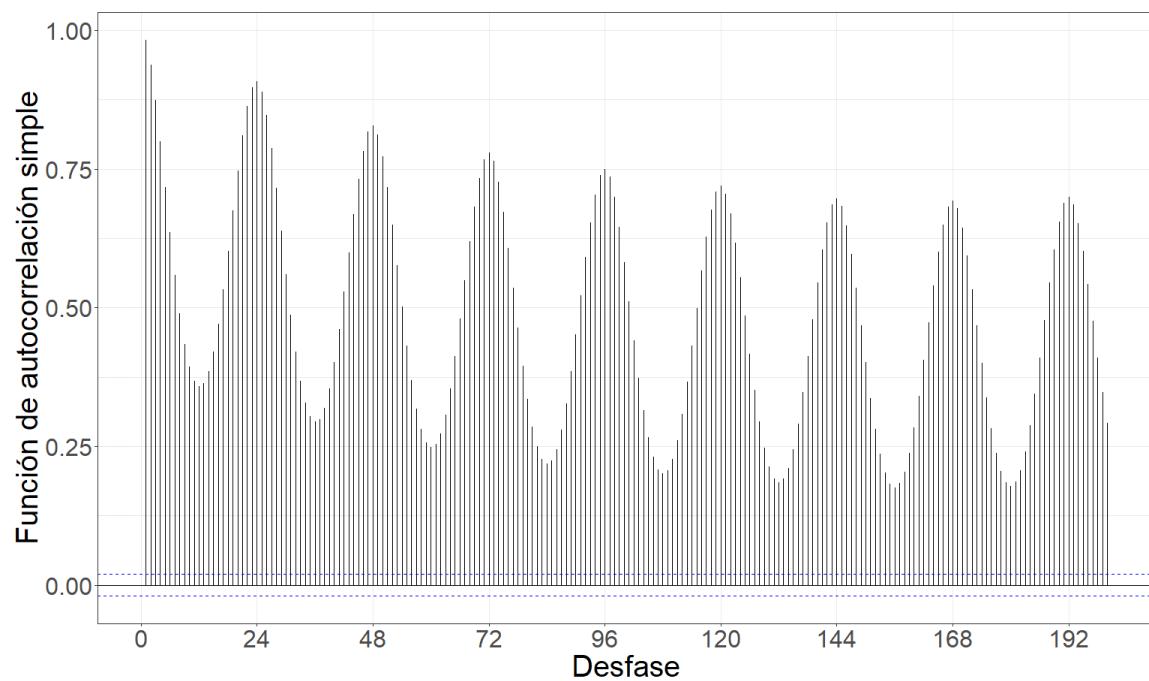


Figura 5.20: Gráfico de la función de autocorrelación simple de la serie.

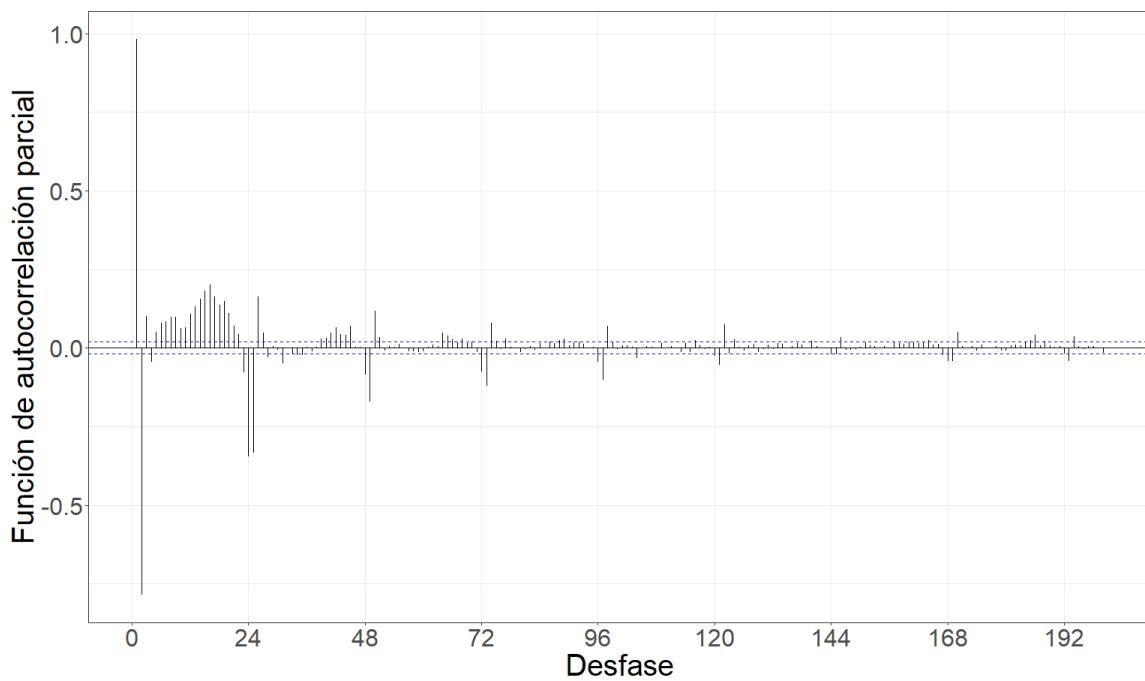


Figura 5.21: Gráfico de la función de autocorrelación parcial de la serie.

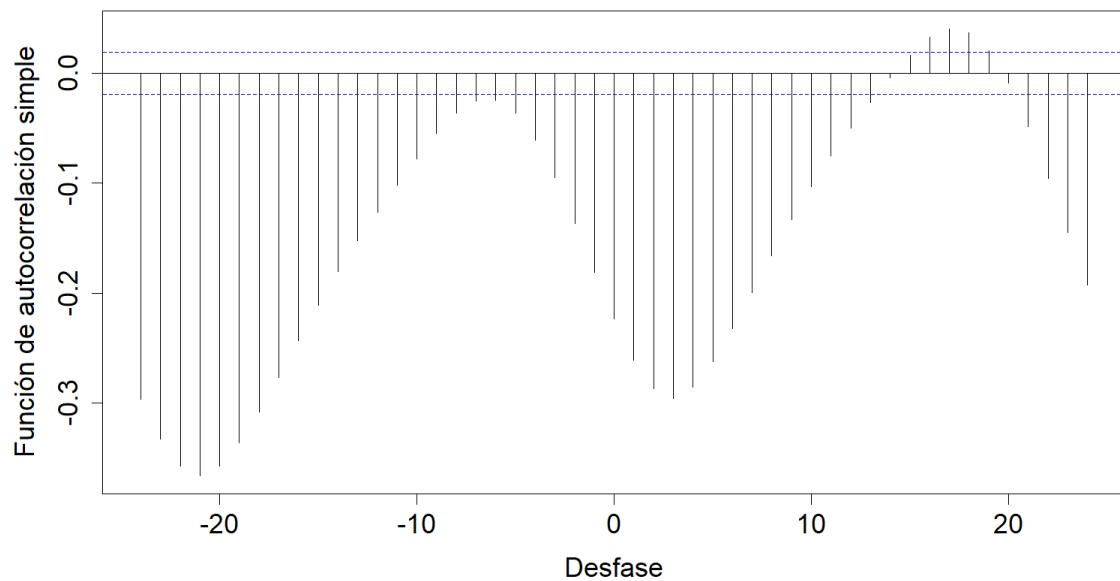


Figura 5.22: Gráfico de la correlación cruzada entre las series previo al filtrado por el modelo teórico.

Tabla 5.4: Modelo teórico ajustado.

Modelo	AIC	AICc	BIC
ARIMA(2,0,1)(2,1,0)[24]	18579.78	18579.79	18579.79

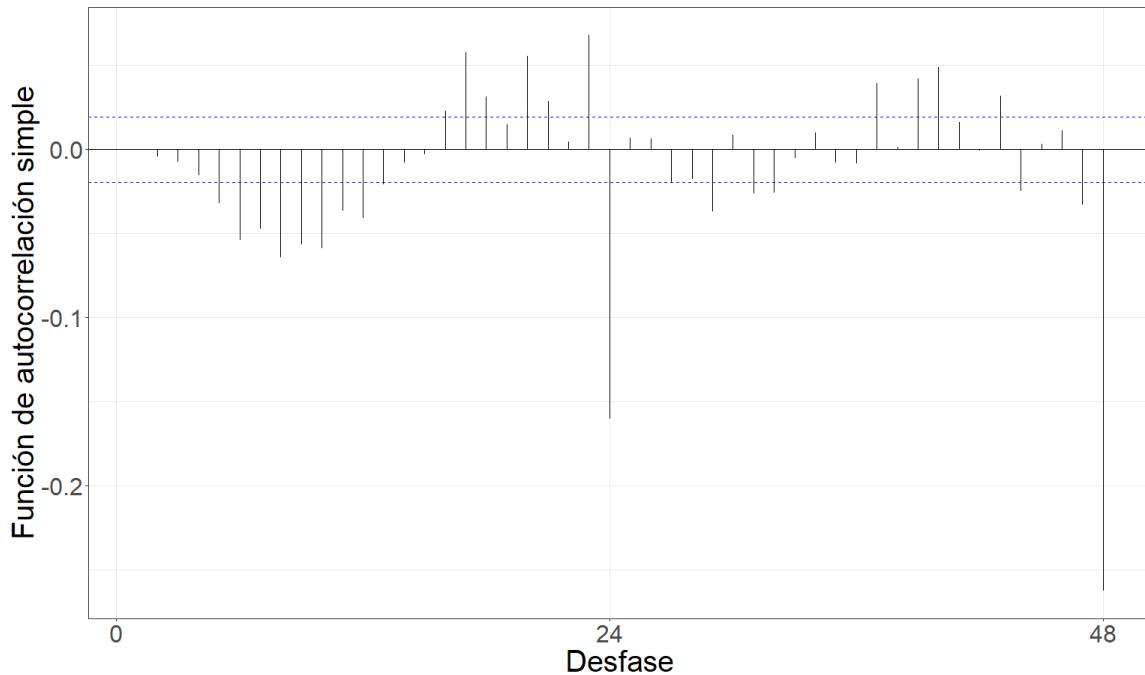


Figura 5.23: Gráfico de la función de autocorrelación simple de la serie filtrada por el modelo teórico ajustado.

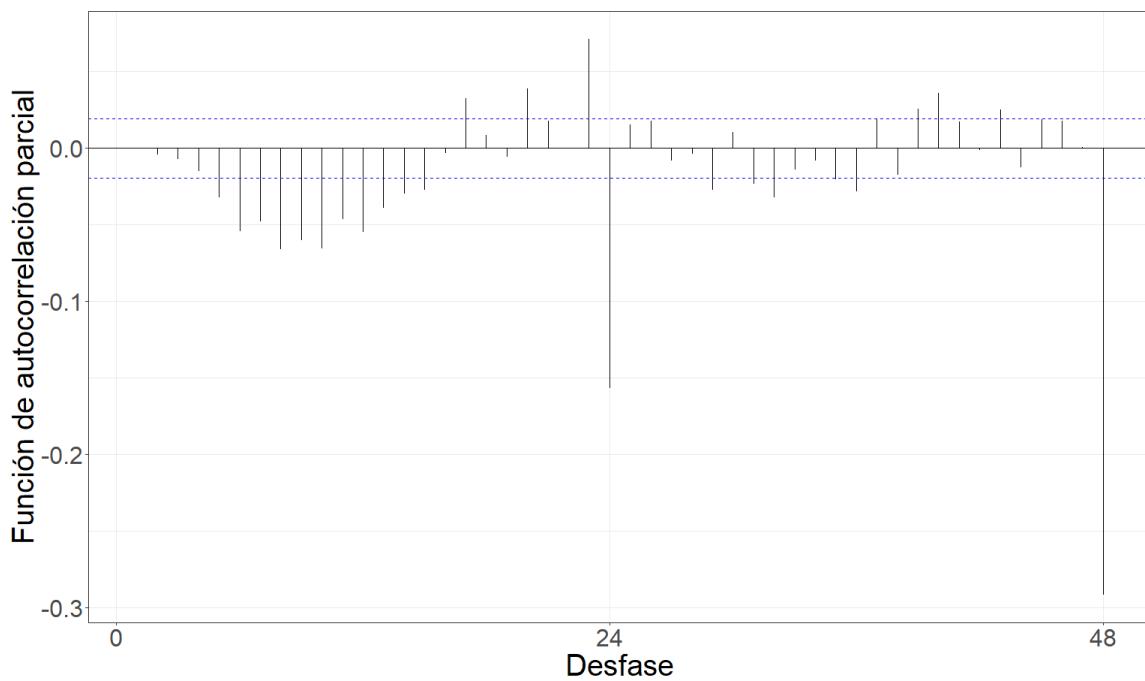


Figura 5.24: Gráfico de la función de autocorrelación parcial de la serie filtrada por el modelo teórico ajustado.

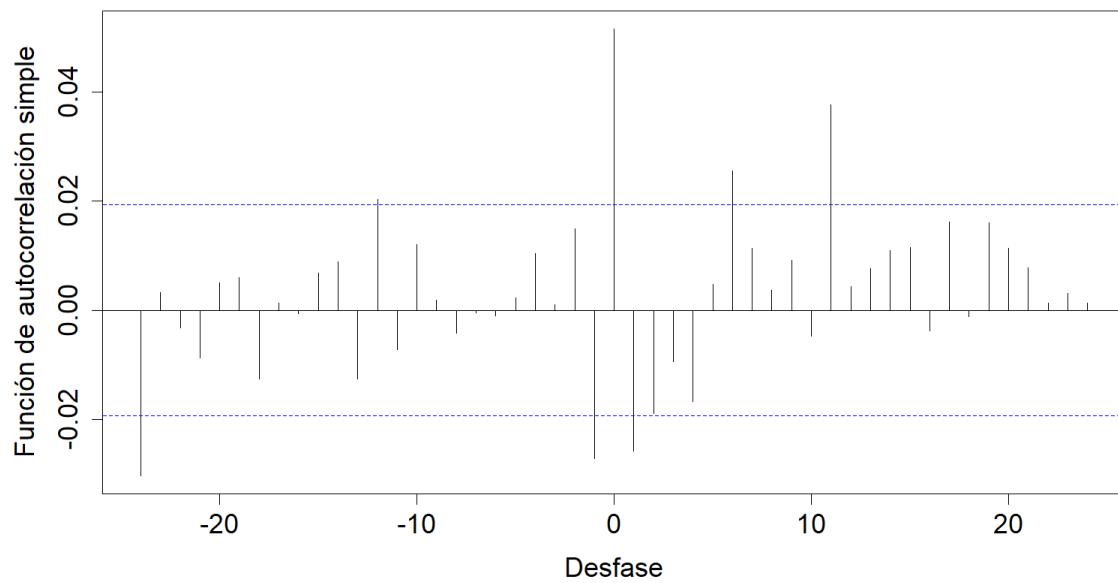


Figura 5.25: Gráfico de la correlación cruzada entre las series filtradas por el modelo teórico ajustado.

Tabla 5.5: Coeficientes estimados de la relación funcional.

Desfasaje	Con diferenciación	Sin diferenciación
t	-0.1573001	-1.04115
t-1	0.2441853	1.44786
t-12	-0.0197213	-0.06497
t-24	-0.2628158	-0.68048

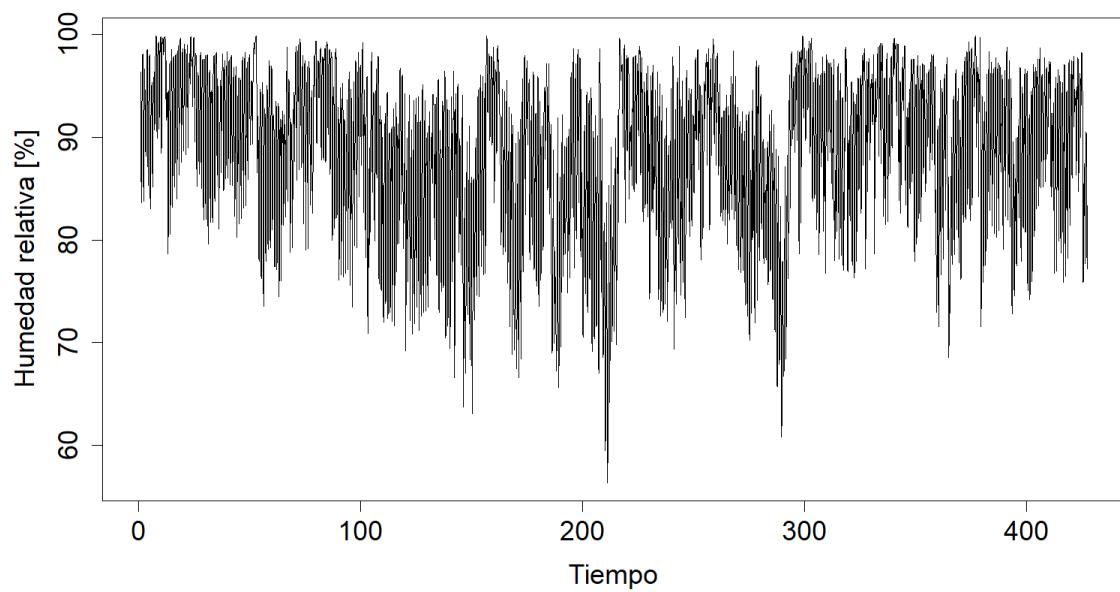
Humedad relativa

Figura 5.26: Gráfico de la serie horaria promedio de humedad relativa.

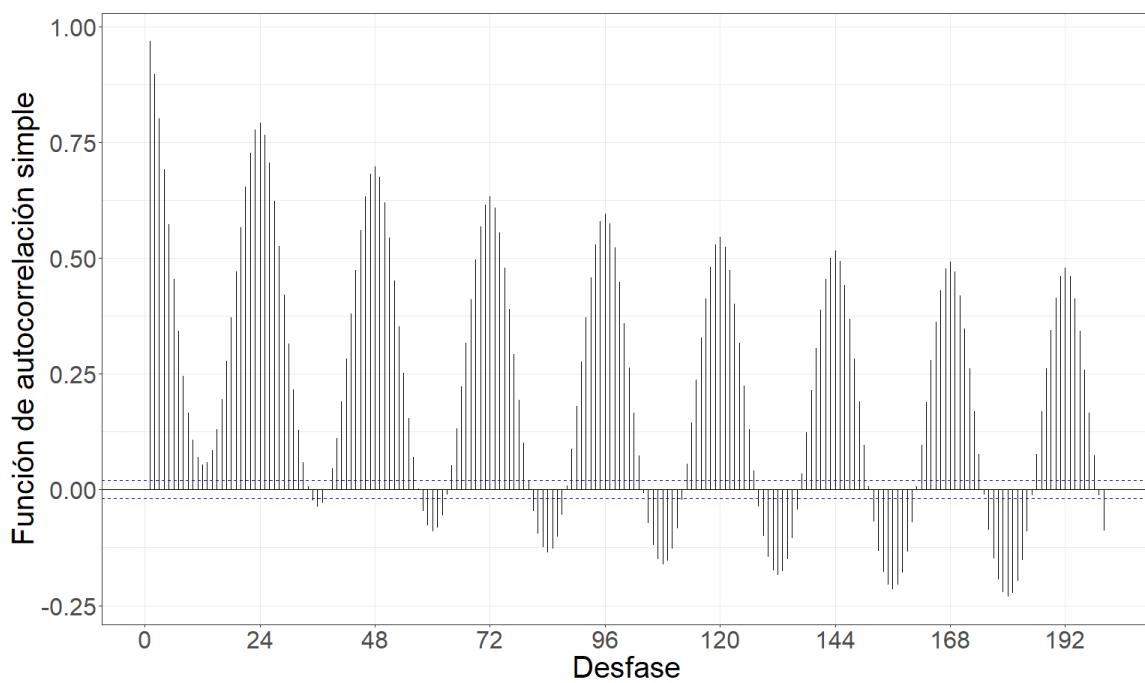


Figura 5.27: Gráfico de la función de autocorrelación simple de la serie.

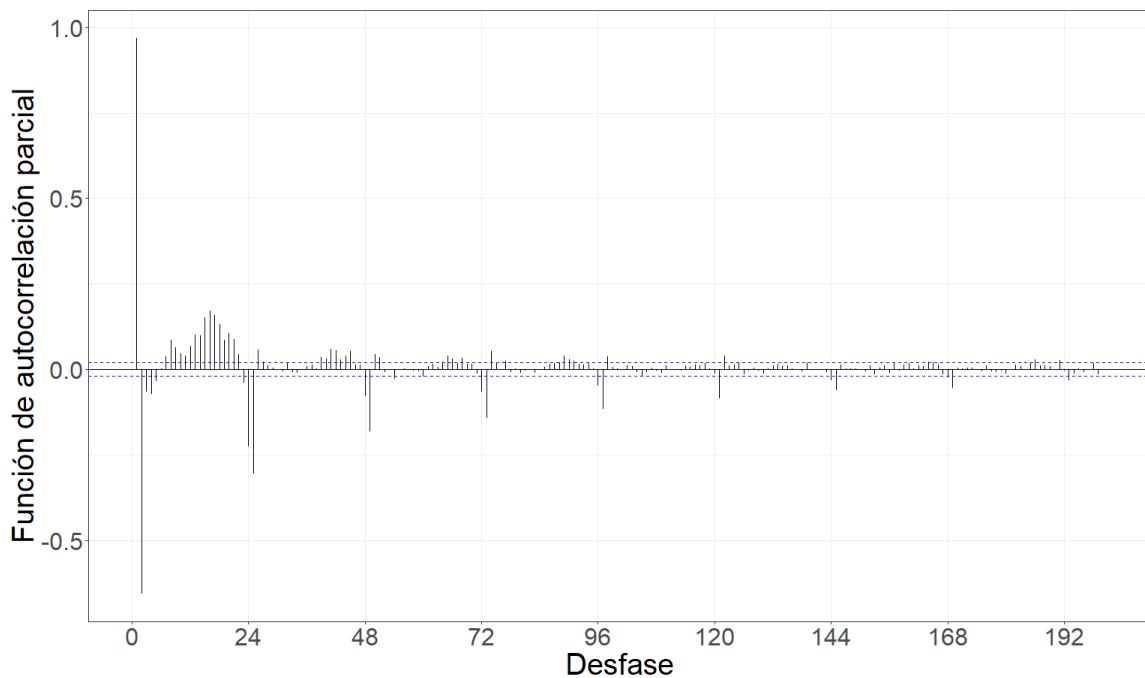


Figura 5.28: Gráfico de la función de autocorrelación parcial de la serie.

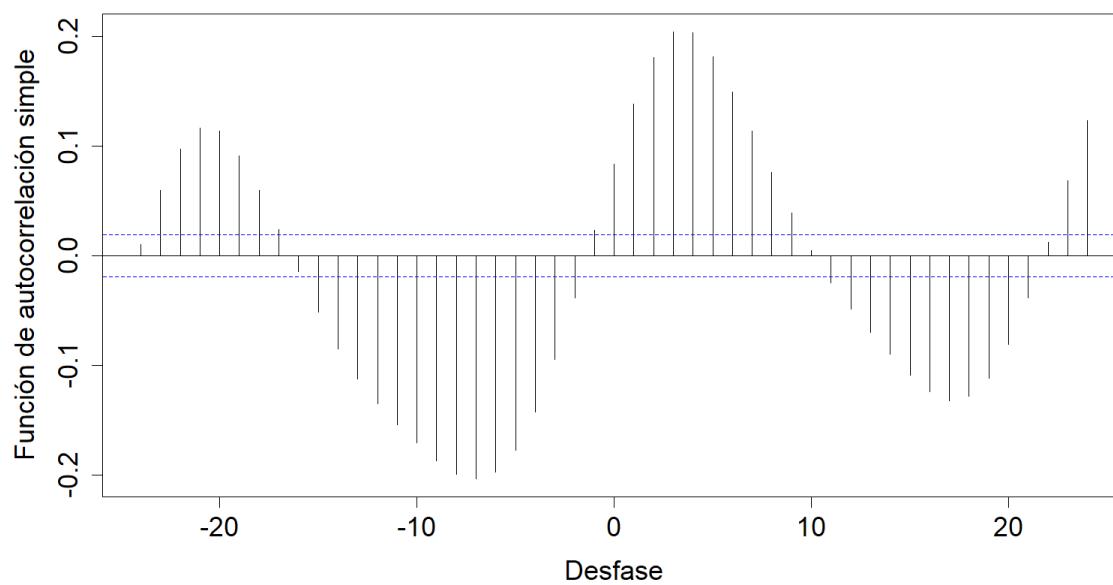


Figura 5.29: Gráfico de la correlación cruzada entre las series previo al filtrado por el modelo teórico ajustado.

Tabla 5.6: Modelo teórico ajustado.

Modelo	AIC	AICc	BIC
ARIMA(5,0,0)(2,1,0)[24]	32693.17	32693.18	32751.02

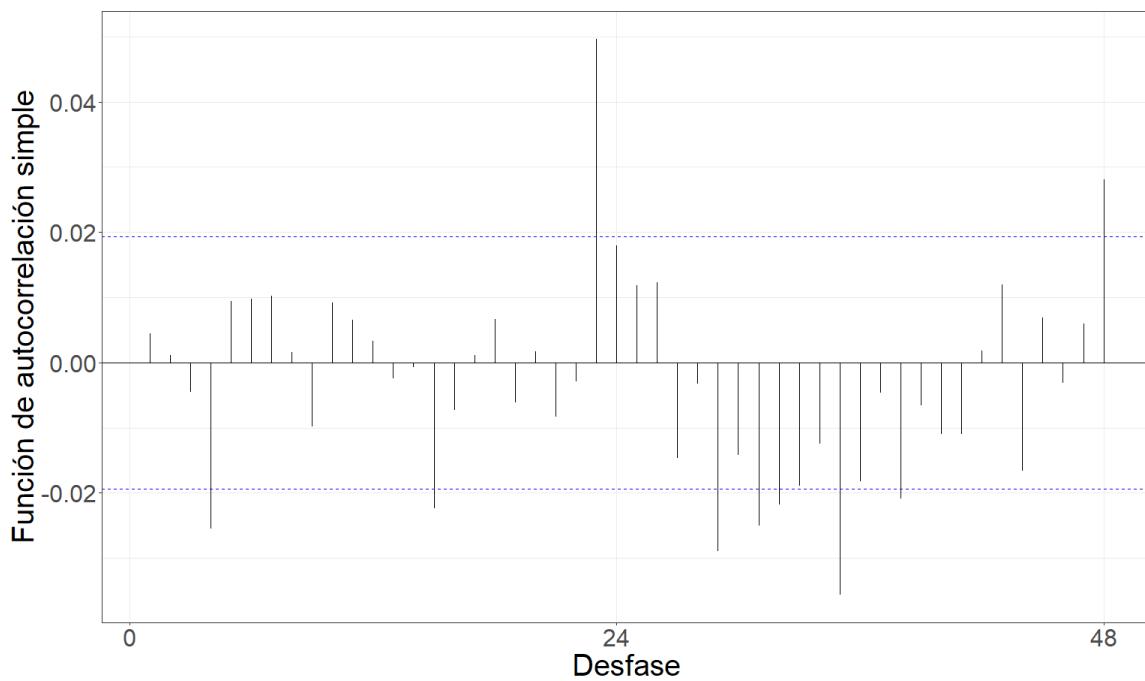


Figura 5.30: Gráfico de la función de autocorrelación simple de la serie filtrada por el modelo teórico ajustado.

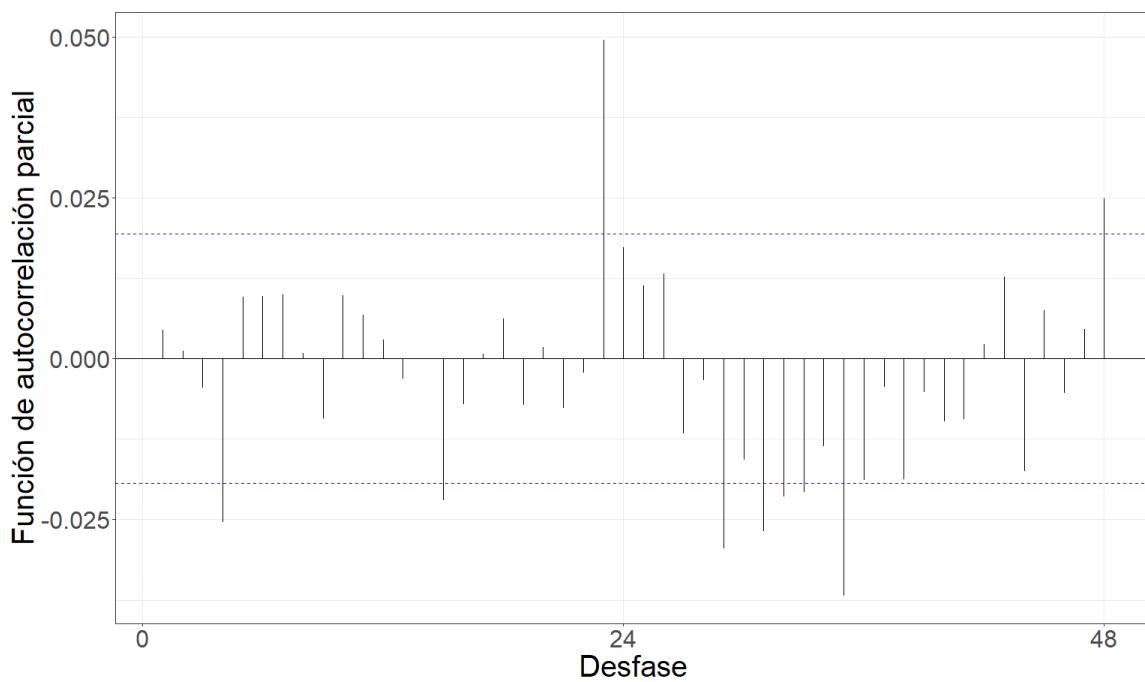


Figura 5.31: Gráfico de la función de autocorrelación parcial de la serie filtrada por el modelo teórico ajustado.

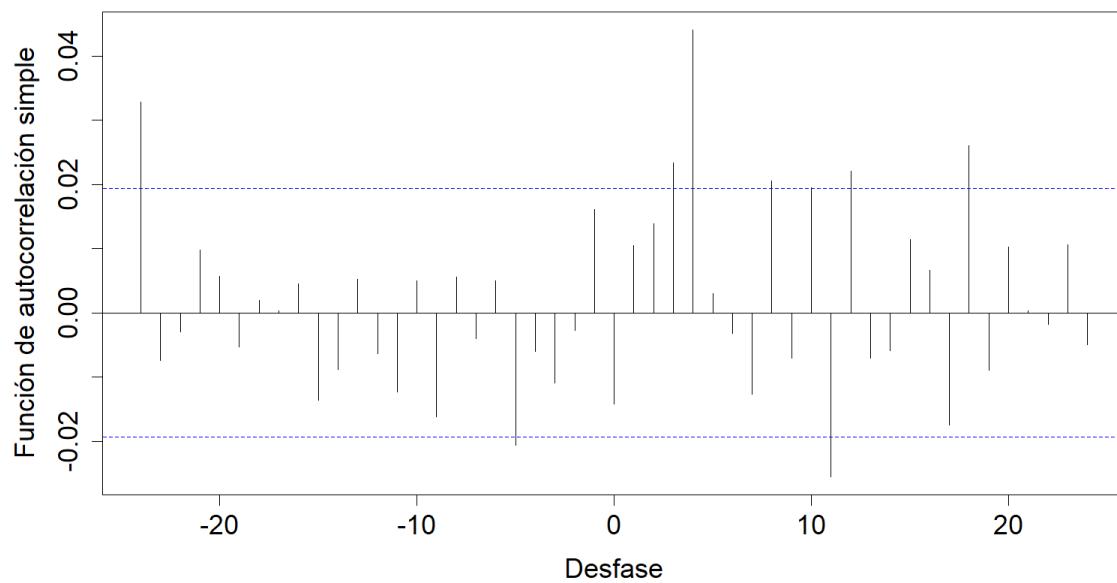


Figura 5.32: Gráfico de la correlación cruzada entre las series filtradas por el modelo teórico ajustado.

Tabla 5.7: Coeficientes estimados de la relación funcional.

Desfasaje	Con diferenciación	Sin diferenciación
t-5	-0.2681536	-0.2548
t-24	0.1367988	0.1309

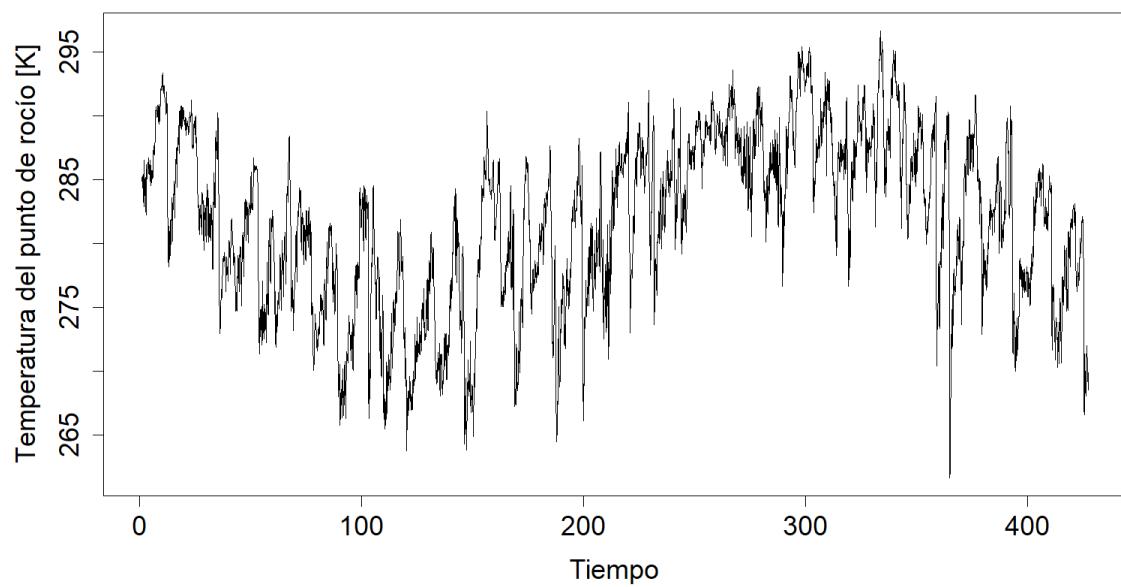
Temperatura del punto de rocío

Figura 5.33: Gráfico de la serie horaria promedio de temperatura del punto de rocío.

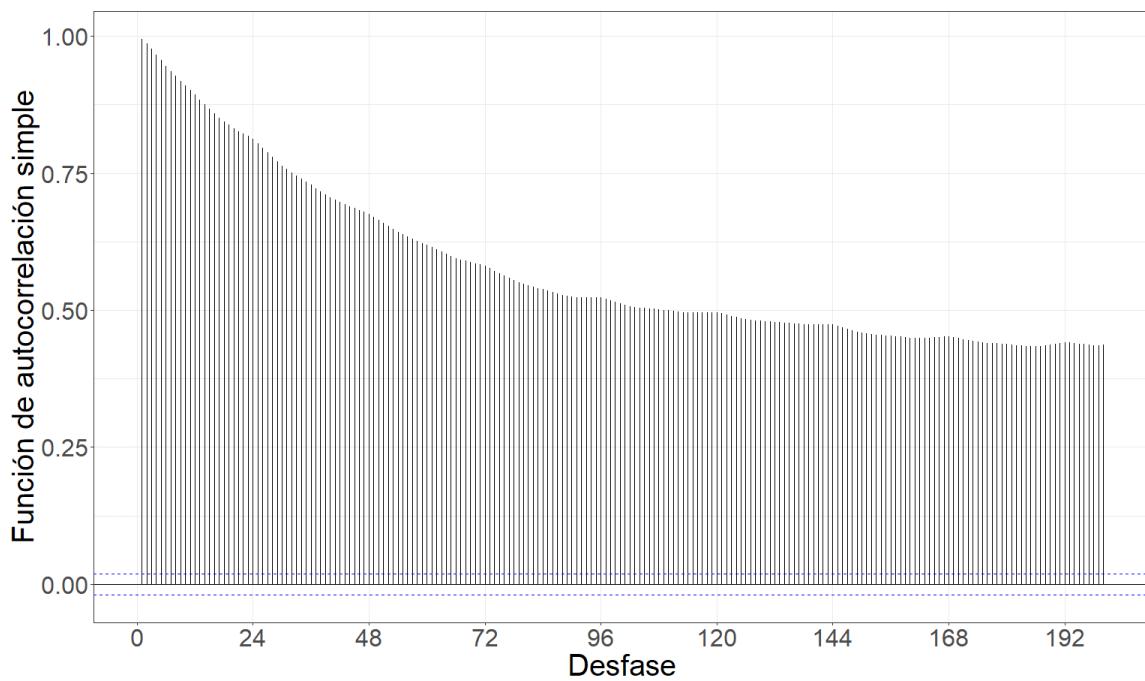


Figura 5.34: Gráfico de la función de autocorrelación simple de la serie.

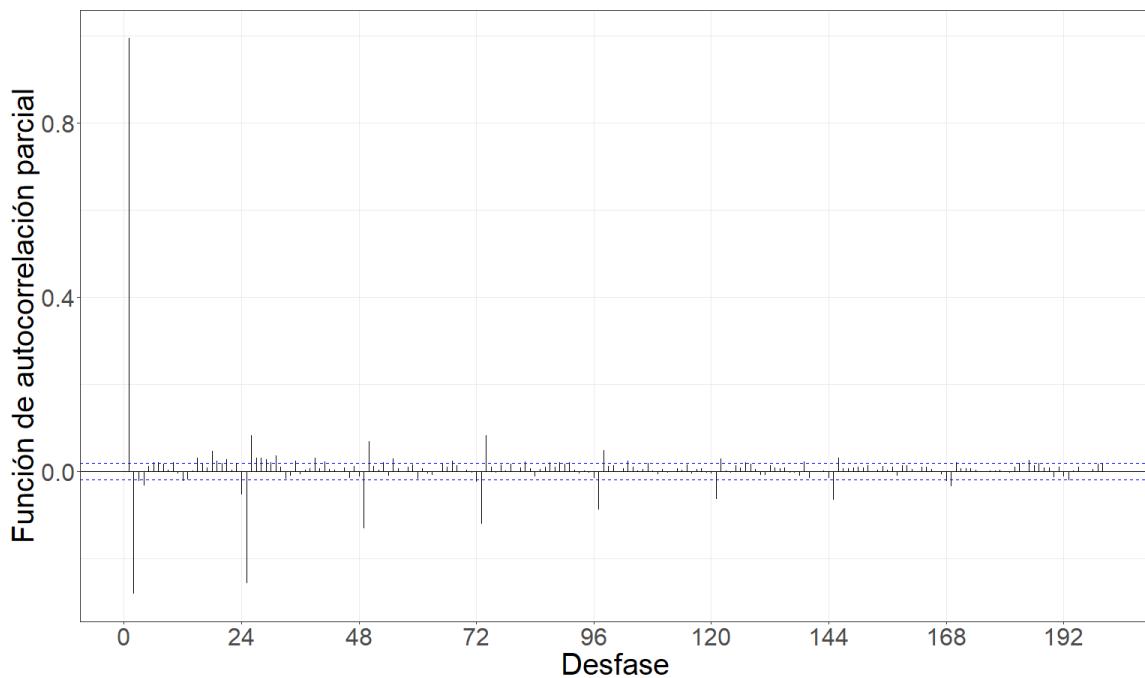


Figura 5.35: Gráfico de la función de autocorrelación parcial de la serie.

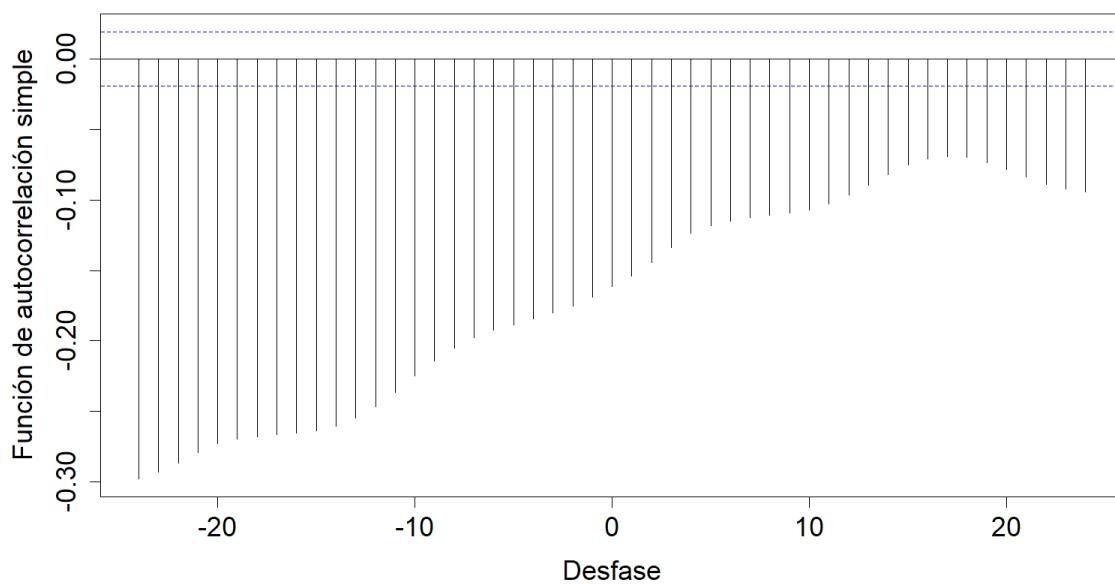


Figura 5.36: Gráfico de la correlación cruzada entre las series previo al filtrado por el modelo teórico ajustado.

Tabla 5.8: Modelo teórico ajustado.

Modelo	AIC	AICc	BIC
ARIMA(3,1,0)(2,0,0)[24]	19288.83	19288.84	19332.23

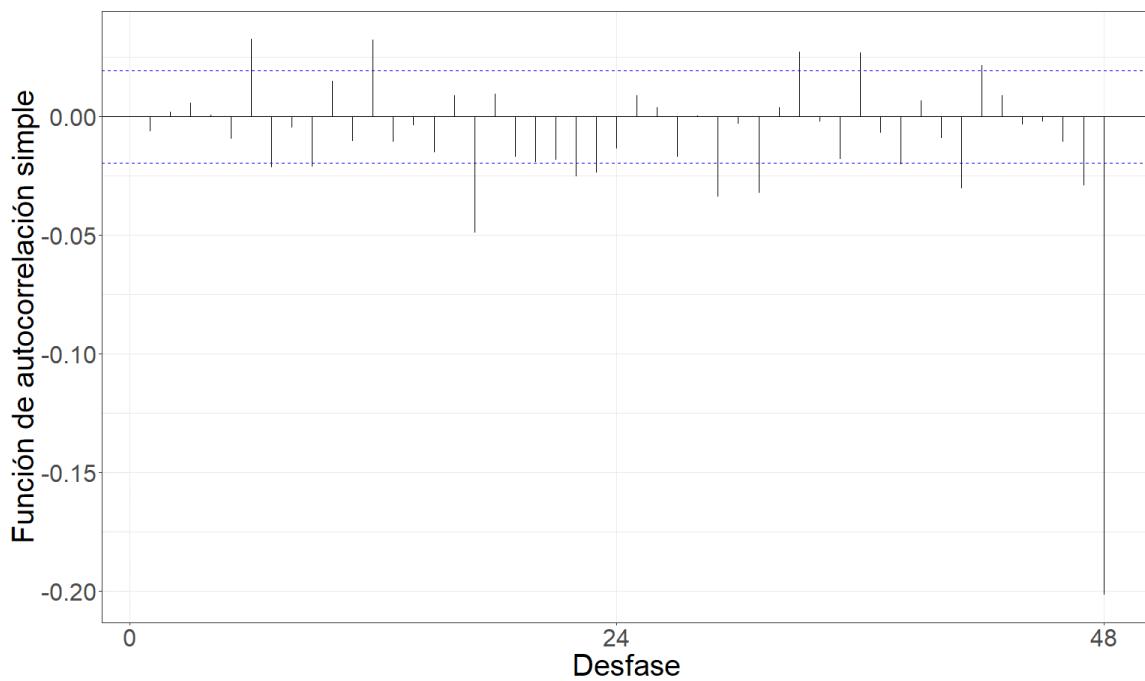


Figura 5.37: Gráfico de la función de autocorrelación simple de la serie filtrada por el modelo teórico ajustado.

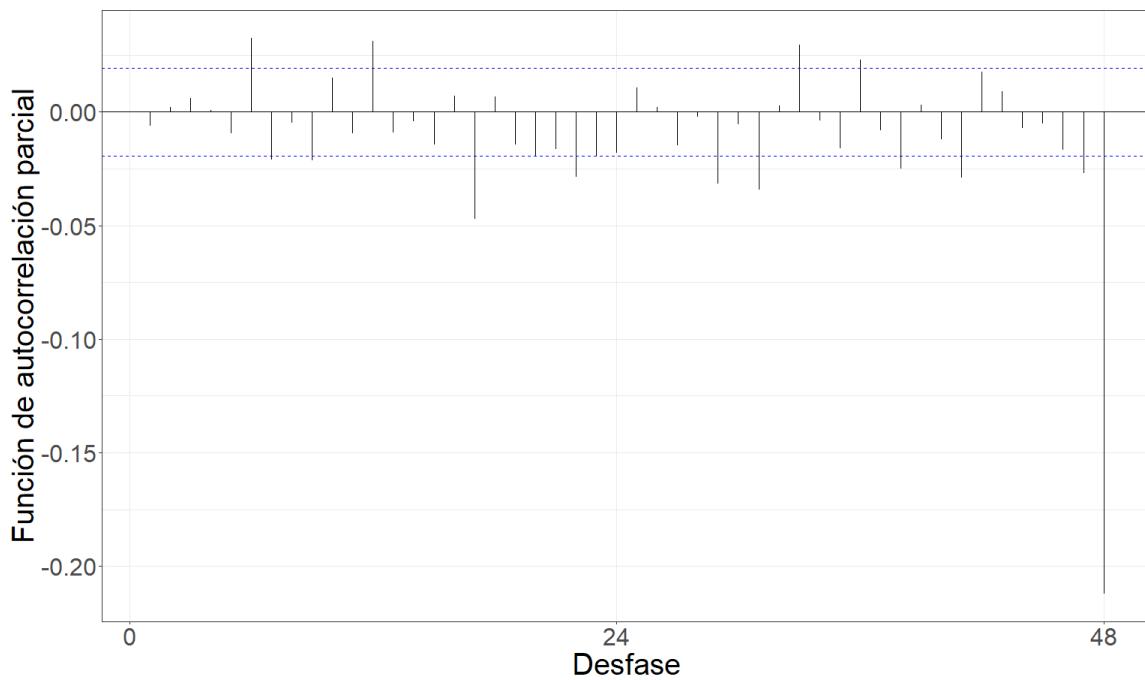


Figura 5.38: Gráfico de la función de autocorrelación parcial de la serie filtrada por el modelo teórico ajustado.

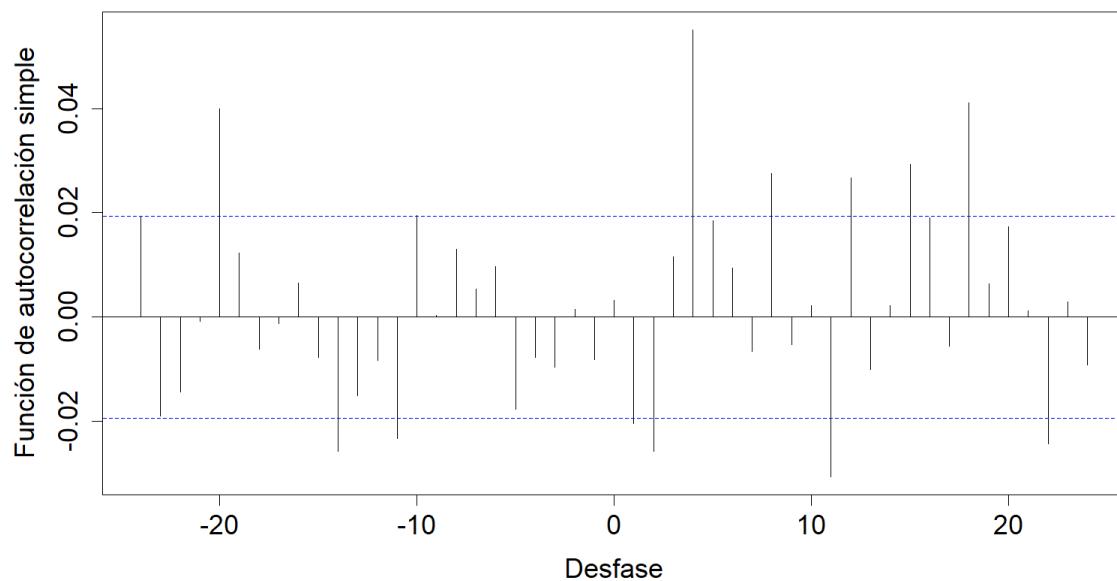


Figura 5.39: Gráfico de la correlación cruzada entre las series filtradas por el modelo teórico ajustado.

Tabla 5.9: Coeficientes estimados de la relación funcional.

Desfasaje	Con diferenciación	Sin diferenciación
t-11	-0.1597575	-0.2921
t-14	-0.3212396	-0.3222
t-20	0.3752728	-0.3377

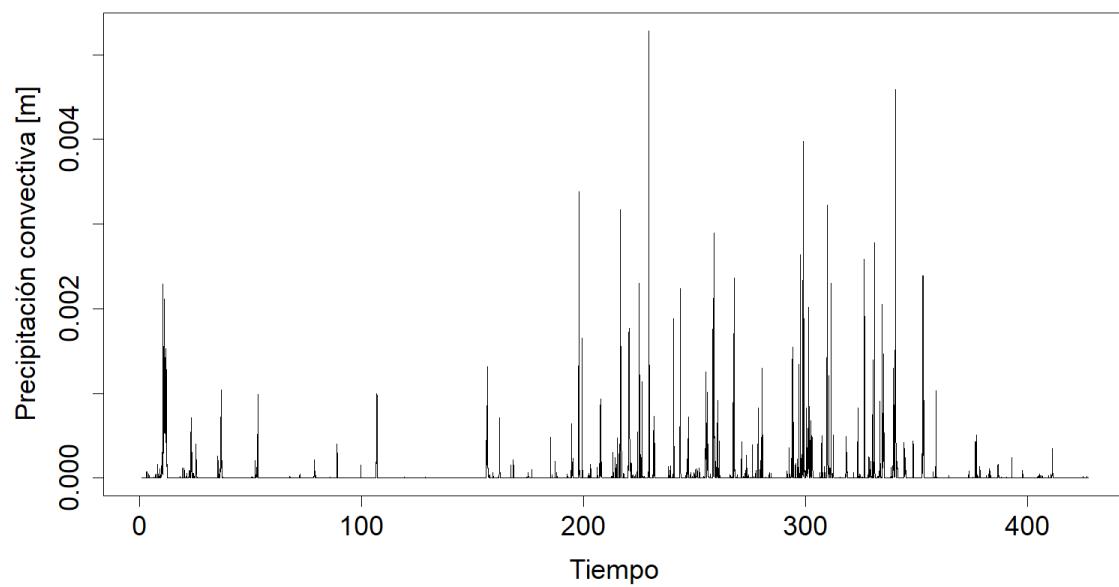
Precipitación convectiva

Figura 5.40: Gráfico de la serie horaria promedio de precipitación convectiva.

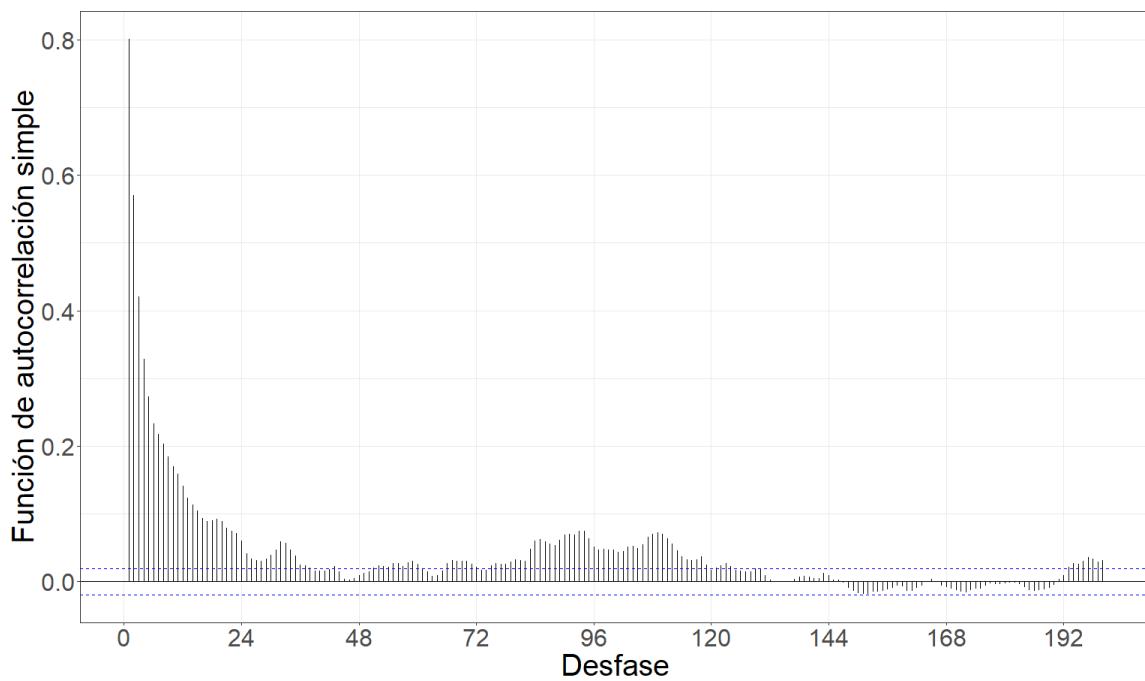


Figura 5.41: Gráfico de la función de autocorrelación simple de la serie.

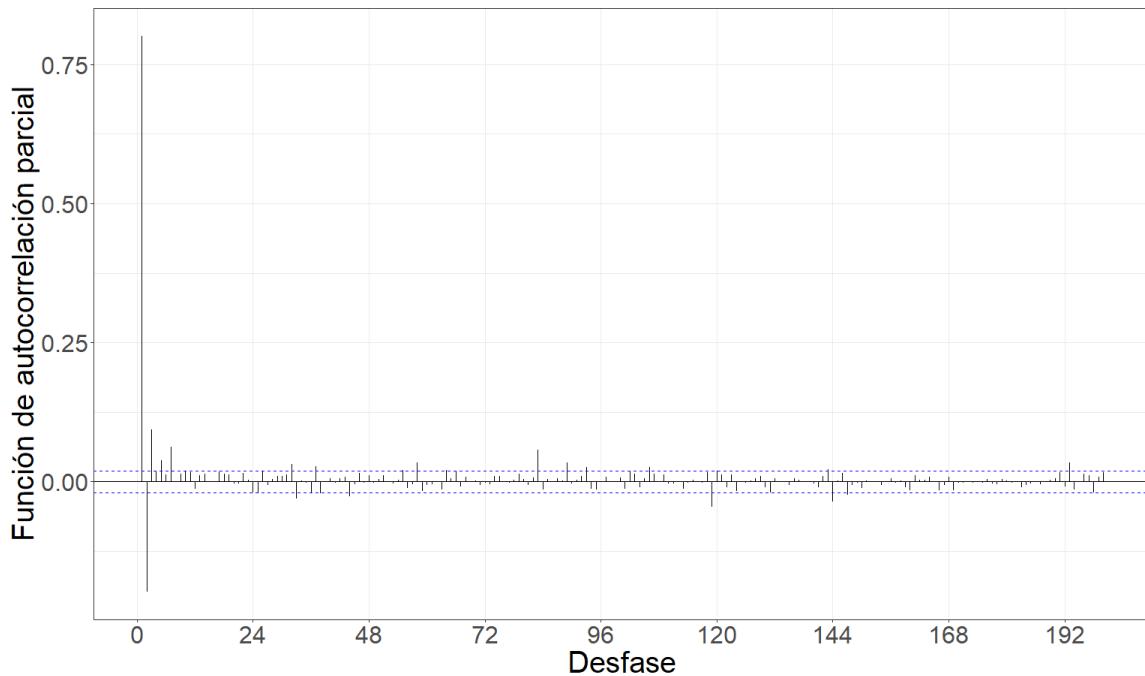


Figura 5.42: Gráfico de la función de autocorrelación parcial de la serie.

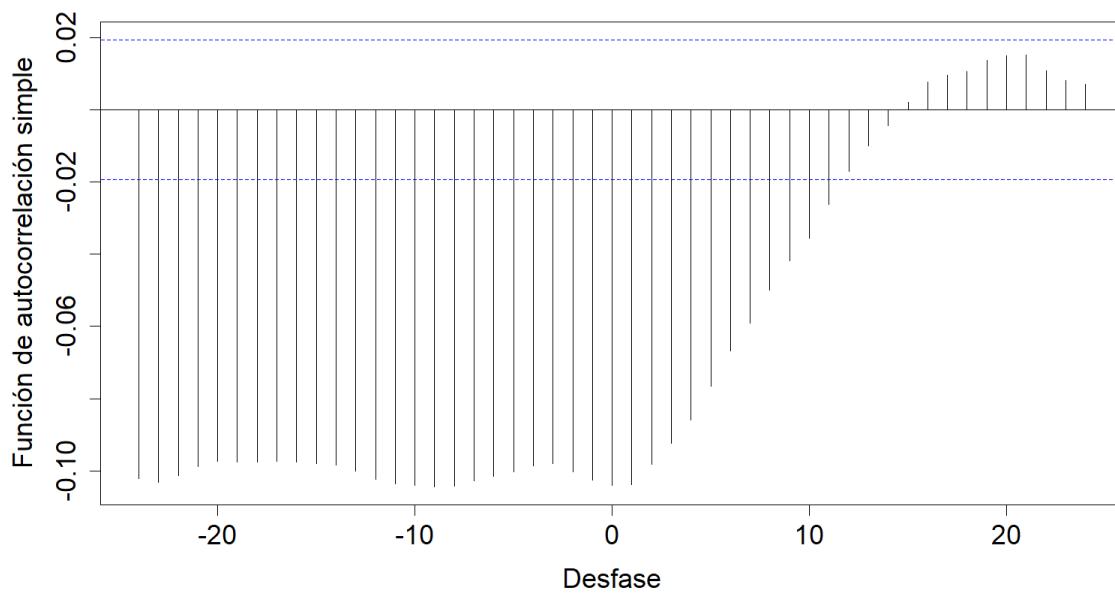


Figura 5.43: Gráfico de la correlación cruzada entre las series previo al filtrado por el modelo teórico ajustado.

Tabla 5.10: Modelo teórico ajustado.

Modelo	AIC	AICc	BIC
ARIMA(0,1,5)(0,0,0)[24]	-152270.8	-152270.8	-152227.4

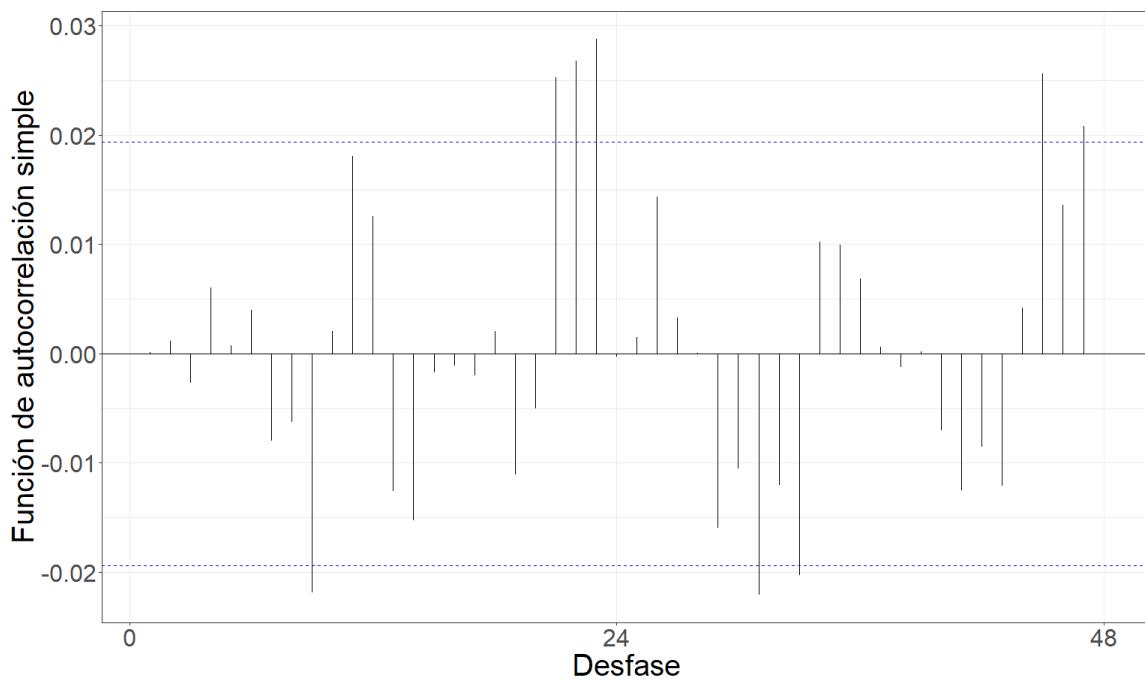


Figura 5.44: Gráfico de la función de autocorrelación simple de la serie filtrada por el modelo teórico ajustado.

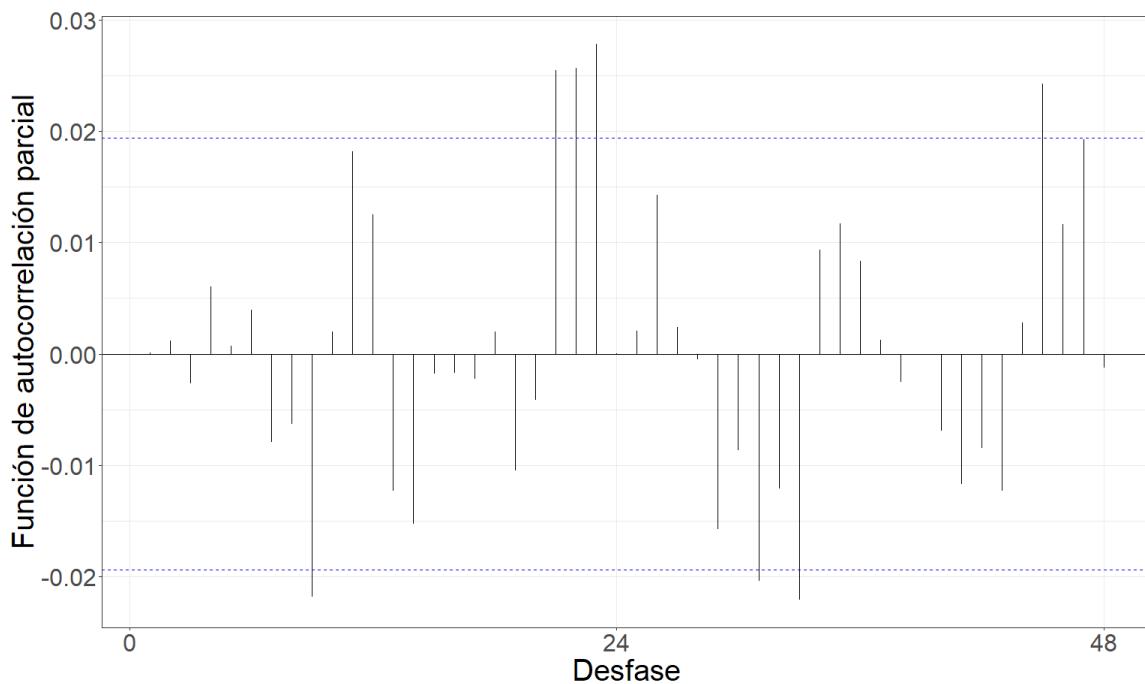


Figura 5.45: Gráfico de la función de autocorrelación parcial de la serie filtrada por el modelo teórico ajustado.

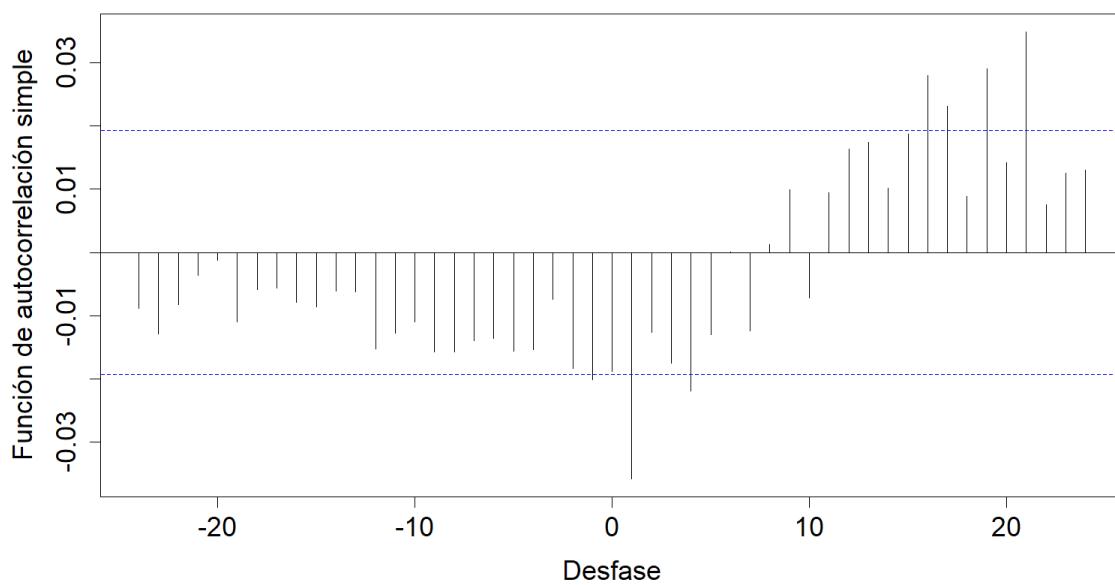


Figura 5.46: Gráfico de la correlación cruzada entre las series filtradas por el modelo teórico ajustado.

Tabla 5.11: Coeficientes estimados de la relación funcional.

Desfasaje	Con diferenciación	Sin diferenciación
t-1	-6.162e+01	-3466.73

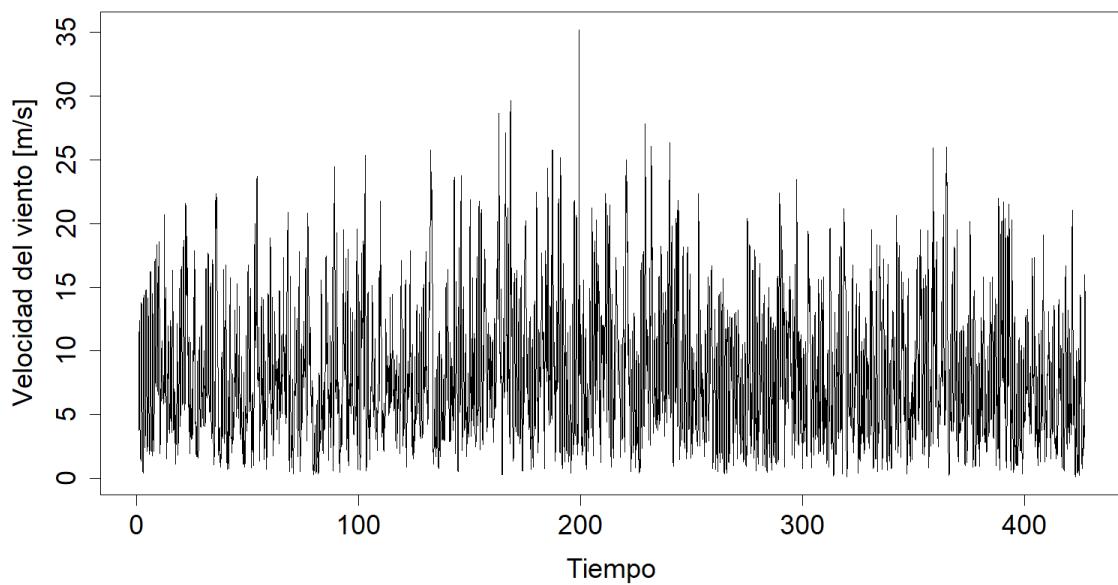
Velocidad del viento

Figura 5.47: Gráfico de la serie horaria promedio de velocidad del viento.

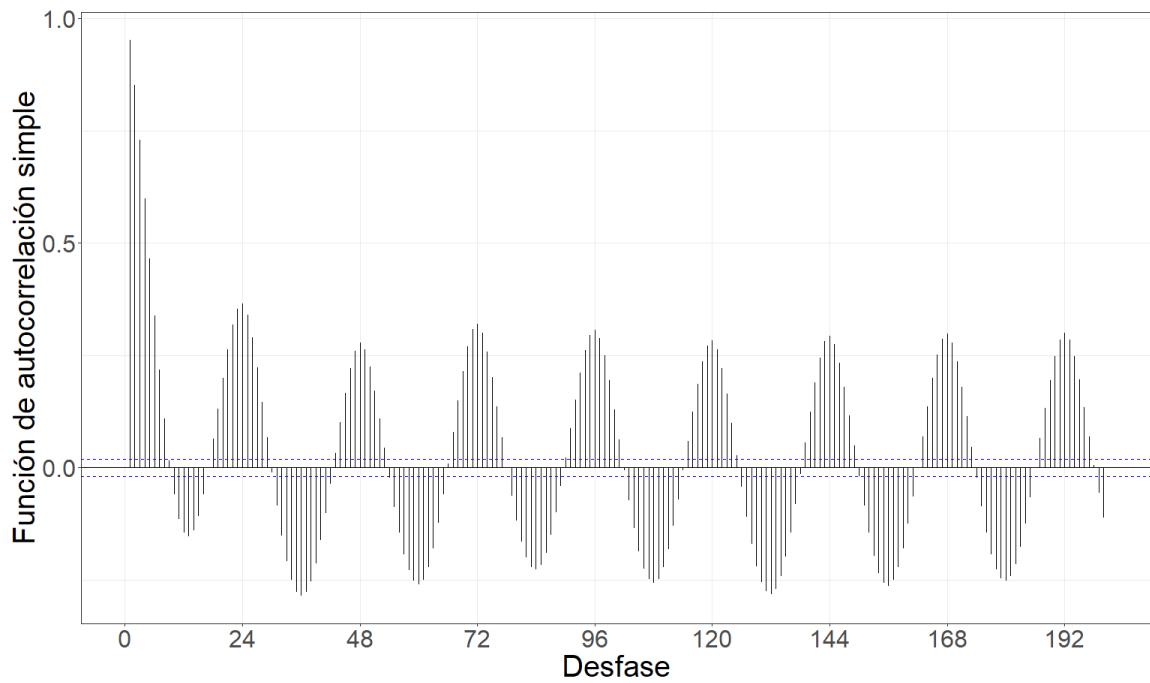


Figura 5.48: Gráfico de la función de autocorrelación simple de la serie.

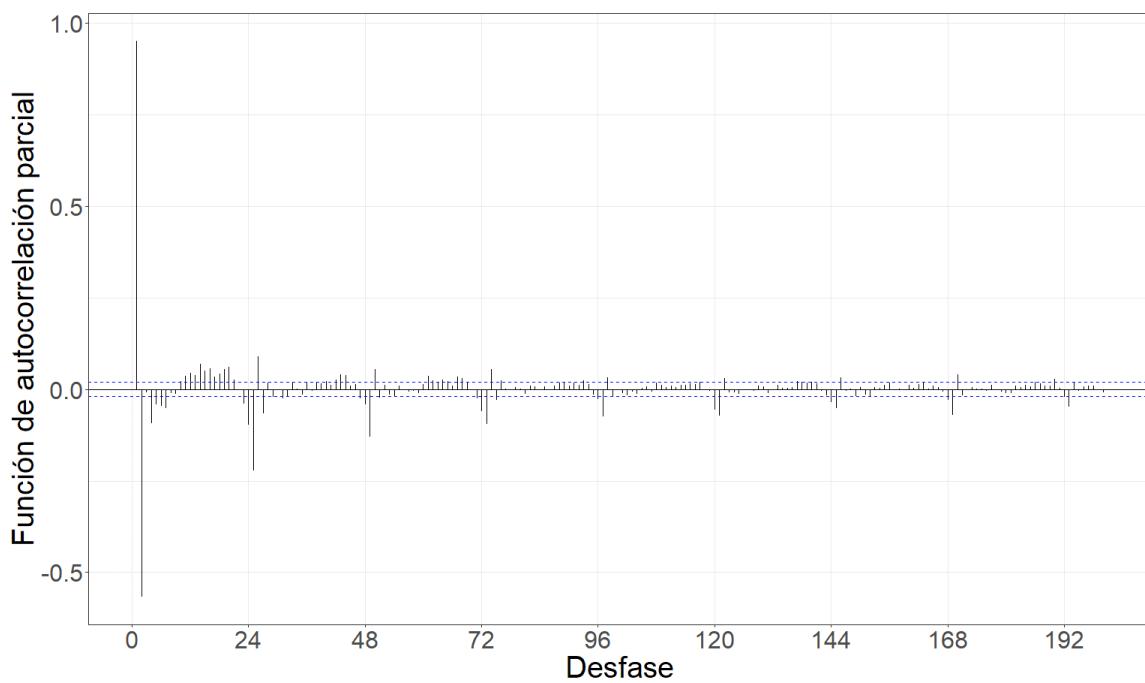


Figura 5.49: Gráfico de la función de autocorrelación parcial de la serie.

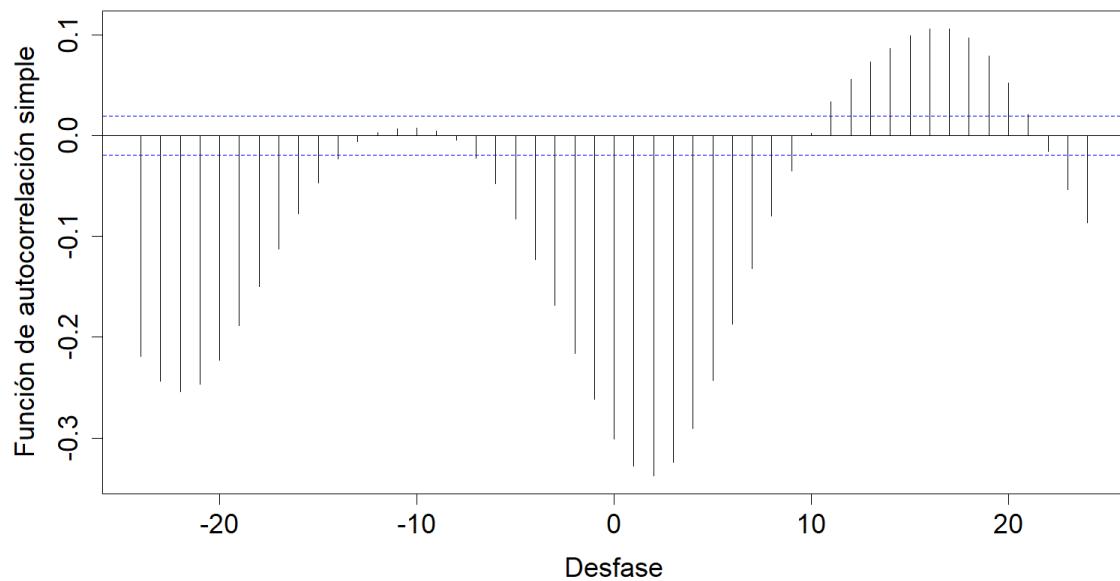


Figura 5.50: Gráfico de la correlación cruzada entre las series previo al filtrado por el modelo teórico ajustado.

Tabla 5.12: Modelo teórico ajustado.

Modelo	AIC	AICc	BIC
ARIMA(4,0,0)(2,1,0)[24]	34503.31	34503.32	34553.93

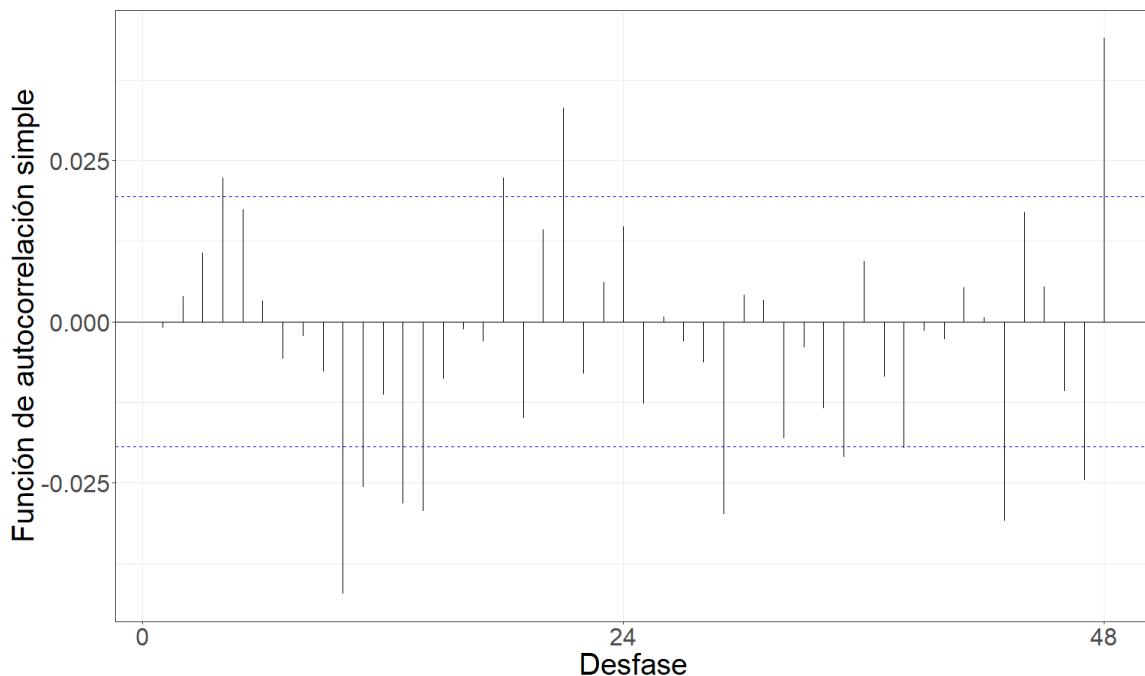


Figura 5.51: Gráfico de la función de autocorrelación simple de la serie filtrada por el modelo teórico ajustado.

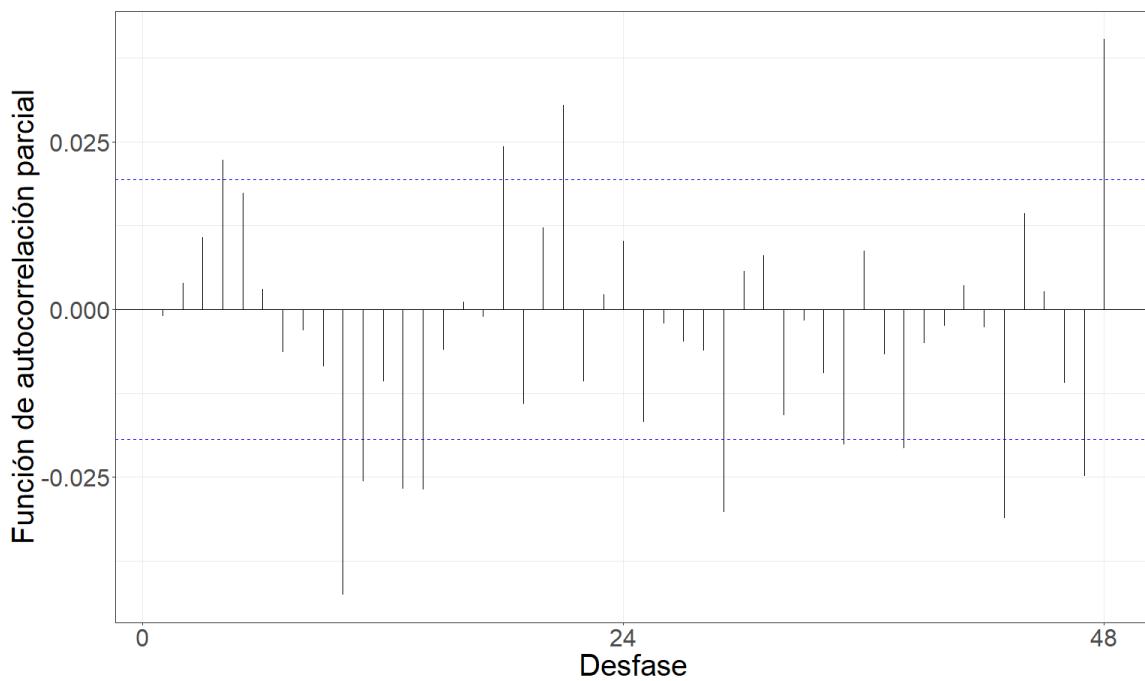


Figura 5.52: Gráfico de la función de autocorrelación parcial de la serie filtrada por el modelo teórico ajustado.

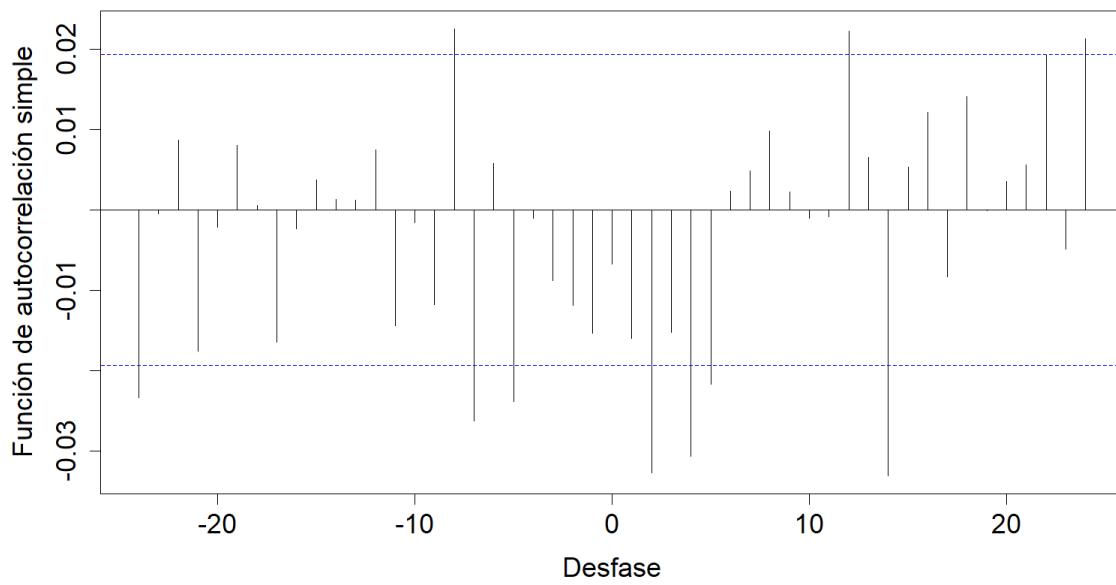


Figura 5.53: Gráfico de la correlación cruzada entre las series filtradas por el modelo teórico ajustado.

Tabla 5.13: Coeficientes estimados de la relación funcional.

Desfasaje	Con diferenciación	Sin diferenciación
t-5	0.1016645	-0.1288
t-7	0.1317611	-0.03429
t-8	0.1445327	-0.006416
t-24	-0.1998146	-0.3433

Presión atmosférica

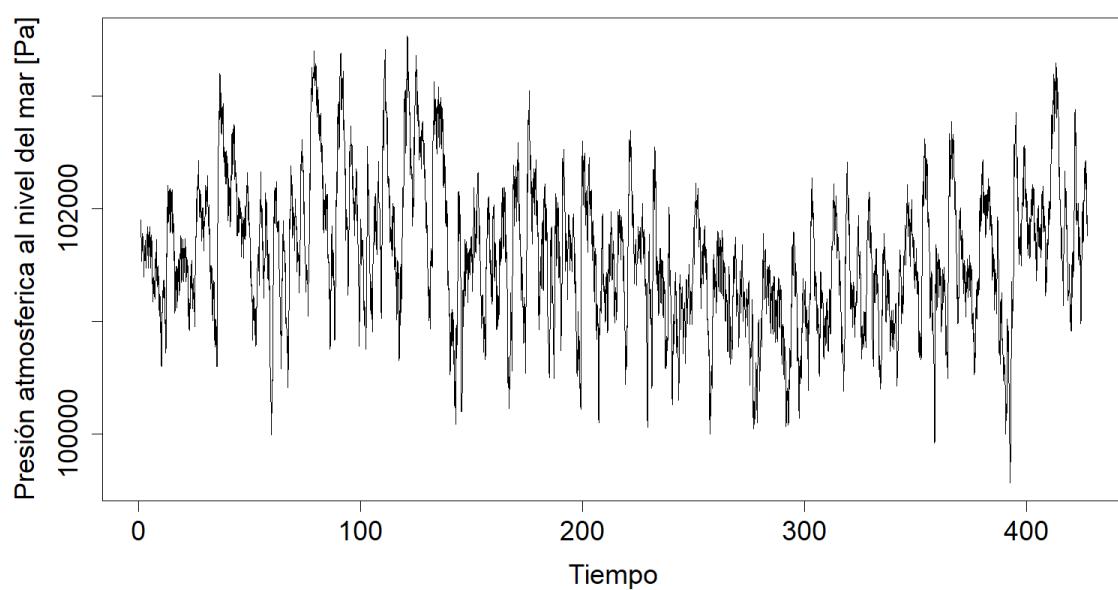


Figura 5.54: Gráfico de la serie horaria promedio de presión atmosférica al nivel del mar.

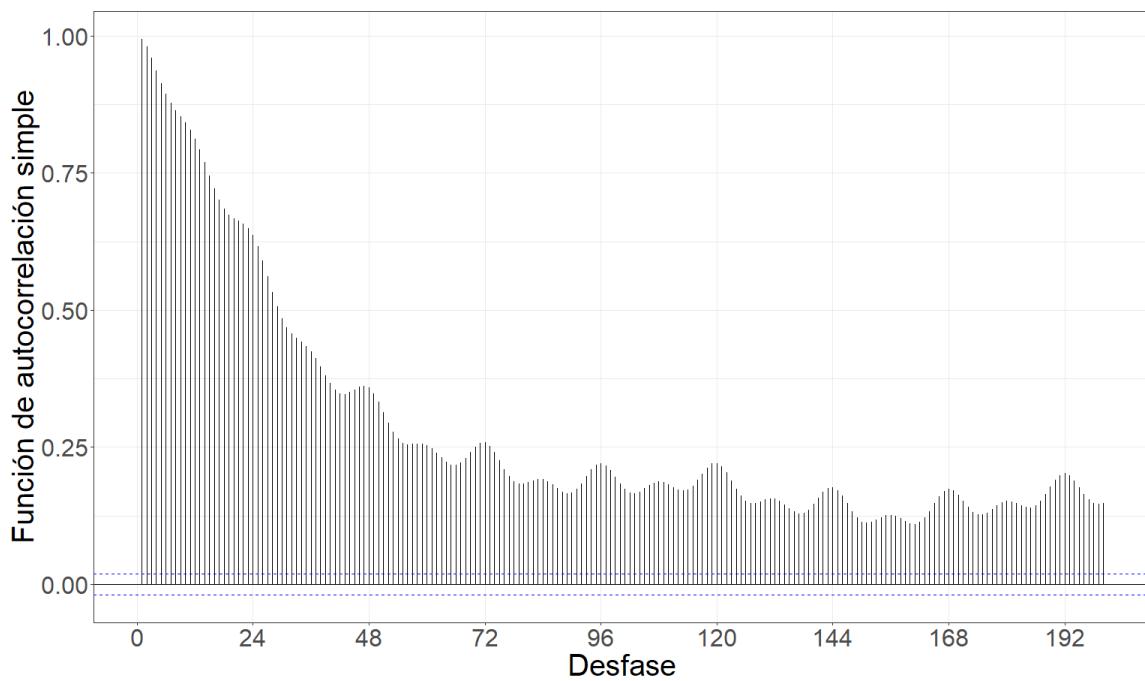


Figura 5.55: Gráfico de la función de autocorrelación simple de la serie.

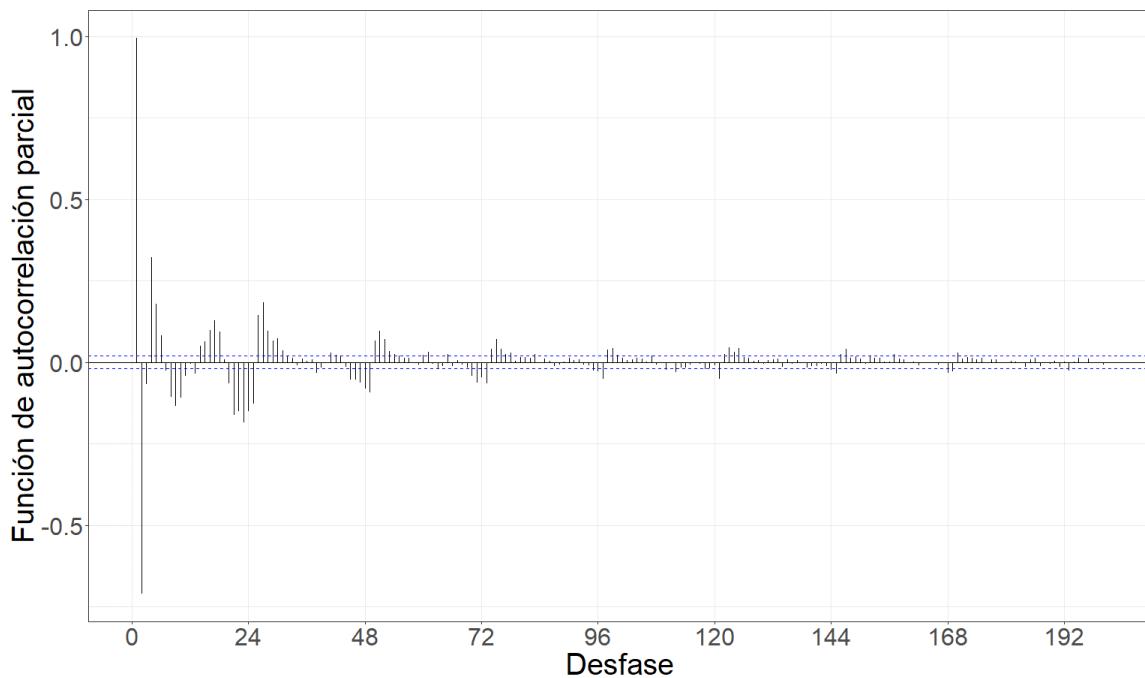


Figura 5.56: Gráfico de la función de autocorrelación parcial de la serie.

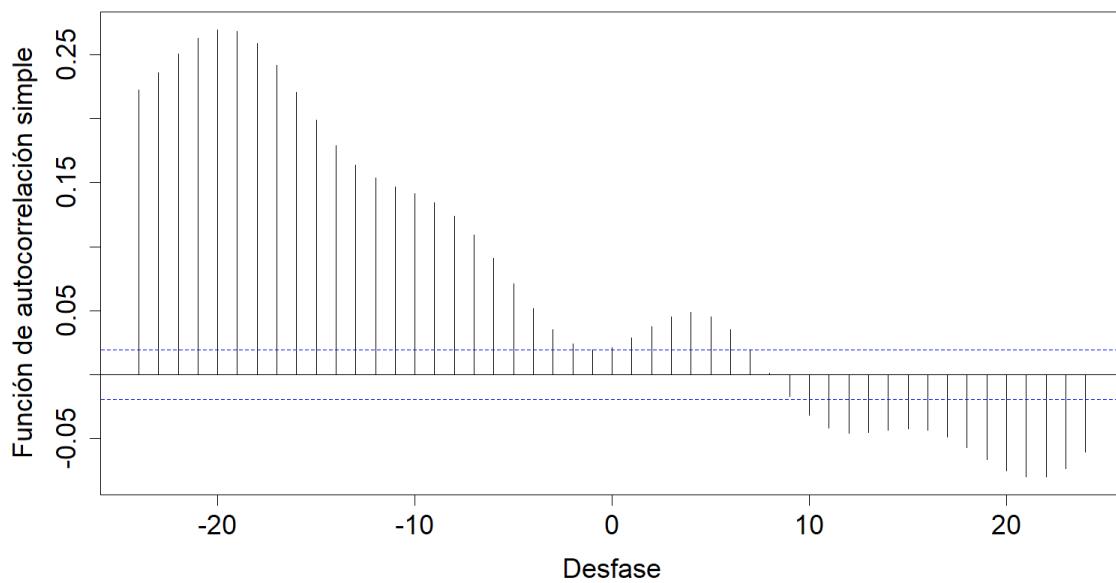


Figura 5.57: Gráfico de la correlación cruzada entre las series previo al filtrado por el modelo teórico ajustado.

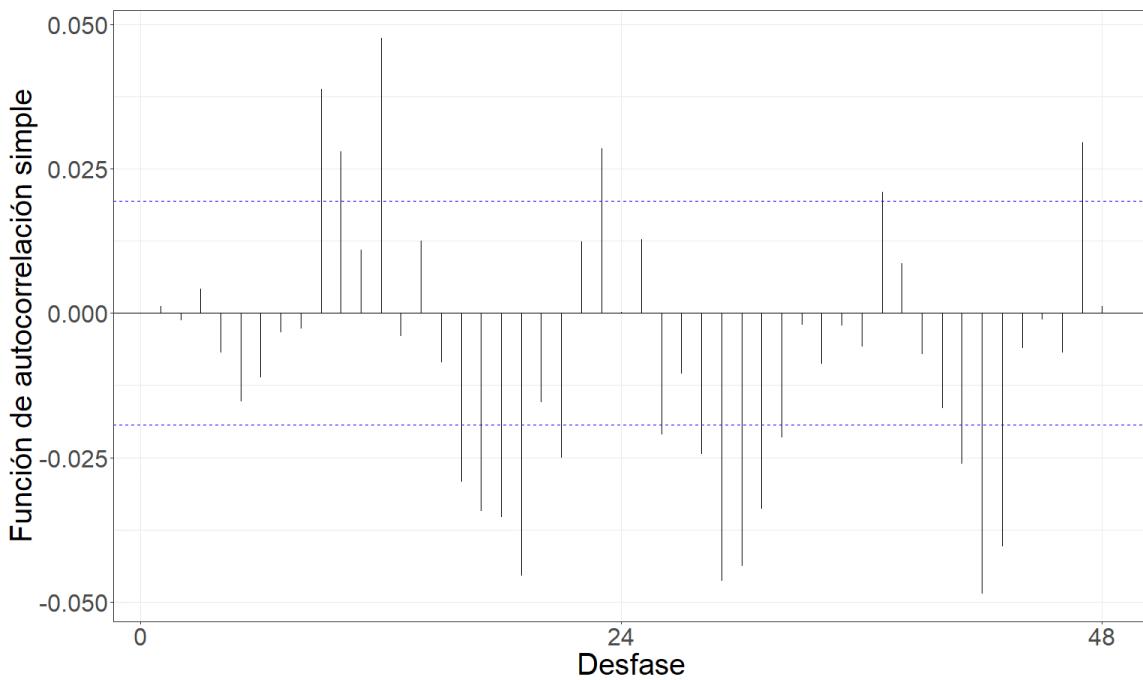


Figura 5.58: Gráfico de la función de autocorrelación simple de la serie filtrada por el modelo teórico ajustado.

Tabla 5.14: Modelo teórico ajustado.

Modelo	AIC	AICc	BIC
ARIMA(4,1,2)(2,1,2)[24]	104424.1	104424.1	104474.7

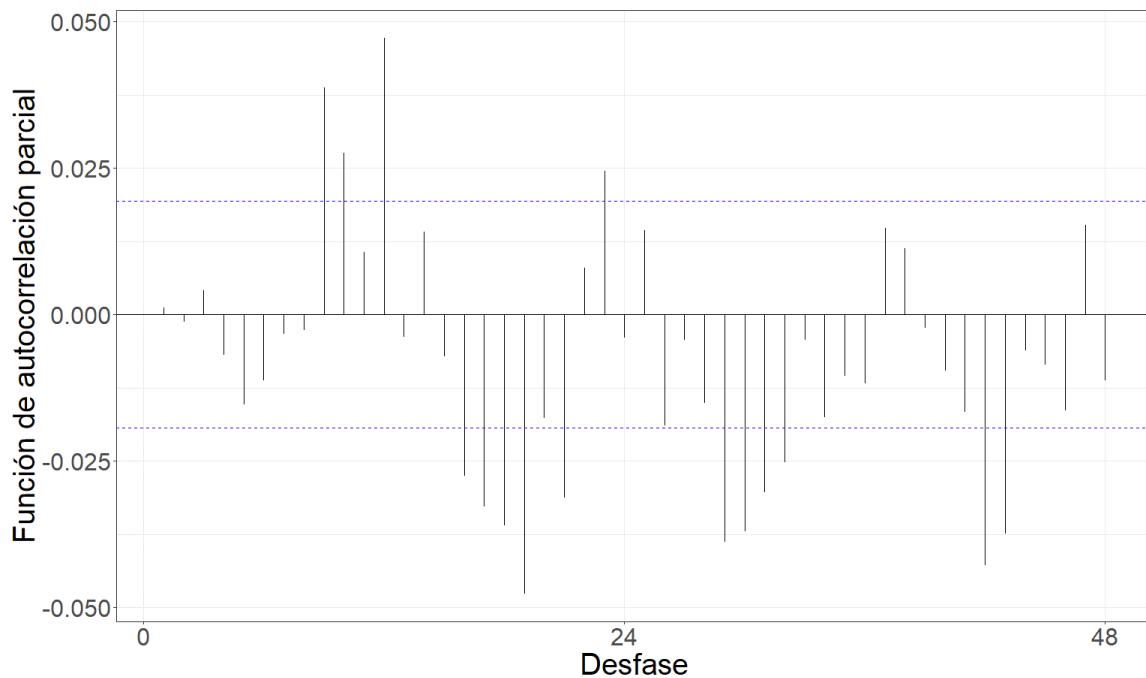


Figura 5.59: Gráfico de la función de autocorrelación parcial de la serie filtrada por el modelo teórico ajustado.

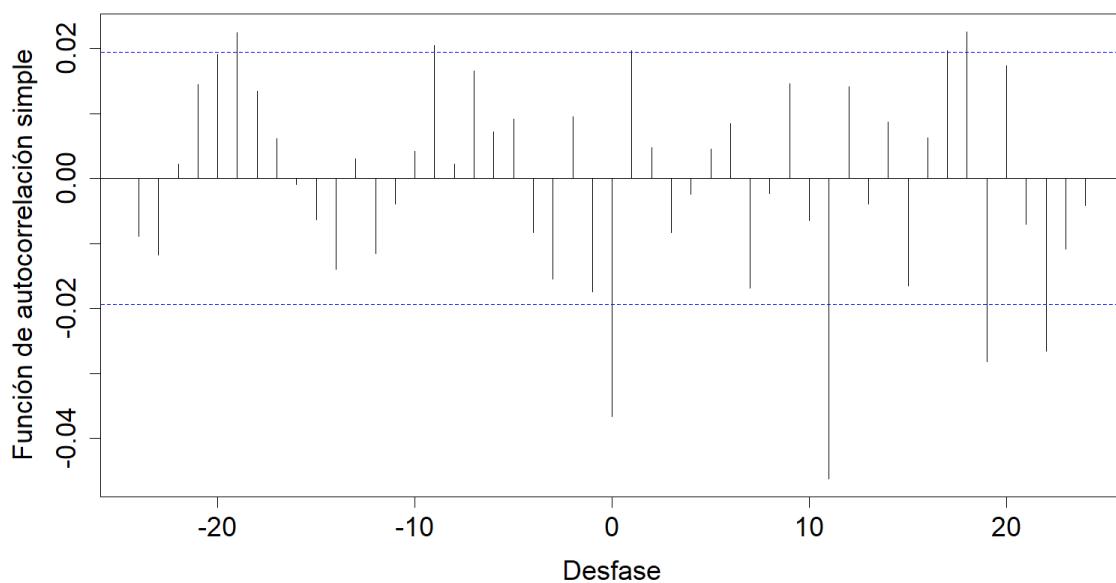


Figura 5.60: Gráfico de la correlación cruzada entre las series filtradas por el modelo teórico ajustado.

Tabla 5.15: Coeficientes estimados de la relación funcional.

Desfasaje	Con diferenciación	Sin diferenciación
t	-0.003024	-3.434e-03
t-9	0.002529	-7.623e-04
t-19	0.008081	6.478e-03

Radiación solar

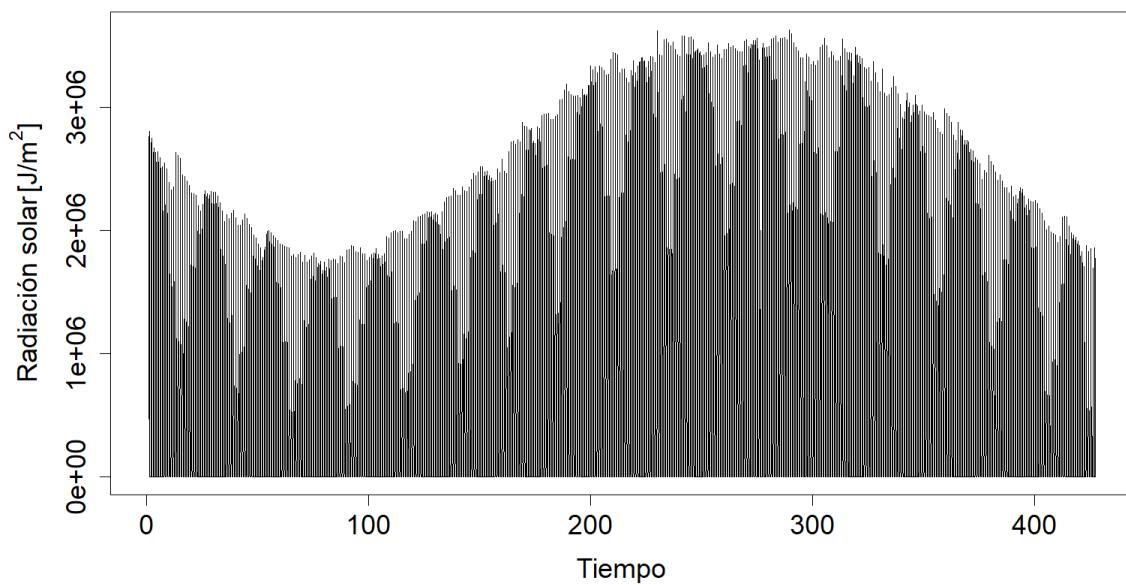


Figura 5.61: Gráfico de la serie horaria promedio de radiación solar.

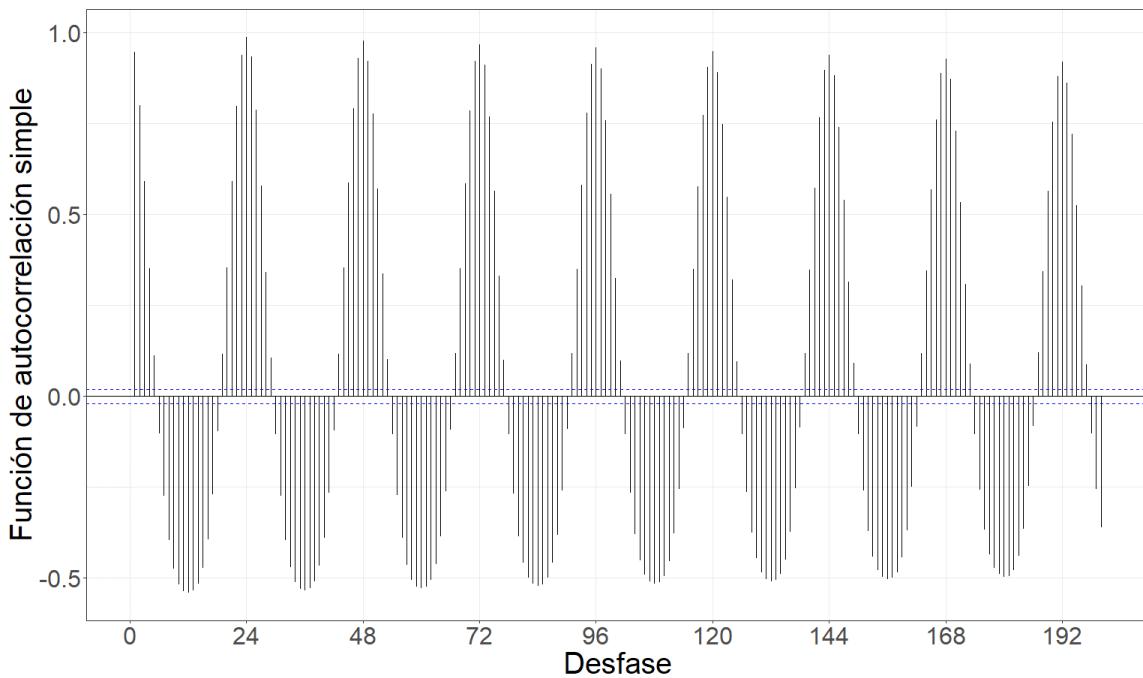


Figura 5.62: Gráfico de la función de autocorrelación simple de la serie.

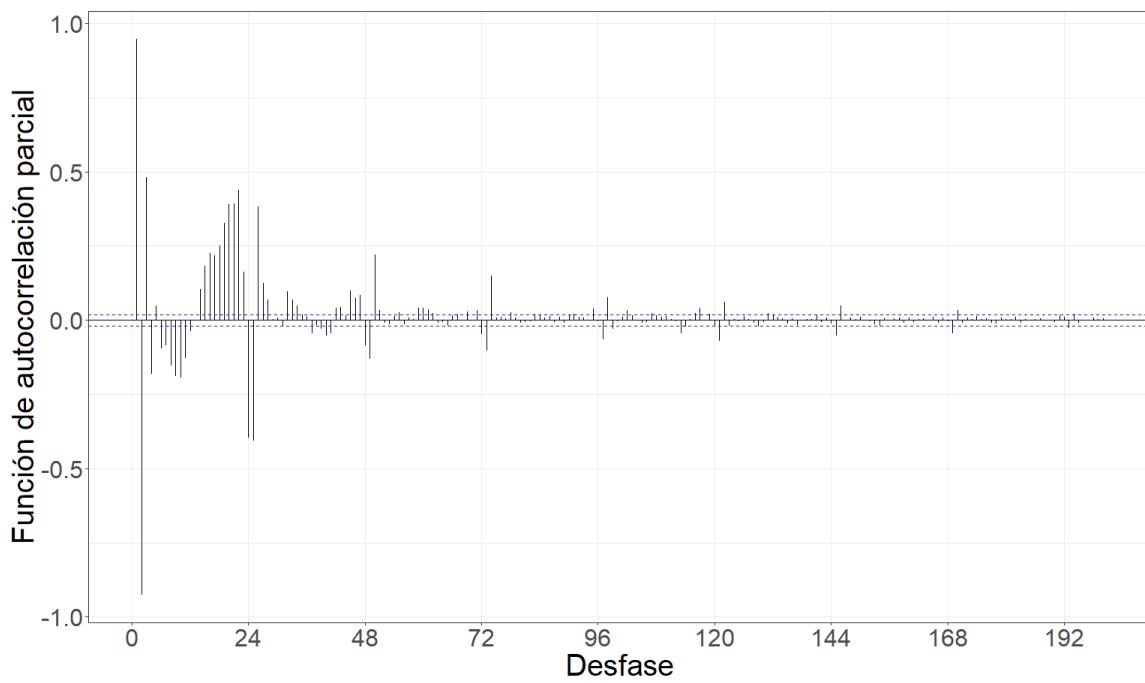


Figura 5.63: Gráfico de la función de autocorrelación parcial de la serie.

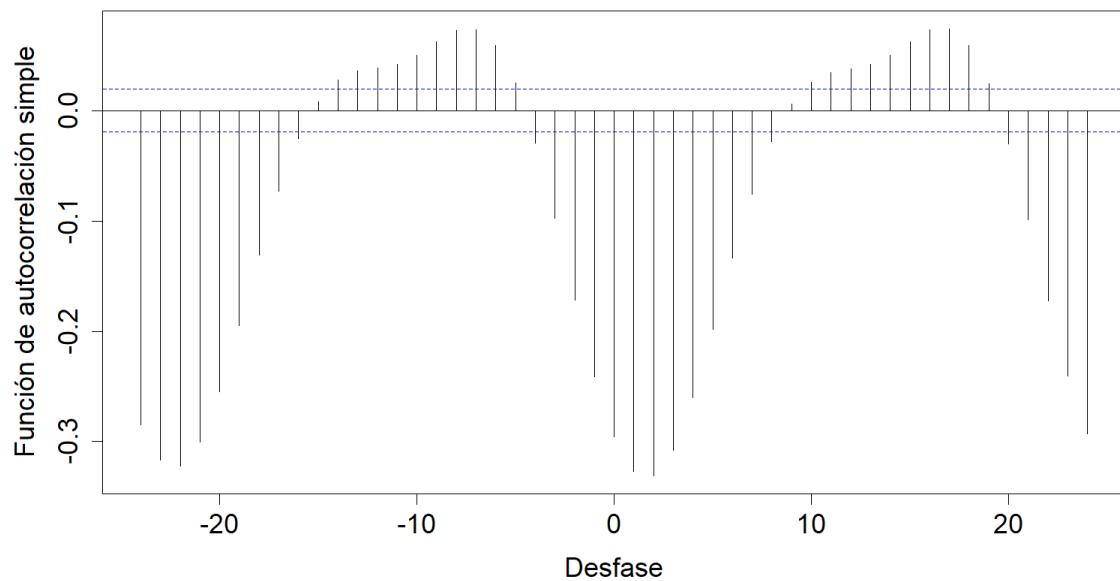


Figura 5.64: Gráfico de la correlación cruzada entre las series previo al filtrado por el modelo teórico ajustado.

Tabla 5.16: Modelo teórico ajustado.

Modelo	AIC	AICc	BIC
ARIMA(2,2,1)(4,3,2)[24]	248262.3	248262.3	248327.4

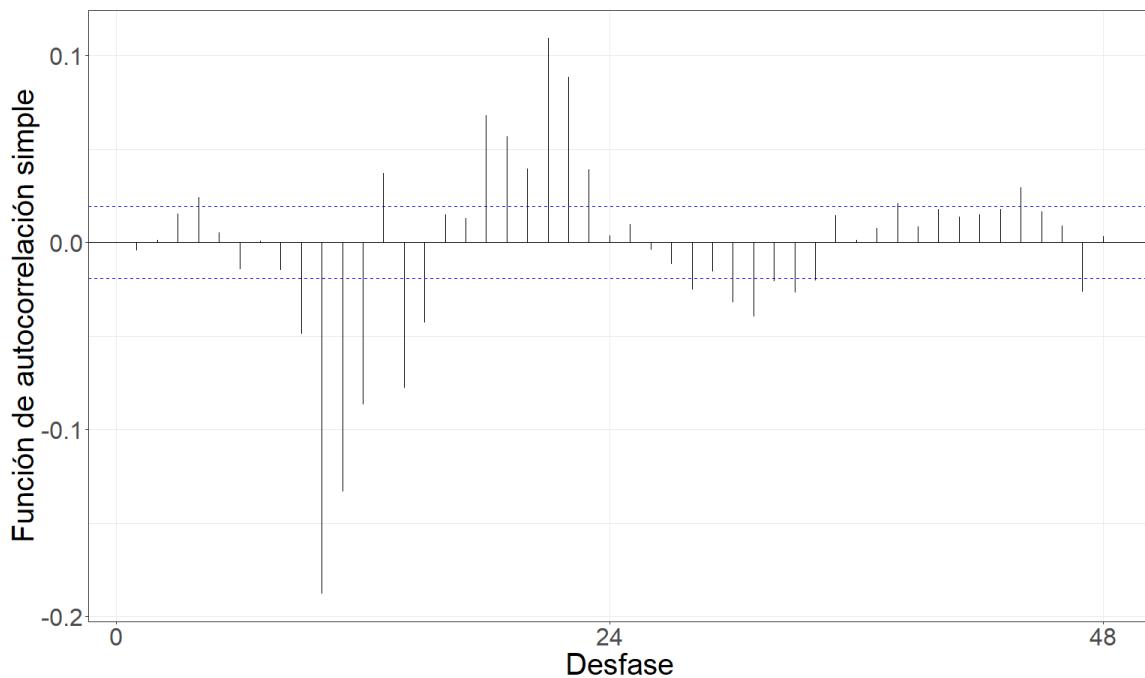


Figura 5.65: Gráfico de la función de autocorrelación simple de la serie filtrada por el modelo teórico ajustado.

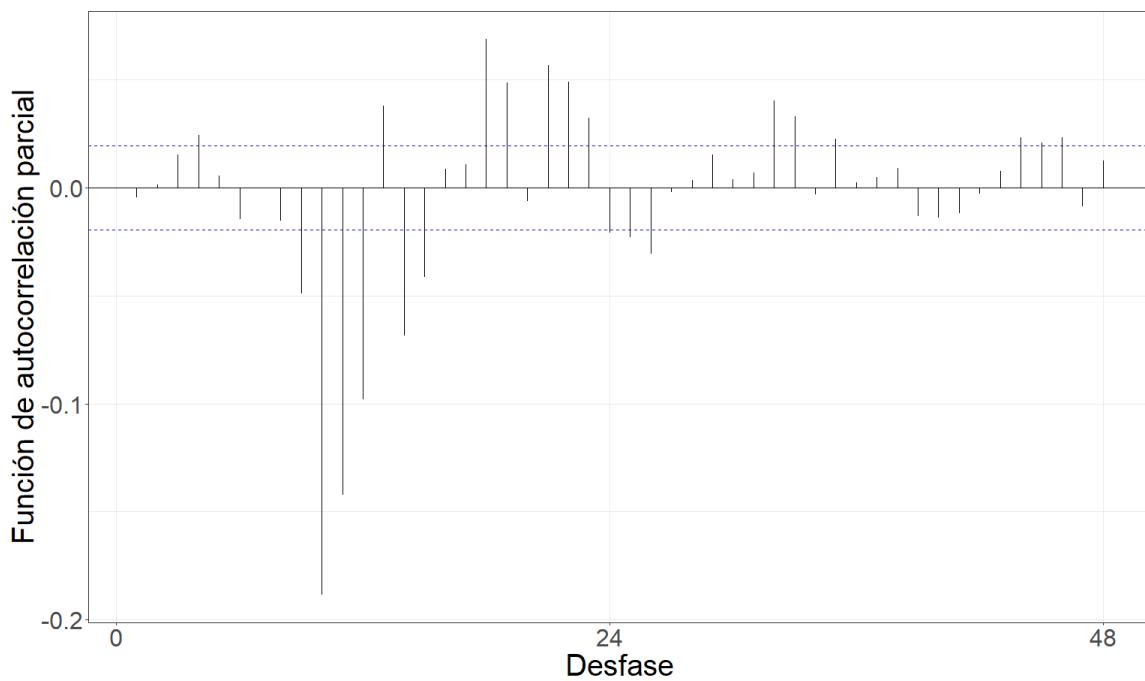


Figura 5.66: Gráfico de la función de autocorrelación parcial de la serie filtrada por el modelo teórico ajustado.

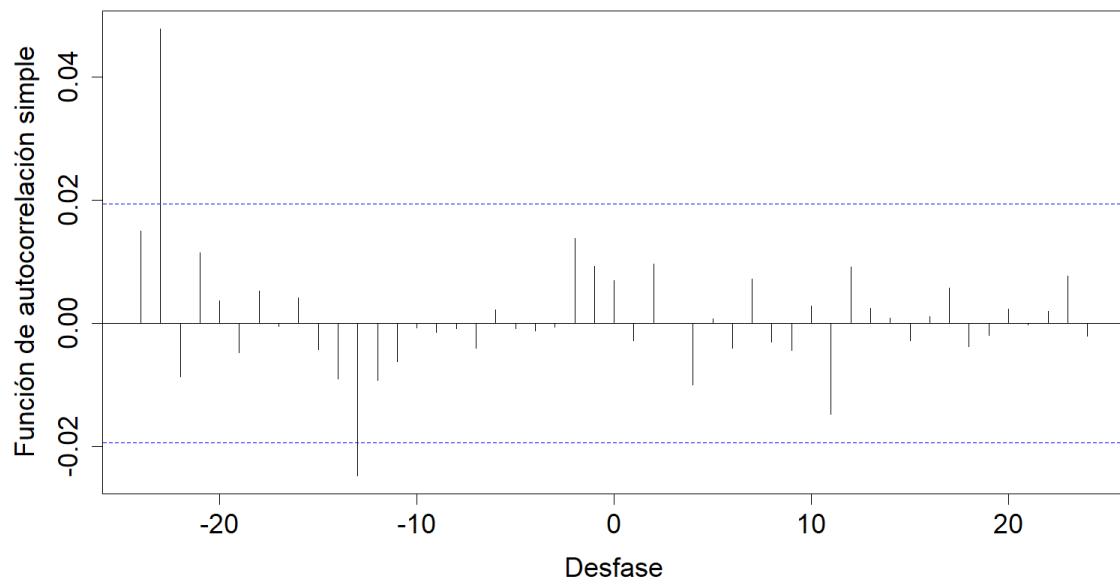


Figura 5.67: Gráfico de la correlación cruzada entre las series filtradas por el modelo teórico ajustado.

Tabla 5.17: Coeficientes estimados de la relación funcional.

Desfasaje	Con diferenciación	Sin diferenciación
t-13	4.023e-07	2.753e-07
t-23	-1.884e-06	-2.413e-06

Profundidad óptica de aerosoles

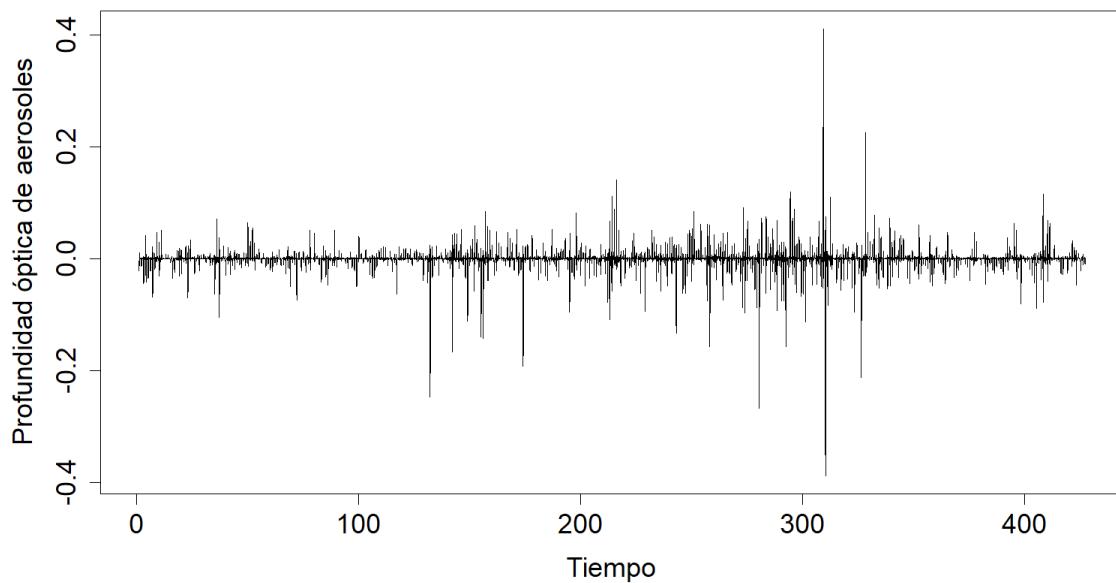


Figura 5.68: Gráfico de la serie horaria promedio de profundidad óptica de aerosoles.

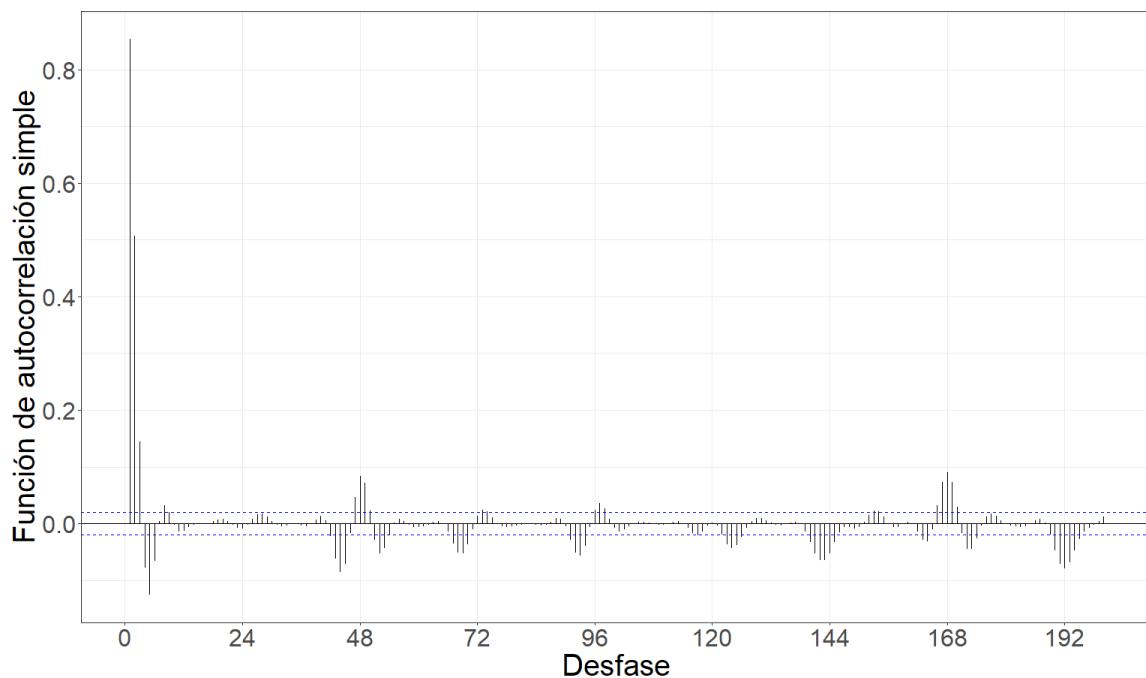


Figura 5.69: Gráfico de la función de autocorrelación simple de la serie.

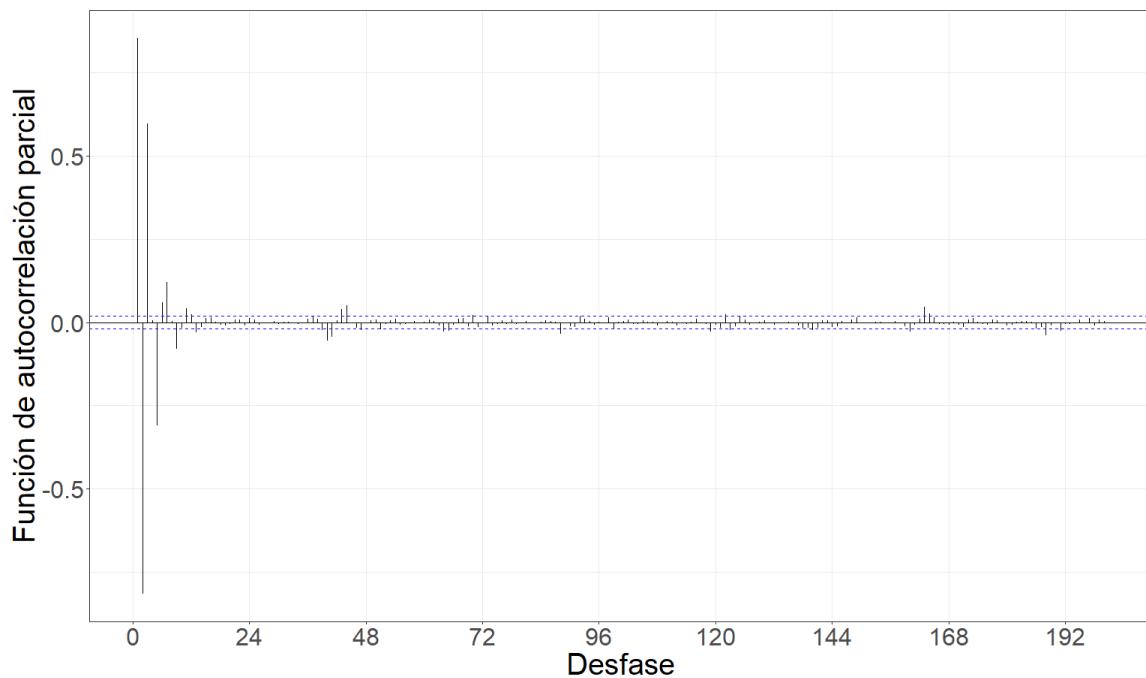


Figura 5.70: Gráfico de la función de autocorrelación parcial de la serie.

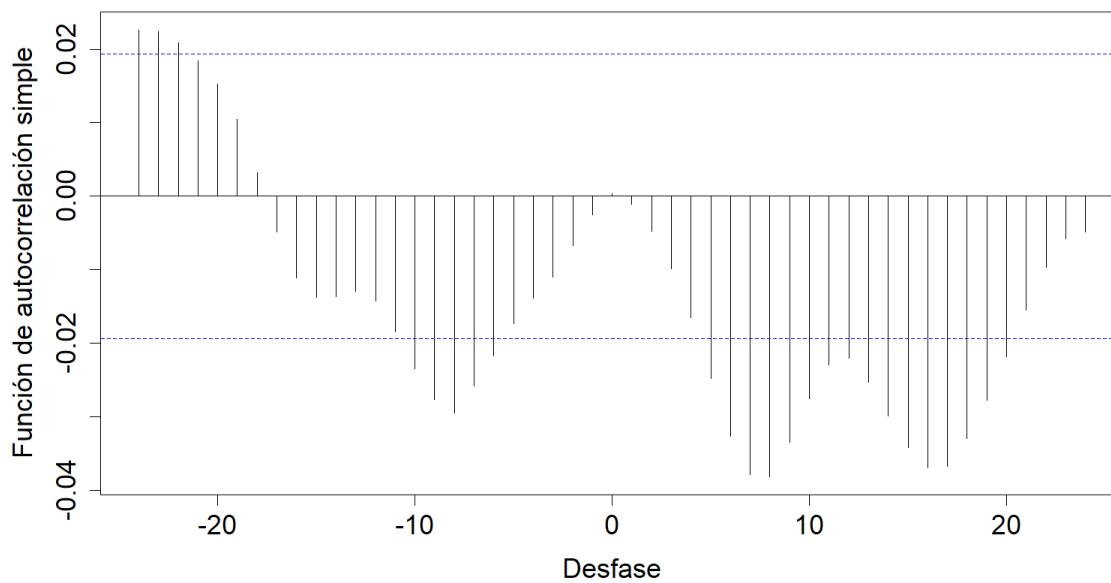


Figura 5.71: Gráfico de la correlación cruzada entre las series previo al filtrado por el modelo teórico ajustado.

Tabla 5.18: Modelo teórico ajustado.

Modelo	AIC	AICc	BIC
ARIMA(3,2,6)(3,1,6)[24]	22579.78	22579.79	22579.79

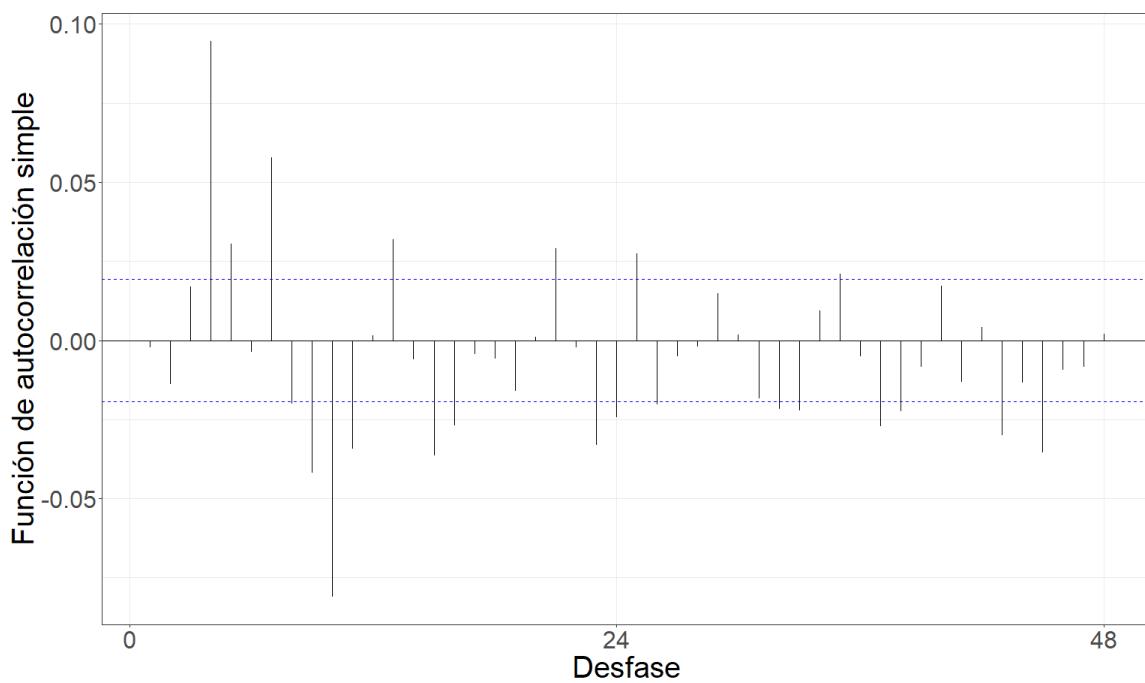


Figura 5.72: Gráfico de la función de autocorrelación simple de la serie filtrada por el modelo teórico ajustado.

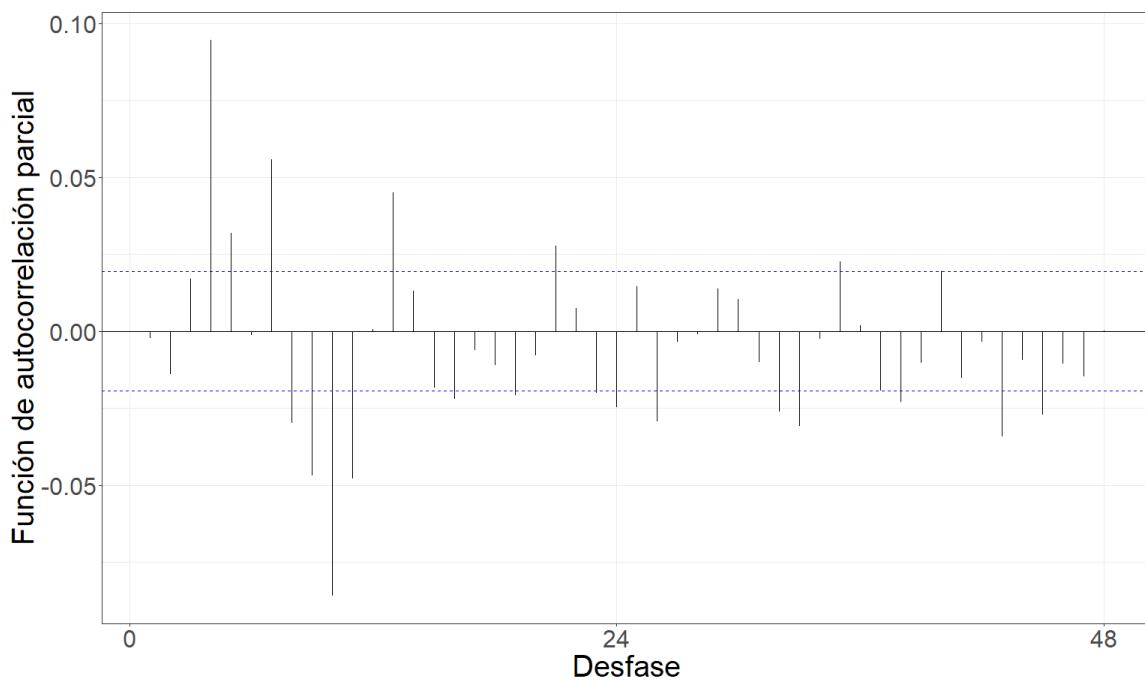


Figura 5.73: Gráfico de la función de autocorrelación parcial de la serie filtrada por el modelo teórico ajustado.

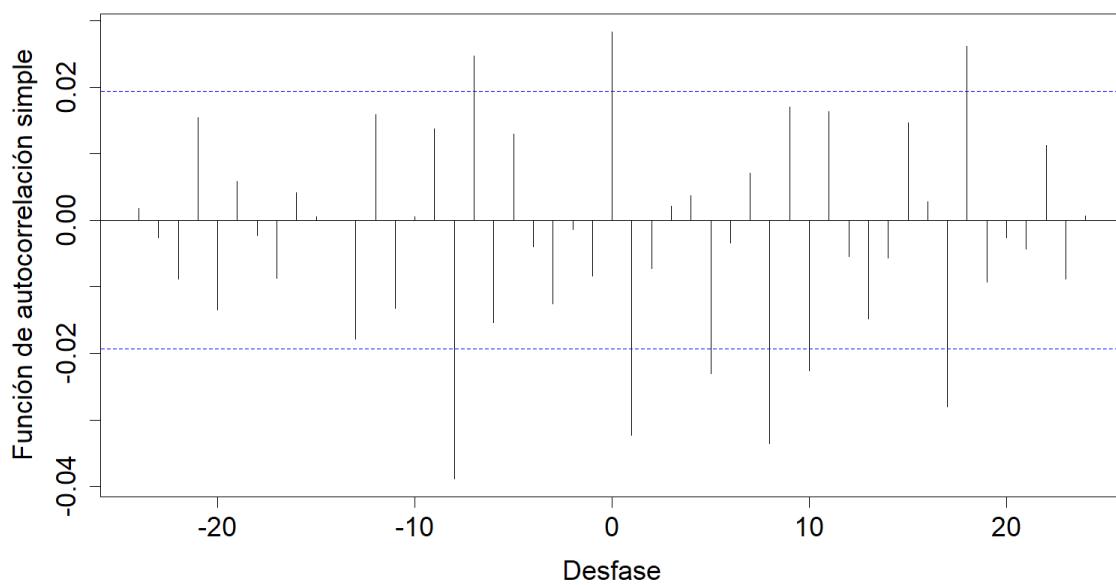


Figura 5.74: Gráfico de la correlación cruzada entre las series filtradas por el modelo teórico ajustado.

Tabla 5.19: Coeficientes estimados de la relación funcional.

Desfasaje	Con diferenciación	Sin diferenciación
t	6.777e+00	0.4331
t-7	7.795e+00	-0.9688
t-8	1.384e+01	-11.15574

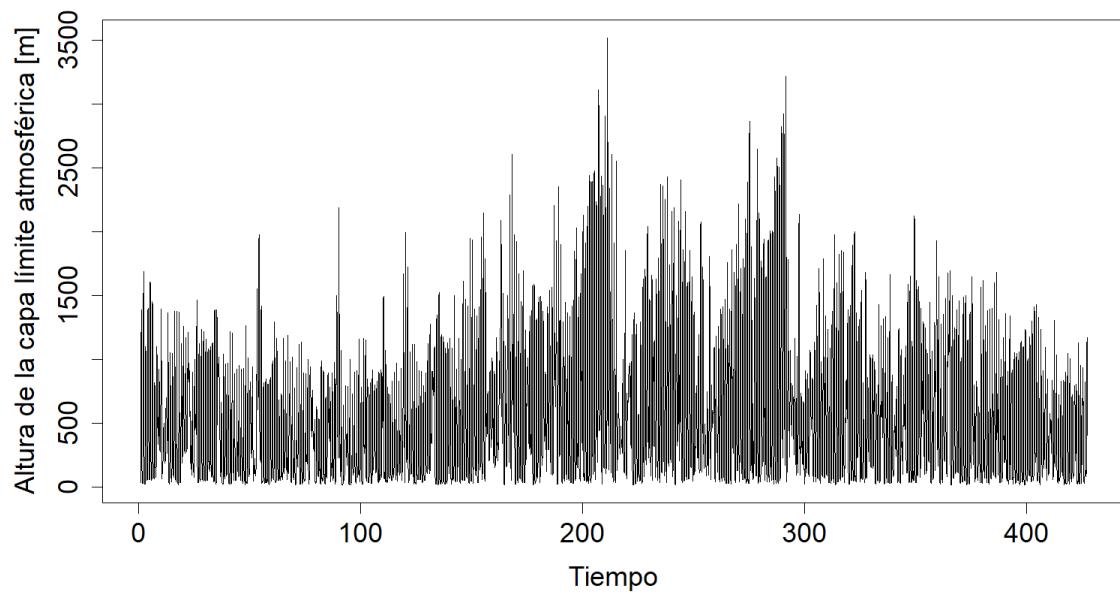
Altura de la capa límite atmosférica

Figura 5.75: Gráfico de la serie horaria promedio de altura de la capa límite atmosférica.

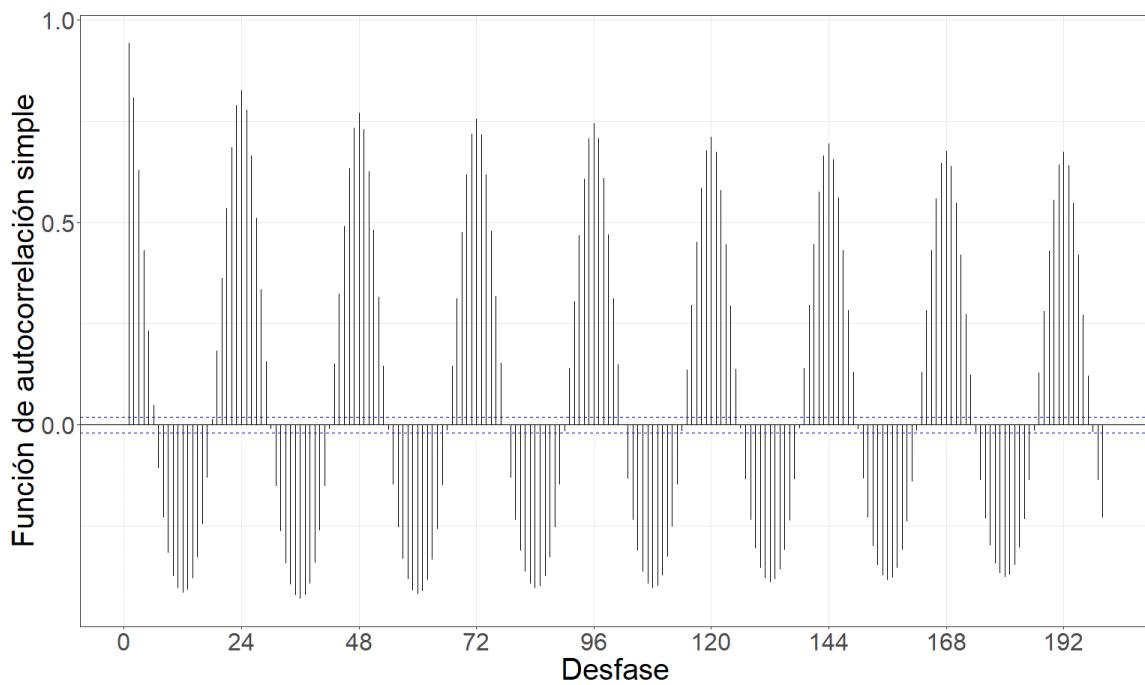


Figura 5.76: Gráfico de la función de autocorrelación simple de la serie.

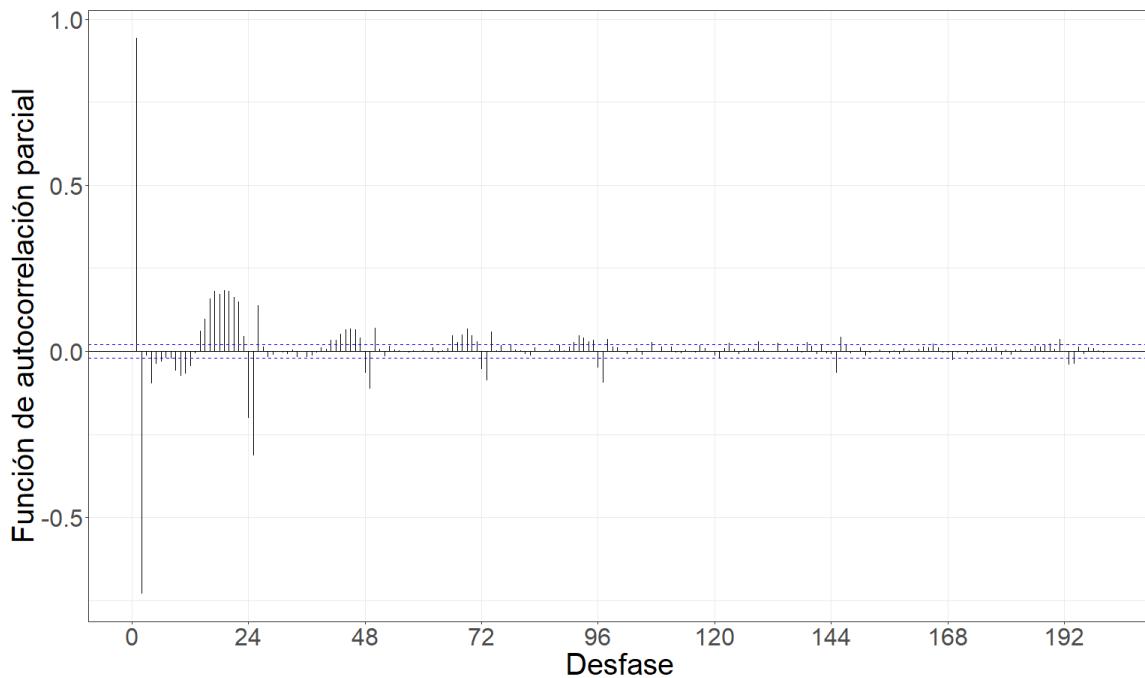


Figura 5.77: Gráfico de la función de autocorrelación parcial de la serie.

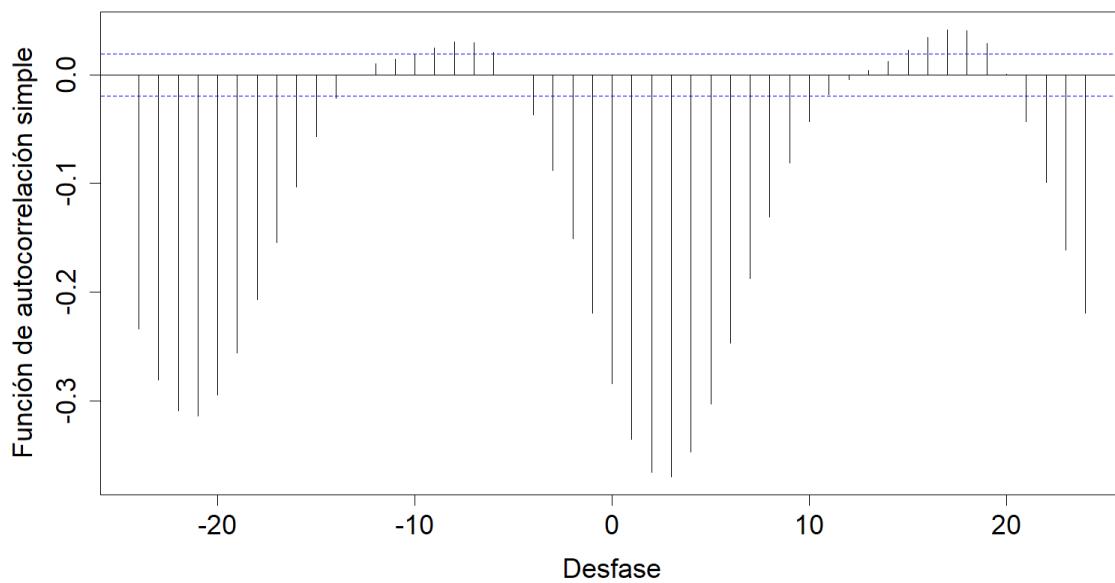


Figura 5.78: Gráfico de la correlación cruzada entre las series previo al filtrado por el modelo teórico ajustado.

Tabla 5.20: Modelo teórico ajustado.

Modelo	AIC	AICc	BIC
ARIMA(2,0,0)(2,1,0)[24]	123028.6	123028.6	123064.7

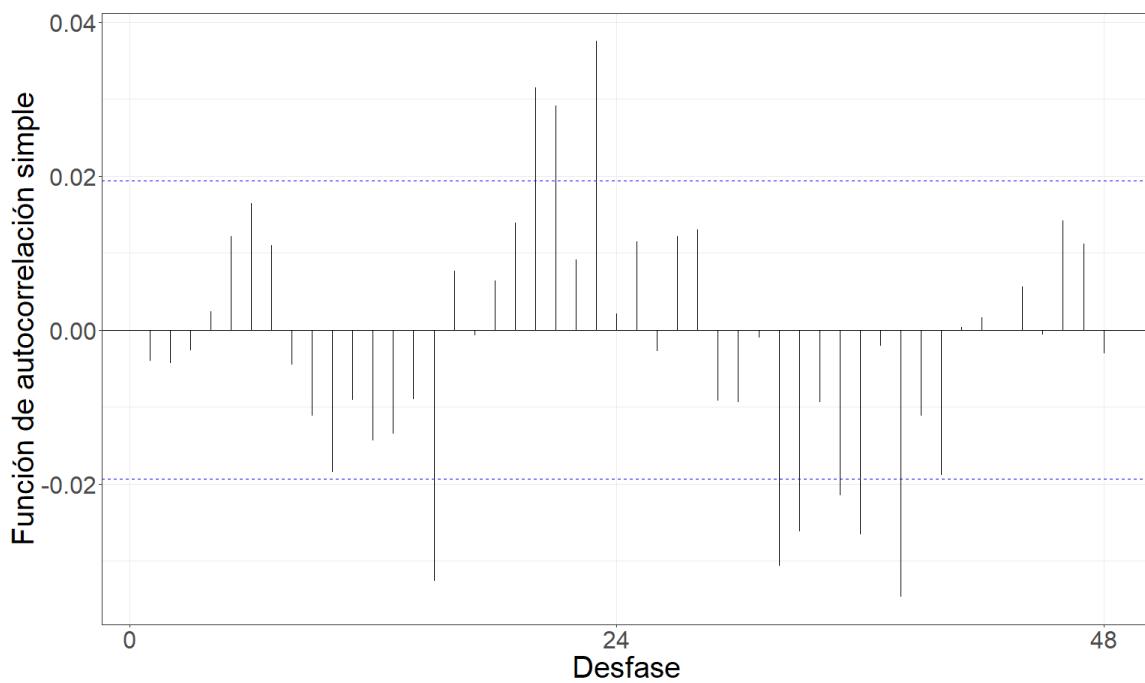


Figura 5.79: Gráfico de la función de autocorrelación simple de la serie filtrada por el modelo teórico ajustado.

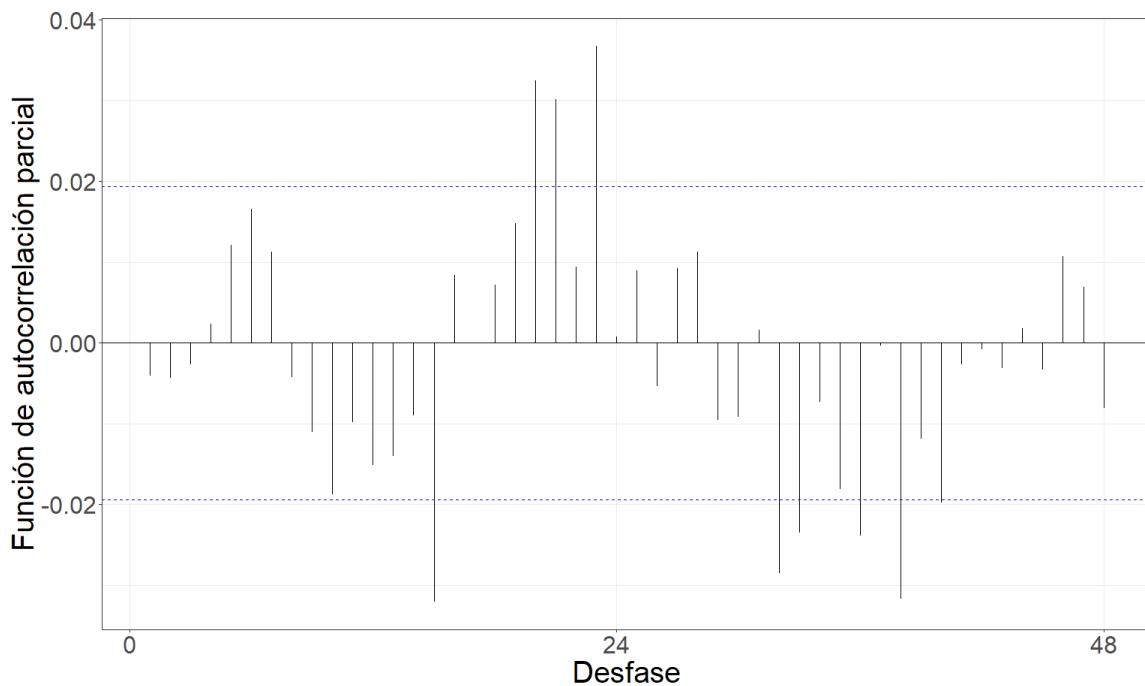


Figura 5.80: Gráfico de la función de autocorrelación parcial de la serie filtrada por el modelo teórico ajustado.

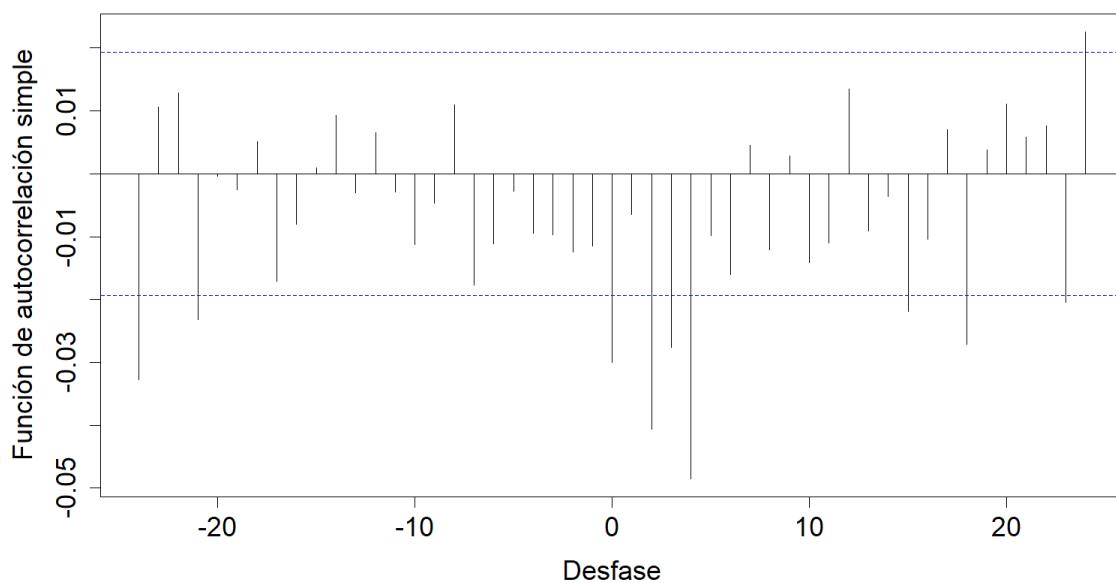


Figura 5.81: Gráfico de la correlación cruzada entre las series filtradas por el modelo teórico ajustado.

Tabla 5.21: Coeficientes estimados de la relación funcional.

Desfasaje	Con diferenciación	Sin diferenciación
t	-0.0019158	-0.00436
t-21	-0.003367	-0.004821
t-24	-1.928e-03	-0.003588

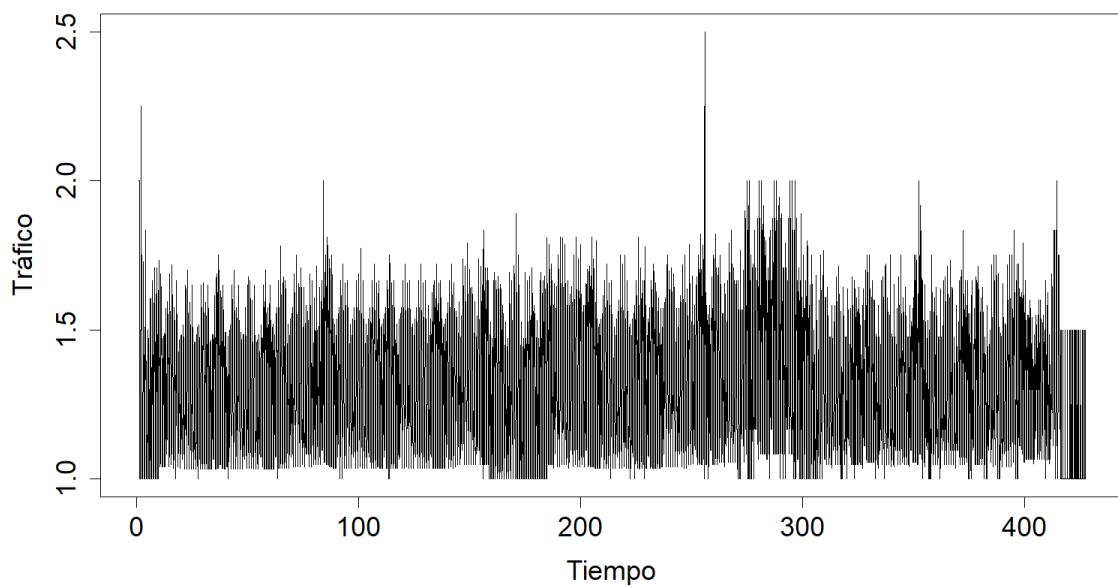
Tráfico

Figura 5.82: Gráfico de la serie horaria promedio de tráfico.

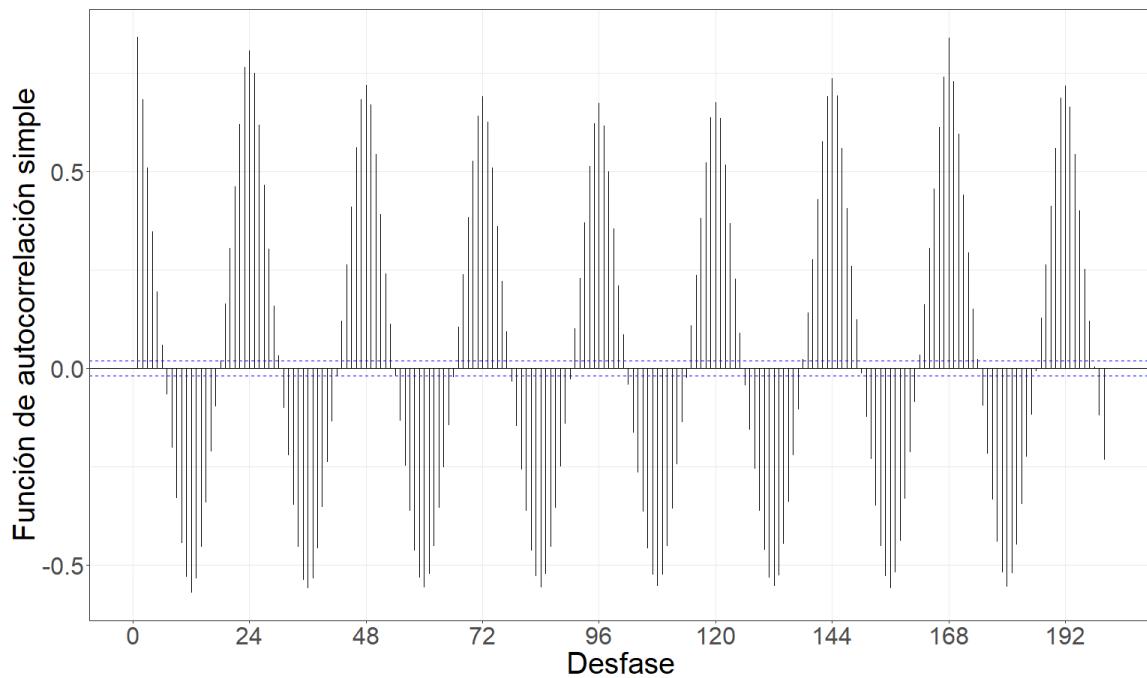


Figura 5.83: Gráfico de la función de autocorrelación simple de la serie.

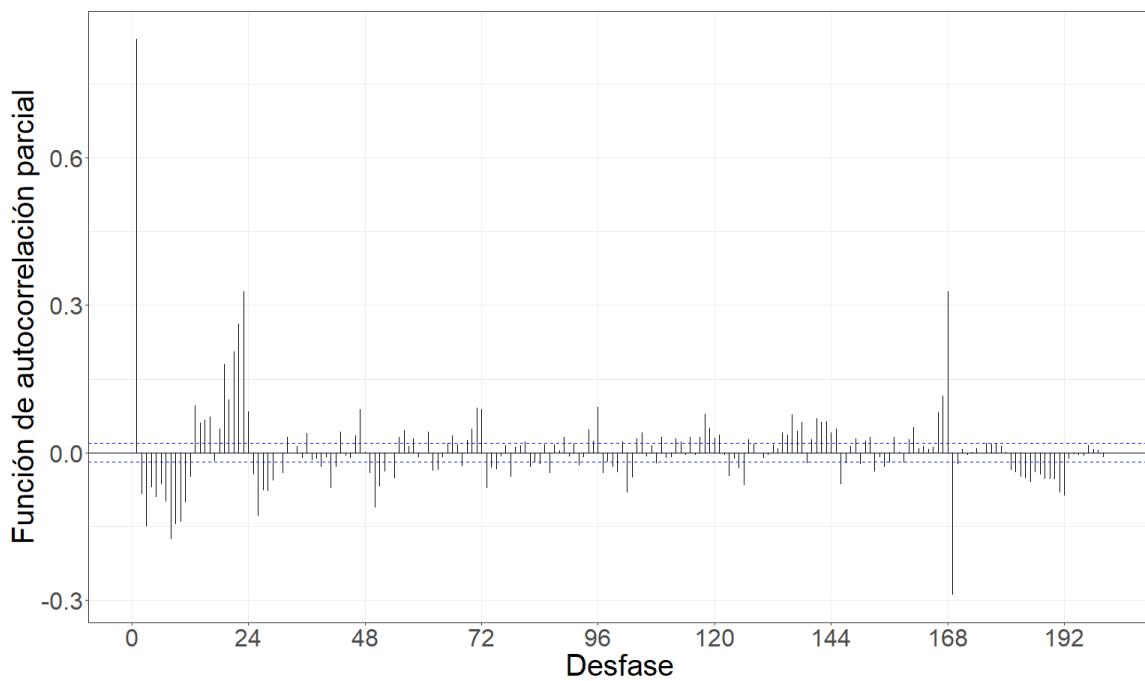


Figura 5.84: Gráfico de la función de autocorrelación parcial de la serie.

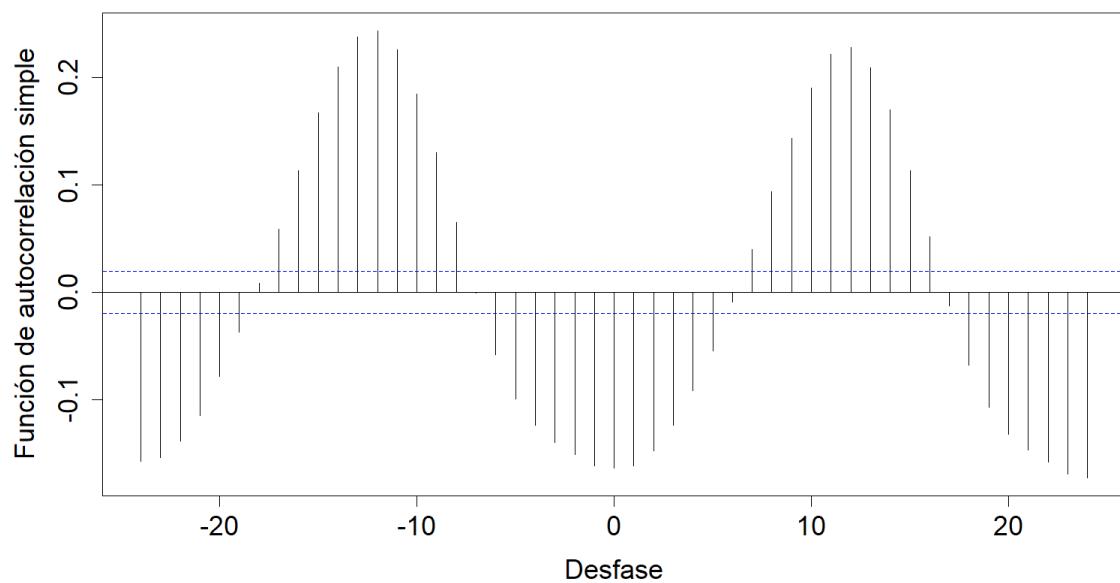


Figura 5.85: Gráfico de la correlación cruzada entre las series previo al filtrado por el modelo teórico ajustado.

Tabla 5.22: Modelo teórico ajustado.

Modelo	AIC	AICc	BIC
ARIMA(3,2,5)(4,2,1)[24]	-17854.73	-17854.72	-17804.12

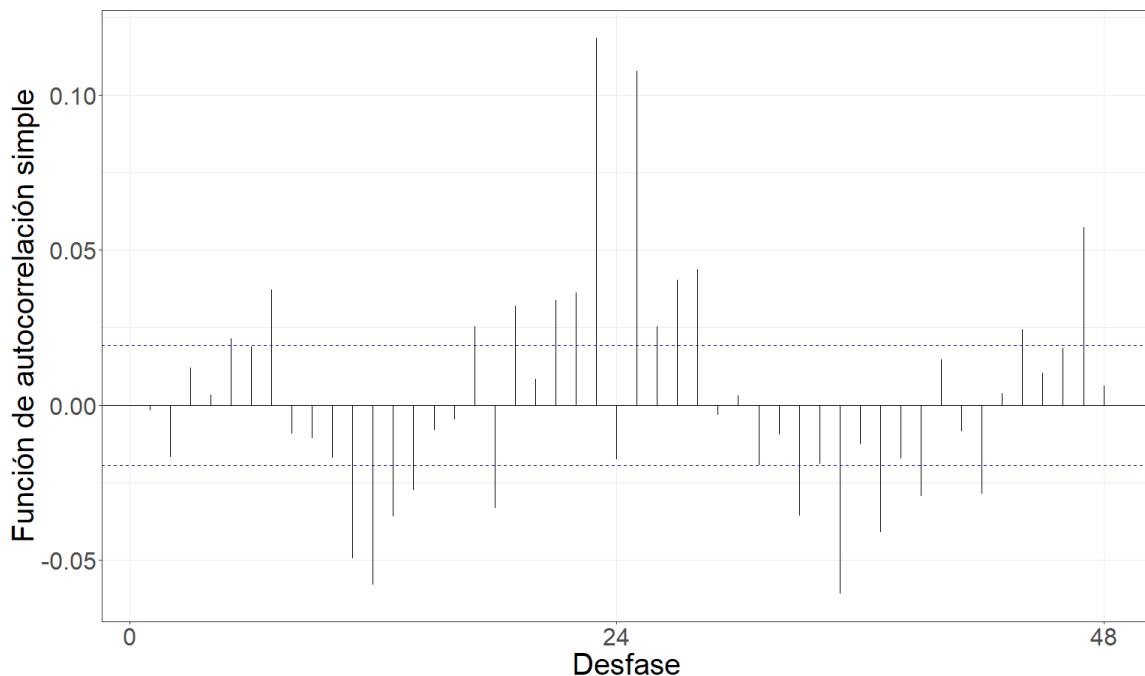


Figura 5.86: Gráfico de la función de autocorrelación simple de la serie filtrada por el modelo teórico ajustado.

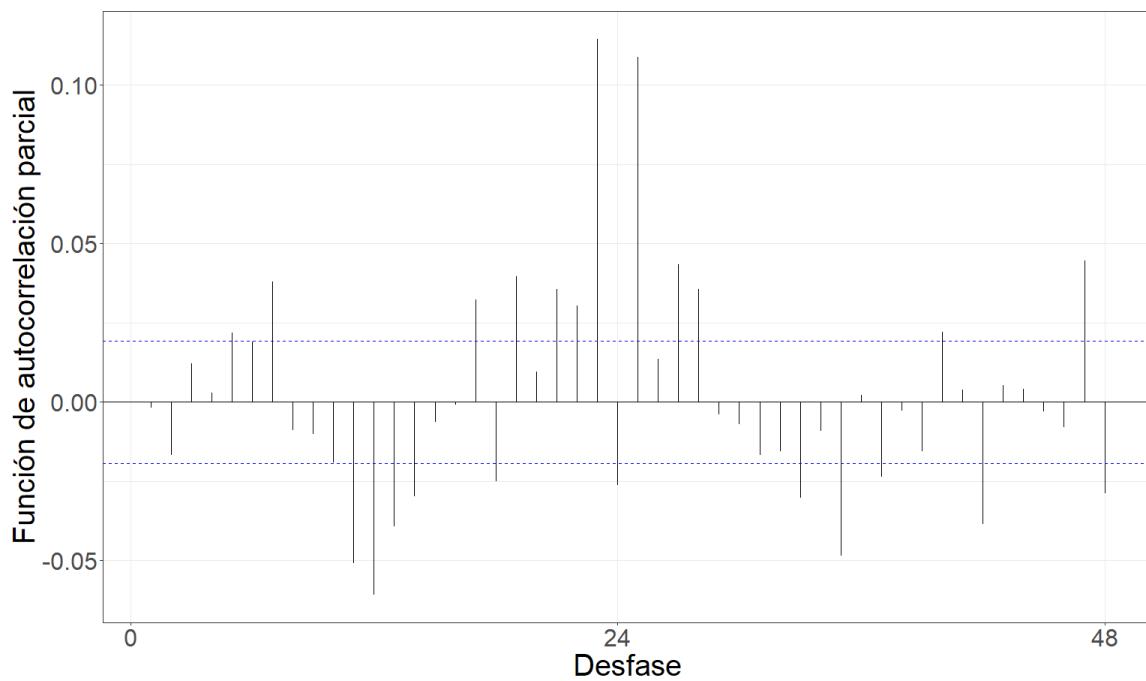


Figura 5.87: Gráfico de la función de autocorrelación parcial de la serie filtrada por el modelo teórico ajustado.

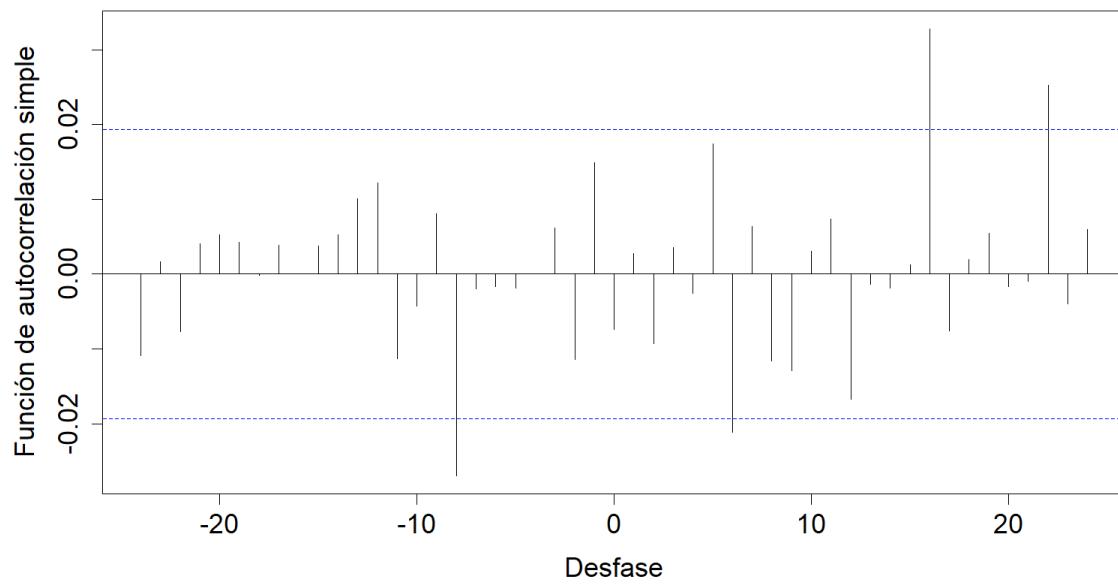


Figura 5.88: Gráfico de la correlación cruzada entre las series filtradas por el modelo teórico ajustado.

Tabla 5.23: Coeficientes estimados de la relación funcional.

Desfasaje	Con diferenciación	Sin diferenciación
t-8	1.085e-01	2.442

Temperatura de la superficie terrestre

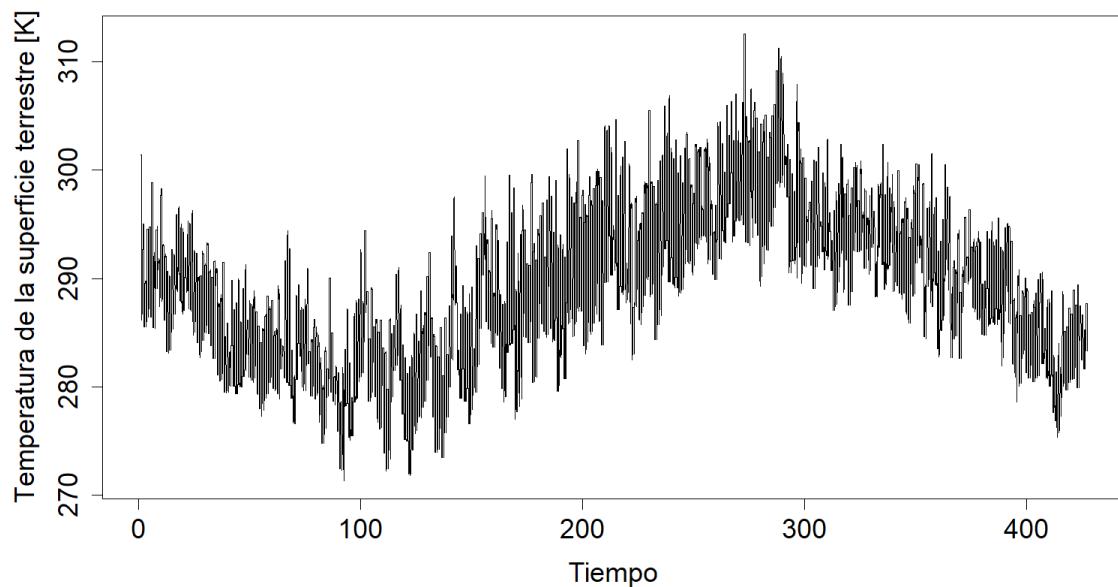


Figura 5.89: Gráfico de la serie horaria promedio de temperatura de la superficie terrestre.

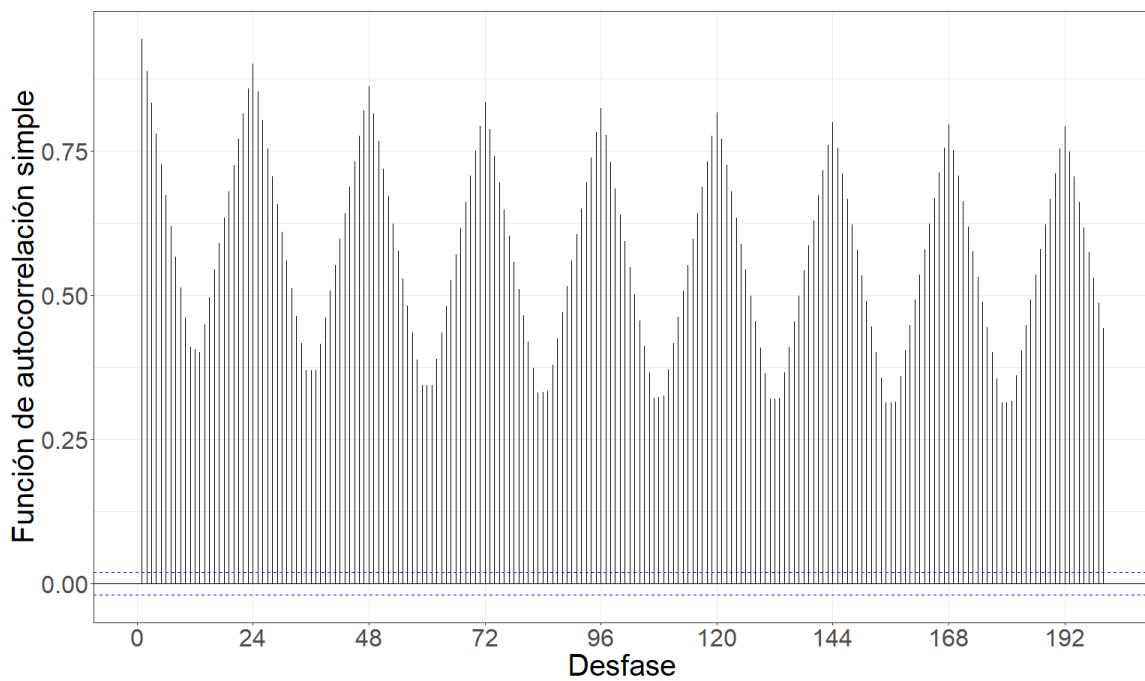


Figura 5.90: Gráfico de la función de autocorrelación simple de la serie.

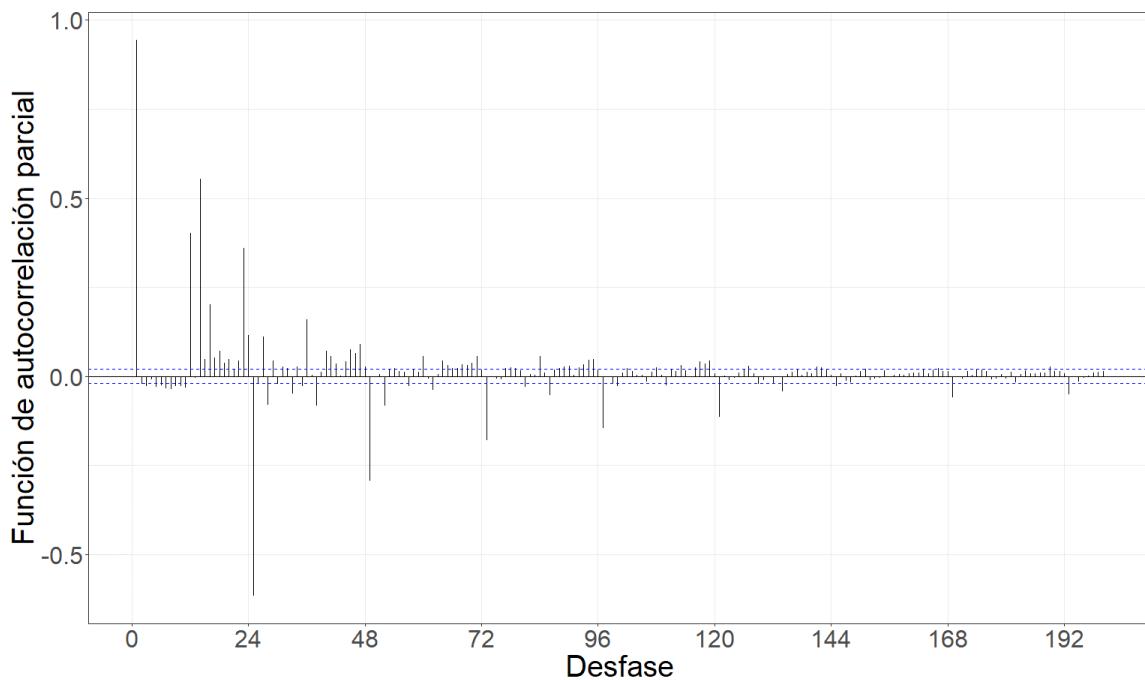


Figura 5.91: Gráfico de la función de autocorrelación parcial de la serie.

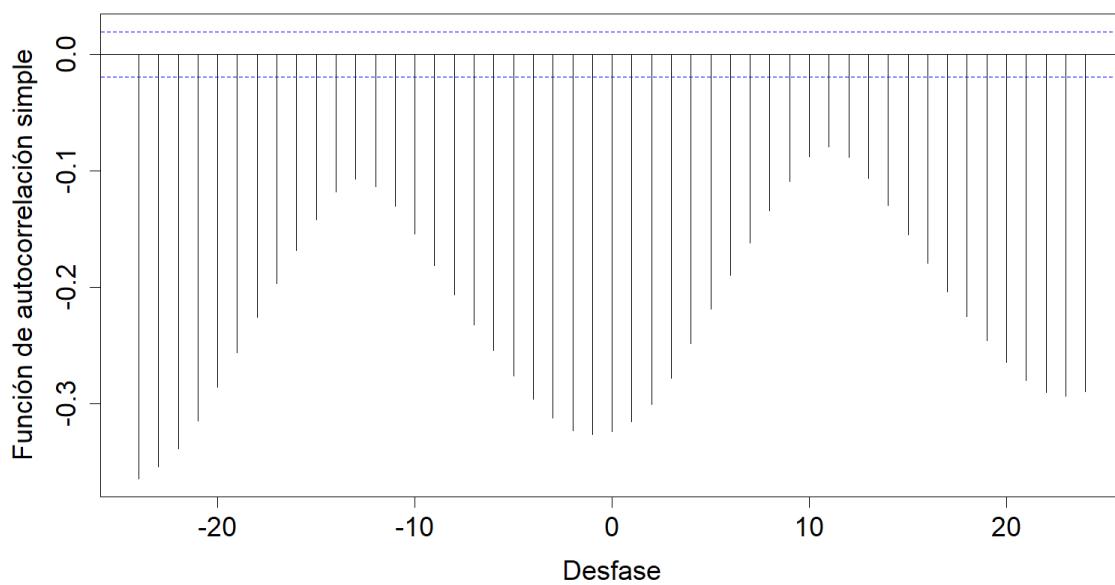


Figura 5.92: Gráfico de la correlación cruzada entre las series previo al filtrado por el modelo teórico ajustado.

Tabla 5.24: Modelo teórico ajustado.

Modelo	AIC	AICc	BIC
ARIMA(6,2,4)(4,1,6)[24]	32396.77	32396.79	32469.08

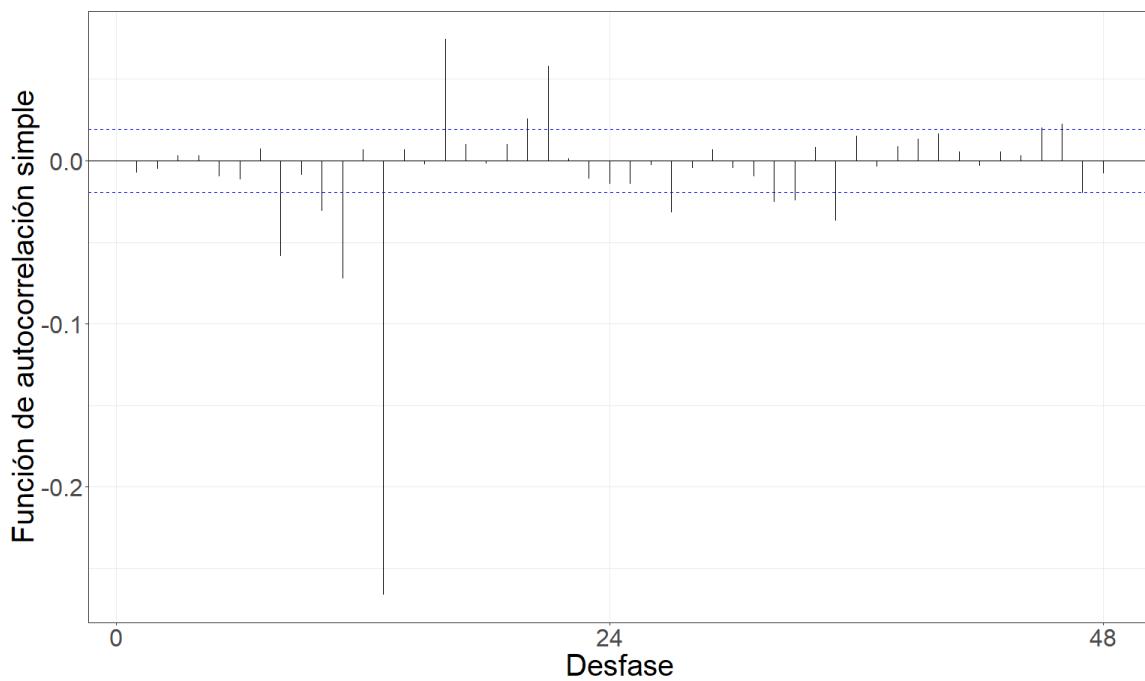


Figura 5.93: Gráfico de la función de autocorrelación simple de la serie filtrada por el modelo teórico ajustado.

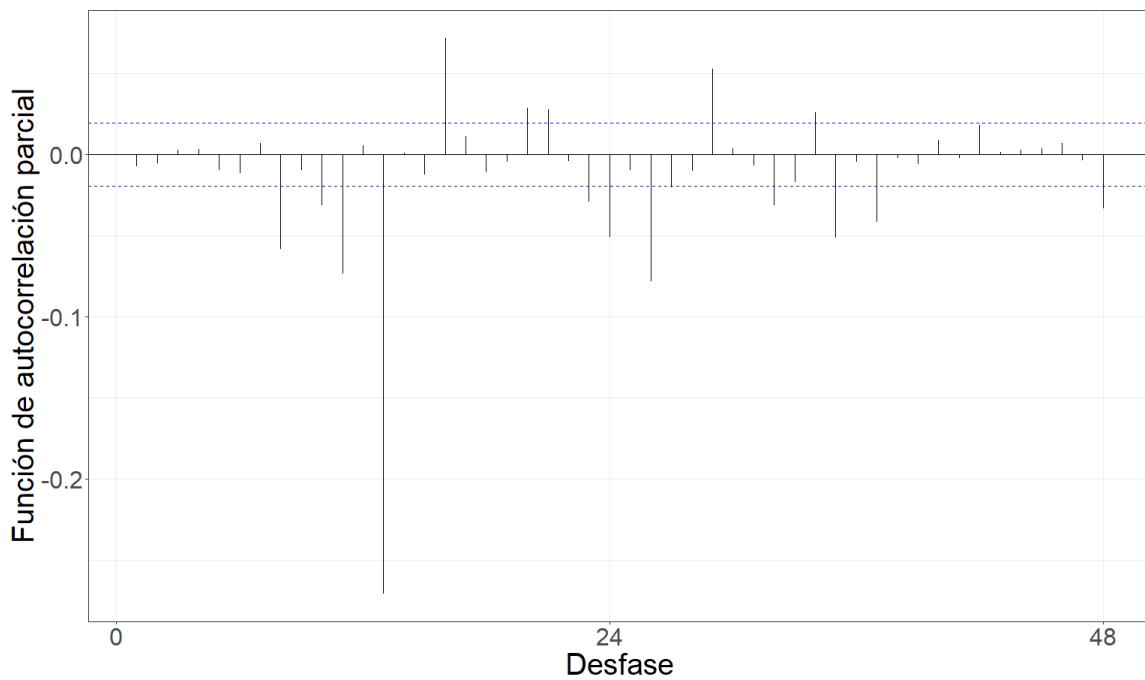


Figura 5.94: Gráfico de la función de autocorrelación parcial de la serie filtrada por el modelo teórico ajustado.

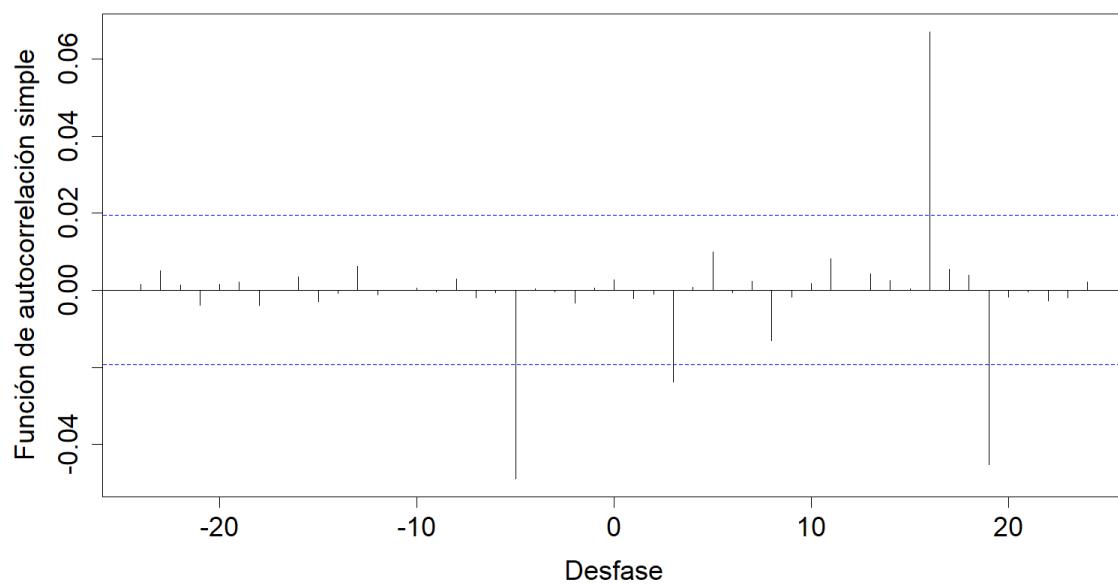


Figura 5.95: Gráfico de la correlación cruzada entre las series filtradas por el modelo teórico ajustado.

Tabla 5.25: Coeficientes estimados de la relación funcional.

Desfasaje	Con diferenciación	Sin diferenciación
t-5	-0.0139078	-0.3025

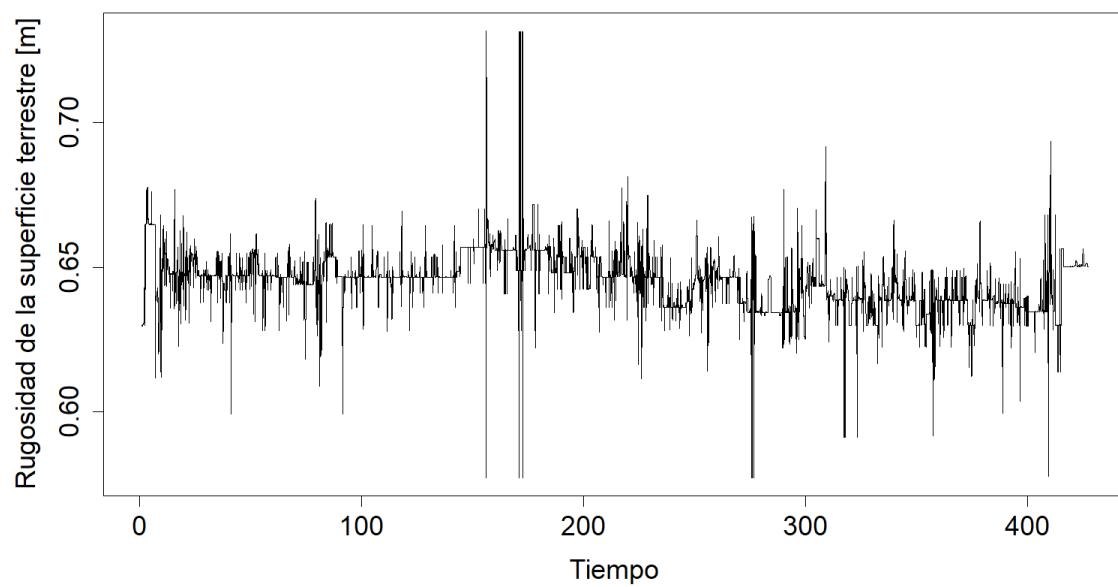
Rugosidad de la superficie terrestre

Figura 5.96: Gráfico de la serie horaria promedio de rugosidad de la superficie terrestre.

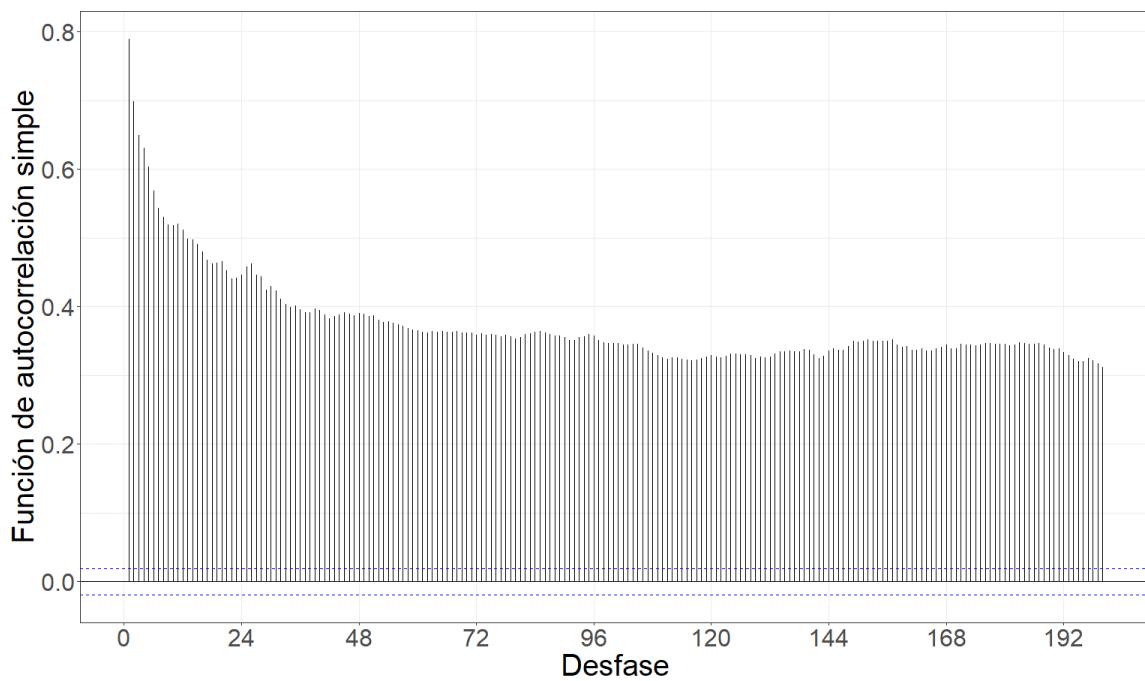


Figura 5.97: Gráfico de la función de autocorrelación simple de la serie.

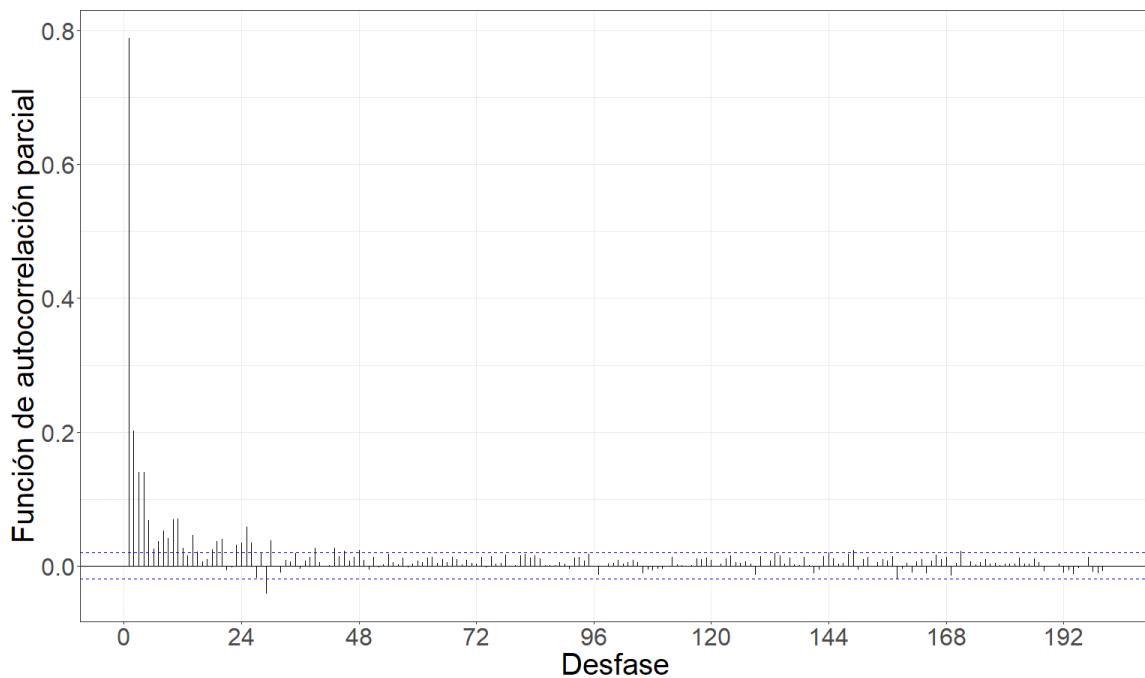


Figura 5.98: Gráfico de la función de autocorrelación parcial de la serie.

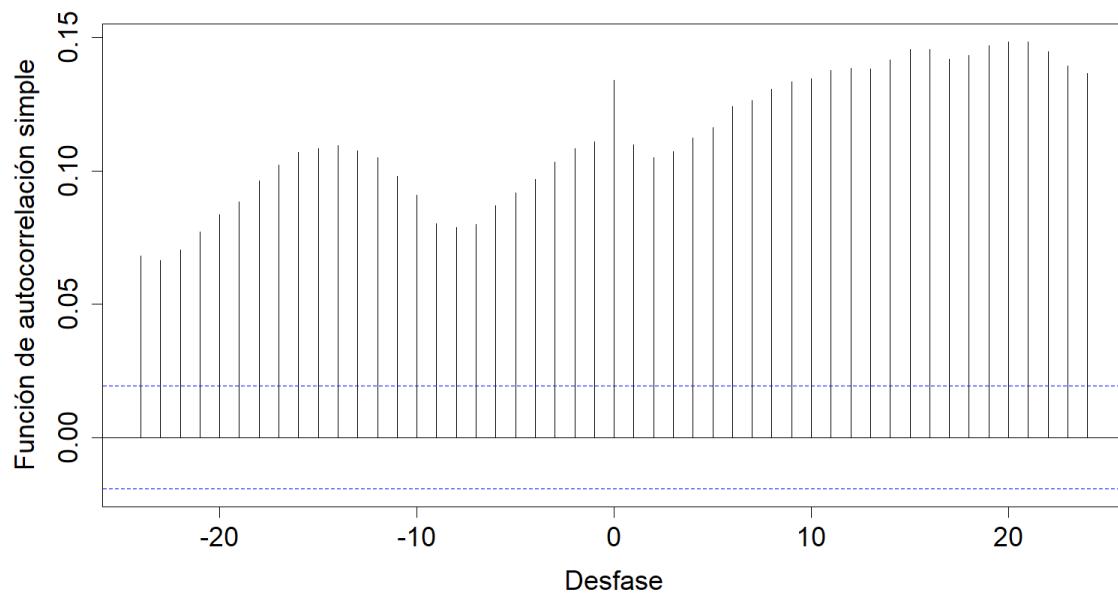


Figura 5.99: Gráfico de la correlación cruzada entre las series previo al filtrado por el modelo teórico ajustado.

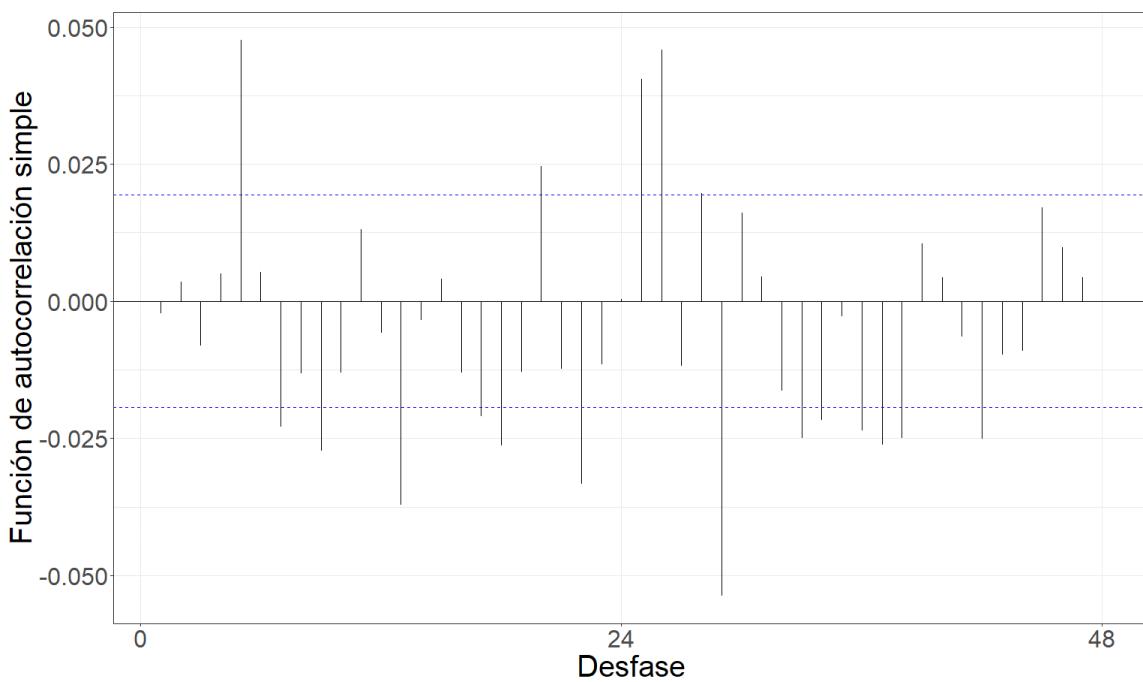


Figura 5.100: Gráfico de la función de autocorrelación simple de la serie filtrada por el modelo teórico ajustado.

Tabla 5.26: Modelo teórico ajustado.

Modelo	AIC	AICc	BIC
ARIMA(4,2,4)(4,1,4)[24]	-76679	-76678.97	-76599.43

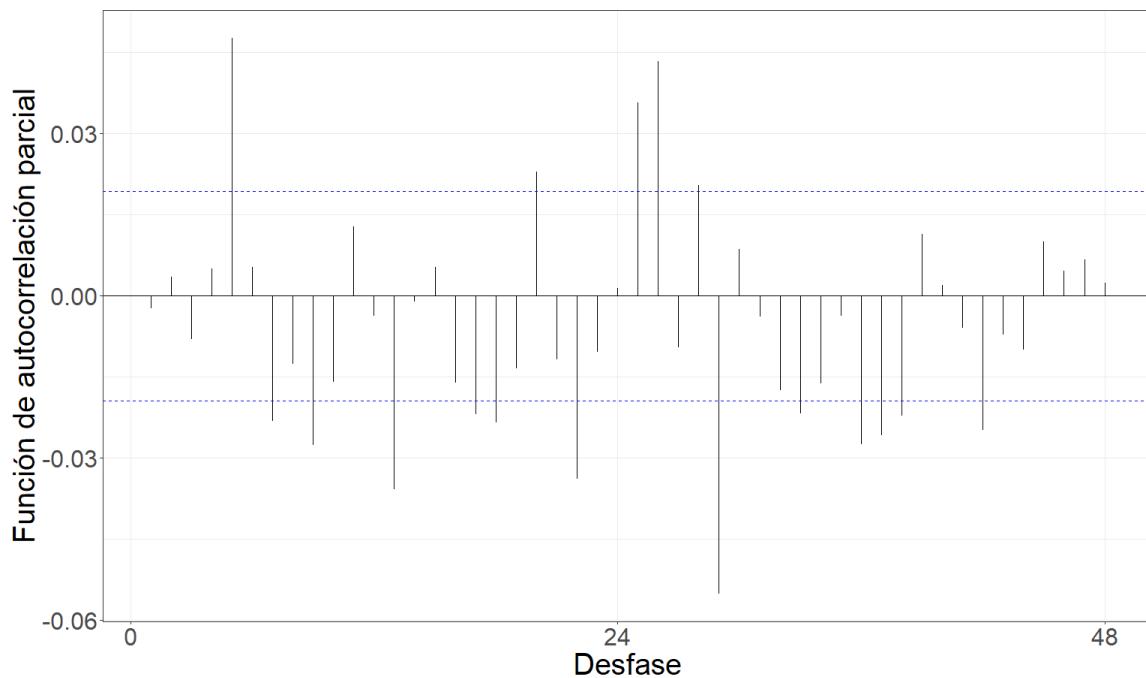


Figura 5.101: Gráfico de la función de autocorrelación parcial de la serie filtrada por el modelo teórico ajustado.

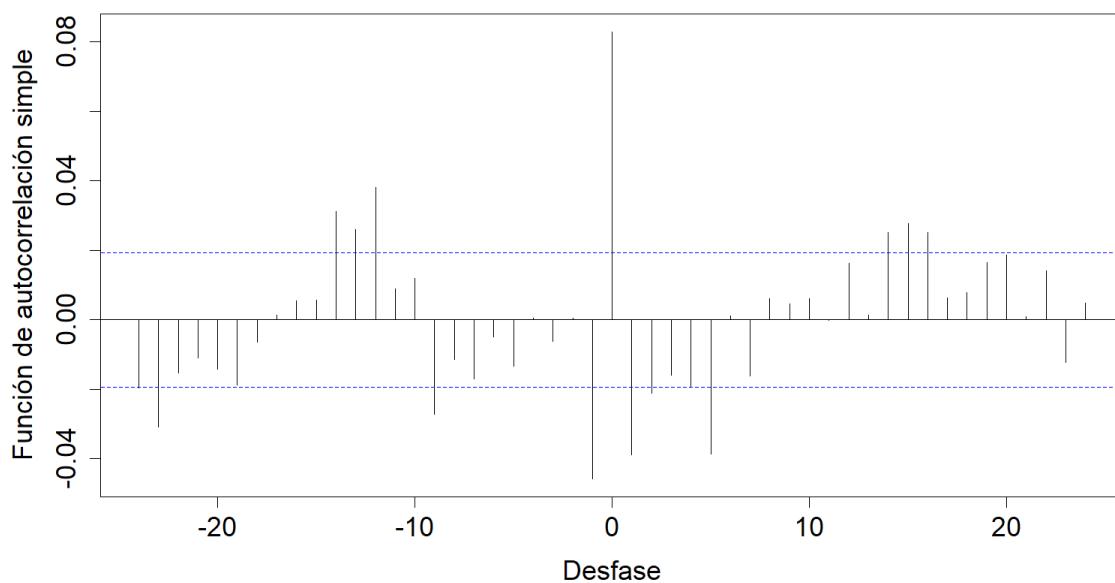


Figura 5.102: Gráfico de la correlación cruzada entre las series filtradas por el modelo teórico ajustado.

Tabla 5.27: Coeficientes estimados de la relación funcional.

Desfasaje	Con diferenciación	Sin diferenciación
t	9.317e+01	111.32
t-1	-4.080e+01	92.03
t-9	-1.868e+01	66.55
t-12	9.1090057	87.25
t-13	0.4461504	89.23
t-14	6.760e+00	91.01
t-23	-1.032e+01	55.22

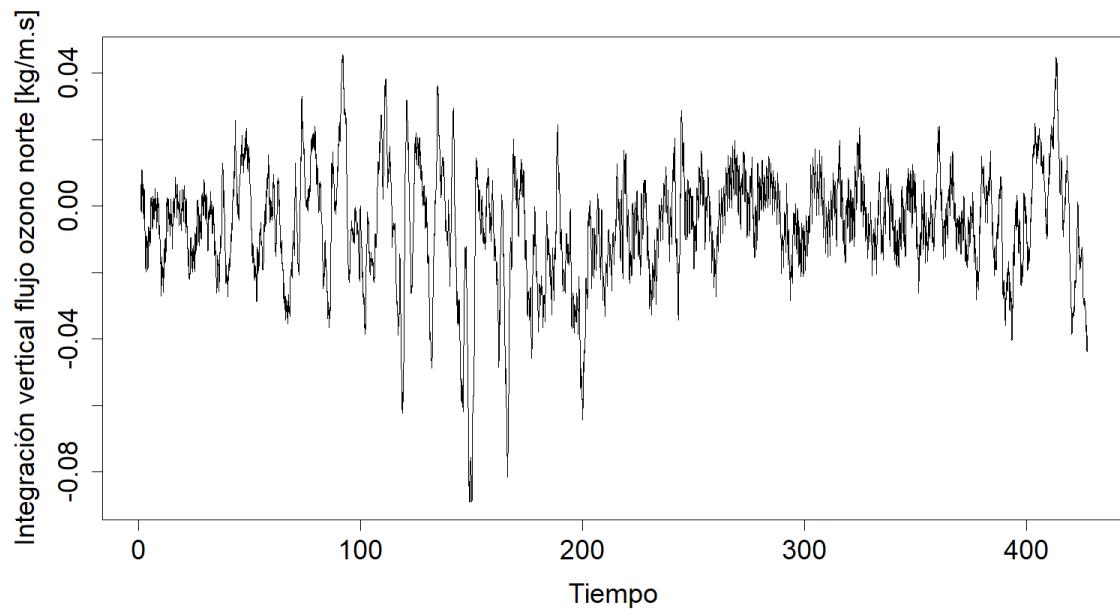
Integración vertical flujo ozono norte

Figura 5.103: Gráfico de la serie horaria promedio de integración vertical flujo ozono norte.

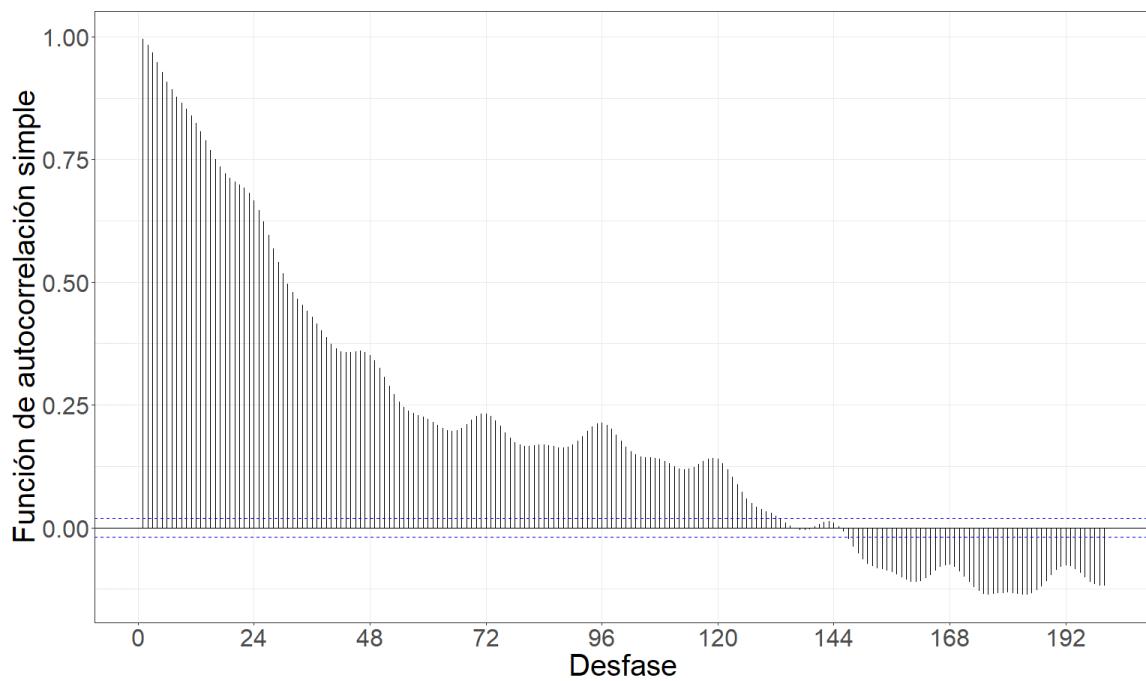


Figura 5.104: Gráfico de la función de autocorrelación simple de la serie.

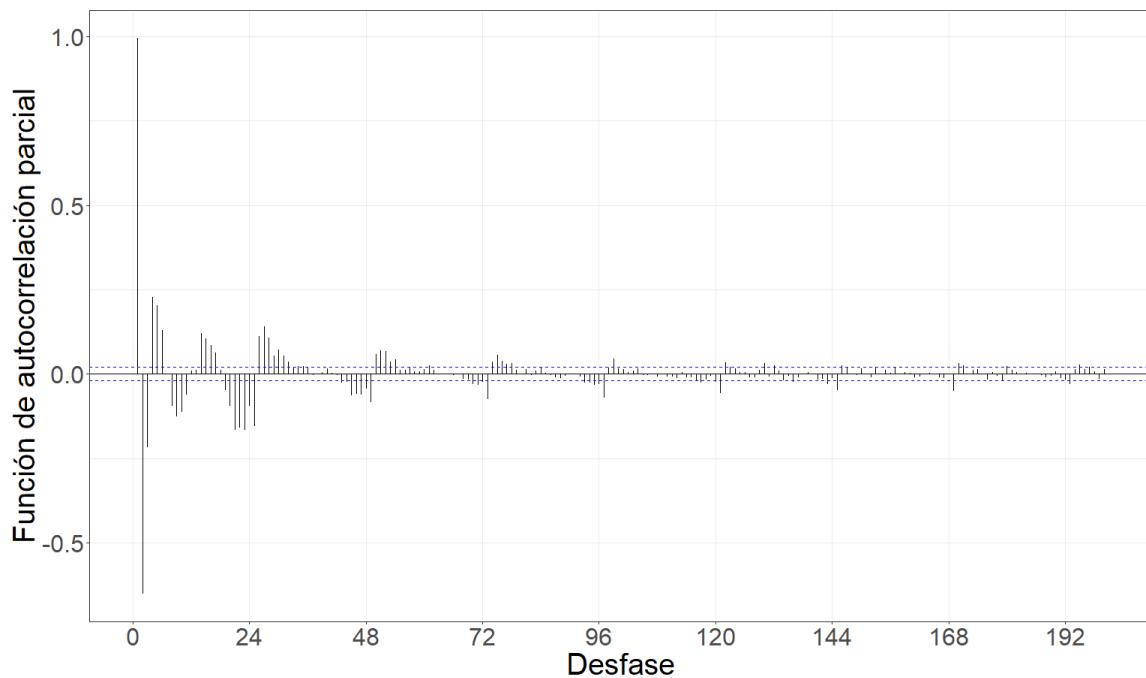


Figura 5.105: Gráfico de la función de autocorrelación parcial de la serie.

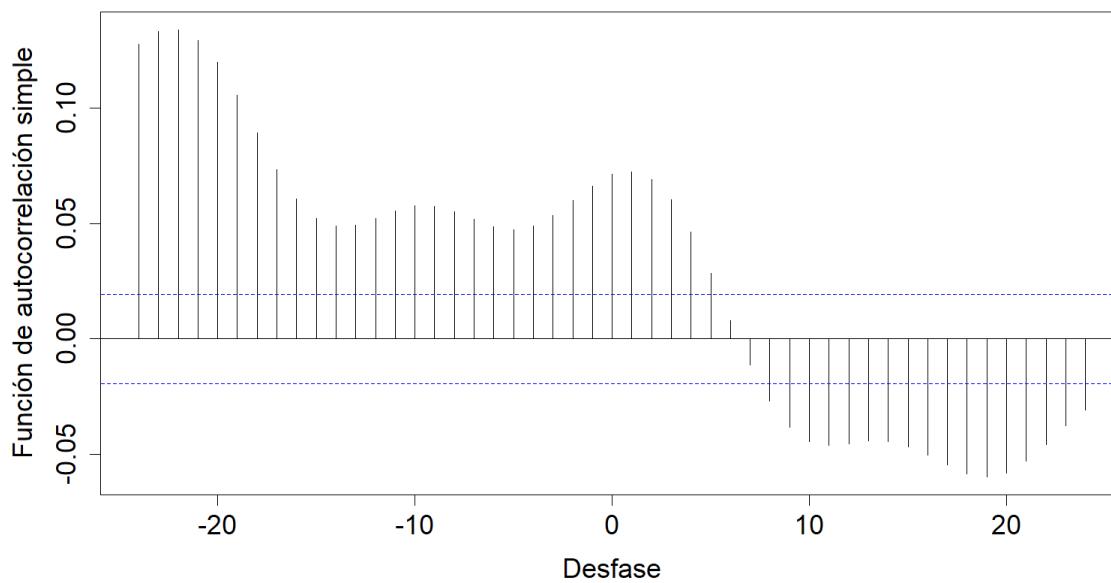


Figura 5.106: Gráfico de la correlación cruzada entre las series previo al filtrado por el modelo teórico ajustado.

Tabla 5.28: Modelo teórico ajustado.

Modelo	AIC	AICc	BIC
ARIMA(4,2,4)(4,2,4)[24]	-72004.18	-72004.18	-71982.48

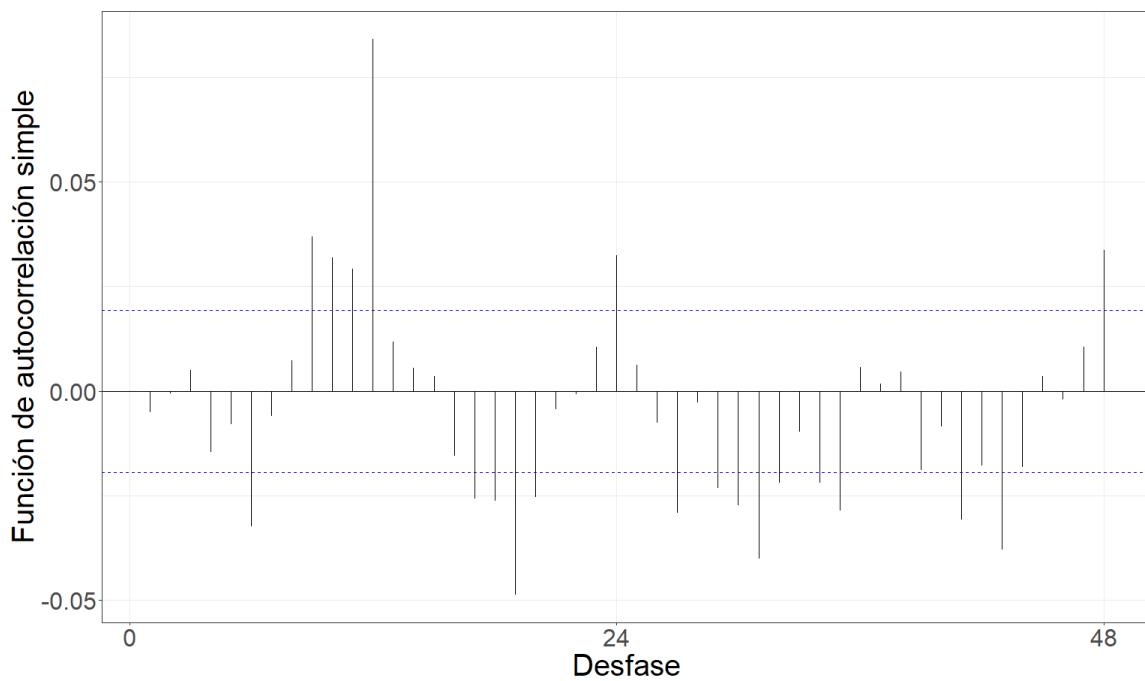


Figura 5.107: Gráfico de la función de autocorrelación simple de la serie filtrada por el modelo teórico ajustado.

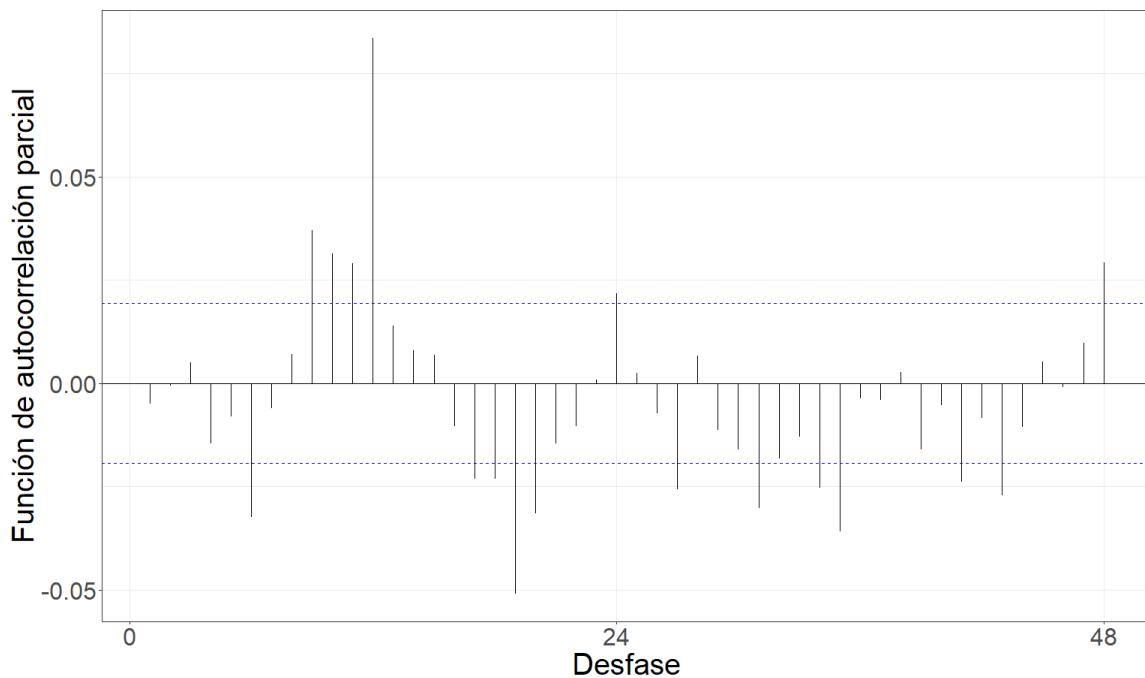


Figura 5.108: Gráfico de la función de autocorrelación parcial de la serie filtrada por el modelo teórico ajustado.

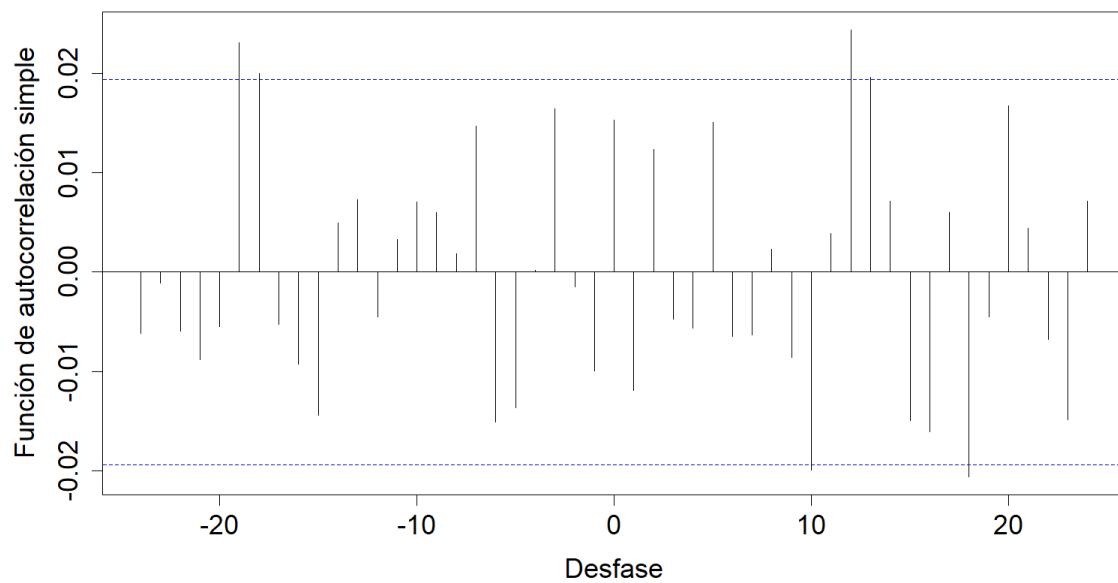


Figura 5.109: Gráfico de la correlación cruzada entre las series filtradas por el modelo teórico ajustado.

Tabla 5.29: Coeficientes estimados de la relación funcional.

Desfasaje	Con diferenciación	Sin diferenciación
t-18	-3.189e+01	43.90
t-19	1.219e+02	52.00

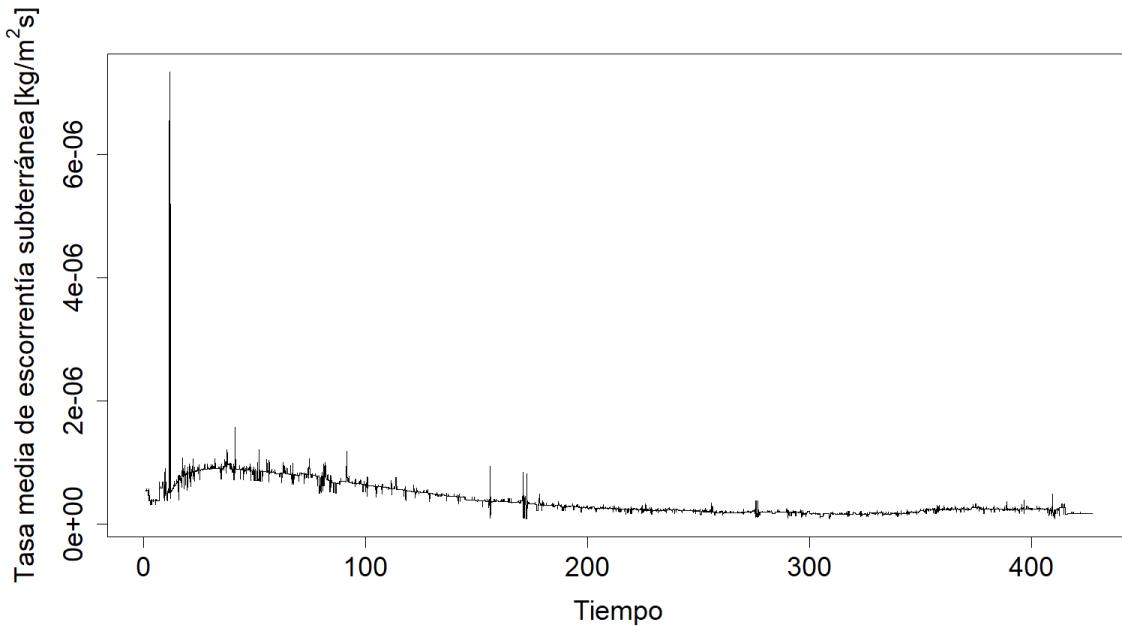
Tasa media de escorrentía subterránea

Figura 5.110: Gráfico de la serie horaria promedio de tasa media de escorrentía subterránea.

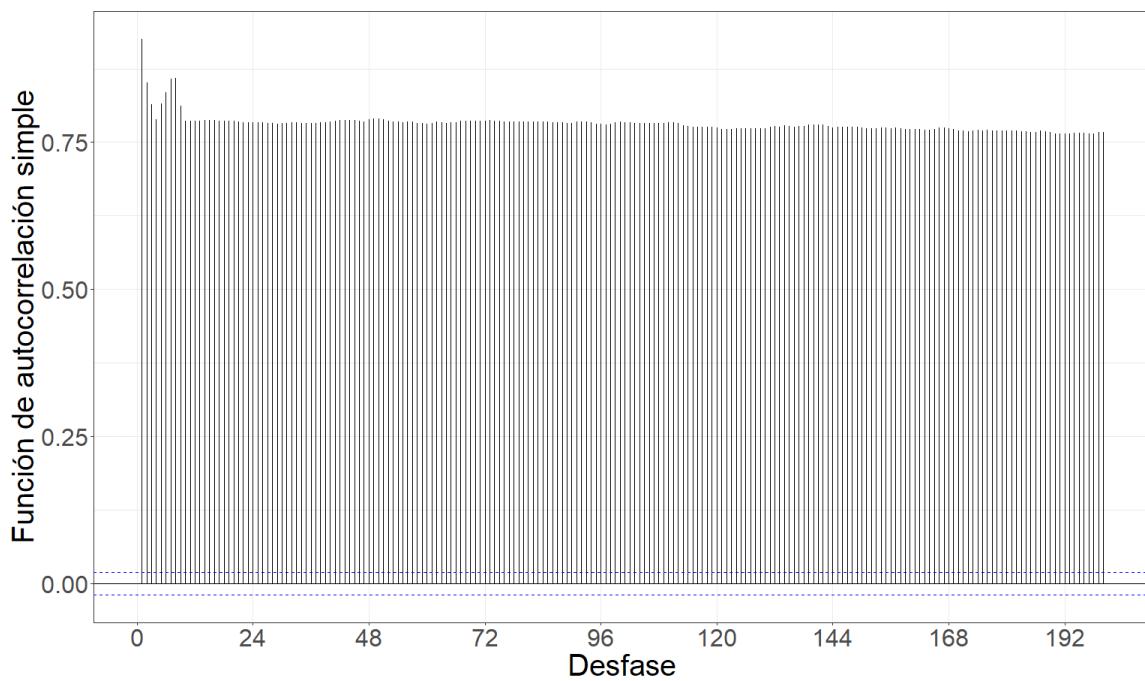


Figura 5.111: Gráfico de la función de autocorrelación simple de la serie.

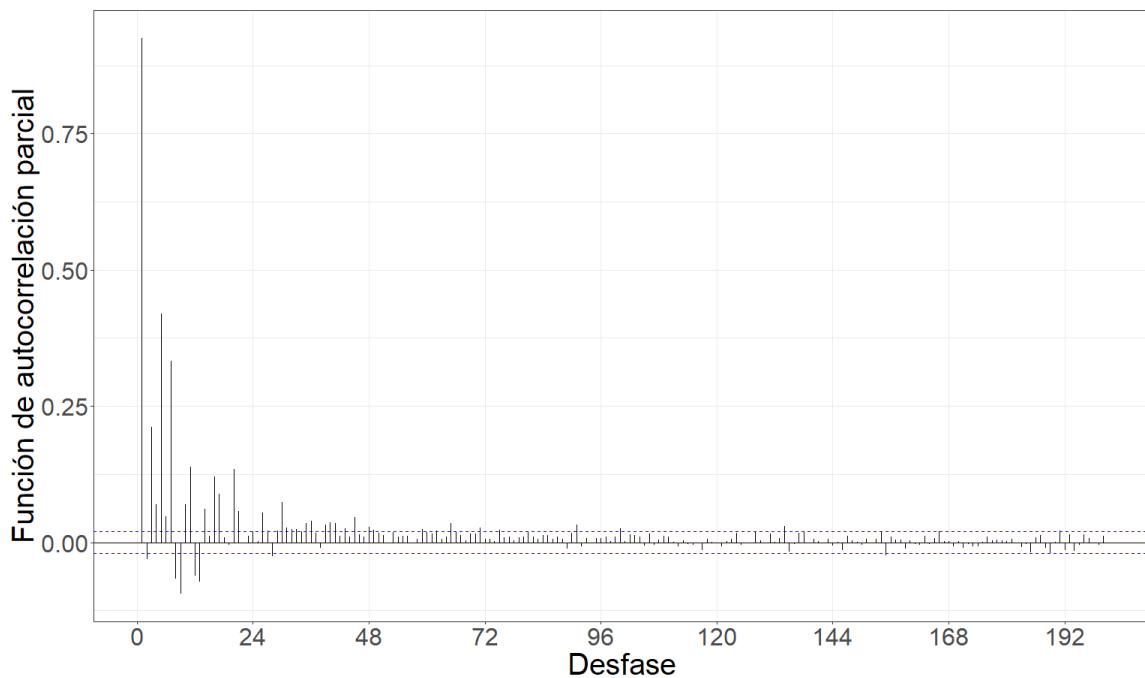


Figura 5.112: Gráfico de la función de autocorrelación parcial de la serie.

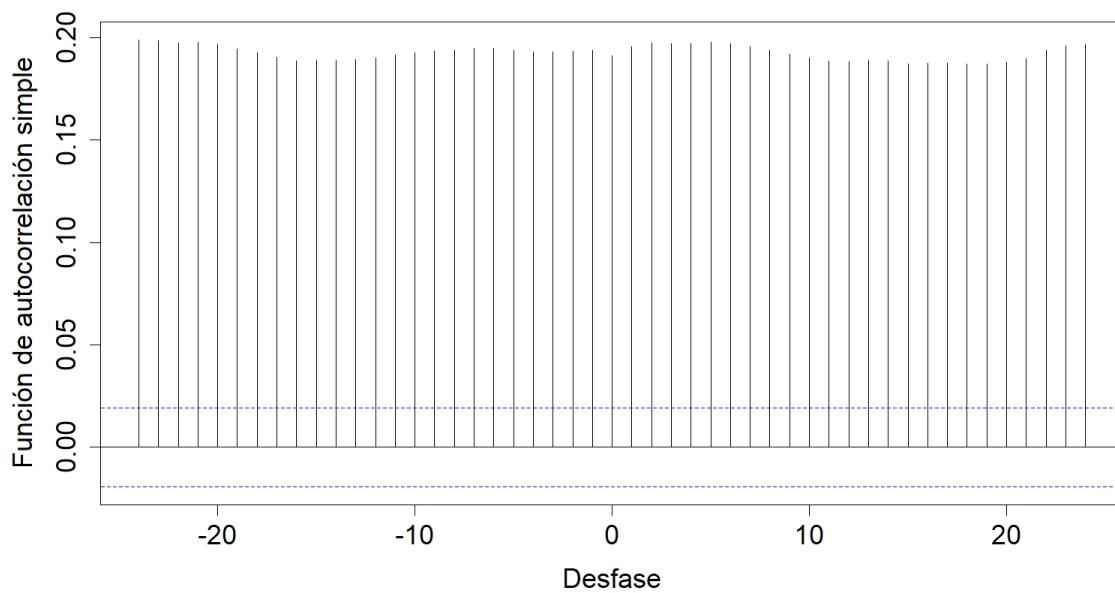


Figura 5.113: Gráfico de la correlación cruzada entre las series previo al filtrado por el modelo teórico ajustado.

Tabla 5.30: Modelo teórico ajustado.

Modelo	AIC	AICc	BIC
ARIMA(1,2,4)(4,2,3)[24]	-299719.1	-299719.1	-299711.8

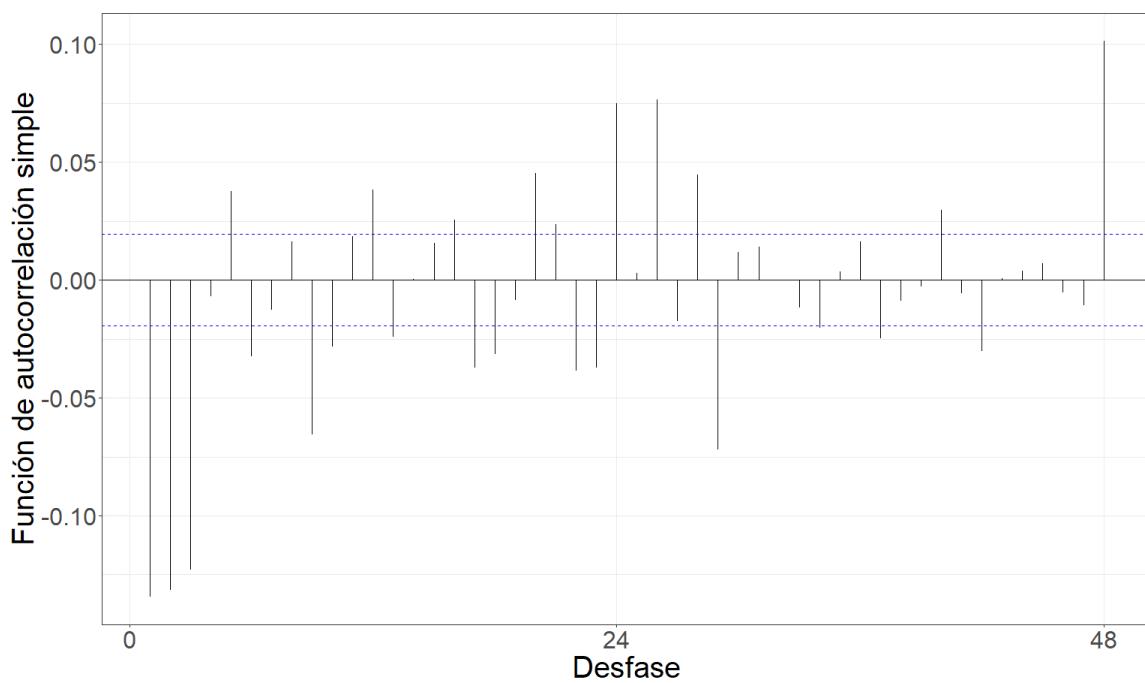


Figura 5.114: Gráfico de la función de autocorrelación simple de la serie filtrada por el modelo teórico ajustado.

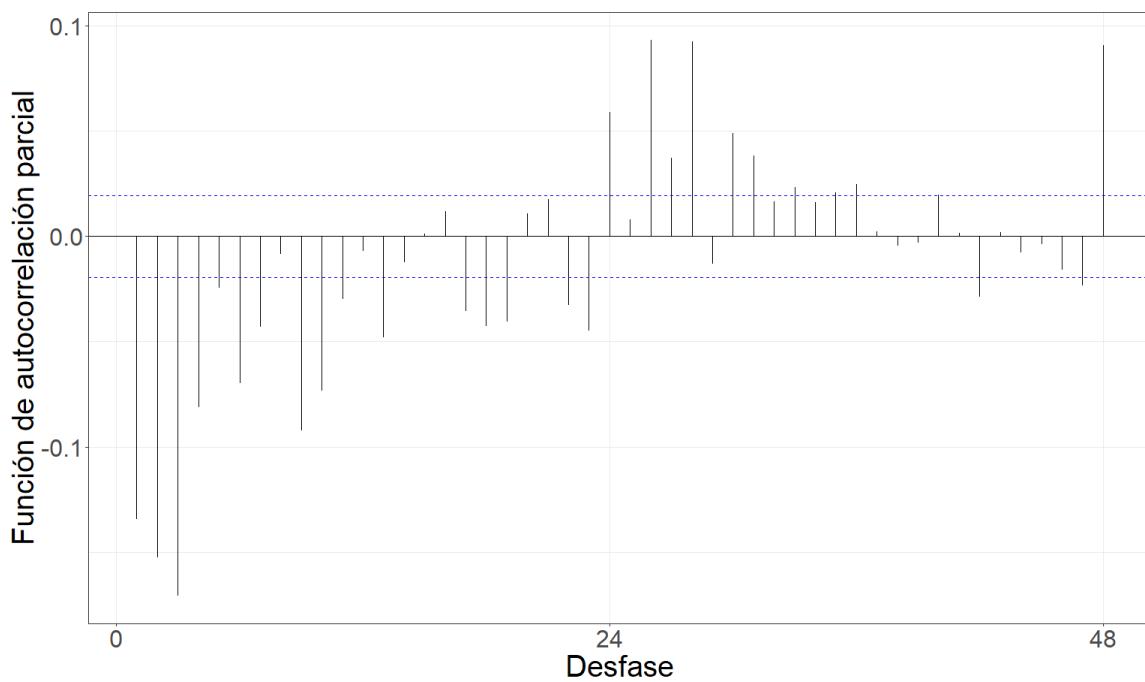


Figura 5.115: Gráfico de la función de autocorrelación parcial de la serie filtrada por el modelo teórico ajustado.

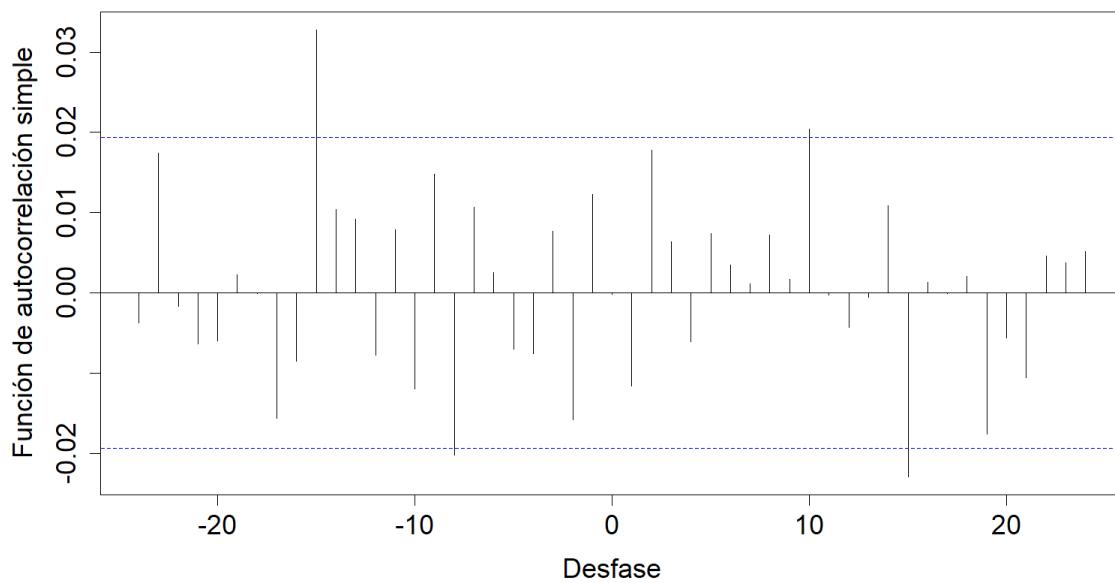


Figura 5.116: Gráfico de la correlación cruzada entre las series filtradas por el modelo teórico ajustado.

Tabla 5.31: Coeficientes estimados de la relación funcional.

Desfasaje	Con diferenciación	Sin diferenciación
t-8	-2.648e+05	5.770e+06
t-15	5.326e+04	5.633e+06

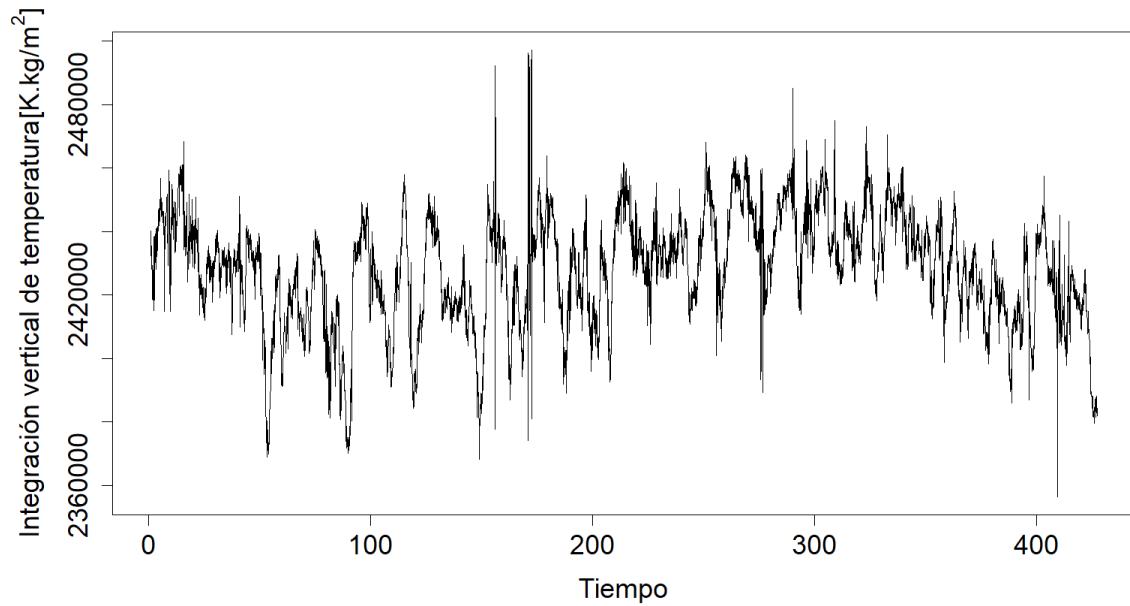
Integración vertical de temperatura

Figura 5.117: Gráfico de la serie horaria promedio de integración vertical de temperatura.

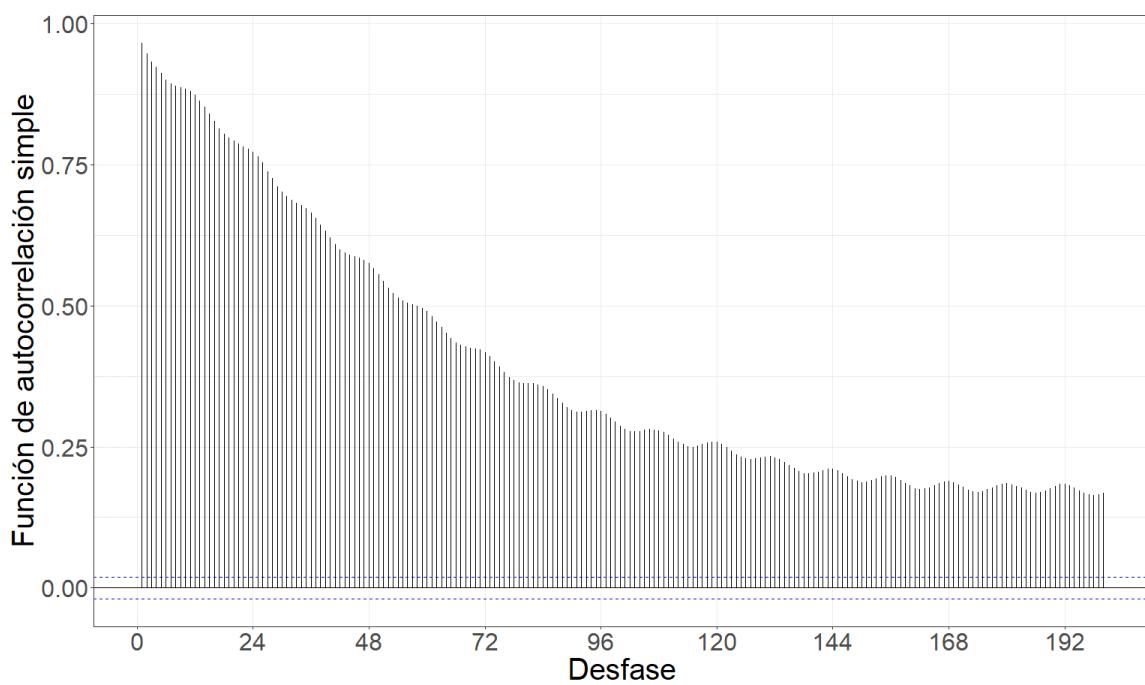


Figura 5.118: Gráfico de la función de autocorrelación simple de la serie.

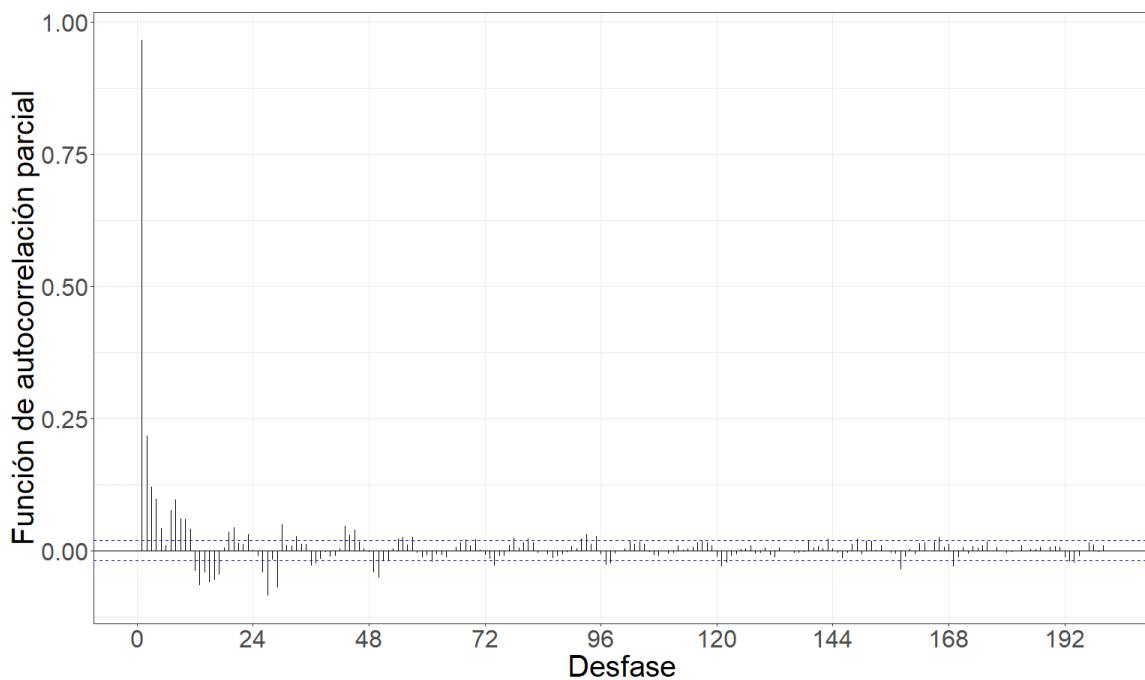


Figura 5.119: Gráfico de la función de autocorrelación parcial de la serie.

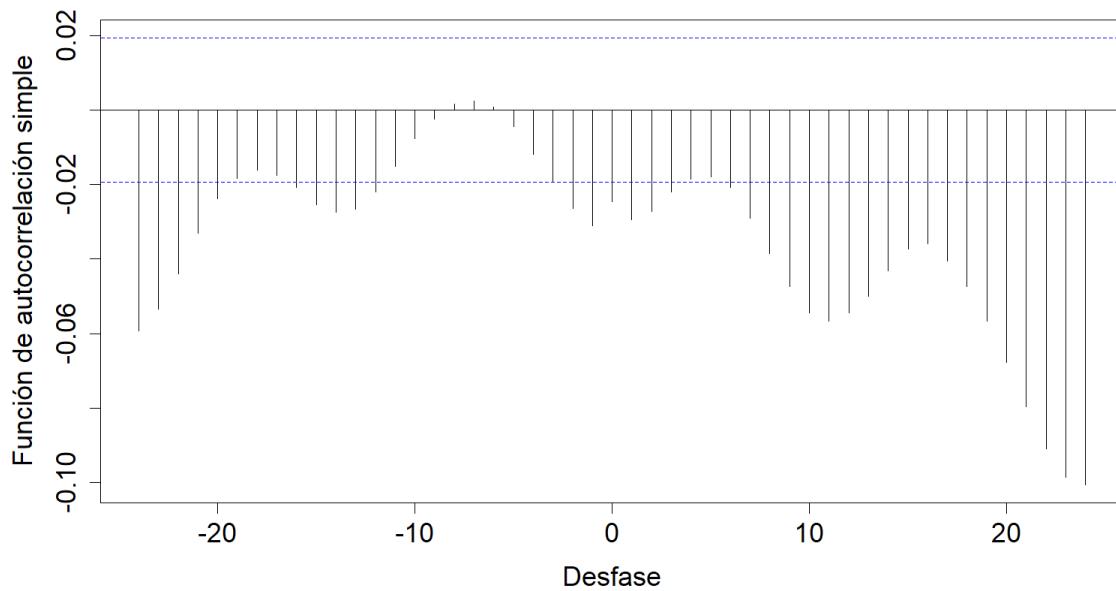


Figura 5.120: Gráfico de la correlación cruzada entre las series previo al filtrado por el modelo teórico ajustado.

Tabla 5.32: Modelo teórico ajustado.

Modelo	AIC	AICc	BIC
ARIMA(4,2,4)(4,2,4)[24]	200568	200568	200611.4

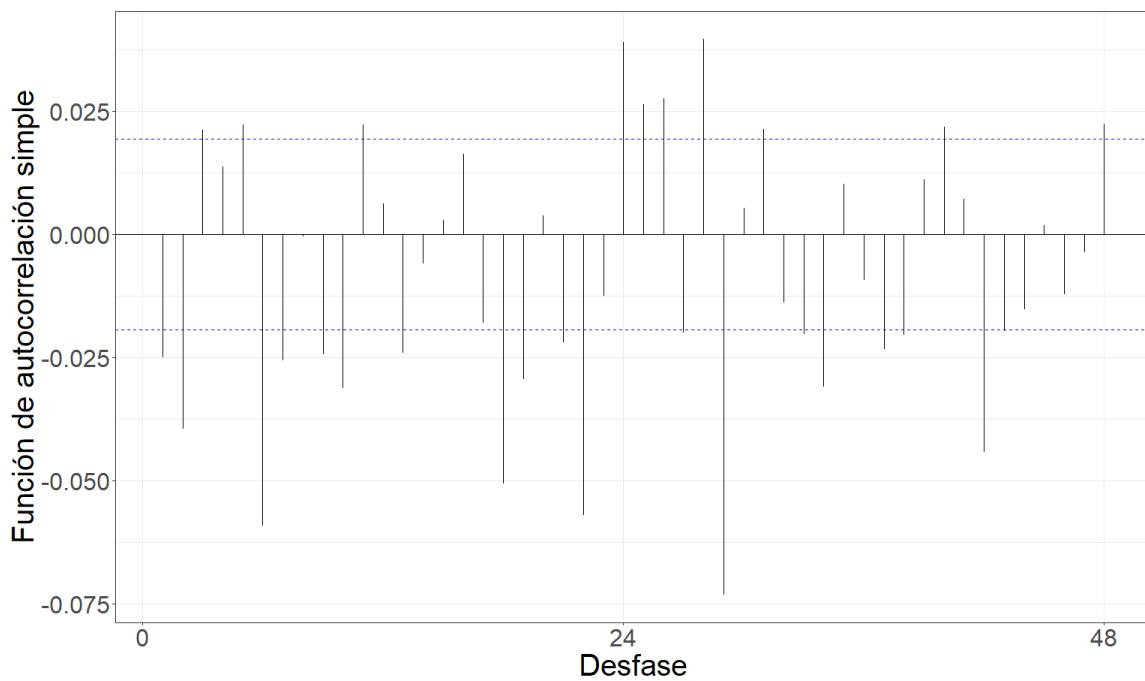


Figura 5.121: Gráfico de la función de autocorrelación simple de la serie filtrada por el modelo teórico ajustado.

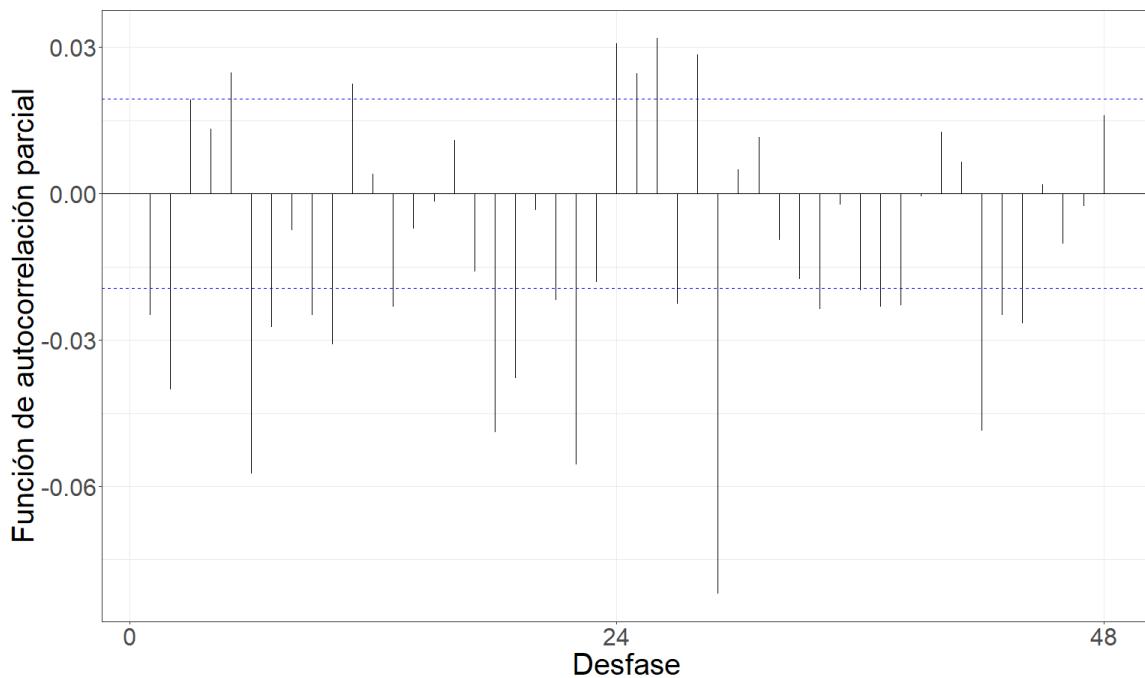


Figura 5.122: Gráfico de la función de autocorrelación parcial de la serie filtrada por el modelo teórico ajustado.

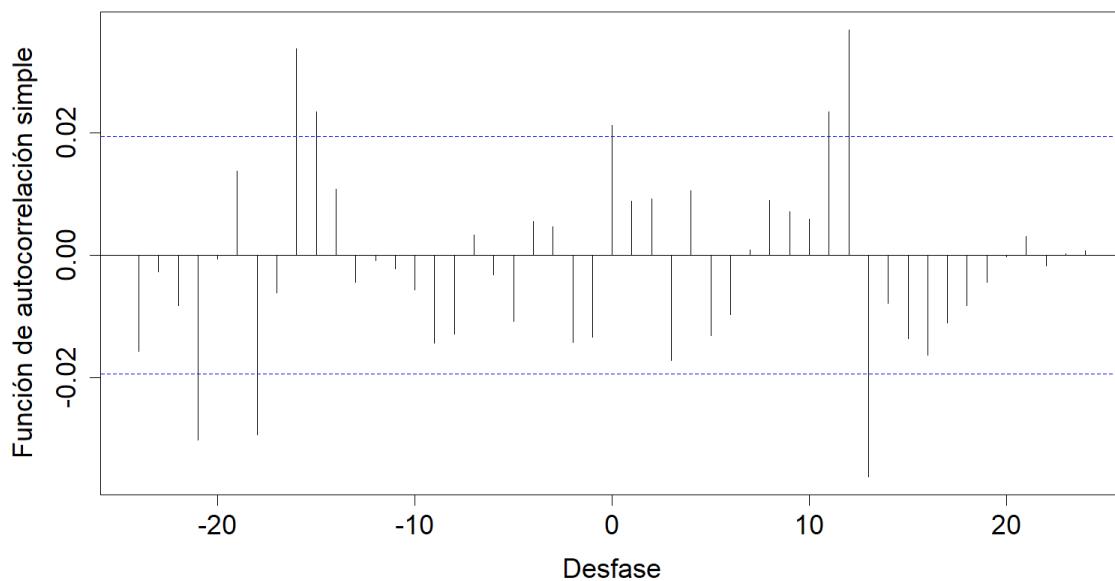


Figura 5.123: Gráfico de la correlación cruzada entre las series filtradas por el modelo teórico ajustado.

Tabla 5.33: Coeficientes estimados de la relación funcional.

Desfasaje	Con diferenciación	Sin diferenciación
t	7.654e-05	-3.442e-06
t-15	-1.411e-05	-3.438e-05
t-16	8.115e-06	2.149e-05
t-18	3.007e-05	6.352e-05
t-21	1.269e-05	-6.070e-05

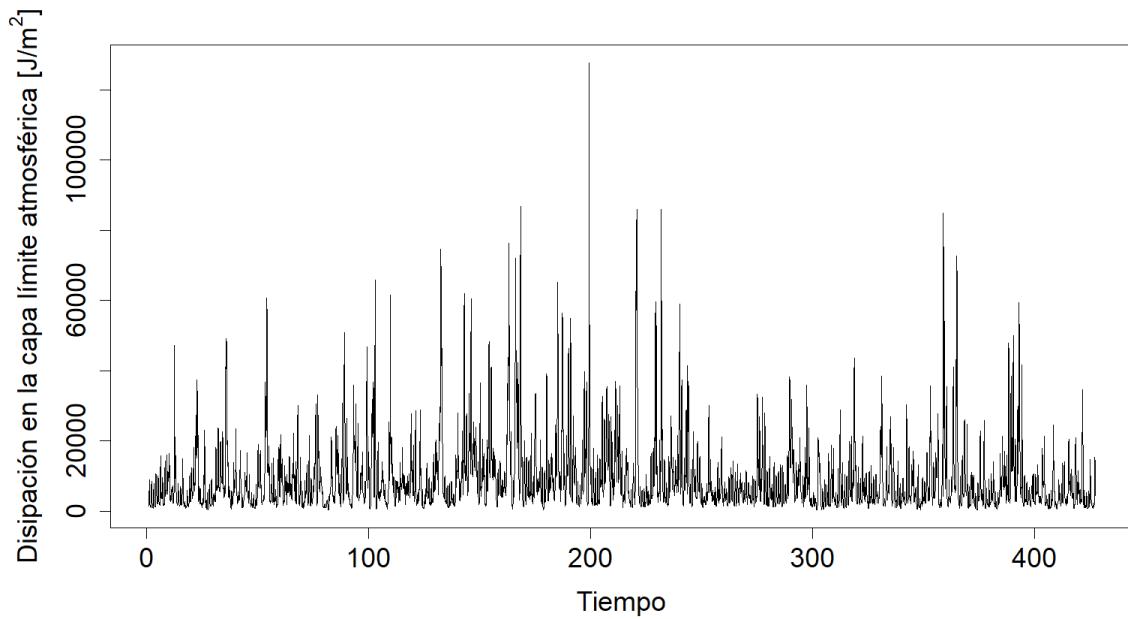
Disipación en la capa límite atmosférica

Figura 5.124: Gráfico de la serie horaria promedio de disipación en la capa límite atmosférica.

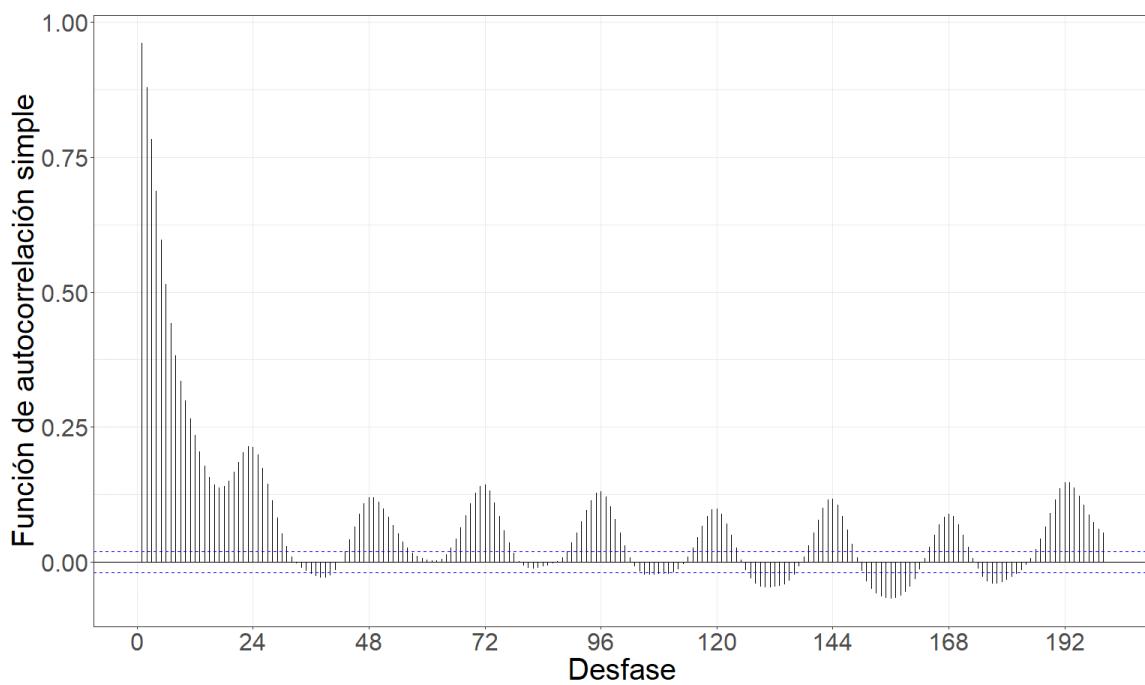


Figura 5.125: Gráfico de la función de autocorrelación simple de la serie.

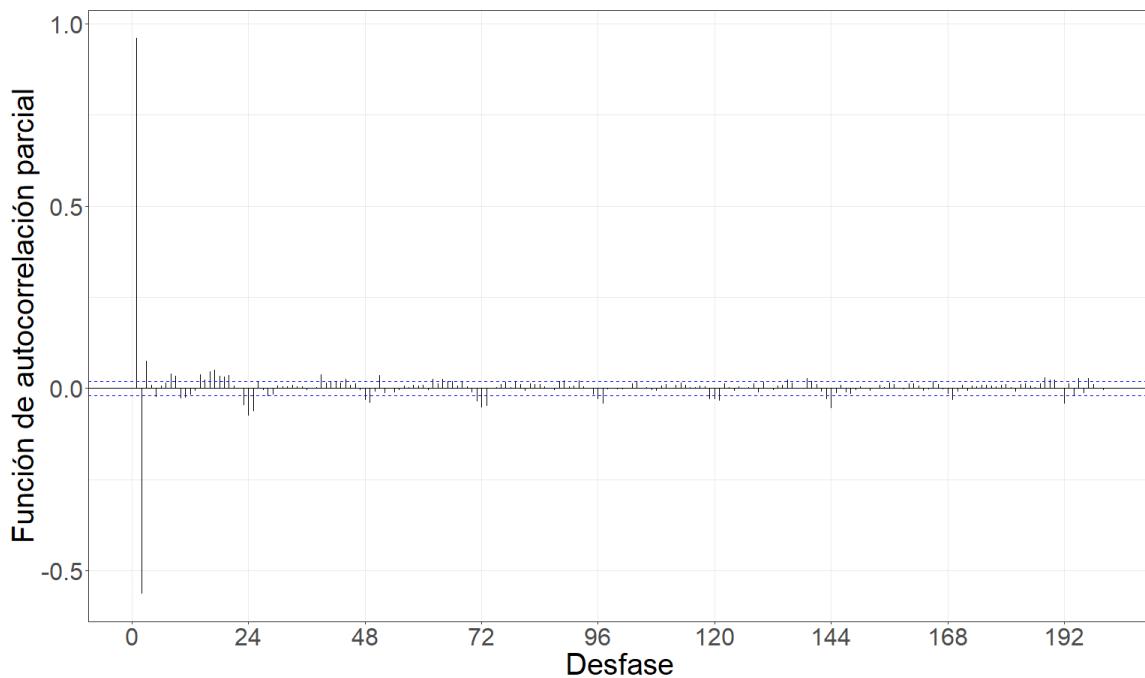


Figura 5.126: Gráfico de la función de autocorrelación parcial de la serie.

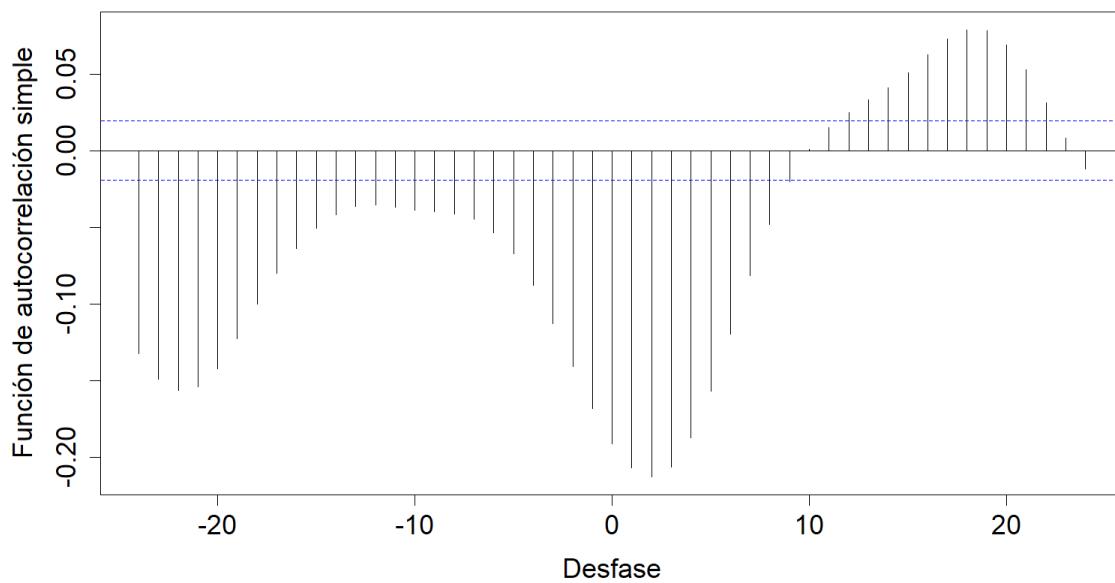


Figura 5.127: Gráfico de la correlación cruzada entre las series previo al filtrado por el modelo teórico ajustado.

Tabla 5.34: Modelo teórico ajustado.

Modelo	AIC	AICc	BIC
ARIMA(2,3,6)(2,2,4)[24]	187961.2	187961.2	188019.1

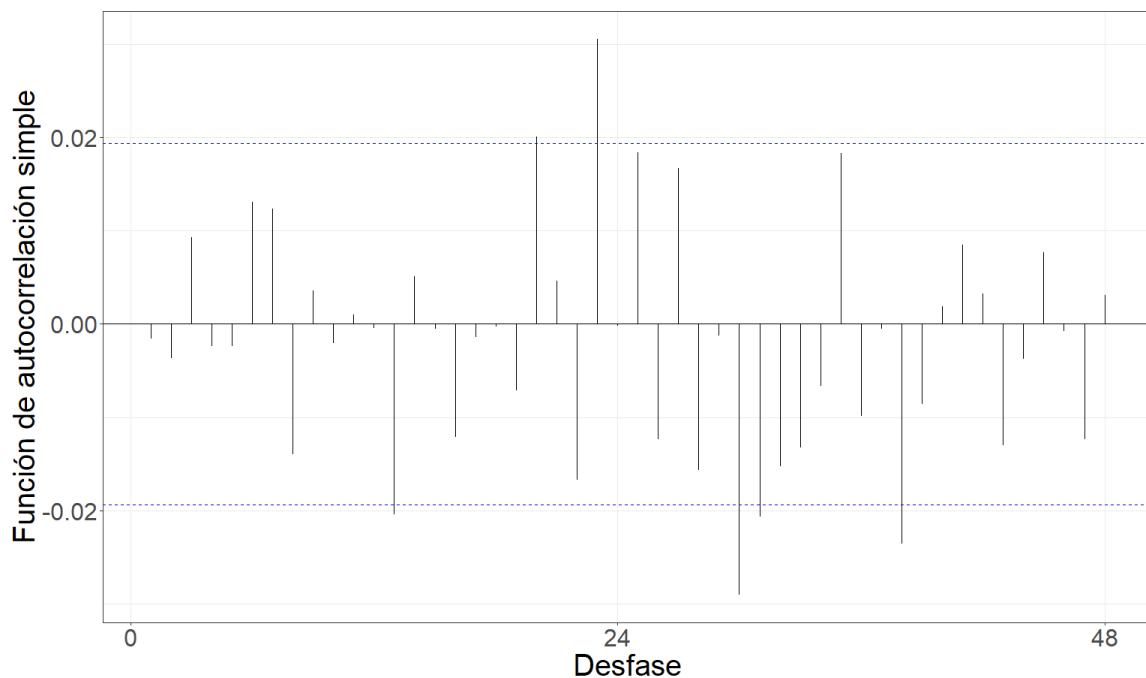


Figura 5.128: Gráfico de la función de autocorrelación simple de la serie filtrada por el modelo teórico ajustado.

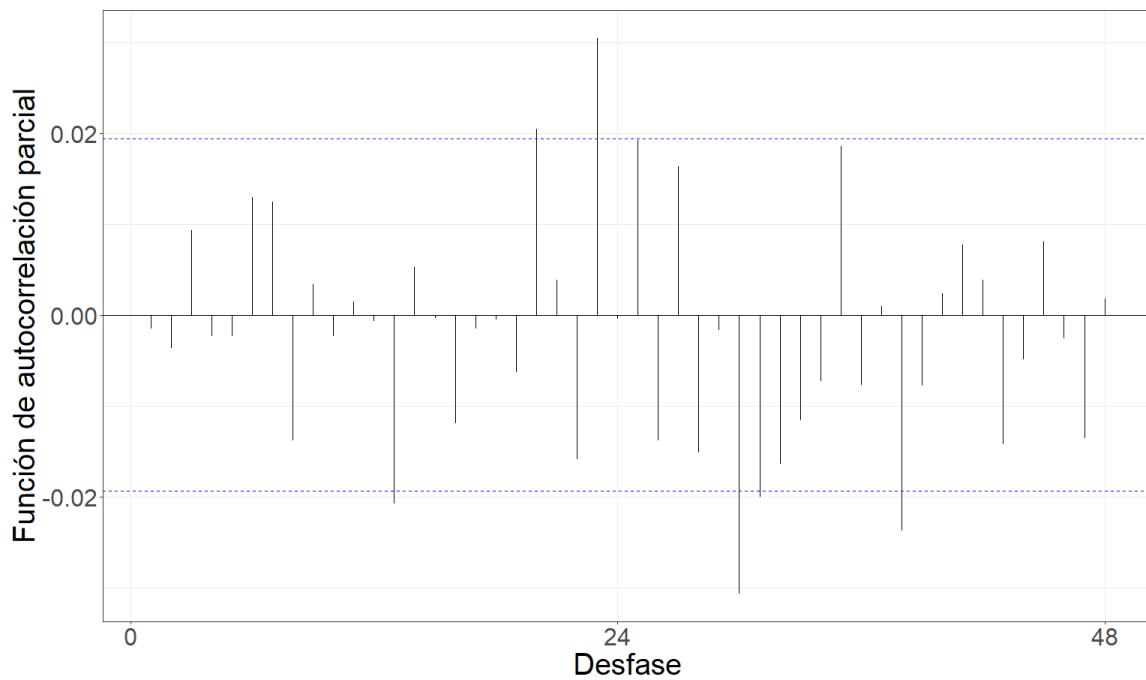


Figura 5.129: Gráfico de la función de autocorrelación parcial de la serie filtrada por el modelo teórico ajustado.

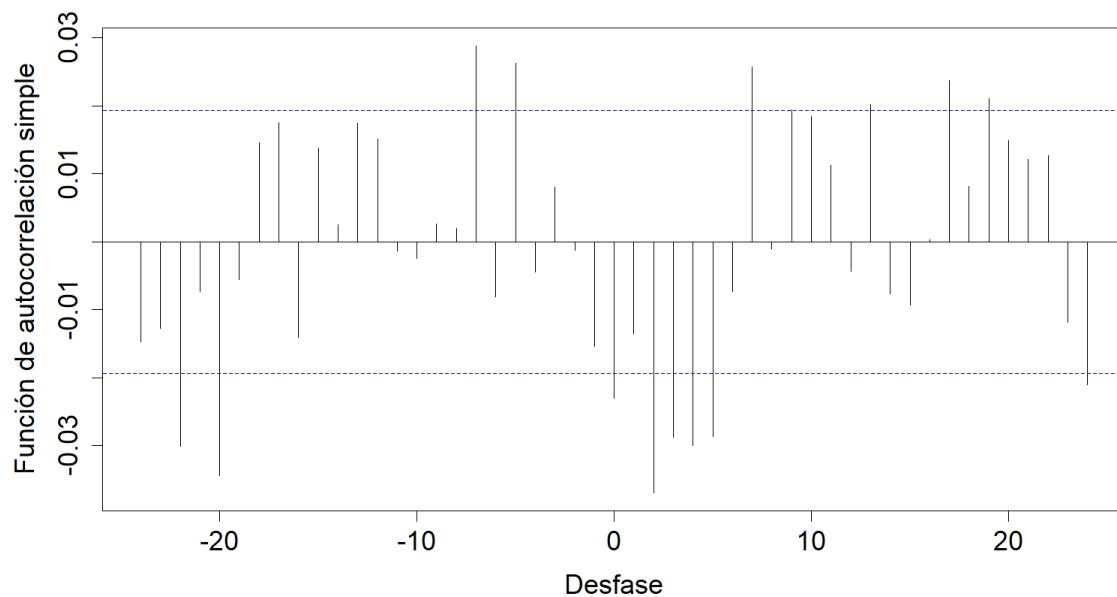


Figura 5.130: Gráfico de la correlación cruzada entre las series filtradas por el modelo teórico ajustado.

Tabla 5.35: Coeficientes estimados de la relación funcional.

Desfasaje	Con diferenciación	Sin diferenciación
t	-7.918e-05	-0.0001536
t-5	6.594e-05	-0.000054
t-7	0.0000533	-3.578e-05
t-20	-8.173e-05	-0.0001142
t-22	-0.0001042	-0.0001257

Resumen de relaciones

Tabla 5.36: Resumen de los desfases temporales con capacidad predictiva para cada una de las variables derivadas a partir de información satelital analizadas.

Variable	Desfase	Dirección	Mayor capacidad predictiva
Temperatura	t	Inversa	*
	t-12	Inversa	
	t-24	Inversa	
Humedad relativa	t-5	Inversa	
	t-24	Directa	*
Temperatura del punto de rocío	t-11	Inversa	
	t-14	Inversa	*
Precipitación convectiva	t-1	Inversa	*
Velocidad del viento	t-24	Inversa	*
Presión atmosférica	t-1	Inversa	
	t-10	Directa	*
Radiación solar	t-13	Directa	
	t-23	Inversa	*
Profundidad óptica de aerosoles	t	Directa	*
Altura de la capa límite atmosférica	t	Inversa	
	t-21	Inversa	
	t-24	Inversa	*
Tráfico	t-8	Directa	*
Temperatura de la superficie terrestre	t-5	Inversa	*
Rugosidad de la superficie terrestre	t	Directa	*
	t-12	Directa	
	t-13	Directa	
	t-14	Directa	

Variable	Desfasaje	Dirección	Mayor capacidad predictiva
Integración vertical flujo ozono norte	t-19	Directa	*
Tasa media de escorrentía subterránea	t-15	Directa	*
Integración vertical de temperatura	t-15	Inversa	
	t-16	Directa	*
	t-18	Directa	
Presión atmosférica al nivel del mar	t	Inversa	*
	t-19	Directa	
Disipación en la capa límite atmosférica	t	Inversa	
	t-20	Inversa	*
	t-22	Inversa	

Anexo 3: Descripción de variables

Las variables que involucran la base de datos empleada en la presente tesis fueron recopiladas a partir de múltiples fuentes de datos. Las meteorológicas, climáticas y ambientales provienen de fuentes de datos obtenidas a partir de la medición de satélites, mientras que las variables temporales fueron derivadas a partir de la fecha en que cada dato fue recopilado. Todas aquellas variables que en su definición no incluyan explícitamente la fuente de datos de la cual fueron obtenidas pertenecen a las derivadas a partir del método de ERA5. A continuación se presenta una descripción detallada de cada una de las variables:

- Hora del día (hora) [adimensional] (temporal): Variable numérica que toma valores en un rango de 1 a 24, indicando la variación de las horas a lo largo del día. Esta variable representa la hora del día en la que cada dato fue recopilado.
- Día de la semana (día) [adimensional] (temporal): Variable numérica que toma valores en un rango de 1 a 7, indicando la variación de los días a lo largo de la semana. Esta variable representa el día de la semana en el que cada dato fue recopilado.
- Mes del año (mes) [adimensional] (temporal): Variable numérica que toma valores en un rango de 1 a 12, indicando la variación de los meses a lo largo del año. Esta variable representa el mes del año en el que cada dato fue recopilado.
- Día del mes [adimensional] (temporal): Variable numérica que toma valores en

un rango de 1 a 31, indicando la variación de los días a lo largo del mes. Esta variable representa el día del mes en el que cada dato fue recopilado.

- Estación del año (estación) [adimensional] (temporal): Variable numérica que toma valores en un rango de 1 a 4, indicando la variación de la estación a lo largo del año. Esta variable representa la estación climática del año en la que cada dato fue recopilado.
- Hora codificación seno [adimensional] (temporal): Codificación seno de la variable hora del día.
- Hora codificación coseno [adimensional] (temporal): Codificación coseno de la variable hora del día.
- Mes codificación seno [adimensional] (temporal): Codificación seno de la variable mes del año.
- Mes codificación coseno [adimensional] (temporal): Codificación coseno de la variable mes del año.
- Día codificación seno [adimensional] (temporal): Codificación seno de la variable día de la semana.
- Día codificación coseno [adimensional] (temporal): Codificación coseno de la variable día de la semana.
- Día del mes codificación seno [adimensional] (temporal): Codificación seno de la variable día del mes.
- Día del mes codificación coseno [adimensional] (temporal): Codificación seno de la variable día del mes.
- Estación codificación seno [adimensional] (temporal): Codificación seno de la variable estación del año.

- Estación codificación coseno [adimensional] (temporal): Codificación coseno de la variable estación del año.
- Profundidad óptica de aerosoles (AOD) [adimensional] (MODIS-MAIAC/MERRA-2): Mide la extinción de la luz debida a su dispersión y absorción por la presencia de partículas (por ejemplo, neblina urbana, humo, polvo desértico, sal marina) distribuidas en la columna de aire desde la superficie terrestre a la parte superior de la atmósfera. Esta variable se obtuvo a partir de dos fuentes, en primer lugar, del algoritmo denominado implementación de corrección atmosférica multi-ángulo (MAIAC) que ha sido recientemente desarrollado para los datos obtenidos con el sensor MODIS (a bordo de los satélites Terra y Aqua/NASA) aplicado a la capa (*Blue band (0.47 μm) aerosol optical depth over land*) (Alexei Lyapustin et al., 2018, 2019). En segundo lugar, la variable se obtuvo del Análisis retrospectivo de la era moderna para investigación y aplicaciones, versión 2 (MERRA-2), el producto empleado fue (M2I3NXGAS) el cual permite obtener 1 dato cada 3 horas con una resolución espacial de 0.625 x 0.5 (Bosilovich et al., 2015).
- Incremento de la profundidad óptica de aerosoles [adimensional] (MERRA-2): Diferencia entre dos mediciones sucesivas de la profundidad óptica de aerosoles. Esta variable se denomina *AODINC (Aerosol Optical Depth Analysis Increment)* y proviene del producto satelital M2I3NXGAS obtenido a partir del análisis de MERRA-2.
- Temperatura de la superficie terrestre (LST) [K] (MODIS): Obtenida a partir del producto MODIS LST y emisividad, LST&E de sus siglas en inglés (MOD21), disponible en la colección 6, el cual utiliza un algoritmo basado en la física para recuperar dinámicamente tanto esta temperatura como la emisividad simultáneamente para las tres bandas infrarrojas térmicas MODIS 29, 31 y

32 cuyos canales infrarrojos se encuentran centrados en $8.55\ \mu m$, $11.03\ \mu m$ y $12.02\ \mu m$, con una resolución espacial de 1 km en el nadir (Hulley et al., 2014; Savtchenko et al., 2004). Las unidades en las que se mide este parámetro son kelvin (K).

- Tráfico de vehículos [adimensional] (Google Maps): Variable obtenida a partir de los datos que se encuentran publicados en Google Maps, en donde puede visualizarse el tráfico promedio usual para las distintas horas en los distintos días de la semana (Y.-J. Wu et al., 2007).
- Componente u del viento a 100 m [m/s] (ERA5): Este parámetro es la componente hacia el este del viento a 100 m de altura. Es la velocidad horizontal del aire moviéndose hacia el este, a una altura de 100 metros sobre la superficie de la Tierra, en metros por segundo. Este parámetro puede combinarse con la componente hacia el norte para dar la velocidad y dirección del viento horizontal de 100 m.
- Componente v del viento a 100 m [m/s] (ERA5): Este parámetro es la componente hacia el norte del viento a 100 m de altura. Es la velocidad horizontal del aire moviéndose hacia el norte, a una altura de 100 metros sobre la superficie de la Tierra, en metros por segundo. Este parámetro se puede combinar con el componente hacia el este para dar la velocidad y dirección del viento horizontal de 100 m.
- Velocidad del viento a 100 m [m/s] (ERA5): Combinación de las componentes del viento horizontal a 100 m de altura.
- Componente u del viento neutro a 10 m [m/s] (ERA5): Este parámetro es la componente hacia el este del “viento neutro”, a una altura de 10 metros sobre la superficie de la Tierra. El viento neutro se calcula a partir de la tensión

superficial y la longitud de rugosidad correspondiente suponiendo que el aire está estratificado de forma neutra. El viento neutral es más lento que el viento real en condiciones estables y más rápido en condiciones inestables. El viento neutro está, por definición, en la dirección de la tensión superficial. El tamaño de la longitud de la rugosidad depende de las propiedades de la superficie terrestre o del estado del mar.

- Componente v de viento neutral a 10 m [m/s] (ERA5): Este parámetro es la componente hacia el norte del “viento neutro”, a una altura de 10 metros sobre la superficie de la Tierra. El viento neutro se calcula a partir de la tensión superficial y la longitud de rugosidad correspondiente suponiendo que el aire está estratificado de forma neutra. El viento neutral es más lento que el viento real en condiciones estables y más rápido en condiciones inestables. El viento neutro está, por definición, en la dirección de la tensión superficial. El tamaño de la longitud de la rugosidad depende de las propiedades de la superficie terrestre o del estado del mar.
- Velocidad del viento neutro a 10 m [m/s] (ERA5): Combinación de las componentes del viento neutro a 10 m de altura.
- Componente u del viento a 10 m [m/s] (ERA5): Este parámetro es la componente hacia el este del viento a 10 m de altura. Es la velocidad horizontal del aire moviéndose hacia el este, a una altura de diez metros sobre la superficie de la Tierra, en metros por segundo. Este parámetro se puede combinar con la componente V del viento de 10 m para dar la velocidad y la dirección del viento horizontal de 10 m.
- Componente del viento a 10 m [m/s] (ERA5): Este parámetro es la componente hacia el norte del viento de 10 m. Es la velocidad horizontal del aire que se desplaza hacia el norte, a una altura de diez metros sobre la superficie de

la Tierra, en metros por segundo. Este parámetro se puede combinar con la componente U del viento de 10 m para dar la velocidad y dirección del viento horizontal de 10 m.

- Velocidad del viento a 10 m [m/s] (ERA5): Combinación de las componentes del viento horizontal a 10 m de altura.
- Ráfaga de viento a 10 m desde el posprocesamiento anterior [m/s] (ERA5): Viento máximo de cada 3 segundos a 10 m de altura según lo definido por la Organización Meteorológica Mundial (OMM). La parametrización representa la turbulencia y convección. La ráfaga de 3 s es calculada en cada paso de tiempo de 3 s y el máximo se mantiene desde el último procesamiento posterior.
- Máscara de fuego VNP14A1 001 (*FireMask-Confidence of fire*) [adimensional] (Suomi-NPP-NOAA): Es una representación espacial que identifica las áreas en las que se ha detectado actividad de fuego. Esta máscara de fuego se genera a partir de la información recopilada a partir del sensor VIIRS a bordo de los satélites Suomi-NPP y NOAA, que permite la detección y monitoreo de incendios forestales y otros eventos relacionados con el fuego a nivel global.
- Temperatura de punto de rocío a 2 m [K] (ERA5): Este parámetro es la temperatura a la que el aire debe enfriarse a presión constante para que alcance la saturación y el vapor de agua en el aire comience a condensarse en forma de gotas de agua o niebla, a 2 metros sobre la superficie de la Tierra, para que se produzca la saturación. Cuando el aire alcanza su temperatura de punto de rocío, está completamente saturado de humedad y cualquier enfriamiento adicional resultará en la formación de condensación, como rocío o niebla. La temperatura del punto de rocío a 2 m se calcula interpolando entre el nivel más bajo del modelo implementado por ERA5 y la superficie de la Tierra, teniendo en cuenta las condiciones atmosféricas. Este parámetro tiene unidades de kelvin (K). La

temperatura medida en kelvin se puede convertir a grados Celsius ($^{\circ}\text{C}$) restando 273,15.

- Humedad relativa a 2 metros por encima de la superficie terrestre [%] (ERA5): establece la relación entre la presión parcial del vapor de agua y la presión de vapor de equilibrio del agua a la temperatura dada a esa altura. Esta variable es calculada a partir de la combinación de los datos de temperatura y presión estimados por el satélite y se expresa en porcentaje ()�.
- Temperatura a 2 m [K] (ERA5): Este parámetro es la temperatura del aire a 2 m sobre la superficie de la tierra, el mar o las aguas interiores. La temperatura a 2 m se calcula interpolando entre el nivel más bajo del modelo y la superficie de la Tierra, teniendo en cuenta las condiciones atmosféricas. Este parámetro tiene unidades de kelvin (K). La temperatura medida en kelvin se puede convertir a grados Celsius ($^{\circ}\text{C}$) restando 273,15.
- Ángulo de orografía a escala inferior a la cuadrícula [radianes] (ERA5): Este parámetro es uno de los cuatro parámetros (los otros son la desviación estándar, la pendiente y la anisotropía) que describen las características de la orografía, las cuales son demasiado pequeñas para ser resueltas por la cuadrícula del modelo. Estos cuatro parámetros se calculan para características orográficas con escalas horizontales comprendidas entre 5 km y la resolución de cuadrícula del modelo, y se derivan de la altura de valles, colinas y montañas con una resolución de aproximadamente 1 km. Se utilizan como entrada para el esquema de orografía de subcuadrícula que representa el bloqueo de bajo nivel y los efectos de ondas de gravedad orográficas. El ángulo de la orografía a escala de subcuadrícula caracteriza la orientación geográfica del terreno en el plano horizontal (a vista de pájaro) en relación con un eje hacia el este. Este parámetro no varía en el tiempo.

- Anisotropía de la orografía a escala inferior a la cuadrícula [adimensional] (ERA5):
Este parámetro es uno de los cuatro parámetros (los otros son la desviación estándar, la pendiente y el ángulo de orografía a escala inferior a la cuadrícula) que describen las características de la orografía que son demasiado pequeñas para ser resueltas por la cuadrícula del modelo. Estos cuatro parámetros se calculan para características orográficas con escalas horizontales comprendidas entre 5 km y la resolución de cuadrícula del modelo, y se derivan de la altura de valles, colinas y montañas con una resolución de aproximadamente 1 km. Se utilizan como entrada para el esquema de orografía de subcuadrícula que representa el bloqueo de bajo nivel y los efectos de ondas de gravedad orográficas. Este parámetro es una medida de cuánto se distorsiona la forma del terreno en el plano horizontal (a vista de pájaro) de un círculo. Un valor de uno es un círculo, menos de uno una elipse y 0 es una cresta. En el caso de una cresta, el viento que sopla paralelo a él no ejerce ningún arrastre sobre el flujo, pero el viento que sopla perpendicular a él ejerce el máximo arrastre. Este parámetro no varía en el tiempo.
- Disipación de la capa límite [J/m^2] Este parámetro es la conversión acumulada de energía cinética en el flujo medio en calor, sobre toda la columna atmosférica, por unidad de área, que se debe a los efectos del estrés asociado con remolinos turbulentos cerca de la superficie y arrastre de forma orográfica turbulenta. Se calcula mediante los esquemas de arrastre de formas orográficas turbulentas y de difusión turbulenta del ECMWF Integrated Forecasting System. Los remolinos turbulentos cerca de la superficie están relacionados con la rugosidad de la superficie. El arrastre de forma orográfica turbulenta es la tensión debida a los valles, colinas y montañas en escalas horizontales por debajo de los 5 km, que se especifican a partir de datos de la superficie terrestre con una resolución de aproximadamente 1 km. (La disipación asociada con las características orográficas

con escalas horizontales entre 5 km y la escala de cuadrícula del modelo se tiene en cuenta mediante el esquema orográfico de subcuadrícula). Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación.

- Altura de la capa límite atmosférica [metro] (MERRA-2/ERA5): refleja la altura del estrato de aire por encima de la superficie terrestre que se ve afectado por la resistencia a la transferencia de momento, calor o humedad, que genera la superficie terrestre. En otras palabras, es la altura del estrato de aire sujeto a la influencia de la superficie terrestre. Esta puede ser tan baja como unas pocas decenas de metros, como sucede con el aire de refrigeración por la noche, o tan alta como varios kilómetros como ocurre sobre el desierto en medio de un día caluroso y soleado. Cuando es baja, pueden desarrollarse mayores concentraciones de contaminantes (emitidos en superficie). Su cálculo se basa en el número de Richardson y la unidad en la que se mide este parámetro son los metros (m).
- Radiación solar directa en cielo despejado en la superficie [J/m^2] (ERA5): Este parámetro es la cantidad de radiación directa del Sol, también conocida como radiación solar o de onda corta (400 a 700 nm), que llega a la superficie de la Tierra, suponiendo condiciones de cielo despejado (sin nubes). Es la cantidad de radiación que pasa a través de un plano horizontal. La radiación solar en la superficie puede ser directa o difusa. La radiación solar puede ser dispersada en todas las direcciones por partículas en la atmósfera, algunas de las cuales alcanzan la superficie (radiación solar difusa). Parte de la radiación solar llega a la superficie sin ser dispersada (radiación solar directa). Las cantidades de

radiación en cielo despejado se calculan exactamente para las mismas condiciones atmosféricas de temperatura, humedad, ozono, gases traza y aerosoles que las cantidades correspondientes del cielo total (nubes incluidas), pero asumiendo que las nubes no están allí. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Las unidades son julios por metro cuadrado [J/m^2]. Para convertir a vatios por metro cuadrado [W/m^2], los valores acumulados deben dividirse por el período de acumulación expresado en segundos. La convención ECMWF para flujos verticales es positiva hacia abajo.

- Altura de la base de la nube [metro] (ERA5): La altura sobre la superficie de la Tierra de la base de la capa de nubes más baja, en el momento especificado. Este parámetro se calcula buscando desde el segundo nivel más bajo del modelo hacia arriba, hasta la altura del nivel donde la fracción de nubes se vuelve mayor al 1% y el contenido de condensado mayor a 1.E-6 kg/kg . La niebla (es decir, la nube en la capa más baja del modelo) no se considera al definir la altura base de la nube.
- Energía potencial convectiva disponible [J/kg] (ERA5): Esta es una indicación de la inestabilidad (o estabilidad) de la atmósfera y se puede utilizar para evaluar el potencial de desarrollo de convección, que puede provocar fuertes lluvias, tormentas eléctricas y otras condiciones meteorológicas adversas. En el Sistema Integrado de Pronóstico (IFS) del ECMWF, el CAPE se calcula considerando parcelas de aire que parten a diferentes niveles del modelo por debajo del nivel de 350 hPa. Si una parcela de aire es más flotante (más cálida y/o con más humedad)

que el entorno que la rodea, seguirá ascendiendo (enfriándose a medida que asciende) hasta que alcance un punto en el que ya no tenga flotabilidad positiva.

CAPE es la energía potencial representada por el exceso de flotabilidad total.

El CAPE máximo producido por las distintas parcelas es el valor retenido. Los valores positivos grandes de CAPE indican que una parcela de aire sería mucho más cálida que su entorno circundante y, por lo tanto, muy boyante. CAPE está relacionado con la máxima velocidad vertical potencial del aire dentro de una corriente ascendente; por lo tanto, los valores más altos indican un mayor potencial para el clima severo. Los valores observados en ambientes tormentosos a menudo pueden exceder los 1000 julios por kilogramo [J/kg], pudiendo en casos extremos superar los 5000 [J/kg]. El cálculo de este parámetro asume: (i) la porción de aire no se mezcla con el aire circundante; (ii) el ascenso es pseudo-adiabático (toda el agua condensada cae) y (iii) otras simplificaciones relacionadas con el calentamiento por condensación de fase mixta.

- Inhibición convectiva [J/kg] (ERA5): Este parámetro es una medida de la cantidad de energía requerida para que comience la convección. Si el valor de este parámetro es demasiado alto, entonces es poco probable que ocurra una convección profunda y húmeda, incluso si la energía potencial convectiva disponible o el corte de la energía potencial convectiva disponible son grandes. Los valores de CIN superiores a 200 [J/kg] se considerarían altos. Una capa atmosférica donde la temperatura aumenta con la altura (lo que se conoce como inversión de temperatura) inhibiría el levantamiento convectivo y es una situación en la que la inhibición convectiva sería grande.
- Precipitación convectiva [metro] (ERA5): Este parámetro es la precipitación acumulada a lo largo de 3 horas previas a la fecha que indica el dato, que cae a la superficie de la Tierra, la cual es generada por el esquema de convección en el

Sistema de Pronóstico Integrado (IFS) del ECMWF. El esquema de convección representa la convección a escalas espaciales más pequeñas que el cuadro de cuadrícula. La precipitación también puede ser generada por el esquema de nubes en el IFS, que representa la formación y disipación de nubes y la precipitación a gran escala debido a cambios en las cantidades atmosféricas (como presión, temperatura y humedad) pronosticadas directamente a escalas espaciales del cuadro de cuadrícula o más grande. En el IFS, la precipitación se compone de lluvia y nieve. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora que finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Las unidades de este parámetro son la profundidad en metros de agua equivalente. Es la profundidad que tendría el agua si se distribuyera uniformemente sobre la caja de rejilla.

- Tasa de lluvia convectiva [$kg/m^2 s$] (ERA5): Este parámetro es la tasa de lluvia (intensidad de lluvia), en la superficie de la Tierra y en el momento especificado, que se genera mediante el esquema de convección en el Sistema Integrado de Pronóstico (IFS) del ECMWF. El esquema de convección representa la convección a escalas espaciales más pequeñas que el cuadro de cuadrícula. Las precipitaciones también pueden ser generadas por el esquema de nubes en el IFS, que representa la formación y disipación de nubes y precipitaciones a gran escala debido a cambios en las cantidades atmosféricas (como presión, temperatura y humedad) pronosticadas directamente en las escalas espaciales de la cuadrícula o más grande. En el IFS, la precipitación se compone de lluvia y nieve. Este parámetro es la tasa que tendría la lluvia si se distribuyera uniformemente sobre el cuadro de la cuadrícula. 1 kg de agua repartida en 1 metro cuadrado de superficie

tiene 1 mm de profundidad (despreciando los efectos de la temperatura sobre la densidad del agua), por lo que las unidades equivalen a mm por segundo.

- Nevadas convectivas [metros de agua equivalente] (ERA5): Este parámetro es la nieve acumulada que cae a la superficie de la Tierra, la cual es generada por el esquema de convección en el Sistema Integrado de Pronóstico (IFS) del ECMWF. El esquema de convección representa la convección a escalas espaciales más pequeñas que el cuadro de cuadrícula. Las nevadas también pueden ser generadas por el esquema de nubes en el IFS, que representa la formación y disipación de nubes y precipitaciones a gran escala debido a cambios en las cantidades atmosféricas (como presión, temperatura y humedad) pronosticadas directamente en las escalas espaciales de la cuadrícula. o más grande En el IFS, la precipitación se compone de lluvia y nieve. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, media del conjunto y dispersión del conjunto, el período de acumulación es durante las 3 horas que terminan en la fecha y hora de validez. Las unidades de este parámetro son la profundidad en metros de agua equivalente. Es la profundidad que tendría el agua si se distribuyera uniformemente sobre la caja de rejilla.
- Tasa de nevadas convectivas equivalente en agua [$kg/m^2 s$] (ERA5): Este parámetro es la tasa de nevadas (intensidad de las nevadas), en la superficie de la Tierra y en el momento especificado, que se genera mediante el esquema de convección en el Sistema Integrado de Pronóstico (IFS) del ECMWF. El esquema de convección representa la convección a escalas espaciales más pequeñas que el cuadro de cuadrícula. Las nevadas también pueden ser generadas por el esquema de nubes en el IFS, que representa la formación y disipación de nubes

y precipitaciones a gran escala debido a cambios en las cantidades atmosféricas (como presión, temperatura y humedad) pronosticadas directamente en las escalas espaciales de la cuadrícula. o más grande En el IFS, la precipitación se compone de lluvia y nieve. Este parámetro es la tasa que tendría la nevada si se distribuyera uniformemente sobre el cuadro de la cuadrícula. Dado que 1 kg de agua esparcida sobre 1 metro cuadrado de superficie tiene 1 mm de espesor (despreciando los efectos de la temperatura sobre la densidad del agua), las unidades son equivalentes a mm (de agua líquida) por segundo.

- Radiación UV hacia abajo en la superficie [J/m^2] (ERA5): Este parámetro es la cantidad de radiación ultravioleta (UV) que llega a la superficie. Es la cantidad de radiación que pasa a través de un plano horizontal. La radiación ultravioleta es parte del espectro electromagnético emitido por el Sol que tiene longitudes de onda más cortas que la luz visible. En el sistema de pronóstico integrado (IFS) del ECMWF se define como radiación con una longitud de onda de 0,20-0,44 [μm] (micras, 1 millonésima de metro). Pequeñas cantidades de UV son esenciales para los organismos vivos, pero la sobre exposición puede provocar daño celular; en humanos esto incluye efectos agudos y crónicos en la salud de la piel, los ojos y el sistema inmunológico. La radiación ultravioleta es absorbida por la capa de ozono, pero parte llega a la superficie. El agotamiento de la capa de ozono está causando preocupación por el aumento de los efectos dañinos de los rayos UV. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Las unidades son julios por metro cuadrado [J/m^2]. Para convertir a vatios por metro cuadrado [W/m^2], los valores acumulados deben dividirse por el período de acumulación

expresado en segundos. La convención ECMWF para flujos verticales es positiva hacia abajo.

- Altura de la base del conducto [metro] (ERA5): Altura de la base del conducto diagnosticada a partir del gradiente vertical de refractividad atmosférica.
- Esfuerzo superficial de ondas de gravedad hacia el este [N/m^2] (ERA5): El aire que fluye sobre una superficie ejerce una tensión (arrastre) que transfiere impulso a la superficie y frena el viento. Este parámetro es el componente de la tensión superficial acumulada en una dirección hacia el este, asociada con el bloqueo orográfico de bajo nivel y las ondas de gravedad orográficas. Se calcula mediante el esquema de orografía de subcuadrícula del Sistema Integrado de Pronóstico del ECMWF, que representa el estrés debido a valles, colinas y montañas no resueltos con escalas horizontales entre 5 km y la escala de cuadrícula del modelo. (La tensión asociada con las características orográficas con escalas horizontales menores de 5 km se explica por el esquema de arrastre de forma orográfica turbulenta). Las ondas de gravedad orográficas son oscilaciones en el flujo mantenidas por la flotabilidad de las parcelas de aire desplazadas, producidas cuando el aire es desviado hacia arriba por colinas y montañas. Este proceso puede crear tensión en la atmósfera en la superficie de la Tierra y en otros niveles de la atmósfera. Los valores positivos (negativos) indican tensión en la superficie de la Tierra en una dirección hacia el este (hacia el oeste). Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros

del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación..

- Tensión superficial turbulenta hacia el este [N/m^2] (ERA5): El aire que fluye sobre una superficie ejerce una tensión (arrastre) que transfiere impulso a la superficie y frena el viento. Este parámetro es el componente de la tensión superficial acumulada en una dirección hacia el este, asociado con remolinos turbulentos cerca de la superficie y arrastre de forma orográfica turbulenta. Se calcula mediante los esquemas de arrastre de formas orográficas turbulentas y de difusión turbulenta del ECMWF Integrated Forecasting System. Los remolinos turbulentos cerca de la superficie están relacionados con la rugosidad de la superficie. El arrastre de forma orográfica turbulenta es la tensión debida a los valles, colinas y montañas en escalas horizontales por debajo de los 5 km, que se especifican a partir de datos de la superficie terrestre con una resolución de aproximadamente 1 km. La tensión asociada con las características orográficas con escalas horizontales entre 5 km y la escala de cuadrícula del modelo se tiene en cuenta mediante el esquema orográfico de subcuadrícula. Los valores positivos (negativos) indican tensión en la superficie de la Tierra en dirección este (oeste). Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación.
- Evaporación [metros de agua equivalente] (ERA5): Este parámetro es la cantidad acumulada de agua que se ha evaporado de la superficie de la Tierra, incluida una representación simplificada de la transpiración (de la vegetación), en forma de vapor en el aire de arriba. Este parámetro se acumula durante un período

de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. La convención del Sistema Integrado de Pronóstico (IFS) del ECMWF es que los flujos descendentes son positivos. Por lo tanto, los valores negativos indican evaporación y los valores positivos indican condensación.

- Pronóstico de albedo [adimensional] (ERA5): Este parámetro es una medida de la reflectividad de la superficie de la Tierra. Es la fracción de radiación de onda corta (solar) reflejada por la superficie de la Tierra, para radiación difusa, asumiendo un espectro fijo de radiación de onda corta hacia abajo en la superficie. Los valores de este parámetro varían entre cero y uno. Por lo general, la nieve y el hielo tienen una alta reflectividad con valores de albedo de 0,8 y superiores, la tierra tiene valores intermedios entre 0,1 y 0,4 y el océano tiene valores bajos de 0,1 o menos. La radiación de onda corta del Sol es parcialmente reflejada de regreso al espacio por las nubes y partículas en la atmósfera (aerosoles) y parte de ella es absorbida. El resto incide sobre la superficie de la Tierra, donde parte de él se refleja. La porción que es reflejada por la superficie de la Tierra depende del albedo. En el Sistema Integrado de Pronóstico (IFS) del ECMWF, se utiliza un albedo de fondo climatológico (valores observados promediados durante un período de varios años), modificado por el modelo sobre agua, hielo y nieve. El albedo a menudo se muestra como un porcentaje (%).
- Logaritmo de pronóstico de la rugosidad de la superficie para el calor [adimensional] (ERA5): Este parámetro es el logaritmo natural de la longitud de rugosidad para el calor. La rugosidad de la superficie para el calor es una medida de la resistencia de la superficie a la transferencia de calor. Este parámetro se

utiliza para determinar la transferencia de calor del aire a la superficie. Para determinadas condiciones atmosféricas, una mayor rugosidad de la superficie para el calor significa que es más difícil que el aire intercambie calor con la superficie. Una menor rugosidad de la superficie para el calor significa que es más fácil que el aire intercambie calor con la superficie. Sobre el océano, la rugosidad de la superficie para el calor depende de las olas. Sobre el hielo marino, tiene un valor constante de 0,001 m. Sobre tierra, se deriva del tipo de vegetación y la capa de nieve.

- Rugosidad de la superficie [metro] (ERA5): Este parámetro es la longitud de rugosidad aerodinámica en metros. Es una medida de la resistencia superficial. Este parámetro se utiliza para determinar la transferencia de cantidad de movimiento del aire a la superficie. Para condiciones atmosféricas dadas, una mayor rugosidad de la superficie provoca una velocidad del viento cerca de la superficie más lenta. Sobre el océano, la rugosidad de la superficie depende de las olas. En tierra, la rugosidad de la superficie se deriva del tipo de vegetación y la capa de nieve.
- Velocidad de fricción [m/s] (ERA5): El aire que fluye sobre una superficie ejerce una tensión que transfiere impulso a la superficie y frena el viento. Este parámetro es una velocidad teórica del viento en la superficie de la Tierra que expresa la magnitud del estrés. Se calcula dividiendo la tensión superficial por la densidad del aire y sacando su raíz cuadrada. Para flujo turbulento, la velocidad de fricción es aproximadamente constante en los pocos metros más bajos de la atmósfera. Este parámetro aumenta con la rugosidad de la superficie. Se utiliza para calcular la forma en que el viento cambia con la altura en los niveles más bajos de la atmósfera.
- Geopotencial [m^2/s^2] (ERA5): Este parámetro es la energía potencial gravi-

tacional de una unidad de masa, en un lugar particular de la superficie de la Tierra, en relación con el nivel medio del mar. También es la cantidad de trabajo que tendría que realizarse, contra la fuerza de la gravedad, para elevar una unidad de masa hasta ese lugar desde el nivel medio del mar. La altura geopotencial (superficial) (orografía) se puede calcular dividiendo el geopotencial (superficial) por la aceleración gravitatoria de la Tierra, g ($=9,80665 [m^2/s^2]$). Este parámetro no varía en el tiempo.

- Disipación de ondas de gravedad [J/m^2] (ERA5): Este parámetro es la conversión acumulada de energía cinética en el flujo medio en calor, sobre toda la columna atmosférica, por unidad de área, que se debe a los efectos del estrés asociado con el bloqueo orográfico de bajo nivel y las ondas de gravedad orográficas. Se calcula mediante el esquema de orografía de subcuadrícula del Sistema Integrado de Pronóstico del ECMWF, que representa el estrés debido a valles, colinas y montañas no resueltas con escalas horizontales entre 5 km y la escala de cuadrícula del modelo. (La disipación asociada con características orográficas con escalas horizontales menores de 5 km se explica por el esquema de arrastre de forma orográfica turbulenta). Las ondas de gravedad orográficas son oscilaciones en el flujo mantenidas por la flotabilidad de las parcelas de aire desplazadas, producidas cuando el aire es desviado hacia arriba por colinas y montañas. Este proceso puede crear tensión en la atmósfera en la superficie de la Tierra y en otros niveles de la atmósfera. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación.
- Nubosidad alta [adimensional] (ERA5): La proporción de una caja de cuadrícula

cubierta por nubes que se producen en los niveles altos de la troposfera. La nube alta es un campo de un solo nivel calculado a partir de la nube que se produce en los niveles del modelo con una presión inferior a 0,45 veces la presión superficial. Por lo tanto, si la presión en la superficie es de 1000 hPa (hectopascal), la nubosidad alta se calcularía utilizando niveles con una presión de menos de 450 hPa (aproximadamente 6 km y más (suponiendo una “atmósfera estándar”)). El parámetro de nubosidad alta se calcula a partir de la nube para los niveles de modelo apropiados, como se describe anteriormente. Se hacen suposiciones sobre el grado de superposición/aleatoriedad entre nubes en diferentes niveles de modelo. Las fracciones de nubes varían de 0 a 1.

- Cobertura vegetal alta [adimensional] (ERA5): Este parámetro es la fracción del cuadro de cuadrícula que está cubierta con vegetación que se clasifica como “alta”. Los valores varían entre 0 y 1 pero no varían en el tiempo. Este es uno de los parámetros del modelo que describe la vegetación de la superficie terrestre. La “vegetación alta” consiste en árboles de hoja perenne, árboles de hoja caduca, bosque mixto/bosque y bosque interrumpido.
- Ráfaga de viento instantánea a 10 m [m/s] (ERA5): Este parámetro es la ráfaga de viento máxima en el momento especificado, a una altura de diez metros sobre la superficie de la Tierra. La OMM define una ráfaga de viento como el máximo del viento promediado en intervalos de 3 segundos. Esta duración es más corta que un paso de tiempo del modelo, por lo que el Sistema Integrado de Pronóstico (IFS) del ECMWF deduce la magnitud de una ráfaga dentro de cada paso de tiempo a partir de la tensión superficial promediada por paso de tiempo, la fricción de la superficie, la cizalladura del viento y la estabilidad.
- Tensión superficial turbulenta instantánea hacia el este [N/m^2] (ERA5): El aire que fluye sobre una superficie ejerce una tensión (arrastre) que transfiere impulso

a la superficie y frena el viento. Este parámetro es el componente de la tensión superficial en el momento especificado, en dirección este, asociado con remolinos turbulentos cerca de la superficie y arrastre de forma orográfica turbulenta. Se calcula mediante los esquemas de arrastre de formas orográficas turbulentas y de difusión turbulenta del ECMWF Integrated Forecasting System. Los remolinos turbulentos cerca de la superficie están relacionados con la rugosidad de la superficie. El arrastre de forma orográfica turbulenta es la tensión debida a los valles, colinas y montañas en escalas horizontales por debajo de los 5 km, que se especifican a partir de datos de la superficie terrestre con una resolución de aproximadamente 1 km.

- Fracción de precipitación superficial instantánea a gran escala [adimensional] (ERA5): Este parámetro es la fracción de la cuadrícula (0-1) cubierta por precipitación a gran escala en el momento especificado. La precipitación a gran escala es lluvia y nieve que cae sobre la superficie de la Tierra y es generada por el esquema de nubes en el Sistema Integrado de Pronóstico (IFS) del ECMWF. El esquema de nubes representa la formación y disipación de nubes y precipitaciones a gran escala debido a cambios en las cantidades atmosféricas (como presión, temperatura y humedad) pronosticadas directamente por el IFS a escalas espaciales de una cuadrícula o más grandes. La precipitación también puede deberse a la convección generada por el esquema de convección en el IFS. El esquema de convección representa la convección a escalas espaciales más pequeñas que el cuadro de cuadrícula.
- Flujo instantáneo de humedad [kg/m^2s] (ERA5): Este parámetro es la tasa neta de intercambio de humedad entre la superficie terrestre/oceánica y la atmósfera, debido a los procesos de evaporación (incluida la evapotranspiración) y condensación, en el momento especificado. Por convención, los flujos descendentes

son positivos, lo que significa que la evaporación se representa con valores negativos y la condensación con valores positivos.

- Tensión superficial turbulenta instantánea hacia el norte [N/m^2] (ERA5): El aire que fluye sobre una superficie ejerce una tensión (arrastre) que transfiere impulso a la superficie y frena el viento. Este parámetro es el componente de la tensión superficial en el momento especificado, en dirección norte, asociado con remolinos turbulentos cerca de la superficie y arrastre de forma orográfica turbulenta. Se calcula mediante los esquemas de arrastre de formas orográficas turbulentas y de difusión turbulenta del ECMWF Integrated Forecasting System. Los remolinos turbulentos cerca de la superficie están relacionados con la rugosidad de la superficie. El arrastre de forma orográfica turbulenta es la tensión debida a los valles, colinas y montañas en escalas horizontales por debajo de los 5 km, que se especifican a partir de datos de la superficie terrestre con una resolución de aproximadamente 1 km.
- Flujo de calor sensible superficial instantáneo [W/m^2] (ERA5): Este parámetro es la transferencia de calor entre la superficie de la Tierra y la atmósfera, en el momento especificado, a través de los efectos del movimiento turbulento del aire (pero excluyendo cualquier transferencia de calor resultante de la condensación o la evaporación). La magnitud del flujo de calor sensible se rige por la diferencia de temperatura entre la superficie y la atmósfera suprayacente, la velocidad del viento y la rugosidad de la superficie. Por ejemplo, el aire frío que recubre una superficie cálida produciría un flujo de calor sensible desde la tierra (o el océano) hacia la atmósfera. La convención ECMWF para flujos verticales es positiva hacia abajo.
- Índice K [k] (ERA5): Este parámetro es una medida del potencial de desarrollo de una tormenta eléctrica, calculado a partir de la temperatura y la temperatura

del punto de rocío en la parte inferior de la atmósfera. El cálculo utiliza la temperatura a 850, 700 y 500 hPa y la temperatura de punto de rocío a 850 y 700 hPa. Los valores más altos de K indican un mayor potencial para el desarrollo de tormentas eléctricas. Este parámetro está relacionado con la probabilidad de ocurrencia de una tormenta: <20 K Sin tormenta, 20-25 K Tormentas aisladas, 26-30 K Tormentas dispersas, 31-35 K Tormentas dispersas, >35 K Numerosas tormentas.

- Temperatura del fondo del lago [k] (ERA5): Este parámetro es la temperatura del agua en el fondo de los cuerpos de agua continentales (lagos, embalses, ríos y aguas costeras). Este parámetro se define en todo el mundo, incluso donde no hay aguas continentales. Las regiones sin aguas continentales se pueden enmascarar considerando solo los puntos de cuadrícula donde la cobertura del lago es superior a 0,0. En mayo de 2015, se implementó un modelo de lago en el Sistema Integrado de Pronóstico (IFS) del ECMWF para representar la temperatura del agua y el hielo del lago de las principales masas de agua continentales del mundo. La profundidad del lago y la fracción de área (cobertura) se mantienen constantes en el tiempo. Este parámetro tiene unidades de kelvin (K). La temperatura medida en kelvin se puede convertir a grados Celsius ($^{\circ}\text{C}$) restando 273,15.
- Cubierta del lago [adimensional] (ERA5): Este parámetro es la proporción de una cuadrícula cubierta por cuerpos de agua continentales (lagos, embalses, ríos y aguas costeras). Los valores varían entre 0: sin agua continental y 1: el cuadro de la cuadrícula está completamente cubierto con agua continental. Este parámetro se especifica a partir de observaciones y no varía en el tiempo. En mayo de 2015, se implementó un modelo de lago en el Sistema Integrado de Pronóstico (IFS) del ECMWF para representar la temperatura del agua y el hielo del lago de las principales masas de agua continentales del mundo.

- Profundidad del lago [metro] (ERA5): Este parámetro es la profundidad media de las masas de agua continentales (lagos, embalses, ríos y aguas costeras). Este parámetro se especifica a partir de mediciones in situ y estimaciones indirectas y no varía en el tiempo. Este parámetro se define en todo el mundo, incluso donde no hay aguas continentales. Las regiones sin aguas continentales se pueden enmascarar considerando solo los puntos de cuadrícula donde la cobertura del lago es superior a 0,0. En mayo de 2015, se implementó un modelo de lago en el Sistema Integrado de Pronóstico (IFS) del ECMWF para representar la temperatura del agua y el hielo del lago de las principales masas de agua continentales del mundo.
- Profundidad del hielo del lago [metro] (ERA5): Este parámetro es el espesor del hielo en las masas de agua continentales (lagos, embalses, ríos y aguas costeras). Este parámetro se define en todo el mundo, incluso donde no hay aguas continentales. Las regiones sin aguas continentales se pueden enmascarar considerando solo los puntos de cuadrícula donde la cobertura del lago es superior a 0,0. En mayo de 2015, se implementó un modelo de lago en el Sistema Integrado de Pronóstico (IFS) del ECMWF para representar la temperatura del agua y el hielo del lago de las principales masas de agua continentales del mundo. La profundidad del lago y la fracción de área (cobertura) se mantienen constantes en el tiempo. Se utiliza una sola capa de hielo para representar la formación y el derretimiento del hielo en masas de agua continentales. Este parámetro es el espesor de esa capa de hielo. temperatura del hielo del lago k Este parámetro es la temperatura de la superficie superior del hielo en cuerpos de agua continentales (lagos, embalses, ríos y aguas costeras). Es la temperatura en la interfaz hielo/atmósfera o hielo/nieve. Este parámetro se define en todo el mundo, incluso donde no hay aguas continentales. Las regiones sin aguas continentales se pueden enmascarar considerando solo los puntos de cuadrícula

donde la cobertura del lago es superior a 0,0. En mayo de 2015, se implementó un modelo de lago en el Sistema Integrado de Pronóstico (IFS) del ECMWF para representar la temperatura del agua y el hielo del lago de las principales masas de agua continentales del mundo. La profundidad del lago y la fracción de área (cobertura) se mantienen constantes en el tiempo. Se utiliza una sola capa de hielo para representar la formación y el derretimiento del hielo en masas de agua continentales. Este parámetro tiene unidades de kelvin (K).

- Profundidad de la capa de mezcla del lago [metro] (ERA5): Este parámetro es el espesor de la capa superior de las masas de agua continentales (lagos, embalses, ríos y aguas costeras) que está bien mezclada y tiene una temperatura casi constante con la profundidad (es decir, una distribución uniforme de la temperatura con la profundidad). La mezcla puede ocurrir cuando la densidad del agua superficial (y cercana a la superficie) es mayor que la del agua debajo. La mezcla también puede ocurrir por la acción del viento en la superficie del agua. Este parámetro se define en todo el mundo, incluso donde no hay aguas continentales. Las regiones sin aguas continentales se pueden enmascarar considerando solo los puntos de cuadrícula donde la cobertura del lago es superior a 0,0. En mayo de 2015, se implementó un modelo de lago en el Sistema Integrado de Pronóstico (IFS) del ECMWF para representar la temperatura del agua y el hielo del lago de las principales masas de agua continentales del mundo. La profundidad del lago y la fracción de área (cobertura) se mantienen constantes en el tiempo. Los cuerpos de agua continentales se representan con dos capas en la vertical, la capa mixta arriba y la termoclina abajo, donde la temperatura cambia con la profundidad. El límite superior de la termoclina se encuentra en el fondo de la capa mixta y el límite inferior de la termoclina en el fondo del lago. Se utiliza una sola capa de hielo para representar la formación y el derretimiento del hielo en masas de agua continentales.

- Temperatura de la capa de mezcla del lago [k] (ERA5): Este parámetro es la temperatura de la capa superior de las masas de agua continentales (lagos, embalses, ríos y aguas costeras) que está bien mezclada y tiene una temperatura casi constante con la profundidad (es decir, una distribución uniforme de la temperatura con la profundidad). La mezcla puede ocurrir cuando la densidad del agua superficial (y cercana a la superficie) es mayor que la del agua debajo. La mezcla también puede ocurrir por la acción del viento en la superficie del agua. Este parámetro se define en todo el mundo, incluso donde no hay aguas continentales. Las regiones sin aguas continentales se pueden enmascarar considerando solo los puntos de cuadrícula donde la cobertura del lago es superior a 0,0. En mayo de 2015, se implementó un modelo de lago en el Sistema Integrado de Pronóstico (IFS) del ECMWF para representar la temperatura del agua y el hielo del lago de las principales masas de agua continentales del mundo. La profundidad del lago y la fracción de área (cobertura) se mantienen constantes en el tiempo. Los cuerpos de agua continentales se representan con dos capas en la vertical, la capa mixta arriba y la termoclina abajo, donde la temperatura cambia con la profundidad. El límite superior de la termoclina se encuentra en el fondo de la capa mixta y el límite inferior de la termoclina en el fondo del lago. Se utiliza una sola capa de hielo para representar la formación y el derretimiento del hielo en masas de agua continentales. Este parámetro tiene unidades de kelvin (K). La temperatura medida en kelvin se puede convertir a grados Celsius ($^{\circ}\text{C}$) restando 273,15. El límite superior de la termoclina se encuentra en el fondo de la capa mixta y el límite inferior de la termoclina en el fondo del lago. Se utiliza una sola capa de hielo para representar la formación y el derretimiento del hielo en masas de agua continentales. Este parámetro tiene unidades de kelvin (K). La temperatura medida en kelvin se puede convertir a grados Celsius ($^{\circ}\text{C}$) restando 273,15. El límite superior de la termoclina se

encuentra en el fondo de la capa mixta y el límite inferior de la termoclina en el fondo del lago. Se utiliza una sola capa de hielo para representar la formación y el derretimiento del hielo en masas de agua continentales. Este parámetro tiene unidades de kelvin (K). La temperatura medida en kelvin se puede convertir a grados Celsius ($^{\circ}\text{C}$) restando 273,15.

- Factor de forma del lago [adimensional] (ERA5): Este parámetro describe la forma en que la temperatura cambia con la profundidad en la capa de termoclina de los cuerpos de agua continentales (lagos, embalses, ríos y aguas costeras), es decir, describe la forma del perfil de temperatura vertical. Se utiliza para calcular la temperatura del fondo del lago y otros parámetros relacionados con el lago. Este parámetro se define en todo el mundo, incluso donde no hay aguas continentales. Las regiones sin aguas continentales se pueden enmascarar considerando solo los puntos de cuadrícula donde la cobertura del lago es superior a 0,0. En mayo de 2015, se implementó un modelo de lago en el Sistema Integrado de Pronóstico (IFS) del ECMWF para representar la temperatura del agua y el hielo del lago de las principales masas de agua continentales del mundo. La profundidad del lago y la fracción de área (cobertura) se mantienen constantes en el tiempo. Los cuerpos de agua continentales se representan con dos capas en la vertical, la capa mixta arriba y la termoclina abajo, donde la temperatura cambia con la profundidad. El límite superior de la termoclina se encuentra en el fondo de la capa mixta y el límite inferior de la termoclina en el fondo del lago. Se utiliza una sola capa de hielo para representar la formación y el derretimiento del hielo en masas de agua continentales.
- Temperatura total de la capa del lago [k] (ERA5): Este parámetro es la temperatura media de la columna de agua total en cuerpos de agua continentales (lagos, embalses, ríos y aguas costeras). Este parámetro se define en todo el

mundo, incluso donde no hay aguas continentales. Las regiones sin aguas continentales se pueden enmascarar considerando solo los puntos de cuadrícula donde la cobertura del lago es superior a 0,0. En mayo de 2015, se implementó un modelo de lago en el Sistema Integrado de Pronóstico (IFS) del ECMWF para representar la temperatura del agua y el hielo del lago de las principales masas de agua continentales del mundo. La profundidad del lago y la fracción de área (cobertura) se mantienen constantes en el tiempo. Los cuerpos de agua continentales se representan con dos capas en la vertical, la capa mixta arriba y la termoclina abajo, donde la temperatura cambia con la profundidad. Este parámetro es la temperatura media sobre las dos capas. El límite superior de la termoclina se encuentra en el fondo de la capa mixta y el límite inferior de la termoclina en el fondo del lago. Se utiliza una sola capa de hielo para representar la formación y el derretimiento del hielo en masas de agua continentales. Este parámetro tiene unidades de kelvin (K). La temperatura medida en kelvin se puede convertir a grados Celsius ($^{\circ}\text{C}$) restando 273,15.

- Tasa de lluvia a gran escala [$\text{kg}/\text{m}^2\text{s}$] (ERA5): Este parámetro es la tasa de lluvia (intensidad de lluvia), en la superficie de la Tierra y en el momento especificado, que genera el esquema de nubes en el Sistema Integrado de Pronóstico (IFS) del ECMWF. El esquema de nubes representa la formación y disipación de nubes y la precipitación a gran escala debido a cambios en las cantidades atmosféricas (como presión, temperatura y humedad) pronosticadas directamente a escalas espaciales de la cuadrícula o mayores. La lluvia también puede ser generada por el esquema de convección en el IFS, que representa la convección a escalas espaciales más pequeñas que la cuadrícula. En el IFS, la precipitación se compone de lluvia y nieve. Este parámetro es la tasa que tendría la lluvia si se distribuyera uniformemente sobre el cuadro de la cuadrícula. Dado que 1 kg de agua repartida en 1 metro cuadrado de superficie tiene 1 mm de profundidad (despreciando

los efectos de la temperatura sobre la densidad del agua), las unidades son equivalentes a mm por segundo.

- Tasa de nevadas a gran escala equivalente en agua [kg/m^2s] (ERA5): Este parámetro es la tasa de nevadas (intensidad de las nevadas), en la superficie de la Tierra y en el momento especificado, que genera el esquema de nubes en el Sistema Integrado de Pronóstico (IFS) del ECMWF. El esquema de nubes representa la formación y disipación de nubes y la precipitación a gran escala debido a cambios en las cantidades atmosféricas (como presión, temperatura y humedad) pronosticadas directamente a escalas espaciales de la cuadrícula o mayores. Las nevadas también pueden ser generadas por el esquema de convección en el IFS, que representa la convección a escalas espaciales más pequeñas que la cuadrícula. En el IFS, la precipitación se compone de lluvia y nieve. Este parámetro es la tasa que tendría la nevada si se distribuyera uniformemente sobre el cuadro de la cuadrícula. Dado que 1 kg de agua repartida en 1 metro cuadrado de superficie tiene 1 mm de profundidad (despreciando los efectos de la temperatura sobre la densidad del agua), las unidades equivalen a mm (de agua líquida) por segundo.
- Precipitación a gran escala [metro] (ERA5): Este parámetro es la precipitación acumulada que cae a la superficie de la Tierra, que es generada por el esquema de nubes en el Sistema Integrado de Pronóstico (IFS) del ECMWF. El esquema de nubes representa la formación y disipación de nubes y la precipitación a gran escala debido a cambios en las cantidades atmosféricas (como presión, temperatura y humedad) pronosticadas directamente a escalas espaciales de la cuadrícula o mayores. La precipitación también puede ser generada por el esquema de convección en el IFS, que representa la convección a escalas espaciales más pequeñas que la cuadrícula. En el IFS, la precipitación se compone de lluvia

y nieve. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, media del conjunto y dispersión del conjunto, el período de acumulación es durante las 3 horas que terminan en la fecha y hora de validez. Las unidades de este parámetro son la profundidad en metros de agua equivalente. Es la profundidad que tendría el agua si se distribuyera uniformemente sobre la caja de rejilla.

- Fracción de precipitación a gran escala [s] (ERA5): Este parámetro es la acumulación de la fracción de la cuadrícula (0-1) que está cubierta por precipitaciones a gran escala. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación.
- Nevadas a gran escala [metros de agua equivalente] (ERA5): Este parámetro es la nieve acumulada que cae a la superficie de la Tierra, que es generada por el esquema de nubes en el Sistema Integrado de Pronóstico (IFS) del ECMWF. El esquema de nubes representa la formación y disipación de nubes y la precipitación a gran escala debido a cambios en las cantidades atmosféricas (como presión, temperatura y humedad) pronosticadas directamente a escalas espaciales de la cuadrícula o mayores. Las nevadas también pueden ser generadas por el esquema de convección en el IFS, que representa la convección a escalas espaciales más pequeñas que la cuadrícula. En el IFS, la precipitación se compone de lluvia y nieve. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es

superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, media del conjunto y dispersión del conjunto, el período de acumulación es durante las 3 horas que terminan en la fecha y hora de validez. Las unidades de este parámetro son la profundidad en metros de agua equivalente. Es la profundidad que tendría el agua si se distribuyera uniformemente sobre la caja de rejilla.

- Índice de área foliar, vegetación alta [m^2/m^2] (ERA5): Este parámetro es la superficie de una cara de todas las hojas que se encuentran sobre una superficie de terreno para vegetación clasificada como “alta”. Este parámetro tiene valor 0 sobre suelo desnudo o sin hojas. Se puede calcular diariamente a partir de datos satelitales. Es importante para pronosticar, por ejemplo, cuánta agua de lluvia será interceptada por el dosel vegetal, en lugar de caer al suelo. Este es uno de los parámetros del modelo que describe la vegetación de la superficie terrestre. La “vegetación alta” consiste en árboles de hoja perenne, árboles de hoja caduca, bosque mixto/bosque y bosque interrumpido.
- Índice de área foliar, vegetación baja [m^2/m^2] (ERA5): Este parámetro es la superficie de una cara de todas las hojas que se encuentran sobre una superficie de terreno para vegetación clasificada como “baja”. Este parámetro tiene valor 0 sobre suelo desnudo o sin hojas. Se puede calcular diariamente a partir de datos satelitales. Es importante para pronosticar, por ejemplo, cuánta agua de lluvia será interceptada por el dosel vegetal, en lugar de caer al suelo. Este es uno de los parámetros del modelo que describe la vegetación de la superficie terrestre. La “vegetación baja” consiste en cultivos y agricultura mixta, cultivos de regadío, pastos cortos, pastos altos, tundra, semidesierto, pantanos y marismas, arbustos de hoja perenne, arbustos de hoja caduca y mezclas de agua y tierra.
- Nubosidad baja [adimensional] (ERA5): Este parámetro es la proporción de

una cuadrícula cubierta por nubes que se encuentran en los niveles inferiores de la troposfera. La nube baja es un campo de un solo nivel calculado a partir de nubes que se producen en los niveles del modelo con una presión superior a 0,8 veces la presión superficial. Por lo tanto, si la presión en la superficie es de 1000 hPa (hectopascal), la nube baja se calcularía utilizando niveles con una presión superior a 800 hPa (por debajo de aproximadamente 2 km (suponiendo una “atmósfera estándar”)). Se hacen suposiciones sobre el grado de superposición/aleatoriedad entre nubes en diferentes niveles de modelo. Este parámetro tiene valores de 0 a 1.

- Cobertura vegetal baja [adimensional] (ERA5): Este parámetro es la fracción del cuadro de cuadrícula que está cubierta con vegetación que se clasifica como “baja”. Los valores varían entre 0 y 1 pero no varían en el tiempo. Este es uno de los parámetros del modelo que describe la vegetación de la superficie terrestre. La “vegetación baja” consiste en cultivos y agricultura mixta, cultivos de regadío, pastos cortos, pastos altos, tundra, semidesierto, pantanos y marismas, arbustos de hoja perenne, arbustos de hoja caduca y mezclas de agua y tierra.
- Temperatura máxima de 2 m desde el posprocesamiento anterior [k] (ERA5): Este parámetro es la temperatura más alta del aire a 2 m sobre la superficie de la tierra, el mar o las aguas continentales desde que se archivó el parámetro por última vez en un pronóstico particular. La temperatura a 2 m se calcula interpolando entre el nivel más bajo del modelo y la superficie de la Tierra, teniendo en cuenta las condiciones atmosféricas. Este parámetro tiene unidades de kelvin (K). La temperatura medida en kelvin se puede convertir a grados Celsius ($^{\circ}\text{C}$) restando 273,15.
- Tasa de precipitación total máxima desde el posprocesamiento anterior [$\text{kg}/\text{m}^2\text{s}$] (ERA5): La precipitación total se calcula a partir de las tasas combinadas de

lluvia y nevadas a gran escala y convectivas en cada paso de tiempo y se mantiene el máximo desde el último procesamiento posterior.

- Disipación media de la capa límite [W/m^2] (ERA5): Este parámetro es la tasa media de conversión de energía cinética en el flujo medio en calor, en toda la columna atmosférica, por unidad de área, que se debe a los efectos de la tensión asociada con remolinos turbulentos cerca de la superficie y arrastre de forma orográfica turbulenta. Se calcula mediante los esquemas de arrastre de formas orográficas turbulentas y de difusión turbulenta del ECMWF Integrated Forecasting System. Los remolinos turbulentos cerca de la superficie están relacionados con la rugosidad de la superficie. El arrastre de forma orográfica turbulenta es la tensión debida a los valles, colinas y montañas en escalas horizontales por debajo de los 5 km, que se especifican a partir de datos de la superficie terrestre con una resolución de aproximadamente 1 km. (La disipación asociada con las características orográficas con escalas horizontales entre 5 km y la escala de cuadrícula del modelo se tiene en cuenta mediante el esquema orográfico de subcuadrícula). Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación.
- Tasa media de precipitación convectiva [kg/m^2s] (ERA5): Este parámetro es la tasa de precipitación en la superficie de la Tierra, que es generada por el esquema de convección en el Sistema Integrado de Pronóstico (IFS) del ECMWF. El esquema de convección representa la convección a escalas espaciales más pequeñas que el cuadro de cuadrícula. La precipitación también puede ser generada por el esquema de nubes en el IFS, que representa la formación y disipación de nubes

y la precipitación a gran escala debido a cambios en las cantidades atmosféricas (como presión, temperatura y humedad) pronosticadas directamente a escalas espaciales del cuadro de cuadrícula. o más grande En el IFS, la precipitación se compone de lluvia y nieve. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Es la tasa que tendría la precipitación si se distribuyera uniformemente sobre la cuadrícula. 1 kg de agua repartida en 1 metro cuadrado de superficie tiene 1 mm de profundidad (despreciando los efectos de la temperatura sobre la densidad del agua), por lo que las unidades equivalen a mm (de agua líquida) por segundo.

- Tasa media de nevadas convectivas [kg/m^2s] (ERA5): Este parámetro es la tasa de nevadas (intensidad de las nevadas) en la superficie de la Tierra, que es generada por el esquema de convección en el Sistema Integrado de Pronóstico (IFS) del ECMWF. El esquema de convección representa la convección a escalas espaciales más pequeñas que el cuadro de cuadrícula. Las nevadas también pueden ser generadas por el esquema de nubes en el IFS, que representa la formación y disipación de nubes y precipitaciones a gran escala debido a cambios en las cantidades atmosféricas (como presión, temperatura y humedad) pronosticadas directamente en las escalas espaciales de la cuadrícula. o más grande En el IFS, la precipitación se compone de lluvia y nieve. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Es la tasa que tendría la nevada si se distribuyera

uniformemente sobre la cuadrícula. Dado que 1 kg de agua esparcida sobre 1 metro cuadrado de superficie tiene 1 mm de espesor (despreciando los efectos de la temperatura sobre la densidad del agua), las unidades son equivalentes a mm (de agua líquida) por segundo.

- Tensión superficial media de la onda de gravedad hacia el este [N/m^2] (ERA5): El aire que fluye sobre una superficie ejerce una tensión (arrastre) que transfiere impulso a la superficie y frena el viento. Este parámetro es el componente de la tensión superficial media en una dirección hacia el este, asociado con el bloqueo orográfico de bajo nivel y las ondas de gravedad orográficas. Se calcula mediante el esquema de orografía de subcuadrícula del Sistema Integrado de Pronóstico del ECMWF, que representa el estrés debido a valles, colinas y montañas no resueltos con escalas horizontales entre 5 km y la escala de cuadrícula del modelo. (La tensión asociada con las características orográficas con escalas horizontales menores de 5 km se explica por el esquema de arrastre de forma orográfica turbulenta). Las ondas de gravedad orográficas son oscilaciones en el flujo mantenidas por la flotabilidad de las parcelas de aire desplazadas, producidas cuando el aire es desviado hacia arriba por colinas y montañas. Este proceso puede crear tensión en la atmósfera en la superficie de la Tierra y en otros niveles de la atmósfera. Los valores positivos (negativos) indican tensión en la superficie de la Tierra en una dirección hacia el este (hacia el oeste). Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y

finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación.

- Tensión superficial turbulenta media hacia el este [N/m^2] (ERA5): El aire que fluye sobre una superficie ejerce una tensión (arrastre) que transfiere impulso a la superficie y frena el viento. Este parámetro es el componente de la tensión superficial media en dirección este, asociado con remolinos turbulentos cerca de la superficie y arrastre de forma orográfica turbulenta. Se calcula mediante los esquemas de arrastre de formas orográficas turbulentas y de difusión turbulenta del ECMWF Integrated Forecasting System. Los remolinos turbulentos cerca de la superficie están relacionados con la rugosidad de la superficie. El arrastre de forma orográfica turbulenta es la tensión debida a los valles, colinas y montañas en escalas horizontales por debajo de los 5 km, que se especifican a partir de datos de la superficie terrestre con una resolución de aproximadamente 1 km. (La tensión asociada con las características orográficas con escalas horizontales entre 5 km y la escala de cuadrícula del modelo se tiene en cuenta mediante el esquema orográfico de subcuadrícula). Los valores positivos (negativos) indican tensión en la superficie de la Tierra en dirección este (oeste). Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación.

- Tasa media de evaporación [$kg/m^2 s$] (ERA5): Este parámetro es la cantidad de agua que se ha evaporado de la superficie de la Tierra, incluida una representación simplificada de la transpiración (de la vegetación), en forma de vapor en el aire de arriba. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. La convención del Sistema Integrado de Pronóstico (IFS) del ECMWF es que los flujos descendentes son positivos. Por lo tanto, los valores negativos indican evaporación y los valores positivos indican condensación.
- Disipación media de ondas de gravedad [W/m^2] (ERA5): Este parámetro es la tasa media de conversión de la energía cinética en el flujo medio en calor, en toda la columna atmosférica, por unidad de área, que se debe a los efectos de la tensión asociada con el bloqueo orográfico de bajo nivel y las ondas de gravedad orográficas. Se calcula mediante el esquema de orografía de subcuadrícula del Sistema Integrado de Pronóstico del ECMWF, que representa el estrés debido a valles, colinas y montañas no resueltas con escalas horizontales entre 5 km y la escala de cuadrícula del modelo. (La disipación asociada con características orográficas con escalas horizontales menores de 5 km se explica por el esquema de arrastre de forma orográfica turbulenta). Las ondas de gravedad orográficas son oscilaciones en el flujo mantenidas por la flotabilidad de las parcelas de aire desplazadas, producidas cuando el aire es desviado hacia arriba por colinas y montañas. Este proceso puede crear tensión en la atmósfera en la superficie de la Tierra y en otros niveles de la atmósfera. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza

en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación.

- Fracción media de precipitación a gran escala [adimensional] (ERA5): Este parámetro es la media de la fracción de la cuadrícula (0-1) que está cubierta por precipitaciones a gran escala. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación.
- Tasa de precipitación media a gran escala [kg/m^2s] (ERA5): Este parámetro es la tasa de precipitación en la superficie de la Tierra, que es generada por el esquema de nubes en el Sistema Integrado de Pronóstico (IFS) del ECMWF. El esquema de nubes representa la formación y disipación de nubes y la precipitación a gran escala debido a cambios en las cantidades atmosféricas (como presión, temperatura y humedad) pronosticadas directamente a escalas espaciales de la cuadrícula o mayores. La precipitación también puede ser generada por el esquema de convección en el IFS, que representa la convección a escalas espaciales más pequeñas que la cuadrícula. En el IFS, la precipitación se compone de lluvia y nieve. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, media del conjunto y dispersión del conjunto, el período de procesamiento es de más de 3 horas y finaliza en la fecha y hora de validez. Es la tasa que tendría la precipitación si se distribuyera uniformemente

sobre la cuadrícula. Dado que 1 kg de agua repartida en 1 metro cuadrado de superficie tiene 1 mm de profundidad (despreciando los efectos de la temperatura sobre la densidad del agua), las unidades equivalen a mm (de agua líquida) por segundo.

- Tasa media de nevadas a gran escala [kg/m^2s] (ERA5): Este parámetro es la tasa de nevadas (intensidad de las nevadas) en la superficie de la Tierra, que es generada por el esquema de nubes en el Sistema Integrado de Pronóstico (IFS) del ECMWF. El esquema de nubes representa la formación y disipación de nubes y la precipitación a gran escala debido a cambios en las cantidades atmosféricas (como presión, temperatura y humedad) pronosticadas directamente a escalas espaciales de la cuadrícula o mayores. Las nevadas también pueden ser generadas por el esquema de convección en el IFS, que representa la convección a escalas espaciales más pequeñas que la cuadrícula. En el IFS, la precipitación se compone de lluvia y nieve. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Es la tasa que tendría la nevada si se distribuyera uniformemente sobre la cuadrícula. Dado que 1 kg de agua repartida en 1 metro cuadrado de superficie tiene 1 mm de profundidad (despreciando los efectos de la temperatura sobre la densidad del agua), las unidades equivalen a mm (de agua líquida) por segundo.
- Tensión superficial media de la onda de gravedad hacia el norte [N/m^2] (ERA5): El aire que fluye sobre una superficie ejerce una tensión (arrastre) que transfiere impulso a la superficie y frena el viento. Este parámetro es el componente de la tensión superficial media en una dirección hacia el norte, asociado con

el bloqueo orográfico de bajo nivel y las ondas de gravedad orográficas. Se calcula mediante el esquema de orografía de subcuadrícula del Sistema Integrado de Pronóstico del ECMWF, que representa el estrés debido a valles, colinas y montañas no resueltos con escalas horizontales entre 5 km y la escala de cuadrícula del modelo. (La tensión asociada con las características orográficas con escalas horizontales menores de 5 km se explica por el esquema de arrastre de forma orográfica turbulenta). Las ondas de gravedad orográficas son oscilaciones en el flujo mantenidas por la flotabilidad de las parcelas de aire desplazadas, producidas cuando el aire es desviado hacia arriba por colinas y montañas. Este proceso puede crear tensión en la atmósfera en la superficie de la Tierra y en otros niveles de la atmósfera. Los valores positivos (negativos) indican tensión en la superficie de la Tierra en una dirección hacia el norte (hacia el sur). Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación.

- Tensión superficial turbulenta media hacia el norte [N/m^2] (ERA5): El aire que

fluye sobre una superficie ejerce una tensión (arrastre) que transfiere impulso a la superficie y frena el viento. Este parámetro es el componente de la tensión superficial media en dirección norte, asociado con remolinos turbulentos cerca de la superficie y arrastre de forma orográfica turbulenta. Se calcula mediante los esquemas de arrastre de formas orográficas turbulentas y de difusión turbulenta del ECMWF Integrated Forecasting System. Los remolinos turbulentos cerca de la superficie están relacionados con la rugosidad de la superficie. El arrastre de forma orográfica turbulenta es la tensión debida a los valles, colinas y montañas en escalas horizontales por debajo de los 5 km, que se especifican a partir de datos de la superficie terrestre con una resolución de aproximadamente 1 km. (La tensión asociada con las características orográficas con escalas horizontales entre 5 km y la escala de cuadrícula del modelo se tiene en cuenta mediante el esquema orográfico de subcuadrícula). Los valores positivos (negativos) indican tensión en la superficie de la Tierra en una dirección hacia el norte (hacia el sur). Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación.

- Tasa de evaporación potencial media [$kg/m^2 s$] (ERA5): Este parámetro es una medida de la medida en que las condiciones atmosféricas cercanas a la superficie conducen al proceso de evaporación. Generalmente se considera que es la cantidad de evaporación, bajo las condiciones atmosféricas existentes, de una superficie de agua pura que tiene la temperatura de la capa más baja de la atmósfera y da una indicación de la evaporación máxima posible. La evaporación potencial en el actual Sistema Integrado de Pronóstico (IFS) del ECMWF se basa en los cálculos del balance de energía de la superficie con los parámetros de vegetación

establecidos en “cultivos/agricultura mixta” y asumiendo que “no hay estrés por la humedad del suelo”. En otras palabras, la evaporación se calcula para la tierra agrícola como si estuviera bien regada y suponiendo que la atmósfera no se ve afectada por esta condición de superficie artificial. Esto último puede no ser siempre realista. Aunque la evaporación potencial está destinada a proporcionar una estimación de los requisitos de riego, el método puede dar resultados poco realistas en condiciones áridas debido a una evaporación demasiado fuerte forzada por el aire seco. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación.

- Tasa media de escorrentía [kg/m^2s] (ERA5): Parte del agua de la lluvia, de la nieve que se derrite o de lo profundo del suelo, permanece almacenada en el suelo. De lo contrario, el agua se escurre, ya sea sobre la superficie (escorrentía superficial) o bajo tierra (escorrentía subterránea) y la suma de estos dos se denomina escorrentía. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Es la tasa que tendría la escorrentía si se distribuyera uniformemente sobre la

cuadrícula. Se debe tener cuidado al comparar los parámetros del modelo con las observaciones, porque las observaciones suelen ser locales a un punto en particular en lugar de promediar sobre un cuadro de cuadrícula.

- Tasa media de evaporación de la nieve [kg/m^2s] (ERA5): Este parámetro es la tasa promedio de evaporación de la nieve desde el área cubierta de nieve de una caja de rejilla hasta convertirse en vapor en el aire de arriba. El Sistema Integrado de Pronóstico (IFS) del ECMWF representa la nieve como una sola capa adicional sobre el nivel superior del suelo. La nieve puede cubrir todo o parte de la caja de rejilla. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Es la tasa que tendría la evaporación de la nieve si se distribuyera uniformemente sobre la caja de rejilla. 1 kg de agua repartida en 1 metro cuadrado de superficie tiene 1 mm de profundidad (despreciando los efectos de la temperatura sobre la densidad del agua), por lo que las unidades equivalen a mm (de agua líquida) por segundo. La convención IFS es que los flujos descendentes son positivos. Por lo tanto, los valores negativos indican evaporación y los valores positivos indican deposición.
- Tasa media de nevadas [kg/m^2s] (ERA5): Este parámetro es la tasa de nevadas en la superficie de la Tierra. Es la suma de las nevadas a gran escala y convectivas. Las nevadas a gran escala son generadas por el esquema de nubes en el Sistema Integrado de Pronóstico (IFS) del ECMWF. El esquema de nubes representa la formación y disipación de nubes y la precipitación a gran escala debido a cambios en las cantidades atmosféricas (como presión, temperatura y humedad)

pronosticadas directamente a escalas espaciales de la cuadrícula o mayores. La nevada convectiva es generada por el esquema de convección en el IFS, que representa la convección a escalas espaciales más pequeñas que la cuadrícula. En el IFS, la precipitación se compone de lluvia y nieve. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora que finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Es la tasa que tendría la nevada si se distribuyera uniformemente sobre la cuadrícula. 1 kg de agua repartida en 1 metro cuadrado de superficie tiene 1 mm de profundidad (despreciando los efectos de la temperatura sobre la densidad del agua), por lo que las unidades equivalen a mm (de agua líquida) por segundo.

- Tasa media de deshielo [$kg/m^2 s$] (ERA5): Este parámetro es la tasa de derretimiento de la nieve en el área cubierta de nieve de una cuadrícula. El Sistema Integrado de Pronóstico (IFS) del ECMWF representa la nieve como una sola capa adicional sobre el nivel superior del suelo. La nieve puede cubrir todo o parte de la caja de rejilla. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Es la velocidad que tendría la fusión si se distribuyera uniformemente sobre la caja de rejilla. 1 kg de agua repartida en 1 metro cuadrado de superficie tiene 1 mm de profundidad (despreciando los efectos de la temperatura sobre la densidad del agua),

- Tasa media de escorrentía subterránea [$kg/m^2 s$] (ERA5): Parte del agua de la lluvia, de la nieve que se derrite o de lo profundo del suelo, permanece almacenada en el suelo. De lo contrario, el agua se escurre, ya sea sobre la superficie (escorrentía superficial) o bajo tierra (escorrentía subterránea) y la suma de estos dos se denomina escorrentía. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Es la tasa que tendría la escorrentía si se distribuyera uniformemente sobre la cuadrícula.
- Flujo medio de radiación directa de onda corta en la superficie [W/m^2] Este parámetro es la cantidad de radiación solar directa (también conocida como radiación de onda corta) que llega a la superficie de la Tierra. Es la cantidad de radiación que pasa a través de un plano horizontal. La radiación solar en la superficie puede ser directa o difusa. La radiación solar puede ser dispersada en todas las direcciones por partículas en la atmósfera, algunas de las cuales alcanzan la superficie (radiación solar difusa). Parte de la radiación solar llega a la superficie sin ser dispersada (radiación solar directa). Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación.
- Flujo medio de radiación directa de onda corta en superficie, cielo despejado [W/m^2] (ERA5): Este parámetro es la cantidad de radiación directa del Sol

(también conocida como radiación solar o de onda corta) que llega a la superficie de la Tierra, suponiendo condiciones de cielo despejado (sin nubes). Es la cantidad de radiación que pasa a través de un plano horizontal. La radiación solar en la superficie puede ser directa o difusa. La radiación solar puede ser dispersada en todas las direcciones por partículas en la atmósfera, algunas de las cuales alcanzan la superficie (radiación solar difusa). Parte de la radiación solar llega a la superficie sin ser dispersada (radiación solar directa). Las cantidades de radiación en cielo despejado se calculan exactamente para las mismas condiciones atmosféricas de temperatura, humedad, ozono, gases traza y aerosoles que las cantidades correspondientes del cielo total (nubes incluidas), pero asumiendo que las nubes no están allí. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. La convención ECMWF para flujos verticales es positiva hacia abajo.

- Flujo de radiación UV superficial medio hacia abajo [W/m^2] (ERA5): Este parámetro es la cantidad de radiación ultravioleta (UV) que llega a la superficie. Es la cantidad de radiación que pasa a través de un plano horizontal. La radiación ultravioleta es parte del espectro electromagnético emitido por el Sol que tiene longitudes de onda más cortas que la luz visible. En el sistema de pronóstico integrado (IFS) del ECMWF se define como radiación con una longitud de onda de 0,20-0,44 [μm] (micras, 1 millonésima de metro). Pequeñas cantidades de UV son esenciales para los organismos vivos, pero la sobreexposición puede provocar daño celular; en humanos esto incluye efectos agudos y crónicos en la salud de la piel, los ojos y el sistema inmunológico. La radiación ultravioleta es absorbida por la capa de ozono, pero parte llega a la superficie. El agotamiento de la capa

de ozono está causando preocupación por el aumento de los efectos dañinos de los rayos UV. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. La convención ECMWF para flujos verticales es positiva hacia abajo.

- Flujo medio de radiación de onda larga hacia abajo en la superficie [W/m^2] (ERA5): Este parámetro es la cantidad de radiación térmica (también conocida como de onda larga o terrestre) emitida por la atmósfera y las nubes que alcanza un plano horizontal en la superficie de la Tierra. La superficie de la Tierra emite radiación térmica, parte de la cual es absorbida por la atmósfera y las nubes. La atmósfera y las nubes también emiten radiación térmica en todas las direcciones, parte de la cual alcanza la superficie (representada por este parámetro). Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. La convención ECMWF para flujos verticales es positiva hacia abajo.
- Flujo medio de radiación de onda larga hacia abajo en la superficie, cielo despejado [W/m^2] (ERA5): Este parámetro es la cantidad de radiación térmica (también conocida como de onda larga o terrestre) emitida por la atmósfera que alcanza un plano horizontal en la superficie de la Tierra, suponiendo condiciones de cielo despejado (sin nubes). La superficie de la Tierra emite radiación térmica, parte de la cual es absorbida por la atmósfera y las nubes. La atmósfera y las

nubes también emiten radiación térmica en todas las direcciones, parte de la cual alcanza la superficie. Las cantidades de radiación en cielo despejado se calculan exactamente para las mismas condiciones atmosféricas de temperatura, humedad, ozono, gases traza y aerosoles que las cantidades correspondientes del cielo total (nubes incluidas), pero asumiendo que las nubes no están allí. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora que finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. La convención ECMWF para flujos verticales es positiva hacia abajo.

- Flujo medio de radiación de onda corta hacia abajo en la superficie [W/m^2] (ERA5): Este parámetro es la cantidad de radiación solar (también conocida como radiación de onda corta) que llega a un plano horizontal en la superficie de la Tierra. Este parámetro comprende tanto la radiación solar directa como la difusa. La radiación del Sol (radiación solar o de onda corta) es parcialmente reflejada de regreso al espacio por las nubes y partículas en la atmósfera (aerosoles) y parte de ella es absorbida. El resto incide sobre la superficie terrestre (representada por este parámetro). En una aproximación razonablemente buena, este parámetro es el modelo equivalente a lo que mediría un piranómetro (un instrumento utilizado para medir la radiación solar) en la superficie. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. La convención ECMWF para flujos verticales es positiva hacia abajo.

- Flujo medio de radiación de onda corta hacia abajo en la superficie, cielo despejado [W/m^2] (ERA5): Este parámetro es la cantidad de radiación solar (también conocida como radiación de onda corta) que alcanza un plano horizontal en la superficie de la Tierra, suponiendo condiciones de cielo despejado (sin nubes). Este parámetro comprende tanto la radiación solar directa como la difusa. La radiación del Sol (radiación solar o de onda corta) es parcialmente reflejada de regreso al espacio por las nubes y partículas en la atmósfera (aerosoles) y parte de ella es absorbida. El resto incide en la superficie de la Tierra. Las cantidades de radiación en cielo despejado se calculan exactamente para las mismas condiciones atmosféricas de temperatura, humedad, ozono, gases traza y aerosoles que las cantidades correspondientes del cielo total (nubes incluidas), pero asumiendo que las nubes no están allí. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. La convención ECMWF para flujos verticales es positiva hacia abajo.
- Flujo de calor latente superficial medio [W/m^2] (ERA5): Este parámetro es la transferencia de calor latente (resultado de los cambios de fase del agua, como la evaporación o la condensación) entre la superficie de la Tierra y la atmósfera a través de los efectos del movimiento turbulento del aire. La evaporación desde la superficie de la Tierra representa una transferencia de energía desde la superficie hacia la atmósfera. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden

a un período de acumulación de 3 horas desde la fecha y hora de validación. La convención ECMWF para flujos verticales es positiva hacia abajo.

- Flujo de radiación de onda larga neto superficial medio [W/m^2] (ERA5): La radiación térmica (también conocida como radiación de onda larga o terrestre) se refiere a la radiación emitida por la atmósfera, las nubes y la superficie de la Tierra. Este parámetro es la diferencia entre la radiación térmica descendente y ascendente en la superficie de la Tierra. Es la cantidad de radiación que pasa a través de un plano horizontal. La atmósfera y las nubes emiten radiación térmica en todas las direcciones, parte de la cual llega a la superficie como radiación térmica descendente. La radiación térmica ascendente en la superficie consiste en la radiación térmica emitida por la superficie más la fracción de radiación térmica descendente reflejada hacia arriba por la superficie. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. La convención ECMWF para flujos verticales es positiva hacia abajo.
- Flujo de radiación de onda larga neto superficial medio, cielo despejado [W/m^2] (ERA5): La radiación térmica (también conocida como radiación de onda larga o terrestre) se refiere a la radiación emitida por la atmósfera, las nubes y la superficie de la Tierra. Este parámetro es la diferencia entre la radiación térmica descendente y ascendente en la superficie de la Tierra, suponiendo condiciones de cielo despejado (sin nubes). Es la cantidad de radiación que pasa a través de un plano horizontal. Las cantidades de radiación en cielo despejado se calculan exactamente para las mismas condiciones atmosféricas de temperatura, humedad,

ozono, gases traza y aerosoles que las cantidades correspondientes del cielo total (nubes incluidas), pero asumiendo que las nubes no están allí. La atmósfera y las nubes emiten radiación térmica en todas las direcciones, parte de la cual llega a la superficie como radiación térmica descendente. La radiación térmica ascendente en la superficie consiste en la radiación térmica emitida por la superficie más la fracción de radiación térmica descendente reflejada hacia arriba por la superficie. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. La convención ECMWF para flujos verticales es positiva hacia abajo. el período de procesamiento es superior a 1 hora que finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. La convención ECMWF para flujos verticales es positiva hacia abajo. el período de procesamiento es superior a 1 hora que finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. La convención ECMWF para flujos verticales es positiva hacia abajo.

- Flujo de radiación de onda corta neto superficial medio [W/m^2] (ERA5): Este parámetro es la cantidad de radiación solar (también conocida como radiación de onda corta) que alcanza un plano horizontal en la superficie de la Tierra (tanto directa como difusa) menos la cantidad reflejada por la superficie de la Tierra (que se rige por el albedo). La radiación del Sol (radiación solar o de onda corta) es parcialmente reflejada de regreso al espacio por las nubes y partículas en la atmósfera (aerosoles) y parte de ella es absorbida. El resto incide sobre la

superficie de la Tierra, donde parte de él se refleja. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media del conjunto y la dispersión del conjunto, el período de procesamiento es superior a las 3 horas que terminan en la fecha y hora de validez. La convención ECMWF para flujos verticales es positiva hacia abajo.

- Flujo de radiación de onda corta neta superficial medio, cielo despejado [W/m^2] (ERA5): Este parámetro es la cantidad de radiación solar (onda corta) que llega a la superficie de la Tierra (tanto directa como difusa) menos la cantidad reflejada por la superficie de la Tierra (que se rige por el albedo), asumiendo condiciones de cielo despejado (sin nubes). Es la cantidad de radiación que pasa a través de un plano horizontal. Las cantidades de radiación en cielo despejado se calculan exactamente para las mismas condiciones atmosféricas de temperatura, humedad, ozono, gases traza y aerosoles que las cantidades correspondientes del cielo total (nubes incluidas), pero asumiendo que las nubes no están allí. La radiación del Sol (radiación solar o de onda corta) es parcialmente reflejada de regreso al espacio por las nubes y partículas en la atmósfera (aerosoles) y parte de ella es absorbida. El resto incide sobre la superficie de la Tierra, donde parte de él se refleja. La diferencia entre la radiación solar hacia abajo y reflejada es la radiación solar neta superficial. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. La convención ECMWF para flujos verticales es positiva hacia abajo. el período de procesamiento es superior a las 3 horas que terminan en la fecha y

hora de validez. La convención ECMWF para flujos verticales es positiva hacia abajo. el período de procesamiento es superior a las 3 horas que terminan en la fecha y hora de validez. La convención ECMWF para flujos verticales es positiva hacia abajo.

- Tasa media de escorrentía superficial [kg/m^2s] (ERA5): Parte del agua de la lluvia, de la nieve que se derrite o de lo profundo del suelo, permanece almacenada en el suelo. De lo contrario, el agua se escurre, ya sea sobre la superficie (escorrentía superficial) o bajo tierra (escorrentía subterránea) y la suma de estos dos se denomina escorrentía. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Es la tasa que tendría la escorrentía si se distribuyera uniformemente sobre la cuadrícula.
- Flujo de calor sensible superficial medio [W/m^2] (ERA5): Este parámetro es la transferencia de calor entre la superficie de la Tierra y la atmósfera a través de los efectos del movimiento turbulento del aire (pero excluyendo cualquier transferencia de calor resultante de la condensación o la evaporación). La magnitud del flujo de calor sensible se rige por la diferencia de temperatura entre la superficie y la atmósfera suprayacente, la velocidad del viento y la rugosidad de la superficie. Por ejemplo, el aire frío que recubre una superficie cálida produciría un flujo de calor sensible desde la tierra (o el océano) hacia la atmósfera. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros

del conjunto, la media del conjunto y la dispersión del conjunto, el período de procesamiento es superior a las 3 horas que terminan en la fecha y hora de validez. La convención ECMWF para flujos verticales es positiva hacia abajo.

- Flujo medio de radiación de onda corta de arriba hacia abajo [W/m^2] (ERA5): Este parámetro es la radiación solar entrante (también conocida como radiación de onda corta), recibida del Sol, en la parte superior de la atmósfera. Es la cantidad de radiación que pasa a través de un plano horizontal. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. La convención ECMWF para flujos verticales es positiva hacia abajo.
- Flujo de radiación de onda larga neto superior medio [W/m^2] (ERA5): La radiación térmica (también conocida como terrestre o de onda larga) emitida al espacio en la parte superior de la atmósfera se conoce comúnmente como Radiación de onda larga saliente (OLR). La radiación térmica neta superior (este parámetro) es igual al negativo de OLR. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. La convención ECMWF para flujos verticales es positiva hacia abajo.
- Flujo de radiación de onda larga neto superior medio, cielo despejado [W/m^2] (ERA5): Este parámetro es la radiación térmica (también conocida como te-

rrestre o de onda larga) emitida al espacio en la parte superior de la atmósfera, suponiendo condiciones de cielo despejado (sin nubes). Es la cantidad que pasa por un plano horizontal. Tenga en cuenta que la convención ECMWF para flujos verticales es positiva hacia abajo, por lo que un flujo de la atmósfera al espacio será negativo. Las cantidades de radiación en cielo despejado se calculan exactamente para las mismas condiciones atmosféricas de temperatura, humedad, ozono, gases traza y aerosoles que las cantidades de cielo total (nubes incluidas), pero asumiendo que las nubes no están allí. La radiación térmica emitida al espacio en la parte superior de la atmósfera se conoce comúnmente como radiación de onda larga saliente (OLR) (es decir, tomando un flujo de la atmósfera al espacio como positivo). Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación.

- Flujo de radiación de onda corta neto superior medio [W/m^2] (ERA5): Este parámetro es la radiación solar entrante (también conocida como radiación de onda corta) menos la radiación solar saliente en la parte superior de la atmósfera. Es la cantidad de radiación que pasa a través de un plano horizontal. La radiación solar entrante es la cantidad recibida del Sol. La radiación solar saliente es la cantidad reflejada y dispersada por la atmósfera y la superficie de la Tierra. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. La convención ECMWF para

flujos verticales es positiva hacia abajo.

- Flujo de radiación de onda corta neta superior medio, cielo despejado [W/m^2] (ERA5): Este parámetro es la radiación solar entrante (también conocida como radiación de onda corta) menos la radiación solar saliente en la parte superior de la atmósfera, suponiendo condiciones de cielo despejado (sin nubes). Es la cantidad de radiación que pasa a través de un plano horizontal. La radiación solar entrante es la cantidad recibida del Sol. La radiación solar saliente es la cantidad reflejada y dispersada por la atmósfera y la superficie de la Tierra, suponiendo condiciones de cielo despejado (sin nubes). Las cantidades de radiación en cielo despejado se calculan exactamente para las mismas condiciones atmosféricas de temperatura, humedad, ozono, gases traza y aerosoles que las cantidades de cielo total (nubes incluidas), pero asumiendo que las nubes no están allí. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora que finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. La convención ECMWF para flujos verticales es positiva hacia abajo.
- Tasa media de precipitación total [$kg/m^2 s$] (ERA5): Este parámetro es la tasa de precipitación en la superficie de la Tierra. Es la suma de las tasas debidas a la precipitación a gran escala y la precipitación convectiva. La precipitación a gran escala es generada por el esquema de nubes en el Sistema Integrado de Pronóstico (IFS) del ECMWF. El esquema de nubes representa la formación y disipación de nubes y la precipitación a gran escala debido a cambios en las cantidades atmosféricas (como presión, temperatura y humedad) pronosticadas directamente a escalas espaciales de la cuadrícula o mayores. La precipitación

convectiva es generada por el esquema de convección en el IFS, que representa la convección a escalas espaciales más pequeñas que la cuadrícula. En el IFS, la precipitación se compone de lluvia y nieve. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Es la tasa que tendría la precipitación si se distribuyera uniformemente sobre la cuadrícula. 1 kg de agua repartida en 1 metro cuadrado de superficie tiene 1 mm de profundidad (despreciando los efectos de la temperatura sobre la densidad del agua), por lo que las unidades equivalen a mm (de agua líquida) por segundo.

- Gradiente vertical medio de refractividad dentro de la capa de captura [1/metro] (ERA5): Gradiente vertical medio de la refractividad atmosférica dentro de la capa de captura.
- Divergencia media de humedad integrada verticalmente [$kg/m^2 s$] (ERA5): La integral vertical del flujo de humedad es la tasa horizontal de flujo de humedad (vapor de agua, líquido de nube y hielo de nube), por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera, que se extiende a una altitud de aproximadamente 80 kilómetros. Su divergencia horizontal es la tasa de humedad que se esparce hacia afuera desde un punto, por metro cuadrado. Este parámetro es una media durante un período de tiempo particular que depende de los datos extraídos (1 hora). Para el reanálisis, el período de procesamiento es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de

validación. Este parámetro es positivo para la humedad que se esparce o diverge, y negativo para lo contrario, para la humedad que se concentra, o convergente (convergencia). Este parámetro indica si los movimientos atmosféricos actúan para disminuir (por divergencia) o aumentar (por convergencia) la integral vertical de la humedad, durante el período de tiempo. Los valores negativos altos de este parámetro (es decir, una gran convergencia de la humedad) pueden estar relacionados con la intensificación de las precipitaciones y las inundaciones. 1 kg de agua repartida en 1 metro cuadrado de superficie tiene 1 mm de profundidad (despreciando los efectos de la temperatura sobre la densidad del agua), por lo que las unidades equivalen a mm (de agua líquida) por segundo, una gran convergencia de humedad puede estar relacionado con la intensificación de las precipitaciones y las inundaciones.

- Nubosidad media [adimensional] (ERA5): Este parámetro es la proporción de una cuadrícula cubierta por nubes que se encuentran en los niveles medios de la troposfera. La nube media es un campo de un solo nivel calculado a partir de la nube que se produce en los niveles del modelo con una presión entre 0,45 y 0,8 veces la presión superficial. Así, si la presión en superficie es de 1000 hPa (hectopascal), la nube media se calcularía utilizando niveles con una presión menor o igual a 800 hPa y mayor o igual a 450 hPa (entre 2 km y 6 km aproximadamente (suponiendo un “estándar atmósfera”)). El parámetro de nubes medias se calcula a partir de la cobertura de nubes para los niveles de modelo apropiados, como se describe anteriormente. Se hacen suposiciones sobre el grado de superposición/aleatoriedad entre nubes en diferentes niveles de modelo. Las fracciones de nubes varían de 0 a 1.
- Temperatura mínima de 2 m desde el posprocesamiento anterior [K] (ERA5): Este parámetro es la temperatura más baja del aire a 2 m sobre la superficie de la

tierra, el mar o las aguas interiores desde que se archivó el parámetro por última vez en un pronóstico particular. La temperatura a 2 m se calcula interpolando entre el nivel más bajo del modelo y la superficie de la Tierra, teniendo en cuenta las condiciones atmosféricas. Ver más información. Este parámetro tiene unidades de kelvin (K). La temperatura medida en kelvin se puede convertir a grados Celsius ($^{\circ}\text{C}$) restando 273,15.

- Tasa de precipitación total mínima desde el posprocesamiento anterior [$\text{kg}/\text{m}^2\text{s}$] (ERA5): La precipitación total se calcula a partir de las tasas combinadas de lluvia y nevadas a gran escala y convectivas en cada paso de tiempo y se mantiene el mínimo desde el último procesamiento posterior.
- Gradiente vertical mínimo de refractividad dentro de la capa de captura: [metro] (ERA5): Gradiente vertical mínimo de refractividad atmosférica dentro de la capa de captura.
- Albedo IR cercano para radiación difusa [adimensional] (ERA5): El albedo es una medida de la reflectividad de la superficie terrestre. Este parámetro es la fracción de radiación solar difusa (onda corta) con longitudes de onda entre 0,7 y 4 [μm] (micras, 1 millonésima de metro) reflejada por la superficie de la Tierra (solo para superficies terrestres sin nieve). Los valores de este parámetro varían entre 0 y 1. En el Sistema Integrado de Pronóstico (IFS) del ECMWF, el albedo se trata por separado para la radiación solar con longitudes de onda mayores.
- Albedo IR cercano para radiación directa [adimensional] (ERA5): El albedo es una medida de la reflectividad de la superficie terrestre. Este parámetro es la fracción de radiación solar directa (onda corta) con longitudes de onda entre 0,7 y 4 [μm] (micras, 1 millonésima de metro) reflejada por la superficie de la Tierra (solo para superficies terrestres sin nieve). Los valores de este parámetro varían

entre 0 y 1. En el Sistema Integrado de Pronóstico (IFS) del ECMWF, el albedo se trata por separado para la radiación solar con longitudes de onda mayores.

- Esfuerzo superficial de ondas de gravedad hacia el norte [$N/m^2 s$] (ERA5): El aire que fluye sobre una superficie ejerce una tensión (arrastre) que transfiere impulso a la superficie y frena el viento. Este parámetro es el componente de la tensión superficial acumulada en una dirección hacia el norte, asociado con el bloqueo orográfico de bajo nivel y las ondas de gravedad orográficas. Se calcula mediante el esquema de orografía de subcuadrícula del Sistema Integrado de Pronóstico del ECMWF, que representa el estrés debido a valles, colinas y montañas no resueltos con escalas horizontales entre 5 km y la escala de cuadrícula del modelo. (La tensión asociada con las características orográficas con escalas horizontales menores de 5 km se explica por el esquema de arrastre de forma orográfica turbulenta). Las ondas de gravedad orográficas son oscilaciones en el flujo mantenidas por la flotabilidad de las parcelas de aire desplazadas, producidas cuando el aire es desviado hacia arriba por colinas y montañas. Este proceso puede crear tensión en la atmósfera en la superficie de la Tierra y en otros niveles de la atmósfera. Los valores positivos (negativos) indican tensión en la superficie de la Tierra en una dirección hacia el norte (hacia el sur). Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de

validación. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos.

- Tensión superficial turbulenta hacia el norte [$N/m^2 s$] (ERA5): El aire que fluye sobre una superficie ejerce una tensión (arrastre) que transfiere impulso a la superficie y frena el viento. Este parámetro es el componente de la tensión superficial acumulada en dirección norte, asociado con remolinos turbulentos cerca de la superficie y arrastre de forma orográfica turbulenta. Se calcula mediante los esquemas de arrastre de formas orográficas turbulentas y de difusión turbulenta del ECMWF Integrated Forecasting System. Los remolinos turbulentos cerca de la superficie están relacionados con la rugosidad de la superficie. El arrastre de forma orográfica turbulenta es la tensión debida a los valles, colinas y montañas en escalas horizontales por debajo de los 5 km, que se especifican a partir de datos de la superficie terrestre con una resolución de aproximadamente 1 km. (La tensión asociada con las características orográficas con escalas horizontales entre 5 km y la escala de cuadrícula del modelo se tiene en cuenta mediante el esquema orográfico de subcuadrícula). Los valores positivos (negativos) indican tensión en la superficie de la Tierra en una dirección hacia el norte (hacia el sur). Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación.
- Evaporación potencial [metro] (ERA5): Este parámetro es una medida de la medida en que las condiciones atmosféricas cercanas a la superficie conducen al proceso de evaporación. Generalmente se considera que es la cantidad de evaporación, bajo las condiciones atmosféricas existentes, de una superficie de

agua pura que tiene la temperatura de la capa más baja de la atmósfera y da una indicación de la evaporación máxima posible. La evaporación potencial en el actual Sistema Integrado de Pronóstico (IFS) del ECMWF se basa en los cálculos del balance de energía de la superficie con los parámetros de vegetación establecidos en “cultivos/agricultura mixta” y asumiendo que “no hay estrés por la humedad del suelo”. En otras palabras, la evaporación se calcula para la tierra agrícola como si estuviera bien regada y suponiendo que la atmósfera no se ve afectada por esta condición de superficie artificial. Esto último puede no ser siempre realista. Aunque la evaporación potencial está destinada a proporcionar una estimación de los requisitos de riego, el método puede dar resultados poco realistas en condiciones áridas debido a una evaporación demasiado fuerte forzada por el aire seco. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. el período de acumulación es superior a 1 hora que finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. el período de acumulación es superior a 1 hora que finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación.

- Tipo de precipitación [adimensional] (ERA5): Este parámetro describe el tipo de precipitación en la superficie, en el momento especificado. Se asigna un tipo de precipitación siempre que haya un valor de precipitación distinto de cero. En el Sistema Integrado de Pronóstico (IFS) del ECMWF solo hay dos variables de precipitación pronosticadas: lluvia y nieve. El tipo de precipitación

se deriva de estas dos variables pronosticadas en combinación con las condiciones atmosféricas, como la temperatura. Valores del tipo de precipitación definidos en el IFS: 0: Sin precipitación, 1: Lluvia, 3: Lluvia helada (es decir, gotas de lluvia superenfriadas que se congelan al contacto con el suelo y otras superficies), 5: Nieve, 6: Nieve húmeda (es decir, partículas de nieve que empiezan a derretirse); 7: Mezcla de lluvia y nieve, 8: Hielo granulado. Estos tipos de precipitación son consistentes con la Tabla de cifrado 4.201 de la OMM. Otros tipos en esta tabla de la OMM no están definidos en el IFS.

- Escapada [metro] (ERA5): Parte del agua de la lluvia, de la nieve que se derrite o de lo profundo del suelo, permanece almacenada en el suelo. De lo contrario, el agua se escurre, ya sea sobre la superficie (escorrentía superficial) o bajo tierra (escorrentía subterránea) y la suma de estos dos se denomina escorrentía. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Las unidades de escorrentía son la profundidad en metros de agua. Esta es la profundidad que tendría el agua si se distribuyera uniformemente sobre la caja de rejilla. Se debe tener cuidado al comparar los parámetros del modelo con las observaciones, porque las observaciones son a menudo locales a un punto en particular en lugar de promediar sobre un cuadro de cuadrícula. Las observaciones también se toman a menudo en diferentes unidades, como mm/día, en lugar de los metros acumulados producidos aquí. La escorrentía es una medida de la disponibilidad de agua en el suelo y puede, por ejemplo, utilizarse como indicador de sequía o inundación.
- Contenido del reservorio de la piel (*Skin reservoir content*) [metros de agua]

equivalente] (ERA5): Este parámetro es la cantidad de agua en el dosel de la vegetación y/o en una capa delgada sobre el suelo. Representa la cantidad de lluvia interceptada por el follaje y el agua del rocío. La cantidad máxima de “contenido de depósito de piel” que puede contener una caja de cuadrícula depende del tipo de vegetación y puede ser cero. El agua sale del “depósito de la piel” por evaporación.

- Temperatura de la piel (*Skin temperature*) [K] (ERA5): Este parámetro es la temperatura de la superficie de la Tierra. La temperatura de la piel es la temperatura teórica que se requiere para satisfacer el balance de energía superficial. Representa la temperatura de la capa superficial superior, que no tiene capacidad calorífica y, por lo tanto, puede responder instantáneamente a los cambios en los flujos superficiales. La temperatura de la piel se calcula de manera diferente sobre la tierra y el mar. Este parámetro tiene unidades de kelvin (K). La temperatura medida en kelvin se puede convertir a grados Celsius (°C) restando 273,15.
- Pendiente de la orografía a escala inferior a la cuadrícula [adimensional] (ERA5): Este parámetro es uno de los cuatro parámetros (los otros son la desviación estándar, el ángulo y la anisotropía) que describen las características de la orografía que son demasiado pequeñas para ser resueltas por la cuadrícula del modelo. Estos cuatro parámetros se calculan para características orográficas con escalas horizontales comprendidas entre 5 km y la resolución de cuadrícula del modelo, y se derivan de la altura de valles, colinas y montañas con una resolución de aproximadamente 1 km. Se utilizan como entrada para el esquema de orografía de subcuadrícula que representa el bloqueo de bajo nivel y los efectos de ondas de gravedad orográficas. Este parámetro representa la pendiente de los valles, colinas y montañas de la subcuadrícula. Una superficie plana tiene un

valor de 0 y una pendiente de 45 grados tiene un valor de 0,5. Este parámetro no varía en el tiempo.

- Albedo de nieve [adimensional] (ERA5): Este parámetro es una medida de la reflectividad de la parte cubierta de nieve de la caja de rejilla. Es la fracción de radiación solar (onda corta) reflejada por la nieve en todo el espectro solar. El Sistema Integrado de Pronóstico (IFS) del ECMWF representa la nieve como una sola capa adicional sobre el nivel superior del suelo. La nieve puede cubrir todo o parte de la caja de rejilla. Este parámetro cambia con la edad de la nieve y también depende de la altura de la vegetación. Tiene un rango de valores entre 0 y 1. Para vegetación baja, oscila entre 0,52 para nieve vieja y 0,88 para nieve fresca. Para vegetación alta con nieve debajo, depende del tipo de vegetación y tiene valores entre 0,27 y 0,38. Este parámetro se define en todo el mundo, incluso donde no hay nieve.
- Densidad de la nieve [kg/m^3] (ERA5): Este parámetro es la masa de nieve por metro cúbico en la capa de nieve. El Sistema Integrado de Pronóstico (IFS) del ECMWF representa la nieve como una sola capa adicional sobre el nivel superior del suelo. La nieve puede cubrir todo o parte de la caja de rejilla. Este parámetro se define en todo el mundo, incluso donde no hay nieve. Las regiones sin nieve se pueden enmascarar considerando solo los puntos de la cuadrícula donde la profundidad de la nieve ([metros de agua equivalente] es mayor que 0,0.
- Profundidad de la nieve [metros de agua equivalente] (ERA5): Este parámetro es la cantidad de nieve del área cubierta de nieve de una cuadrícula. Sus unidades son metros de agua equivalente, por lo que es la profundidad que tendría el agua si la nieve se derritiera y se repartiera uniformemente por toda la cuadrícula. El Sistema Integrado de Pronóstico (IFS) del ECMWF representa la nieve como una sola capa adicional sobre el nivel superior del suelo. La nieve puede cubrir

todo o parte de la caja de rejilla.

- Evaporación de nieve [metros de agua equivalente] (ERA5): Este parámetro es la cantidad acumulada de agua que se ha evaporado de la nieve del área cubierta de nieve de una caja de rejilla en forma de vapor en el aire de arriba. El Sistema Integrado de Pronóstico (IFS) del ECMWF representa la nieve como una sola capa adicional sobre el nivel superior del suelo. La nieve puede cubrir todo o parte de la caja de rejilla. Este parámetro es la profundidad del agua que habría si la nieve evaporada (del área cubierta de nieve de una caja de rejilla) fuera líquida y se distribuyera uniformemente por toda la caja de rejilla. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. La convención IFS es que los flujos descendentes son positivos. Por lo tanto, los valores negativos indican evaporación y los valores positivos indican deposición.
- Nevada [metros de agua equivalente] (ERA5): Este parámetro es la nieve acumulada que cae a la superficie terrestre. Es la suma de las nevadas a gran escala y las nevadas convectivas. Las nevadas a gran escala son generadas por el esquema de nubes en el Sistema Integrado de Pronóstico (IFS) del ECMWF. El esquema de nubes representa la formación y disipación de nubes y la precipitación a gran escala debido a cambios en las cantidades atmosféricas (como presión, temperatura y humedad) pronosticadas directamente a escalas espaciales de la cuadrícula o mayores. La nevada convectiva es generada por el esquema de convección en el IFS, que representa la convección a escalas espaciales más pequeñas que la cuadrícula. En el IFS, la precipitación se compone de lluvia y

nieve. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora que finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Las unidades de este parámetro son la profundidad en metros de agua equivalente. Es la profundidad que tendría el agua si se distribuyera uniformemente sobre la caja de rejilla. Es la profundidad que tendría el agua si se distribuyera uniformemente sobre la caja de rejilla. Es la profundidad que tendría el agua si se distribuyera uniformemente sobre la caja de rejilla.

- Deshielo [metros de agua equivalente] (ERA5): Este parámetro es la cantidad acumulada de agua que se ha derretido de la nieve en el área cubierta de nieve de una caja de rejilla. El Sistema Integrado de Pronóstico (IFS) del ECMWF representa la nieve como una sola capa adicional sobre el nivel superior del suelo. La nieve puede cubrir todo o parte de la caja de rejilla. Este parámetro es la profundidad del agua que habría si la nieve derretida (del área cubierta de nieve de un cuadro de cuadrícula) se distribuyera uniformemente sobre todo el cuadro de cuadrícula. Por ejemplo, si la mitad de la caja de la cuadrícula estuviera cubierta de nieve con una profundidad equivalente al agua de 0,02 m, este parámetro tendría un valor de 0,01 m. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media del conjunto y la dispersión del conjunto,
- Nivel de temperatura del suelo 1 [k] (ERA5): Este parámetro es la temperatura del suelo en el nivel 1 (en el medio de la capa 1). El Sistema Integrado de

Pronóstico (IFS) del ECMWF tiene una representación del suelo de cuatro capas, donde la superficie está a 0 cm: Capa 1: 0 - 7 cm, Capa 2: 7 - 28 cm, Capa 3: 28 - 100 cm, Capa 4: 100 - 289 cm. La temperatura del suelo se establece en el medio de cada capa y la transferencia de calor se calcula en las interfaces entre ellas. Se supone que no hay transferencia de calor desde el fondo de la capa más baja. La temperatura del suelo se define en todo el mundo, incluso sobre el océano. Las regiones con una superficie de agua se pueden enmascarar considerando solo los puntos de cuadrícula donde la máscara tierra^mar tiene un valor superior a 0,5. Este parámetro tiene unidades de kelvin (K). La temperatura medida en kelvin se puede convertir a grados Celsius ($^{\circ}\text{C}$) restando 273,15.

- Nivel de temperatura del suelo 2 [k] (ERA5): Este parámetro es la temperatura del suelo en el nivel 2 (en el medio de la capa 2). El Sistema Integrado de Pronóstico (IFS) del ECMWF tiene una representación del suelo de cuatro capas, donde la superficie está a 0 cm: Capa 1: 0 - 7 cm, Capa 2: 7 - 28 cm, Capa 3: 28 - 100 cm, Capa 4: 100 - 289 cm. La temperatura del suelo se establece en el medio de cada capa y la transferencia de calor se calcula en las interfaces entre ellas. Se supone que no hay transferencia de calor desde el fondo de la capa más baja. La temperatura del suelo se define en todo el mundo, incluso sobre el océano. Las regiones con una superficie de agua se pueden enmascarar considerando solo los puntos de cuadrícula donde la máscara tierra^mar tiene un valor superior a 0,5. Este parámetro tiene unidades de kelvin (K). La temperatura medida en kelvin se puede convertir a grados Celsius ($^{\circ}\text{C}$) restando 273,15.
- Temperatura del suelo nivel 3 [k] (ERA5): Este parámetro es la temperatura del suelo en el nivel 3 (en el medio de la capa 3). El Sistema Integrado de Pronóstico (IFS) del ECMWF tiene una representación del suelo en cuatro capas, donde la superficie está a 0 cm: Capa 1: 0 - 7 cm, Capa 2: 7 - 28 cm, Capa 3: 28 - 100

cm, Capa 4: 100 - 289 cm. La temperatura del suelo se establece en el medio de cada capa y la transferencia de calor se calcula en las interfaces entre ellas. Se supone que no hay transferencia de calor desde el fondo de la capa más baja. La temperatura del suelo se define en todo el mundo, incluso sobre el océano. Las regiones con una superficie de agua se pueden enmascarar considerando solo los puntos de cuadrícula donde la máscara tierra^mar tiene un valor superior a 0,5. Este parámetro tiene unidades de kelvin (K). La temperatura medida en kelvin se puede convertir a grados Celsius ($^{\circ}\text{C}$) restando 273,15.

- Nivel de temperatura del suelo 4 [k] (ERA5): Este parámetro es la temperatura del suelo en el nivel 4 (en el medio de la capa 4). El Sistema Integrado de Pronóstico (IFS) del ECMWF tiene una representación del suelo en cuatro capas, donde la superficie está a 0 cm: Capa 1: 0 - 7 cm, Capa 2: 7 - 28 cm, Capa 3: 28 - 100 cm, Capa 4: 100 - 289 cm. La temperatura del suelo se establece en el medio de cada capa y la transferencia de calor se calcula en las interfaces entre ellas. Se supone que no hay transferencia de calor desde el fondo de la capa más baja. La temperatura del suelo se define en todo el mundo, incluso sobre el océano. Las regiones con una superficie de agua se pueden enmascarar considerando solo los puntos de cuadrícula donde la máscara tierra^mar tiene un valor superior a 0,5. Este parámetro tiene unidades de kelvin (K). La temperatura medida en kelvin se puede convertir a grados Celsius ($^{\circ}\text{C}$) restando 273,15.
- Tipo de suelo [adimensional] (ERA5): Este parámetro es la textura (o clasificación) del suelo utilizado por el esquema de superficie terrestre del Sistema Integrado de Pronóstico (IFS) del ECMWF para predecir la capacidad de retención de agua del suelo en los cálculos de humedad y escorrentía del suelo. Se deriva de los datos de la zona de raíces (30-100 cm por debajo de la superficie) del Mapa Digital de Suelos del Mundo de la FAO/UNESCO, DSMW (FAO,

2003), que existe a una resolución de 5' X 5' (alrededor de 10 kilómetros). Los siete tipos de suelo son: 1: Grueso, 2: Medio, 3: Medio fino, 4: Fino, 5: Muy fino, 6: Orgánico, 7: Orgánico tropical. Un valor de 0 indica un punto que no es tierra. Este parámetro no varía en el tiempo.

- Desviación estándar de la orografía de subred filtrada [metro] (ERA5): Parámetro climatológico (se incluyen escalas entre 3 y 22 km aproximadamente). Este parámetro no varía en el tiempo.
- Desviación estándar de la orografía [adimensional] (ERA5): Este parámetro es uno de los cuatro parámetros (los otros son el ángulo de orografía a escala inferior a la cuadrícula, la pendiente y la anisotropía) que describen las características de la orografía que son demasiado pequeñas para ser resueltas por la cuadrícula del modelo. Estos cuatro parámetros se calculan para características orográficas con escalas horizontales comprendidas entre 5 km y la resolución de cuadrícula del modelo, y se derivan de la altura de valles, colinas y montañas con una resolución de aproximadamente 1 km. Se utilizan como entrada para el esquema de orografía de subcuadrícula que representa el bloqueo de bajo nivel y los efectos de ondas de gravedad orográficas. Este parámetro representa la desviación estándar de la altura de los valles, colinas y montañas de la subcuadrícula dentro de un cuadro de cuadrícula. Este parámetro no varía en el tiempo.
- Escorrentía subterránea [metro] (ERA5): Parte del agua de la lluvia, de la nieve que se derrite o de lo profundo del suelo, permanece almacenada en el suelo. De lo contrario, el agua se escurre, ya sea sobre la superficie (escorrentía superficial) o bajo tierra (escorrentía subterránea) y la suma de estos dos se denomina escorrentía. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para

los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Las unidades de escorrentía son la profundidad en metros de agua. Esta es la profundidad que tendría el agua si se distribuyera uniformemente sobre la caja de rejilla. Se debe tener cuidado al comparar los parámetros del modelo con las observaciones, porque las observaciones son a menudo locales a un punto en particular en lugar de promediar sobre un cuadro de cuadrícula. Las observaciones también se toman a menudo en diferentes unidades, como mm/día, en lugar de los metros acumulados producidos aquí. La escorrentía es una medida de la disponibilidad de agua en el suelo y puede, por ejemplo, utilizarse como indicador de sequía o inundación.

- Flujo de calor latente superficial [J/m^2] (ERA5): Este parámetro es la transferencia de calor latente (resultado de los cambios de fase del agua, como la evaporación o la condensación) entre la superficie de la Tierra y la atmósfera a través de los efectos del movimiento turbulento del aire. La evaporación desde la superficie de la Tierra representa una transferencia de energía desde la superficie hacia la atmósfera. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Las unidades son julios por metro cuadrado [J/m^2]. Para convertir a vatios por metro cuadrado [W/m^2], los valores acumulados deben dividirse por el período de acumulación expresado en segundos. La convención ECMWF para flujos verticales es positiva hacia abajo.
- Radiación solar neta superficial [J/m^2] (ERA5): Este parámetro es la cantidad de

radiación solar (también conocida como radiación de onda corta) que alcanza un plano horizontal en la superficie de la Tierra (tanto directa como difusa) menos la cantidad reflejada por la superficie de la Tierra (que se rige por el albedo). La radiación del Sol (radiación solar o de onda corta) es parcialmente reflejada de regreso al espacio por las nubes y partículas en la atmósfera (aerosoles) y parte de ella es absorbida. El resto incide sobre la superficie de la Tierra, donde parte de él se refleja. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media del conjunto y la dispersión del conjunto, el período de acumulación es superior a las 3 horas que terminan en la fecha y hora de validez. Las unidades son julios por metro cuadrado [J/m^2]. Para convertir a vatios por metro cuadrado [W/m^2], los valores acumulados deben dividirse por el período de acumulación expresado en segundos. La convención ECMWF para flujos verticales es positiva hacia abajo.

- Radiación solar neta superficial, cielo despejado [J/m^2] (ERA5): Este parámetro es la cantidad de radiación solar (onda corta) que llega a la superficie de la Tierra (tanto directa como difusa) menos la cantidad reflejada por la superficie de la Tierra (que se rige por el albedo), asumiendo condiciones de cielo despejado (sin nubes). Es la cantidad de radiación que pasa a través de un plano horizontal. Las cantidades de radiación en cielo despejado se calculan exactamente para las mismas condiciones atmosféricas de temperatura, humedad, ozono, gases traza y aerosoles que las cantidades correspondientes del cielo total (nubes incluidas), pero asumiendo que las nubes no están allí. La radiación del Sol (radiación solar o de onda corta) es parcialmente reflejada de regreso al espacio por las nubes y partículas en la atmósfera (aerosoles) y parte de ella es absorbida. El resto incide sobre la superficie de la Tierra, donde parte de él se refleja. La diferencia entre la

radiación solar hacia abajo y reflejada es la radiación solar neta superficial. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Las unidades son julios por metro cuadrado ($[Jm]$) el período de acumulación es superior a las 3 horas que terminan en la fecha y hora de validez. Las unidades son julios por metro cuadrado ($[Jm]$) el período de acumulación es superior a las 3 horas que terminan en la fecha y hora de validez. Las unidades son julios por metro cuadrado $[J/m^2]$. Para convertir a vatios por metro cuadrado $[W/m^2]$, los valores acumulados deben dividirse por el período de acumulación expresado en segundos. La convención ECMWF para flujos verticales es positiva hacia abajo.

- Radiación térmica superficial neta $[J/m^2]$ (ERA5): La radiación térmica (también conocida como radiación de onda larga o terrestre) se refiere a la radiación emitida por la atmósfera, las nubes y la superficie de la Tierra. Este parámetro es la diferencia entre la radiación térmica descendente y ascendente en la superficie de la Tierra. Es la cantidad de radiación que pasa a través de un plano horizontal. La atmósfera y las nubes emiten radiación térmica en todas las direcciones, parte de la cual llega a la superficie como radiación térmica descendente. La radiación térmica ascendente en la superficie consiste en la radiación térmica emitida por la superficie más la fracción de radiación térmica descendente reflejada hacia arriba por la superficie. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Las unidades son

julios por metro cuadrado [J/m^2]. Para convertir a vatios por metro cuadrado [W/m^2], los valores acumulados deben dividirse por el período de acumulación expresado en segundos. La convención ECMWF para flujos verticales es positiva hacia abajo.

- Radiación térmica neta superficial, cielo despejado [J/m^2] (ERA5): La radiación térmica (también conocida como radiación de onda larga o terrestre) se refiere a la radiación emitida por la atmósfera, las nubes y la superficie de la Tierra. Este parámetro es la diferencia entre la radiación térmica descendente y ascendente en la superficie de la Tierra, suponiendo condiciones de cielo despejado (sin nubes). Es la cantidad de radiación que pasa a través de un plano horizontal. Las cantidades de radiación en cielo despejado se calculan exactamente para las mismas condiciones atmosféricas de temperatura, humedad, ozono, gases traza y aerosoles que las cantidades correspondientes del cielo total (nubes incluidas), pero asumiendo que las nubes no están allí. La atmósfera y las nubes emiten radiación térmica en todas las direcciones, parte de la cual llega a la superficie como radiación térmica descendente. La radiación térmica ascendente en la superficie consiste en la radiación térmica emitida por la superficie más la fracción de radiación térmica descendente reflejada hacia arriba por la superficie. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Las unidades son julios por metro cuadrado ($[Jm]$) Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Las unidades son julios por metro cuadrado ($[Jm]$) Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación.

la fecha y hora de validación. Las unidades son julios por metro cuadrado [J/m^2].

Para convertir a vatios por metro cuadrado [W/m^2], los valores acumulados deben dividirse por el período de acumulación expresado en segundos. La convención ECMWF para flujos verticales es positiva hacia abajo.

- Presión superficial [Pa] (ERA5): Este parámetro es la presión (fuerza por unidad de área) de la atmósfera en la superficie de la tierra, el mar y las aguas continentales. Es una medida del peso de todo el aire en una columna verticalmente sobre un punto en la superficie de la Tierra. La presión superficial se usa a menudo en combinación con la temperatura para calcular la densidad del aire. La fuerte variación de la presión con la altitud hace que sea difícil ver los sistemas meteorológicos de baja y alta presión sobre las zonas montañosas, por lo que normalmente se utiliza para este propósito la presión media a nivel del mar, en lugar de la presión superficial. Las unidades de este parámetro son Pascales (Pa). La presión superficial a menudo se mide en hPa y, a veces, se presenta en las antiguas unidades de milibares, mb ($1 \text{ hPa} = 1 \text{ mb} = 100 \text{ Pa}$).
- Escorrentía superficial [metro] (ERA5): Parte del agua de la lluvia, de la nieve que se derrite o de lo profundo del suelo, permanece almacenada en el suelo. De lo contrario, el agua se escurre, ya sea sobre la superficie (escorrentía superficial) o bajo tierra (escorrentía subterránea) y la suma de estos dos se denomina escorrentía. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Las unidades de escorrentía son la profundidad en metros de agua. Esta es la profundidad que tendría el agua si se distribuyera uniformemente sobre la caja de rejilla. Se debe tener

cuidado al comparar los parámetros del modelo con las observaciones, porque las observaciones son a menudo locales a un punto en particular en lugar de promediar sobre un cuadro de cuadrícula. Las observaciones también se toman a menudo en diferentes unidades, como mm/día, en lugar de los metros acumulados producidos aquí. La escorrentía es una medida de la disponibilidad de agua en el suelo y puede, por ejemplo, utilizarse como indicador de sequía o inundación.

- Flujo de calor sensible superficial [J/m^2] (ERA5): Este parámetro es la transferencia de calor entre la superficie de la Tierra y la atmósfera a través de los efectos del movimiento turbulento del aire (pero excluyendo cualquier transferencia de calor resultante de la condensación o la evaporación). La magnitud del flujo de calor sensible se rige por la diferencia de temperatura entre la superficie y la atmósfera suprayacente, la velocidad del viento y la rugosidad de la superficie. Por ejemplo, el aire frío que recubre una superficie cálida produciría un flujo de calor sensible desde la tierra (o el océano) hacia la atmósfera. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Para convertir a vatios por metro cuadrado [W/m^2], los valores acumulados deben dividirse por el período de acumulación expresado en segundos. La convención ECMWF para flujos verticales es positiva hacia abajo.
- Radiación solar superficial hacia abajo, cielo despejado [J/m^2] (ERA5): Este parámetro es la cantidad de radiación solar (también conocida como radiación de onda corta) que alcanza un plano horizontal en la superficie de la Tierra, suponiendo condiciones de cielo despejado (sin nubes). Este parámetro comprende

tanto la radiación solar directa como la difusa. La radiación del Sol (radiación solar o de onda corta) es parcialmente reflejada de regreso al espacio por las nubes y partículas en la atmósfera (aerosoles) y parte de ella es absorbida. El resto incide en la superficie de la Tierra. Las cantidades de radiación en cielo despejado se calculan exactamente para las mismas condiciones atmosféricas de temperatura, humedad, ozono, gases traza y aerosoles que las cantidades correspondientes del cielo total (nubes incluidas), pero asumiendo que las nubes no están allí. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora que finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Las unidades son julios por metro cuadrado [J/m^2]. Para convertir a vatios por metro cuadrado [W/m^2], los valores acumulados deben dividirse por el período de acumulación expresado en segundos. La convención ECMWF para flujos verticales es positiva hacia abajo.

- Radiación solar superficial hacia abajo [J/m^2] (ERA5): Este parámetro es la cantidad de radiación solar (también conocida como radiación de onda corta) que llega a un plano horizontal en la superficie de la Tierra. Este parámetro comprende tanto la radiación solar directa como la difusa. La radiación del Sol (radiación solar o de onda corta) es parcialmente reflejada de regreso al espacio por las nubes y partículas en la atmósfera (aerosoles) y parte de ella es absorbida. El resto incide sobre la superficie terrestre (representada por este parámetro). En una aproximación razonablemente buena, este parámetro es el modelo equivalente a lo que mediría un piranómetro (un instrumento utilizado para medir la radiación solar) en la superficie. Sin embargo, se debe tener cuidado al comparar los parámetros del modelo con las observaciones, porque las observaciones a menudo son locales en un punto particular en el espacio

y el tiempo, en lugar de representar promedios sobre un cuadro de cuadrícula modelo. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Las unidades son julios por metro cuadrado [J/m^2]. Para convertir a vatios por metro cuadrado [W/m^2], los valores acumulados deben dividirse por el período de acumulación expresado en segundos. La convención ECMWF para flujos verticales es positiva hacia abajo.

- Radiación térmica superficial hacia abajo, cielo despejado [J/m^2] (ERA5): Este parámetro es la cantidad de radiación térmica (también conocida como de onda larga o terrestre) emitida por la atmósfera que alcanza un plano horizontal en la superficie de la Tierra, suponiendo condiciones de cielo despejado (sin nubes). La superficie de la Tierra emite radiación térmica, parte de la cual es absorbida por la atmósfera y las nubes. La atmósfera y las nubes también emiten radiación térmica en todas las direcciones, parte de la cual alcanza la superficie. Las cantidades de radiación en cielo despejado se calculan exactamente para las mismas condiciones atmosféricas de temperatura, humedad, ozono, gases traza y aerosoles que las cantidades correspondientes del cielo total (nubes incluidas), pero asumiendo que las nubes no están allí. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora que finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Las unidades son julios por metro cuadrado [J/m^2]. Para convertir a vatios por metro cuadrado [W/m^2], los valores acumulados deben dividirse por el período de acumulación expresado en segundos. La convención ECMWF para

flujos verticales es positiva hacia abajo.

- Radiación térmica superficial hacia abajo [J/m^2] (ERA5): Este parámetro es la cantidad de radiación térmica (también conocida como de onda larga o terrestre) emitida por la atmósfera y las nubes que alcanza un plano horizontal en la superficie de la Tierra. La superficie de la Tierra emite radiación térmica, parte de la cual es absorbida por la atmósfera y las nubes. La atmósfera y las nubes también emiten radiación térmica en todas las direcciones, parte de la cual alcanza la superficie (representada por este parámetro). Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Las unidades son julios por metro cuadrado ($[J/m^2]$). Para convertir a vatios por metro cuadrado [W/m^2], los valores acumulados deben dividirse por el período de acumulación expresado en segundos. La convención ECMWF para flujos verticales es positiva hacia abajo.
- TOA radiación solar incidente [J/m^2] (ERA5): Este parámetro es la radiación solar entrante (también conocida como radiación de onda corta), recibida del Sol, en la parte superior de la atmósfera. Es la cantidad de radiación que pasa a través de un plano horizontal. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Las unidades son julios por metro cuadrado ($[J/m^2]$). Para convertir a vatios por metro cuadrado [W/m^2], los valores acumulados deben dividirse por el período

de acumulación expresado en segundos. La convención ECMWF para flujos verticales es positiva hacia abajo.

- Temperatura de la capa de nieve [K] (ERA5): Este parámetro da la temperatura de la capa de nieve desde el suelo hasta la interfaz nieve^aire. El Sistema Integrado de Pronóstico (IFS) del ECMWF representa la nieve como una sola capa adicional sobre el nivel superior del suelo. La nieve puede cubrir todo o parte de la caja de rejilla. Este parámetro se define en todo el mundo, incluso donde no hay nieve. Las regiones sin nieve se pueden enmascarar considerando solo los puntos de la cuadrícula donde la profundidad de la nieve ([metros de agua equivalente] es mayor que 0,0. Este parámetro tiene unidades de kelvin (K). La temperatura medida en kelvin se puede convertir a grados Celsius ($^{\circ}\text{C}$) restando 273,15.
- Radiación solar neta superior [J/m^2] (ERA5): Este parámetro es la radiación solar entrante (también conocida como radiación de onda corta) menos la radiación solar saliente en la parte superior de la atmósfera. Es la cantidad de radiación que pasa a través de un plano horizontal. La radiación solar entrante es la cantidad recibida del Sol. La radiación solar saliente es la cantidad reflejada y dispersada por la atmósfera y la superficie de la Tierra. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Las unidades son julios por metro cuadrado ($[J/m^2]$). Para convertir a vatios por metro cuadrado [W/m^2], los valores acumulados deben dividirse por el período de acumulación expresado en segundos. La convención ECMWF para flujos verticales es positiva hacia abajo.

- Radiación solar neta superior, cielo despejado [J/m^2] (ERA5): Este parámetro es la radiación solar entrante (también conocida como radiación de onda corta) menos la radiación solar saliente en la parte superior de la atmósfera, suponiendo condiciones de cielo despejado (sin nubes). Es la cantidad de radiación que pasa a través de un plano horizontal. La radiación solar entrante es la cantidad recibida del Sol. La radiación solar saliente es la cantidad reflejada y dispersada por la atmósfera y la superficie de la Tierra, suponiendo condiciones de cielo despejado (sin nubes). Las cantidades de radiación en cielo despejado se calculan exactamente para las mismas condiciones atmosféricas de temperatura, humedad, ozono, gases traza y aerosoles que las cantidades de cielo total (nubes incluidas), pero asumiendo que las nubes no están allí. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora que finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Las unidades son julios por metro cuadrado [J/m^2]. Para convertir a vatios por metro cuadrado [W/m^2], los valores acumulados deben dividirse por el período de acumulación expresado en segundos. La convención ECMWF para flujos verticales es positiva hacia abajo.
- Radiación térmica neta superior [J/m^2] (ERA5): La radiación térmica (también conocida como terrestre o de onda larga) emitida al espacio en la parte superior de la atmósfera se conoce comúnmente como Radiación de onda larga saliente (OLR). La radiación térmica neta superior (este parámetro) es igual al negativo de OLR. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación

de 3 horas desde la fecha y hora de validación. Las unidades son julios por metro cuadrado [J/m^2]. Para convertir a vatios por metro cuadrado [W/m^2], los valores acumulados deben dividirse por el período de acumulación expresado en segundos. La convención ECMWF para flujos verticales es positiva hacia abajo.

- Radiación térmica neta superior, cielo despejado [J/m^2] (ERA5): Este parámetro es la radiación térmica (también conocida como terrestre o de onda larga) emitida al espacio en la parte superior de la atmósfera, suponiendo condiciones de cielo despejado (sin nubes). Es la cantidad que pasa por un plano horizontal. Tenga en cuenta que la convención ECMWF para flujos verticales es positiva hacia abajo, por lo que un flujo de la atmósfera al espacio será negativo. Las cantidades de radiación en cielo despejado se calculan exactamente para las mismas condiciones atmosféricas de temperatura, humedad, ozono, gases traza y aerosoles que las cantidades de cielo total (nubes incluidas), pero asumiendo que las nubes no están allí. La radiación térmica emitida al espacio en la parte superior de la atmósfera se conoce comúnmente como radiación de onda larga saliente (OLR) (es decir, tomando un flujo de la atmósfera al espacio como positivo). Tenga en cuenta que OLR generalmente se muestra en unidades de vatios por metro cuadrado (W/m^2). Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Las unidades son julios por metro cuadrado [J/m^2]. Para convertir a vatios por metro cuadrado [W/m^2], los valores acumulados deben dividirse por el período de acumulación expresado en segundos.
- Nubosidad total [adimensional] (ERA5): Este parámetro es la proporción de

una cuadrícula cubierta por nubes. La cobertura total de nubes es un campo de un solo nivel calculado a partir de la nube que se produce en diferentes niveles del modelo a través de la atmósfera. Se hacen suposiciones sobre el grado de superposición/aleatoriedad entre las nubes a diferentes alturas. Las fracciones de nubes varían de 0 a 1.

- Columna total nube agua helada [kg/m^2] (ERA5): Este parámetro es la cantidad de hielo contenido dentro de las nubes en una columna que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. La nieve (cristales de hielo agregados) no está incluida en este parámetro. Este parámetro representa el valor promedio del área para un cuadro de cuadrícula modelo. Las nubes contienen un continuo de gotas de agua y partículas de hielo de diferentes tamaños. El esquema de nubes del Sistema Integrado de Pronóstico (IFS) del ECMWF simplifica esto para representar una cantidad de gotas/partículas de nubes discretas que incluyen: gotas de agua de nubes, gotas de lluvia, cristales de hielo y nieve (cristales de hielo agregados). Los procesos de formación de gotas, transición de fase y agregación también están muy simplificados en el IFS.
- Columna total nube agua líquida [kg/m^2] (ERA5): Este parámetro es la cantidad de agua líquida contenida en las gotas de las nubes en una columna que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Las gotas de agua de lluvia, que son mucho más grandes en tamaño (y masa), no se incluyen en este parámetro. Este parámetro representa el valor promedio del área para un cuadro de cuadrícula modelo. Las nubes contienen un continuo de gotas de agua y partículas de hielo de diferentes tamaños. El esquema de nubes del Sistema Integrado de Pronóstico (IFS) del ECMWF simplifica esto para representar una cantidad de gotas/partículas de nubes discretas que incluyen: gotas de agua de nubes, gotas de lluvia, cristales de hielo y nieve (cristales

de hielo agregados). Los procesos de formación de gotas, transición de fase y agregación también están muy simplificados en el IFS.

- Columna de ozono total [kg/m^2] (ERA5): Este parámetro es la cantidad total de ozono en una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Este parámetro también puede denominarse ozono total u ozono integrado verticalmente. Los valores están dominados por el ozono dentro de la estratosfera. En el Sistema Integrado de Pronóstico (IFS) del ECMWF, existe una representación simplificada de la química del ozono (incluida la representación de la química que ha causado el agujero en la capa de ozono). El ozono también se transporta en la atmósfera a través del movimiento del aire. El ozono natural en la estratosfera ayuda a proteger a los organismos en la superficie de la Tierra de los efectos nocivos de la radiación ultravioleta (UV) del sol. El ozono cerca de la superficie, a menudo producido por la contaminación, es dañino para los organismos. En el SFI, las unidades para el ozono total son kilogramos por metro cuadrado, pero antes del 06/12/2001 se usaban unidades dobson. Las unidades Dobson (DU) todavía se utilizan ampliamente para la columna de ozono total. $1\text{ UD} = 2,1415\text{E}^5\text{ [kg/m}^2]$
- Columna total de agua de lluvia [kg/m^2] (ERA5): Este parámetro es la cantidad total de agua en gotas del tamaño de una gota de lluvia (que pueden caer a la superficie como precipitación) en una columna que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Este parámetro representa el valor promedio del área para un cuadro de cuadrícula. Las nubes contienen un continuo de gotas de agua y partículas de hielo de diferentes tamaños. El esquema de nubes del Sistema Integrado de Pronóstico (IFS) del ECMWF simplifica esto para representar una cantidad de gotas/partículas de nubes discretas que

incluyen: gotas de agua de nubes, gotas de lluvia, cristales de hielo y nieve (cristales de hielo agregados). Los procesos de formación, conversión y agregación de gotas también están muy simplificados en el IFS.

- Columna total de agua de nieve [kg/m^2] (ERA5): Este parámetro es la cantidad total de agua en forma de nieve (cristales de hielo agregados que pueden caer a la superficie como precipitación) en una columna que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Este parámetro representa el valor promedio del área para un cuadro de cuadrícula. Las nubes contienen un continuo de gotas de agua y partículas de hielo de diferentes tamaños. El esquema de nubes del Sistema Integrado de Pronóstico (IFS) del ECMWF simplifica esto para representar una cantidad de gotas/partículas de nubes discretas que incluyen: gotas de agua de nubes, gotas de lluvia, cristales de hielo y nieve (cristales de hielo agregados). Los procesos de formación, conversión y agregación de gotas también están muy simplificados en el IFS.
- Columna total de agua líquida sobre enfriada [kg/m^2] (ERA5): Este parámetro es la cantidad de agua líquida contenida en las gotas de las nubes en una columna que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Las gotas de agua de lluvia, que son mucho más grandes en tamaño (y masa), no se incluyen en este parámetro. Este parámetro representa el valor promedio del área para un cuadro de cuadrícula modelo. Las nubes contienen un continuo de gotas de agua y partículas de hielo de diferentes tamaños. El esquema de nubes del Sistema Integrado de Pronóstico (IFS) del ECMWF simplifica esto para representar una cantidad de gotas de nubes discretas.
- Agua de columna total [kg/m^2] (ERA5): Este parámetro es la suma de vapor de agua, agua líquida, hielo de nubes, lluvia y nieve en una columna que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera.

En versiones antiguas del modelo ECMWF (IFS), la lluvia y la nieve no se tenían en cuenta.

- Vapor de agua total de la columna [kg/m^2] (ERA5): Este parámetro es la cantidad total de vapor de agua en una columna que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Este parámetro representa el valor promedio del área para un cuadro de cuadrícula.
- Precipitaciones totales [metro] (ERA5): Este parámetro es el agua líquida y congelada acumulada, compuesta por lluvia y nieve, que cae a la superficie terrestre. Es la suma de la precipitación a gran escala y la precipitación convectiva. La precipitación a gran escala es generada por el esquema de nubes en el Sistema Integrado de Pronóstico (IFS) del ECMWF. El esquema de nubes representa la formación y disipación de nubes y precipitaciones a gran escala debido a cambios en las cantidades atmosféricas (como presión, temperatura y humedad) pronosticadas directamente por el IFS a escalas espaciales de la cuadrícula o mayores. La precipitación convectiva es generada por el esquema de convección en el IFS, que representa la convección a escalas espaciales más pequeñas que la cuadrícula. Este parámetro no incluye niebla, rocío o la precipitación que se evapora en la atmósfera antes de que llegue a la superficie de la Tierra. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Las unidades de este parámetro son la profundidad en metros de agua equivalente. Es la profundidad que tendría el agua si se distribuyera uniformemente sobre la caja de rejilla.
- Radiación solar directa total del cielo en la superficie [J/m^2] (ERA5): Este

parámetro es la cantidad de radiación solar directa (también conocida como radiación de onda corta) que llega a la superficie de la Tierra. Es la cantidad de radiación que pasa a través de un plano horizontal. La radiación solar en la superficie puede ser directa o difusa. La radiación solar puede ser dispersada en todas las direcciones por partículas en la atmósfera, algunas de las cuales alcanzan la superficie (radiación solar difusa). Parte de la radiación solar llega a la superficie sin ser dispersada (radiación solar directa). Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para convertir a vatios por metro cuadrado [W/m^2], los valores acumulados deben dividirse por el período de acumulación expresado en segundos. La convención ECMWF para flujos verticales es positiva hacia abajo.

- Índice de totales [K] (ERA5): Este parámetro da una indicación de la probabilidad de ocurrencia de una tormenta eléctrica y su severidad usando el gradiente vertical de temperatura y humedad. Los valores de este índice indican lo siguiente: <44 Tormentas eléctricas poco probables, 44-50 Tormentas eléctricas probables, 51-52 Tormentas eléctricas severas aisladas, 53-56 Tormentas eléctricas severas ampliamente dispersas, 56-60 Tormentas eléctricas severas dispersas más probables. El índice de totales totales es la diferencia de temperatura entre 850 hPa (cerca de la superficie) y 500 hPa (troposfera media) (tasa de variación) más una medida del contenido de humedad entre 850 hPa y 500 hPa. La probabilidad de convección profunda tiende a aumentar con el aumento de la tasa de caída y el contenido de humedad atmosférica. Hay una serie de limitaciones en este índice. Además, la interpretación del valor del índice varía según la estación y la ubicación.

- Altura base de la capa de captura [metro] (ERA5): Altura de la base de la capa de captura diagnosticada a partir del gradiente vertical de la refractividad atmosférica.
- Altura superior de la capa de captura [metro] (ERA5): Altura superior de la capa de captura diagnosticada a partir del gradiente vertical de la refractividad atmosférica.
- Tipo de vegetación alta [adimensional] (ERA5): Este parámetro indica los 6 tipos de vegetación alta reconocidos por el Sistema Integrado de Pronóstico del ECMWF: 3 = árboles de hojas perennes, 4 = árboles de hoja caduca, 5 = árboles de hoja caduca, 6 = árboles de hoja perenne, 18 = bosque/bosque mixto, 19 = Bosque interrumpido. Un valor de 0 indica un punto sin vegetación alta, incluida una ubicación de aguas continentales o oceánicas. Los tipos de vegetación se utilizan para calcular el balance de energía superficial y el albedo de la nieve. Este parámetro no varía en el tiempo.
- Tipo de vegetación baja [adimensional] (ERA5): Este parámetro indica los 10 tipos de vegetación baja reconocidos por el Sistema Integrado de Pronóstico del ECMWF: 1 = Cultivos, Agricultura mixta, 2 = Hierba, 7 = Hierba alta, 9 = Tundra, 10 = Cultivos de regadío, 11 = Semidesierto, 13 = Ciénagas y pantanos, 16 = Arbustos perennifolios, 17 = Arbustos caducifolios, 20 = Mezclas de agua y tierra. Un valor de 0 indica un punto sin vegetación baja, incluida una ubicación de aguas continentales o oceánicas. Los tipos de vegetación se utilizan para calcular el balance de energía superficial y el albedo de la nieve. Este parámetro no varía en el tiempo.
- Componente u deriva de Stokes [m/s] (ERA5): Este parámetro es el componente hacia el este de la deriva superficial de Stokes. La deriva de Stokes es la velocidad de deriva neta debida a las olas del viento en la superficie. Se limita a los pocos

metros superiores de la columna de agua del océano, con el mayor valor en la superficie. Por ejemplo, una partícula de fluido cerca de la superficie se moverá lentamente en la dirección de propagación de la onda.

- Albedo UV visible para radiación difusa [adimensional] (ERA5): El albedo es una medida de la reflectividad de la superficie terrestre. Este parámetro es la fracción de radiación solar difusa (onda corta) con longitudes de onda entre 0,3 y 0,7 [μm] (micras, 1 millonésima de metro) reflejada por la superficie de la Tierra (solo para superficies terrestres sin nieve). En el Sistema Integrado de Pronóstico (IFS) del ECMWF, el albedo se trata por separado para la radiación solar con longitudes de onda superiores o inferiores a 0,7 [μm] y para la radiación solar directa y difusa (dando 4 componentes al albedo). La radiación solar en la superficie puede ser directa o difusa. La radiación solar puede ser dispersada en todas las direcciones por partículas en la atmósfera, algunas de las cuales alcanzan la superficie (radiación solar difusa). Parte de la radiación solar llega a la superficie sin ser dispersada (radiación solar directa). En el SFI, Se utiliza un albedo de fondo climatológico (valores observados promediados durante un período de varios años) que varía de mes a mes a lo largo del año, modificado por el modelo sobre agua, hielo y nieve. Este parámetro varía entre 0 y 1.
- Albedo UV visible para radiación directa [adimensional] (ERA5): El albedo es una medida de la reflectividad de la superficie terrestre. Este parámetro es la fracción de radiación solar directa (onda corta) con longitudes de onda entre 0,3 y 0,7 [μm] (micras, 1 millonésima de metro) reflejada por la superficie de la Tierra (solo para superficies terrestres sin nieve). En el Sistema Integrado de Pronóstico (IFS) del ECMWF, el albedo se trata por separado para la radiación solar con longitudes de onda superiores o inferiores a 0,7 [μm] y para la radiación solar directa y difusa (dando 4 componentes al albedo). La radiación solar en

la superficie puede ser directa o difusa. La radiación solar puede ser dispersada en todas las direcciones por partículas en la atmósfera, algunas de las cuales alcanzan la superficie (radiación solar difusa). Parte de la radiación solar llega a la superficie sin ser dispersada (radiación solar directa). En el IFS, se utiliza un albedo de fondo climatológico (valores observados promediados durante un período de varios años) que varía de mes a mes a lo largo del año, modificado por el modelo sobre agua, hielo y nieve.

- Componente v deriva de Stokes [m/s] (ERA5): Este parámetro es el componente hacia el norte de la deriva superficial de Stokes. La deriva de Stokes es la velocidad de deriva neta debida a las olas del viento en la superficie. Se limita a los pocos metros superiores de la columna de agua del océano, con el mayor valor en la superficie. Por ejemplo, una partícula de fluido cerca de la superficie se moverá lentamente en la dirección de propagación de la onda.
- Integral vertical de divergencia del flujo de agua congelada en la nube [$kg/m^2 s$] (ERA5): La integral vertical del flujo de agua congelada en la nube es la tasa de flujo horizontal del agua congelada en la nube, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Su divergencia horizontal es la tasa de agua congelada de la nube que se extiende hacia afuera desde un punto, por metro cuadrado. Este parámetro es positivo para agua congelada de nubes que se está dispersando o divergiendo, y negativa para lo contrario, para agua congelada de nubes que se está concentrando o convergiendo (convergencia). Por lo tanto, este parámetro indica si los movimientos atmosféricos actúan para disminuir (por divergencia) o aumentar (por convergencia) la integral vertical del agua congelada en la nube. Tenga en cuenta que “nube de agua congelada” es lo mismo que “nube de agua helada”.

- Integral vertical de divergencia del flujo de agua líquida en la nube [$kg/m^2 s$] (ERA5): La integral vertical del flujo de agua líquida de la nube es la tasa de flujo horizontal del agua líquida de la nube, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Su divergencia horizontal es la tasa de agua líquida de la nube que se extiende hacia afuera desde un punto, por metro cuadrado. Este parámetro es positivo para el agua líquida de las nubes que se está dispersando o divergiendo, y negativo para lo contrario, para el agua líquida de las nubes que se está concentrando o convergiendo (convergencia). Por lo tanto, este parámetro indica si los movimientos atmosféricos actúan para disminuir (por divergencia) o aumentar (por convergencia) la integral vertical del agua líquida de la nube.
- Integral vertical de divergencia del flujo geopotencial [W/m^2] (ERA5): La integral vertical del flujo geopotencial es la tasa de flujo horizontal del geopotencial, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Su divergencia horizontal es la tasa de expansión del geopotencial hacia afuera desde un punto, por metro cuadrado. Este parámetro es positivo para el geopotencial que se está extendiendo o divergiendo, y negativo para el contrario, para el geopotencial que se está concentrando o convergiendo (convergencia). Este parámetro indica si los movimientos atmosféricos actúan para disminuir (por divergencia) o aumentar (por convergencia) la integral vertical del geopotencial. El geopotencial es la energía potencial gravitacional de una unidad de masa, en un lugar particular, en relación con el nivel medio del mar. También es la cantidad de trabajo que habría que hacer, contra la fuerza de la gravedad, para levantar una unidad de masa a ese lugar desde el nivel medio del mar. Este parámetro se puede utilizar para estudiar el balance energético atmosférico.

- Integral vertical de divergencia del flujo de energía cinética [W/m^2] (ERA5): La integral vertical del flujo de energía cinética es la tasa de flujo horizontal de energía cinética, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Su divergencia horizontal es la tasa de energía cinética que se propaga hacia afuera desde un punto, por metro cuadrado. Este parámetro es positivo para energía cinética que se dispersa o diverge, y negativo para lo contrario, para energía cinética que se concentra o converge (convergencia). Este parámetro indica si los movimientos atmosféricos actúan para disminuir (por divergencia) o aumentar (por convergencia) la integral vertical de la energía cinética. La energía cinética atmosférica es la energía de la atmósfera debido a su movimiento. Solo se considera el movimiento horizontal en el cálculo de este parámetro.
- Integral vertical de divergencia del flujo másico [kg/m^2s] (ERA5): La integral vertical del flujo de masa es la tasa horizontal de flujo de masa, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Su divergencia horizontal es la tasa de masa que se extiende hacia afuera desde un punto, por metro cuadrado. Este parámetro es positivo para masa que se dispersa o diverge, y negativo para lo contrario, para masa que se concentra o converge (convergencia). Este parámetro indica si los movimientos atmosféricos actúan para disminuir (por divergencia) o aumentar (por convergencia) la integral vertical de masa. Este parámetro se puede utilizar para estudiar la masa atmosférica y los balances de energía.
- Integral vertical de divergencia del flujo de humedad [kg/m^2s] (ERA5): La integral vertical del flujo de humedad es la tasa horizontal de flujo de humedad, por metro a través del flujo, para una columna de aire que se extiende desde la

superficie de la Tierra hasta la parte superior de la atmósfera. Su divergencia horizontal es la tasa de humedad que se esparce hacia afuera desde un punto, por metro cuadrado. Este parámetro es positivo para la humedad que se dispersa o diverge, y negativo para la humedad que se concentra o converge (convergencia). Este parámetro indica si los movimientos atmosféricos actúan para disminuir (por divergencia) o aumentar (por convergencia) la integral vertical de la humedad. 1 kg de agua repartida en 1 metro cuadrado de superficie tiene 1 mm de profundidad (despreciando los efectos de la temperatura sobre la densidad del agua), por lo que las unidades equivalen a mm (de agua líquida) por segundo.

- Integral vertical de divergencia del flujo de ozono [$kg/m^2 s$] (ERA5): La integral vertical del flujo de ozono es la tasa de flujo horizontal de ozono, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Su divergencia horizontal es la tasa de ozono que se esparce hacia afuera desde un punto, por metro cuadrado. Este parámetro es positivo para el ozono que se dispersa o diverge, y negativo para el ozono que se concentra o converge (convergencia). Este parámetro indica si los movimientos atmosféricos actúan para disminuir (por divergencia) o aumentar (por convergencia) la integral vertical del ozono. En el Sistema Integrado de Pronóstico (IFS) del ECMWF, hay una representación simplificada de la química del ozono (que incluye una representación de la química que ha causado el agujero en la capa de ozono).
- Integral vertical de divergencia del flujo de energía térmica [W/m^2] (ERA5): La integral vertical del flujo de energía térmica es la tasa horizontal de flujo de energía térmica, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Su divergencia horizontal es la tasa de energía térmica que se

propaga hacia afuera desde un punto, por metro cuadrado. Este parámetro es positivo para energía térmica que se dispersa o diverge, y negativo para energía térmica que se concentra o converge (convergencia). Este parámetro indica si los movimientos atmosféricos actúan para disminuir (por divergencia) o aumentar (por convergencia) la integral vertical de la energía térmica. La energía térmica es igual a la entalpía, que es la suma de la energía interna y la energía asociada a la presión del aire sobre su entorno. La energía interna es la energía contenida dentro de un sistema, es decir, la energía microscópica de las moléculas de aire, en lugar de la energía macroscópica asociada con, por ejemplo, el viento o la energía potencial gravitatoria. La energía asociada a la presión del aire sobre su entorno es la energía necesaria para dar cabida al sistema desplazando su entorno y se calcula a partir del producto de la presión por el volumen. Este parámetro se puede utilizar para estudiar el flujo de energía térmica a través del sistema climático y para investigar el balance de energía atmosférica. La energía asociada a la presión del aire sobre su entorno es la energía necesaria para dar cabida al sistema desplazando su entorno y se calcula a partir del producto de la presión por el volumen. Este parámetro se puede utilizar para estudiar el flujo de energía térmica a través del sistema climático y para investigar el balance de energía atmosférica. La energía asociada a la presión del aire sobre su entorno es la energía necesaria para dar cabida al sistema desplazando su entorno y se calcula a partir del producto de la presión por el volumen. Este parámetro se puede utilizar para estudiar el flujo de energía térmica a través del sistema climático y para investigar el balance de energía atmosférica.

- Integral vertical de divergencia del flujo de energía total [W/m^2] (ERA5): La integral vertical del flujo de energía total es la tasa de flujo horizontal de energía total, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Su

divergencia horizontal es la tasa de energía total que se propaga hacia afuera desde un punto, por metro cuadrado. Este parámetro es positivo para la energía total que se está dispersando o divergiendo, y negativo para lo contrario, para la energía total que se está concentrando o convergiendo (convergencia). Este parámetro indica si los movimientos atmosféricos actúan para disminuir (por divergencia) o aumentar (por convergencia) la integral vertical de la energía total. La energía atmosférica total se compone de energía interna, potencial, cinética y latente. Este parámetro se puede utilizar para estudiar el balance energético atmosférico.

- Integral vertical del flujo de agua congelada en la nube hacia el este [kg/ms] (ERA5): Este parámetro es la tasa de flujo horizontal de agua congelada en la nube, en dirección este, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Los valores positivos indican un flujo de oeste a este. Tenga en cuenta que “nube de agua congelada” es lo mismo que “nube de agua helada”.
- Integral vertical del flujo de agua líquida en las nubes hacia el este [kg/ms] (ERA5): Este parámetro es la tasa de flujo horizontal del agua líquida de la nube, en dirección este, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Los valores positivos indican un flujo de oeste a este.
- Integral vertical del flujo geopotencial hacia el este [W/m] (ERA5): Este parámetro es la tasa de flujo horizontal del geopotencial, en dirección este, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Los valores positivos indican un flujo de oeste a este. El geopotencial es la energía potencial gravitacional de una unidad de masa, en un lugar particular, en relación con el nivel medio del

mar. También es la cantidad de trabajo que tendría que realizarse, contra la fuerza de la gravedad, para elevar una unidad de masa hasta ese lugar desde el nivel medio del mar. Este parámetro se puede utilizar para estudiar el balance energético atmosférico.

- Integral vertical del flujo de calor hacia el este [W/m] (ERA5): Este parámetro es la tasa horizontal de flujo de calor en dirección este, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Los valores positivos indican un flujo de oeste a este. El calor (o energía térmica) es igual a la entalpía, que es la suma de la energía interna y la energía asociada con la presión del aire sobre su entorno. La energía interna es la energía contenida dentro de un sistema, es decir, la energía microscópica de las moléculas de aire, en lugar de la energía macroscópica asociada con, por ejemplo, el viento o la energía potencial gravitatoria. La energía asociada a la presión del aire sobre su entorno es la energía necesaria para dar cabida al sistema desplazando su entorno y se calcula a partir del producto de la presión por el volumen.
- Integral vertical del flujo de energía cinética hacia el este [W/m] (ERA5): Este parámetro es la tasa de flujo horizontal de energía cinética, en dirección este, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Los valores positivos indican un flujo de oeste a este. La energía cinética atmosférica es la energía de la atmósfera debido a su movimiento. Solo se considera el movimiento horizontal en el cálculo de este parámetro. Este parámetro se puede utilizar para estudiar el balance energético atmosférico.
- Integral vertical del flujo de masa hacia el este [kg/ms] (ERA5): Este parámetro es la tasa horizontal de flujo de masa, en dirección este, por metro a través del

flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Los valores positivos indican un flujo de oeste a este. Este parámetro se puede utilizar para estudiar la masa atmosférica y los balances de energía.

- Integral vertical del flujo de ozono hacia el este [kg/ms] (ERA5): Este parámetro es la tasa de flujo horizontal de ozono en dirección este, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Los valores positivos denotan un flujo de oeste a este. En el Sistema Integrado de Pronóstico (IFS) del ECMWF, hay una representación simplificada de la química del ozono (que incluye una representación de la química que ha causado el agujero en la capa de ozono). El ozono también se transporta en la atmósfera a través del movimiento del aire.
- Integral vertical del flujo de energía total hacia el este [W/m] (ERA5): Este parámetro es la tasa de flujo horizontal de energía total en dirección este, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Los valores positivos indican un flujo de oeste a este. La energía atmosférica total se compone de energía interna, potencial, cinética y latente. Este parámetro se puede utilizar para estudiar el balance energético atmosférico.
- Integral vertical del flujo de vapor de agua hacia el este [kg/ms] (ERA5): Este parámetro es la tasa de flujo horizontal de vapor de agua, en dirección este, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Los valores positivos indican un flujo de oeste a este.
- Integral vertical de conversión de energía [W/m^2] (ERA5): Este parámetro es una contribución a la cantidad de energía que se convierte entre energía cinética

y energía interna más potencial, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Los valores negativos indican una conversión a energía cinética de potencial más energía interna. Este parámetro se puede utilizar para estudiar el balance energético atmosférico. La circulación de la atmósfera también se puede considerar en términos de conversión de energía.

- Integral vertical de la energía cinética [J/m^2] (ERA5): Este parámetro es la integral vertical de la energía cinética de una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. La energía cinética atmosférica es la energía de la atmósfera debido a su movimiento. Solo se considera el movimiento horizontal en el cálculo de este parámetro. Este parámetro se puede utilizar para estudiar el balance energético atmosférico.
- Integral vertical de la masa de la atmósfera [kg/m^2] (ERA5): Este parámetro es la masa total de aire de una columna que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera, por metro cuadrado. Este parámetro se calcula dividiendo la presión superficial por la aceleración gravitacional de la Tierra, ($g=9,80665 [m^2/s^2]$), y tiene unidades de kilogramos por metro cuadrado. Este parámetro se puede utilizar para estudiar el balance de masa atmosférica.
- Integral vertical de tendencia de masa [kg/m^2s] (ERA5): Este parámetro es la tasa de cambio de la masa de una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Un aumento de la masa de la columna indica un aumento de la presión superficial. Por el contrario, una disminución indica una caída de la presión superficial. La masa de la columna se calcula dividiendo la presión en la superficie de la Tierra por la aceleración gravitatoria, ($g=9,80665 [m^2/s^2]$). Este parámetro se puede utilizar para estudiar la masa atmosférica y los balances de energía.

- Integral vertical del flujo de agua congelada en la nube hacia el norte [kg/ms] (ERA5): Este parámetro es la tasa de flujo horizontal de agua congelada en la nube, en dirección norte, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Los valores positivos indican un flujo de sur a norte. Tenga en cuenta que “nube de agua congelada” es lo mismo que “nube de agua helada”.
- Integral vertical del flujo de agua líquida en las nubes hacia el norte [kg/ms] (ERA5): Este parámetro es la tasa de flujo horizontal del agua líquida de la nube, en dirección al norte, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Los valores positivos indican un flujo de sur a norte.
- Integral vertical del flujo geopotencial hacia el norte [W/m] (ERA5): Este parámetro es la tasa de flujo horizontal de geopotencial en dirección norte, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Los valores positivos indican un flujo de sur a norte. El geopotencial es la energía potencial gravitacional de una unidad de masa, en un lugar particular, en relación con el nivel medio del mar. También es la cantidad de trabajo que tendría que realizarse, contra la fuerza de la gravedad, para elevar una unidad de masa hasta ese lugar desde el nivel medio del mar. Este parámetro se puede utilizar para estudiar el balance energético atmosférico.
- Integral vertical del flujo de calor hacia el norte [W/m]: Este parámetro es la tasa horizontal de flujo de calor en dirección norte, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Los valores positivos indican un flujo de sur a norte. El calor (o energía térmica) es igual a la entalpía, que es la suma de la

energía interna y la energía asociada con la presión del aire sobre su entorno.

La energía interna es la energía contenida dentro de un sistema, es decir, la energía microscópica de las moléculas de aire, en lugar de la energía macroscópica asociada con, por ejemplo, el viento o la energía potencial gravitatoria. La energía asociada a la presión del aire sobre su entorno es la energía necesaria para dar cabida al sistema desplazando su entorno y se calcula a partir del producto de la presión por el volumen.

- Integral vertical del flujo de energía cinética hacia el norte [W/m] (ERA5): Este parámetro es la tasa de flujo horizontal de energía cinética, en dirección norte, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Los valores positivos indican un flujo de sur a norte. La energía cinética atmosférica es la energía de la atmósfera debido a su movimiento. Solo se considera el movimiento horizontal en el cálculo de este parámetro. Este parámetro se puede utilizar para estudiar el balance energético atmosférico.
- Integral vertical del flujo de masa hacia el norte [kg/ms] (ERA5): Este parámetro es la tasa horizontal de flujo de masa, en dirección al norte, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Los valores positivos indican un flujo de sur a norte. Este parámetro se puede utilizar para estudiar la masa atmosférica y los balances de energía.
- Integral vertical del flujo de ozono hacia el norte [kg/ms] (ERA5): Este parámetro es la tasa de flujo horizontal de ozono en dirección norte, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Los valores positivos denotan un flujo de sur a norte. En el Sistema Integrado de Pronóstico (IFS) del ECMWF,

hay una representación simplificada de la química del ozono (que incluye una representación de la química que ha causado el agujero en la capa de ozono). El ozono también se transporta en la atmósfera a través del movimiento del aire.

- Integral vertical del flujo de energía total hacia el norte [W/m] (ERA5): Este parámetro es la tasa de flujo horizontal de energía total en dirección norte, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Los valores positivos indican un flujo de sur a norte. La energía atmosférica total se compone de energía interna, potencial, cinética y latente. Este parámetro se puede utilizar para estudiar el balance energético atmosférico.
- Integral vertical del flujo de vapor de agua hacia el norte [kg/ms] (ERA5): Este parámetro es la tasa de flujo horizontal de vapor de agua, en dirección norte, por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Los valores positivos indican un flujo de sur a norte.
- Integral vertical de energía potencial e interna [J/m^2] (ERA5): Este parámetro es la integral vertical ponderada en masa de la energía potencial e interna de una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. La energía potencial de una parcela de aire es la cantidad de trabajo que tendría que realizarse, contra la fuerza de la gravedad, para elevar el aire hasta ese lugar desde el nivel medio del mar. La energía interna es la energía contenida dentro de un sistema, es decir, la energía microscópica de las moléculas de aire, en lugar de la energía macroscópica asociada con, por ejemplo, el viento o la energía potencial gravitatoria. Este parámetro se puede utilizar para estudiar el balance energético atmosférico. La energía atmosférica total se compone de energía interna, potencial, cinética y latente.

- Integral vertical de energía potencial, interna y latente [J/m^2] (ERA5): Este parámetro es la integral vertical ponderada en masa de la energía potencial, interna y latente de una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. La energía potencial de una parcela de aire es la cantidad de trabajo que tendría que realizarse, contra la fuerza de la gravedad, para elevar el aire hasta ese lugar desde el nivel medio del mar. La energía interna es la energía contenida dentro de un sistema, es decir, la energía microscópica de las moléculas de aire, en lugar de la energía macroscópica asociada con, por ejemplo, el viento o la energía potencial gravitatoria. La energía latente se refiere a la energía asociada con el vapor de agua en la atmósfera y es igual a la energía requerida para convertir el agua líquida en vapor de agua. Este parámetro se puede utilizar para estudiar el balance energético atmosférico.
- Integral vertical de temperatura [kg/m^2] (ERA5): Este parámetro es la integral vertical ponderada en masa de la temperatura para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Este parámetro se puede utilizar para estudiar el balance energético atmosférico.
- Integral vertical de energía térmica [J/m^2] (ERA5): Este parámetro es la integral vertical ponderada en masa de la energía térmica para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. La energía térmica se calcula a partir del producto de la temperatura y la capacidad calorífica específica del aire a presión constante. La energía térmica es igual a la entalpía, que es la suma de la energía interna y la energía asociada a la presión del aire sobre su entorno. La energía interna es la energía contenida dentro de un sistema, es decir, la energía microscópica de las moléculas de aire, en lugar de la energía macroscópica asociada con, por ejemplo, el viento o la energía potencial gravitatoria. La energía asociada a la presión del aire sobre

su entorno es la energía necesaria para dar cabida al sistema desplazando su entorno y se calcula a partir del producto de la presión por el volumen. Este parámetro se puede utilizar para estudiar el balance energético atmosférico. La energía atmosférica total se compone de energía interna, potencial, cinética y latente.

- Integral vertical de energía total [J/m^2] (ERA5): Este parámetro es la integral vertical de la energía total de una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. La energía atmosférica total se compone de energía interna, potencial, cinética y latente. Este parámetro se puede utilizar para estudiar el balance energético atmosférico.
- Divergencia de humedad integrada verticalmente [kg/m^2] (ERA5): La integral vertical del flujo de humedad es la tasa horizontal de flujo de humedad (vapor de agua, líquido de nube y hielo de nube), por metro a través del flujo, para una columna de aire que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Su divergencia horizontal es la tasa de humedad que se esparce hacia afuera desde un punto, por metro cuadrado. Este parámetro se acumula durante un período de tiempo particular que depende de los datos extraídos. Para el reanálisis, el período de acumulación es superior a 1 hora y finaliza en la fecha y hora de validez. Para los miembros del conjunto, la media y su dispersión corresponden a un período de acumulación de 3 horas desde la fecha y hora de validación. Este parámetro es positivo para la humedad que se esparce o diverge, y negativo para lo contrario, para la humedad que se concentra, o convergente (convergencia). Este parámetro indica si los movimientos atmosféricos actúan para disminuir (por divergencia) o aumentar (por convergencia) la integral vertical de la humedad, durante el período de tiempo. Los valores negativos altos de este parámetro (es decir, una gran convergencia de la

humedad) pueden estar relacionados con la intensificación de las precipitaciones y las inundaciones. 1 kg de agua repartida en 1 metro cuadrado de superficie tiene 1 mm de profundidad (despreciando los efectos de la temperatura sobre la densidad del agua), por lo que las unidades equivalen a mm.

- Capa volumétrica de agua del suelo 1 [m^3/m^3] (ERA5): Este parámetro es el volumen de agua en la capa 1 del suelo (0 ^ 7 cm, la superficie está a 0 cm). El Sistema Integrado de Pronóstico (IFS) del ECMWF tiene una representación del suelo en cuatro capas: Capa 1: 0 - 7 cm, Capa 2: 7 - 28 cm, Capa 3: 28 - 100 cm, Capa 4: 100 - 289 cm. El agua del suelo se define en todo el mundo, incluso sobre el océano. Las regiones con una superficie de agua se pueden enmascarar considerando solo los puntos de cuadrícula donde la máscara tierra^mar tiene un valor superior a 0,5. El agua volumétrica del suelo está asociada con la textura del suelo (o clasificación), la profundidad del suelo y el nivel de agua subterránea subyacente.
- Capa volumétrica de agua del suelo 2 [m^3/m^3] (ERA5): Este parámetro es el volumen de agua en la capa 2 del suelo (7 ^ 28 cm, la superficie está a 0 cm). El Sistema Integrado de Pronóstico (IFS) del ECMWF tiene una representación del suelo en cuatro capas: Capa 1: 0 - 7 cm, Capa 2: 7 - 28 cm, Capa 3: 28 - 100 cm, Capa 4: 100 - 289 cm. El agua del suelo se define en todo el mundo, incluso sobre el océano. Las regiones con una superficie de agua se pueden enmascarar considerando solo los puntos de cuadrícula donde la máscara tierra^mar tiene un valor superior a 0,5. El agua volumétrica del suelo está asociada con la textura del suelo (o clasificación), la profundidad del suelo y el nivel de agua subterránea subyacente.
- Capa volumétrica de agua del suelo 3 [m^3/m^3] (ERA5): Este parámetro es el volumen de agua en la capa 3 del suelo (28 ^ 100 cm, la superficie está a 0 cm).

El Sistema Integrado de Pronóstico (IFS) del ECMWF tiene una representación del suelo en cuatro capas: Capa 1: 0 - 7 cm, Capa 2: 7 - 28 cm, Capa 3: 28 - 100 cm, Capa 4: 100 - 289 cm. El agua del suelo se define en todo el mundo, incluso sobre el océano. Las regiones con una superficie de agua se pueden enmascarar considerando solo los puntos de cuadrícula donde la máscara tierra^mar tiene un valor superior a 0,5. El agua volumétrica del suelo está asociada con la textura del suelo (o clasificación), la profundidad del suelo y el nivel de agua subterránea subyacente.

- Capa volumétrica de agua del suelo 4 [m^3/m^3] (ERA5): Este parámetro es el volumen de agua en la capa 4 del suelo (100 ^ 289 cm, la superficie está a 0 cm). El Sistema Integrado de Pronóstico (IFS) del ECMWF tiene una representación del suelo en cuatro capas: Capa 1: 0 - 7 cm, Capa 2: 7 - 28 cm, Capa 3: 28 - 100 cm, Capa 4: 100 - 289 cm. El agua del suelo se define en todo el mundo, incluso sobre el océano. Las regiones con una superficie de agua se pueden enmascarar considerando solo los puntos de cuadrícula donde la máscara tierra^mar tiene un valor superior a 0,5. El agua volumétrica del suelo está asociada con la textura del suelo (o clasificación), la profundidad del suelo y el nivel de agua subterránea subyacente.
- Ancho direccional espectral de onda para ondas de viento [adimensional] (ERA5): Este parámetro indica si las olas generadas por el viento local provienen de direcciones similares o de una amplia gama de direcciones. El campo de ondas de la superficie del océano/mar consiste en una combinación de ondas con diferentes alturas, longitudes y direcciones (conocido como espectro de onda bidimensional). El espectro de olas se puede descomponer en olas de viento-mar, que se ven directamente afectadas por los vientos locales, y se hinchan, las olas que fueron generadas por el viento en un lugar y tiempo diferentes. Este parámetro sólo

tiene en cuenta el oleaje viento-mar. Muchos parámetros de onda ECMWF (como el período medio de onda) brindan información promediada sobre todas las frecuencias y direcciones de onda, por lo que no brindan información sobre la distribución de la energía de onda a través de frecuencias y direcciones. Este parámetro da más información sobre la naturaleza del espectro de onda bidimensional. Este parámetro es una medida del rango de direcciones de onda para cada frecuencia integrada en el espectro bidimensional. Este parámetro toma valores entre 0 y la raíz cuadrada de 2. Donde 0 corresponde a un espectro unidireccional (es decir, todas las frecuencias de onda de la misma dirección) y la raíz cuadrada de 2 indica un espectro uniforme (es decir, todas las frecuencias de onda de una dirección diferente).

- Nivel de cero grados [metro] (ERA5): La altura sobre la superficie de la Tierra donde la temperatura pasa de valores positivos a negativos, correspondientes a la parte superior de una capa cálida, en el momento especificado. Este parámetro se puede utilizar para ayudar a pronosticar la nieve. Si se encuentra más de una capa cálida, entonces el nivel de cero grados corresponde a la parte superior de la segunda capa atmosférica. Este parámetro se establece en cero cuando la temperatura en toda la atmósfera es inferior a 0°C.

Anexo 4: Optimización de los modelos predictivos

Tabla 5.37: Hiperparámetros optimizados para los algoritmos en el uso del suelo urbano sobre la base de datos reducida.

Modelo	Fuente	Hiperparámetro	Valor
GBM	<code>lightgbm</code>	<i>n_estimators</i>	350
		<i>learning_rate</i>	0.112
		<i>num_leaves</i>	300
		<i>max_depth</i>	12
		<i>min_data_in_leaf</i>	400
		<i>lambda_l1</i>	6
		<i>lambda_l2</i>	6
		<i>min_gain_to_split</i>	0.010
		<i>bagging_fraction</i>	0.8
		<i>bagging_freq</i>	40
		<i>feature_fraction</i>	0.600
KNN	<code>sklearn</code>	<i>n_neighbors</i>	7
		<i>weights</i>	<i>uniform</i>
		<i>algorithm</i>	<i>ball tree</i>

Modelo	Fuente	Hiperparámetro	Valor
SVM	sklearn	p	1
		<i>leaf_size</i>	95
Bosques aleatorios	sklearn	<i>epsilon</i>	0.9
		C	2
Regresión SGD	sklearn	<i>n_estimators</i>	150
		<i>max_depth</i>	48
		<i>min_samples_leaf</i>	50
		<i>min_samples_split</i>	92
		<i>max_features</i>	10
Regresión Ridge	sklearn	<i>alpha</i>	0.598
		<i>l1_ratio</i>	0.5
Regresión de red elástica	sklearn	<i>alpha</i>	0.184
Regresión LASSO	sklearn	<i>alpha</i>	0.507
		<i>l1_ratio</i>	0.2
LSTM	tensorflow	<i>Units first layer</i>	200
GRU	tensorflow	<i>Units second layer</i>	100
		<i>Units third layer</i>	100
		<i>Units fourth layer</i>	100
		<i>Units fifth layer</i>	100
RNN	tensorflow	<i>Units first layer</i>	200
		<i>Units second layer</i>	100
		<i>Units third layer</i>	100

Modelo	Fuente	Hiperparámetro	Valor
		<i>Units third layer</i>	100

Modelo, biblioteca-paquete-fuente del algoritmo, nombre del hiperparámetro y valor del hiperparámetro.

Tabla 5.38: Hiperparámetros optimizados para los algoritmos en el uso del suelo industrial sobre la base de datos reducida.

Modelo	Fuente	Hiperparámetro	Valor
GBM	lightgbm	<i>n_estimators</i>	400
		<i>learning_rate</i>	0.215
		<i>num_leaves</i>	200
		<i>max_depth</i>	16
		<i>min_data_in_leaf</i>	300
		<i>lambda_l1</i>	3
		<i>lambda_l2</i>	3
		<i>min_gain_to_split</i>	0.010
		<i>bagging_fraction</i>	0.7
		<i>bagging_freq</i>	35
KNN	sklearn	<i>feature_fraction</i>	0.500
		<i>n_neighbors</i>	10
		<i>weights</i>	<i>uniform</i>
		<i>algorithm</i>	ball tree
		<i>p</i>	2
SVM	sklearn	<i>leaf_size</i>	85
		<i>epsilon</i>	0.7
		<i>C</i>	2
Bosques aleatorios	sklearn	<i>n_estimators</i>	180

Modelo	Fuente	Hiperparámetro	Valor
		<i>max_depth</i>	54
		<i>min_samples_leaf</i>	40
		<i>min_samples_split</i>	65
		<i>max_features</i>	12
Regresión SGD	sklearn	<i>alpha</i>	0.615
		<i>l1_ratio</i>	0.5
Regresión Ridge	sklearn	<i>alpha</i>	0.212
Regresión de red elástica		<i>alpha</i>	0.701
		<i>l1_ratio</i>	0.2
Regresión LASSO	sklearn	<i>alpha</i>	0.376
LSTM	tensorflow	<i>Units first layer</i>	200
		<i>Units second layer</i>	100
GRU	tensorflow	<i>Units first layer</i>	200
		<i>Units second layer</i>	100
RNN	tensorflow	<i>Units first layer</i>	200
		<i>Units second layer</i>	100

Modelo, biblioteca-paquete-fuente del algoritmo, nombre del hiperparámetro y valor del hiperparámetro.

Tabla 5.39: Hiperparámetros optimizados para los algoritmos en el uso del suelo agrícola y áreas verdes sobre la base de datos reducida.

Modelo	Fuente	Hiperparámetro	Valor
GBM	lightgbm	<i>n_estimators</i>	310
		<i>learning_rate</i>	0.108
		<i>num_leaves</i>	220
		<i>max_depth</i>	9

Modelo	Fuente	Hiperparámetro	Valor
		<i>min_data_in_leaf</i>	300
		<i>lambda_l1</i>	2
		<i>lambda_l2</i>	5
		<i>min_gain_to_split</i>	0.008
		<i>bagging_fraction</i>	0.8
		<i>bagging_freq</i>	40
		<i>feature_fraction</i>	0.400
KNN	sklearn	<i>n_neighbors</i>	8
		<i>weights</i>	<i>uniform</i>
		<i>algorithm</i>	<i>auto</i>
		<i>p</i>	1
		<i>leaf_size</i>	50
SVM	sklearn	<i>epsilon</i>	0.2
		<i>C</i>	3
Bosques aleatorios	sklearn	<i>n_estimators</i>	500
		<i>max_depth</i>	40
		<i>min_samples_leaf</i>	50
		<i>min_samples_split</i>	32
		<i>max_features</i>	10
Regresión SGD	sklearn	<i>alpha</i>	0.341
		<i>l1_ratio</i>	1.0
Regresión Ridge	sklearn	<i>alpha</i>	0.291
Regresión de red elástica		<i>alpha</i>	0.470
		<i>l1_ratio</i>	0.8
Regresión LASSO	sklearn	<i>alpha</i>	0.236

Modelo	Fuente	Hiperparámetro	Valor
LSTM	tensorflow	<i>Units first layer</i>	200
		<i>Units second layer</i>	100
		<i>Units third layer</i>	100
		<i>Units fourth layer</i>	100
GRU	tensorflow	<i>Units first layer</i>	200
		<i>Units second layer</i>	100
		<i>Units third layer</i>	100
RNN	tensorflow	<i>Units first layer</i>	200
		<i>Units second layer</i>	100
		<i>Units third layer</i>	100
		<i>Units fourth layer</i>	100

Modelo, biblioteca-paquete-fuente del algoritmo, nombre del hiperparámetro y valor del hiperparámetro.

Anexo 5: Variables no citadas en la literatura como relacionadas a la concentración de PM_{2.5}

Rugosidad de la superficie terrestre

La rugosidad de la superficie terrestre es la altura equivalente en metros de la rugosidad aerodinámica que posee la superficie del terreno. Esta es una medida de la resistencia que ejerce la superficie al flujo de aire, particularmente este parámetro se utiliza para determinar la transferencia de momento del aire a la superficie. Una mayor rugosidad de la superficie genera una reducción de la velocidad del viento cerca de la superficie. Esta variable se deriva de la cobertura, tipo de vegetación y presencia de nieve en la superficie terrestre y se expresa en metros (*m*).

La rugosidad de la superficie terrestre es una variable que, en la escala temporal horaria, no presenta una variación significativa debido a que su cálculo implica la consideración de fenómenos cuyos rangos de variación y ocurrencia superan esta escala temporal. En consecuencia, durante el transcurso de las horas del día, su valor tiende a permanecer constante y solo varía en escalas temporales de mayor magnitud, como los meses del año. Por esta razón, se ha decidido excluir el análisis exploratorio en la escala temporal horaria.

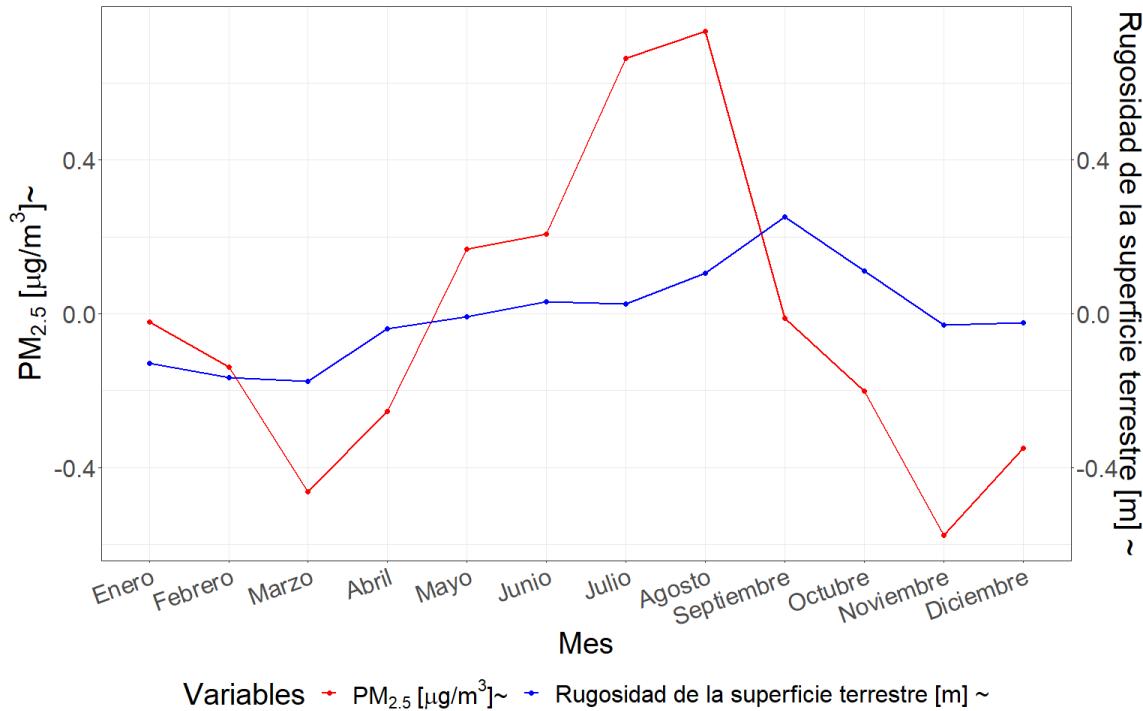


Figura 5.131: Perfil de variación conjunto de los promedios mensuales del PM_{2.5} y la rugosidad de la superficie terrestre a lo largo del año.

La Figura 5.131 presenta el patrón de variación conjunto que describen los promedios mensuales de la rugosidad de la superficie terrestre y la concentración de PM_{2.5} a lo largo del año. Los mayores valores de la variable se registran durante la primavera, específicamente en el mes de septiembre, mientras que el mínimo se da durante el verano, en el mes de marzo. Estos resultados sugieren que existe una relación directa entre ambas variables, donde aumentos en la concentración del contaminante se ven favorecidos por mayores valores de rugosidad de la superficie terrestre.

Tabla 5.40: Coeficientes estimados de la relación funcional.

Desfasaje	Con diferenciación	Sin diferenciación
t	9.317e+01	111.32
t-1	-4.080e+01	92.03
t-9	-1.868e+01	66.55

	Desfasaje	Con diferenciación	Sin diferenciación
t-12	9.1090057	87.25	
t-13	0.4461504	89.23	
t-14	6.760e+00	91.01	
t-23	-1.032e+01	55.22	

Las correlaciones significativas que perduran luego del procedimiento de filtrado de las series se observaron en los tiempos t, t-1, t-9, t-12, t-13, t-14 y t-23, cuatro de los coeficientes demostraron una relación directa estadísticamente significativa (t-0, t-12, t-13 y t-14), mientras que los tres restantes se vieron influenciados por una relación aleatoria (t-1, t-9 y t-23).

Disipación en la capa límite atmosférica

La disipación en la capa límite atmosférica, indica la conversión acumulada de la energía cinética de flujo medio en calor en toda la columna atmosférica, por unidad de superficie, como consecuencia de los efectos de la tensión asociada a remolinos turbulentos cerca de la superficie y al arrastre de forma orográfica turbulenta. Los remolinos turbulentos cerca de la superficie están relacionados con la rugosidad superficial, mientras que el arrastre de forma orográfica turbulenta es la tensión debida a los valles, colinas y montañas en escalas horizontales inferiores a 5 km. Se espera que incrementos de la disipación en la capa límite atmosférica promueva reducciones de la concentración de partículas finas, como consecuencia de la influencia en la disipación de las mismas, por otra parte, reducciones de la disipación no favorecería el fenómeno de dispersión resultando en mayores concentraciones de PM_{2.5}. Las unidades en las que se mide este parámetro son julios por metro cuadrado ($\frac{J}{m^2}$).

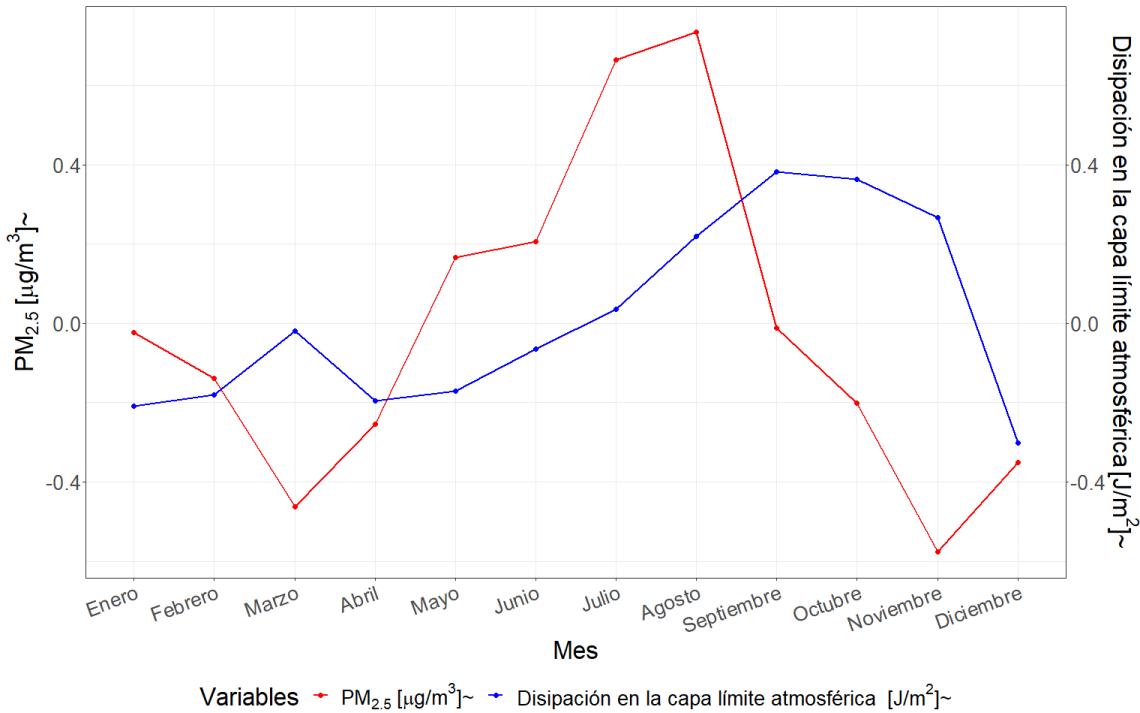


Figura 5.132: Perfil de variación conjunto de los promedios mensuales del $\text{PM}_{2.5}$ y la disipación en la capa límite atmosférica a lo largo del año.

A partir de la Figura 5.132 puede visualizarse el patrón de variación conjunto que describen los promedios mensuales de la disipación en la capa límite atmosférica y la concentración de $\text{PM}_{2.5}$ a lo largo del año. Puede notarse que los valores de menor magnitud de disipación registrados ocurren durante el verano en los meses de enero y diciembre, mientras que los de mayor magnitud ocurren durante la primavera en los meses de septiembre y octubre. A pesar de que no logra apreciarse una clara relación entre las variables en esta escala temporal, la tendencia podría asumirse que es del tipo directa.

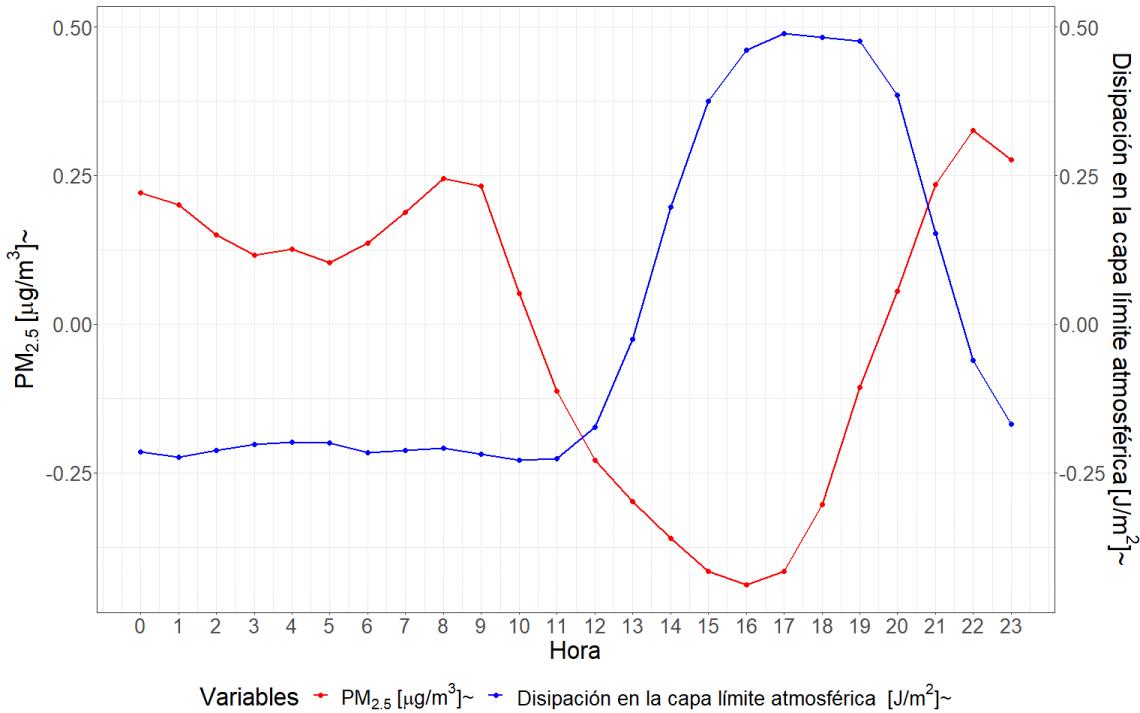


Figura 5.133: Perfil de variación conjunto de los promedios horarios del PM_{2.5} y la disipación en la capa límite atmosférica a lo largo del día.

La Figura 5.133 presenta el patrón de variación conjunto que describen los promedios de la disipación en la capa límite atmosférica y la concentración de PM_{2.5} a lo largo de las horas del día. Puede notarse que ambas variables muestran una relación de vinculación del tipo inversa, en la cual valores de baja magnitud de la variable se condicen con elevadas concentraciones del contaminante, mientras que elevados valores de la variable ocurren en simultaneidad con reducidas concentraciones.

Tabla 5.41: Coeficientes estimados de la relación funcional.

	Desfasaje	Con diferenciación	Sin diferenciación
t		-7.918e-05	-0.0001536
t-5		6.594e-05	-0.000054
t-7		0.0000533	-3.578e-05
t-20		-8.173e-05	-0.0001142

Desfasaje	Con diferenciación	Sin diferenciación
t-22	-0.0001042	-0.0001257

Las correlaciones significativas que perduran luego del procedimiento de filtrado de las series se observaron en los tiempos t, t-5, t-7, t-20 y t-22, dos de los cinco coeficientes demostraron estar influenciados por una relación aleatoria (t-5 y t-7), mientras que los tres restantes exhibieron una relación inversa estadísticamente significativa (t, t-20 y t-22) entre la concentración de PM_{2.5} y la disipación en la capa límite atmosférica.

Integración vertical flujo ozono norte

La variable integración vertical de flujo de ozono norte refiere a la tasa horizontal de flujo de ozono en dirección norte, por metro a través del flujo, para una columna de aire que se extiende desde la superficie hasta la parte superior de la atmósfera. Los valores positivos denotan un flujo de sur a norte. Las unidades en las que se mide este parámetro son kilogramos por metro por segundo ($\frac{Kg}{m.s}$).

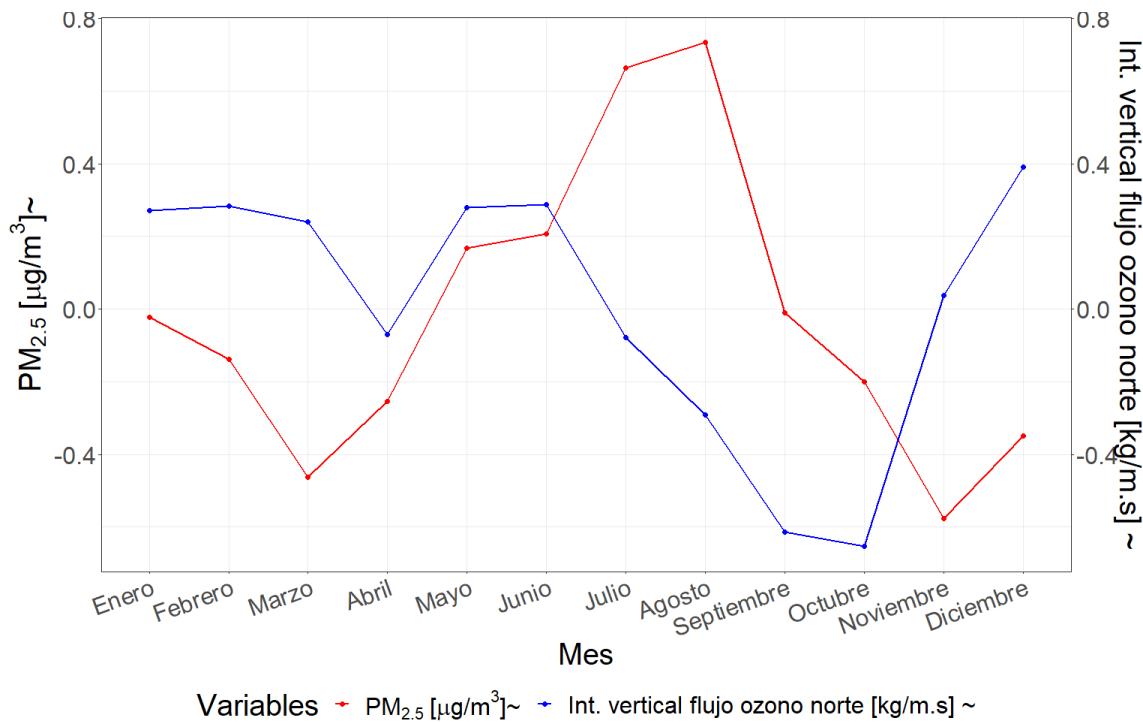


Figura 5.134: Perfil de variación conjunto de los promedios mensuales del PM_{2.5} y la integración vertical del flujo de ozono proveniente del norte a lo largo del año.

La Figura 5.134 presenta el patrón de variación conjunto que describen los promedios mensuales de la integración vertical del flujo de ozono norte y la concentración de PM_{2.5} a lo largo del año. A partir de esta puede notarse que los valores de menor magnitud registrados para la variable ocurren durante la primavera en los meses de septiembre y octubre, mientras que los de mayor magnitud fueron registrados durante el invierno, específicamente en los meses de mayo y junio. Esto sugiere que la relación de vinculación entre las variables es del tipo directa, evidenciando que bajas concentraciones se ven favorecidas por disminuciones de los valores que toma la variable, mientras que elevadas concentraciones se ven impulsadas por aumentos de la misma.

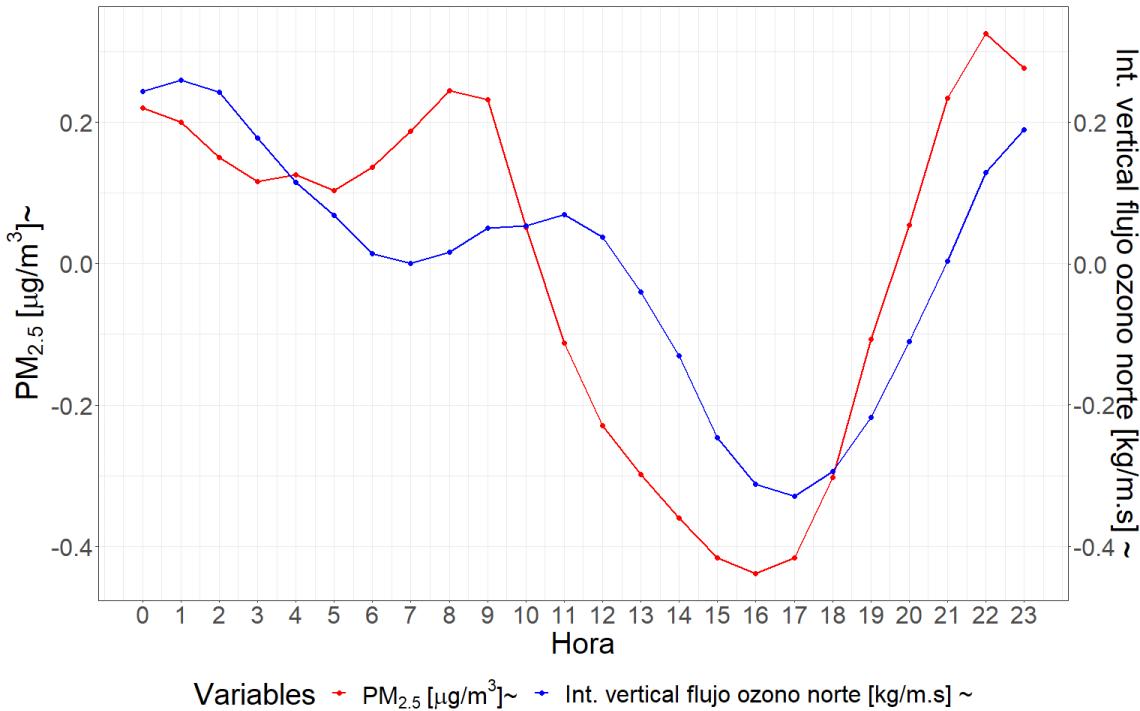


Figura 5.135: Perfil de variación conjunto de los promedios horarios del PM_{2.5} y a lo largo del día.

A partir de la imagen 5.135, puede apreciarse el patrón de variación de la integración vertical del flujo de ozono proveniente del norte promedio a lo largo de las horas del día. En este, puede notarse que los valores más elevados se registran durante la madrugada, específicamente entre las 0 y las 2 horas, experimentando un descenso gradual a lo largo de las horas del día hasta alcanzar su mínimo a las 17 horas. De esta forma, puede apreciarse que la relación de variación conjunta entre la concentración de PM_{2.5} y la integración vertical del flujo de ozono proveniente del norte es del tipo directa.

Tabla 5.42: Coeficientes estimados de la relación funcional.

Desfasaje	Con diferenciación	Sin diferenciación
t-18	-3.189e+01	43.90
t-19	1.219e+02	52.00

Las correlaciones significativas que perduran luego del procedimiento de filtrado de

las series se observaron en los tiempos t-18 y t-19. El *lag* temporal de 5 horas resultó en una relación inversa estadísticamente significativa, mientras que el *lag* de 24 horas lo hizo en una directa estadísticamente significativa, entre la humedad relativa y la concentración de PM_{2.5}. Luego de la estimación de la relación funcional se determinó que uno de los coeficientes demostró una relación aleatoria (t-18), mientras que el restante una relación directa estadísticamente significativa (t-19).

Tasa media de escorrentía subterránea

La tasa media de escorrentía subterránea, es la parte del agua procedente de las precipitaciones, del deshielo o de las napas, que escurre por debajo de la superficie terrestre. El agua que no se almacena escurre, esto puede ser por la superficie (escorrentía superficial) o por debajo del suelo (escorrentía subsuperficial), la suma de ambas se denomina escorrentía. Este parámetro refiere a la tasa que tendría la escorrentía si se repartiera uniformemente por la cuadrícula de resolución espacial. La escorrentía es una medida de la disponibilidad de agua en el suelo y puede utilizarse, por ejemplo, como indicador de sequía o inundación. Se espera que la influencia de la variable en la concentración de PM_{2.5}, sea a través del proceso físico de refrigeración de la atmósfera, influyendo en los niveles de la capa límite atmosférica. Las unidades en las que se mide este parámetro son kilogramos por metro cuadrado por segundo ($\frac{Kg}{m^2.s}$).

Es importante mencionar que el análisis de la tasa media de escorrentía subterránea en la escala temporal horaria no proporciona información relevante dado a su estrecha relación con la precipitación. Cuando no ocurren precipitaciones, la tasa de escorrentía subterránea se mantiene constante a lo largo de las horas del día y, por lo tanto, la interpretación de sus valores podrían dirigirnos a conclusiones erradas. Como consecuencia de esto se decidió excluir el análisis exploratorio en esta escala temporal.

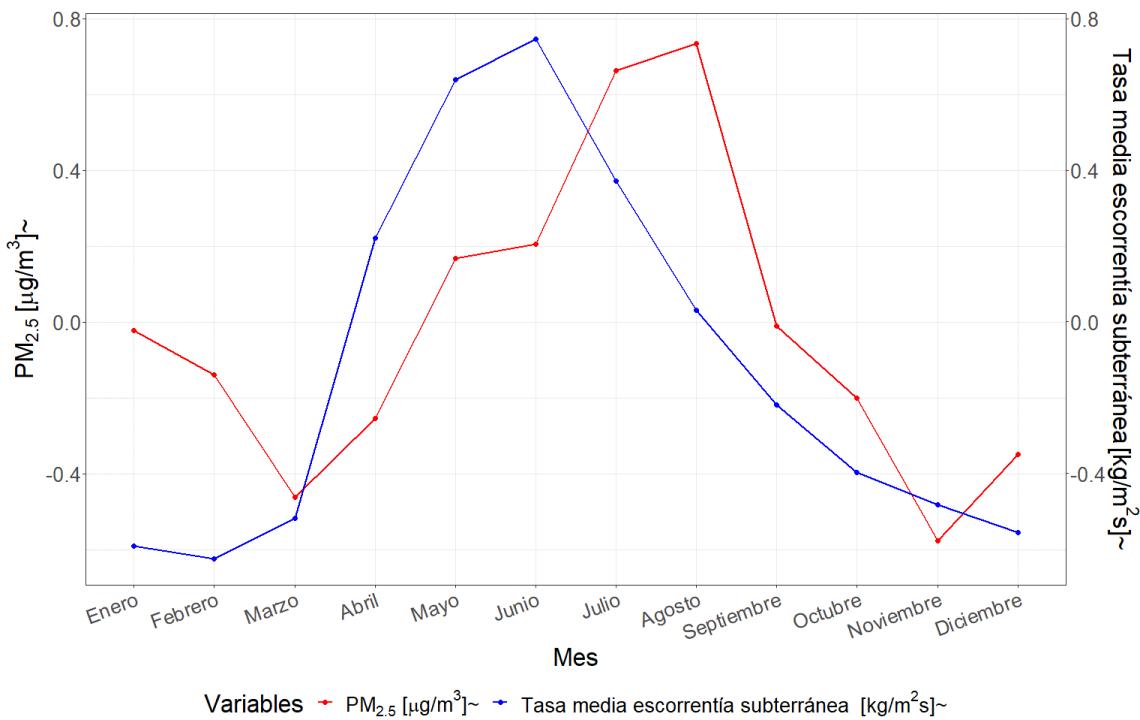


Figura 5.136: Perfil de variación conjunto de los promedios mensuales del $\text{PM}_{2.5}$ y la tasa media de escorrentía subterránea a lo largo del año.

En la Figura 5.136 puede apreciarse el patrón de variación conjunto que describen los promedios mensuales de la tasa media de escorrentía subterránea y la concentración de $\text{PM}_{2.5}$ a lo largo del año. Puede notarse que los valores de menor magnitud registrados para esta tasa ocurren en el verano durante los meses de diciembre, enero y febrero, mientras que los de mayor magnitud se registraron en el invierno durante el mes de junio. Puede notarse que la relación de vinculación entre ambas variables es del tipo directa, demostrando que bajas concentraciones se ven favorecidas por disminuciones de los valores que toma la variable, mientras que elevadas concentraciones se ven favorecidas por aumentos de la tasa media de escorrentía subterránea.

Tabla 5.43: Coeficientes estimados de la relación funcional.

Desfasaje	Con diferenciación	Sin diferenciación
t-8	-2.648e+05	5.770e+06

	Desfasaje	Con diferenciación	Sin diferenciación
t-15		5.326e+04	5.633e+06

Las correlaciones significativas que perduran luego del procedimiento de filtrado de las series se observaron en los tiempos t-8 y t-15, uno de los coeficientes demostró una relación aleatoria (t-8), mientras que el restante reveló una relación de vinculación directa estadísticamente significativa (t-15).

Integración vertical de temperatura

La integración vertical de temperatura es la integral vertical de la temperatura para una columna de aire ponderada por su masa, que se extiende desde la superficie de la Tierra hasta la parte superior de la atmósfera. Este parámetro puede utilizarse para estudiar el balance energético atmosférico. Se espera que esta variable se vincule con la concentración de PM_{2.5}, por medio de la relación existente entre la variación de la temperatura lo largo del espesor de la atmósfera y las estratificaciones que en esta se generan, resultando en aumentos o disminuciones de la concentración PM_{2.5}, en función de la altura a la cual ocurre la estratificación. Este fenómeno se encuentra vinculado también con la variación de la altura de la capa límite atmosférica. Las unidades en las que se mide este parámetro son en kilogramos por Kelvin por metro cuadrado ($\frac{Kg.K}{m^2}$).

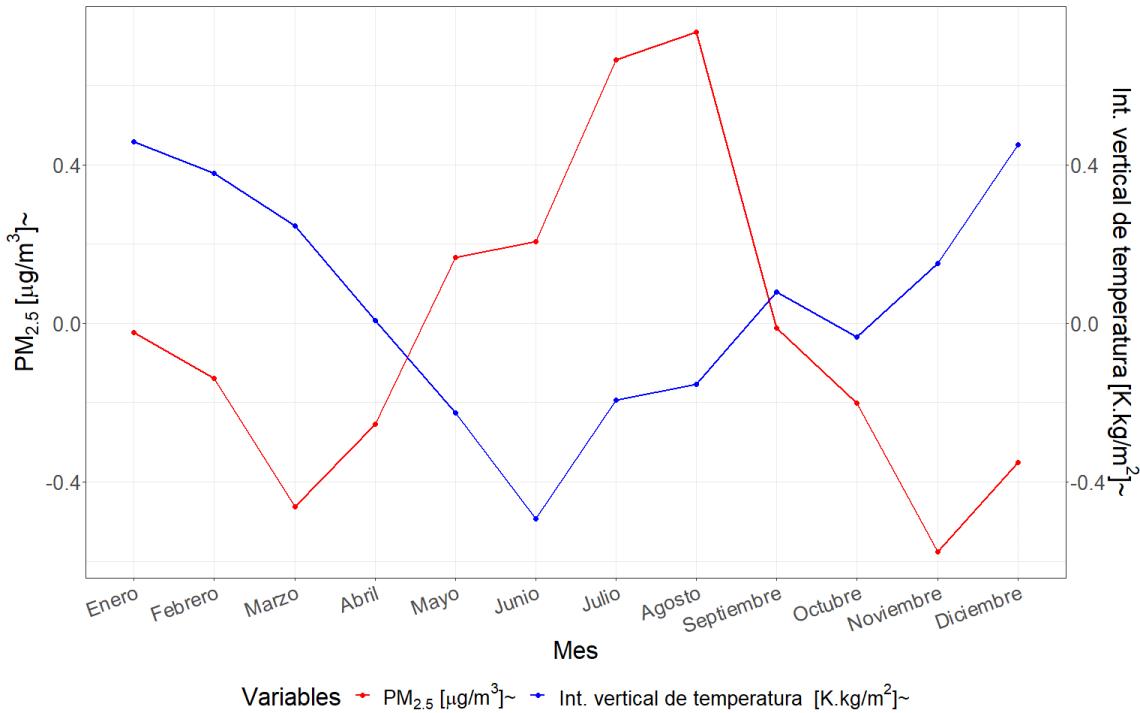


Figura 5.137: Perfil de variación conjunto de los promedios mensuales del $\text{PM}_{2.5}$ y la integración vertical de temperatura a lo largo del año.

En la Figura 5.137 puede visualizarse el patrón de variación conjunto que describen los promedios mensuales de la integración vertical de temperatura y la concentración de $\text{PM}_{2.5}$ a lo largo del año. Puede notarse que los valores de baja magnitud de la variable fueron registrados durante el invierno, específicamente en el mes de junio, mientras que los mayores durante el verano en los meses de diciembre y enero. Estos hallazgos sugieren una relación de vinculación entre las variables del tipo inversa, lo que implica que aumentos en la magnitud de la variable se asocian con disminuciones en las concentraciones de $\text{PM}_{2.5}$, mientras que disminuciones en la variable se vinculan con aumentos en las concentraciones del contaminante.

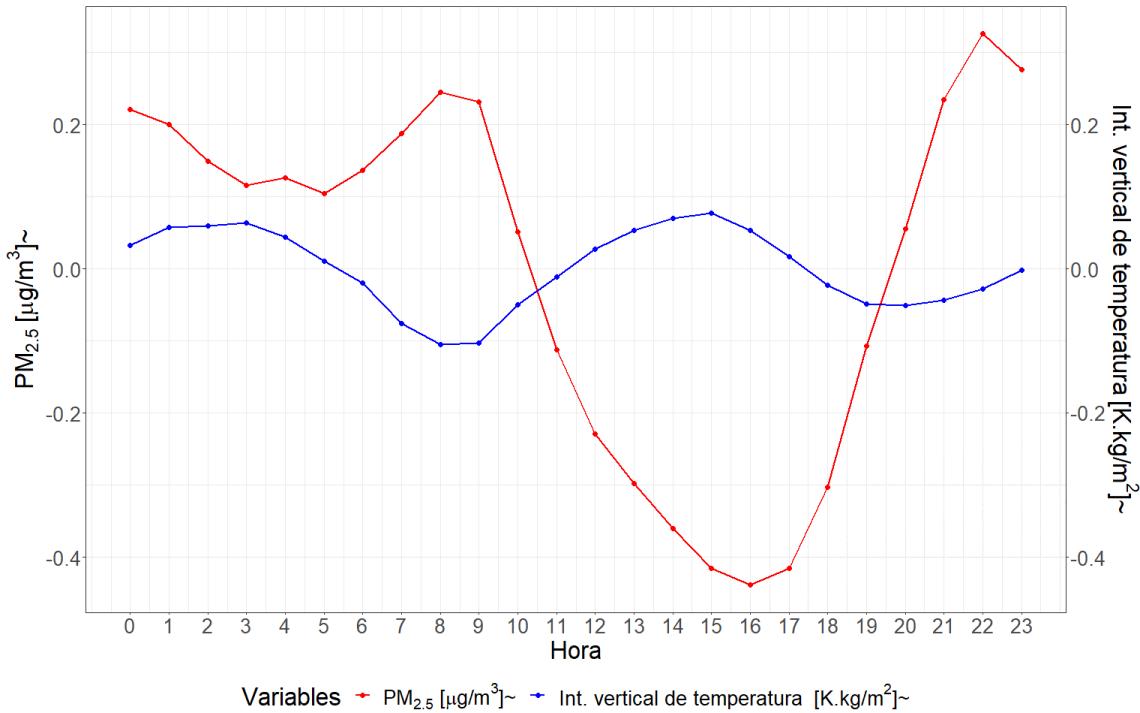


Figura 5.138: Perfil de variación conjunto de los promedios horarios del PM_{2.5} y la integración vertical de temperatura a lo largo del día.

A partir de la Figura 5.138 puede apreciarse el patrón de variación conjunto de la integración vertical de temperatura y la concentración de PM_{2.5} promedio a lo largo de las horas del día. Puede notarse que la temperatura presenta un perfil de variación opuesto al de la concentración de PM_{2.5}, evidenciando una relación de vinculación inversa. De esta forma, valores de baja magnitud de la variable se condicen con aumentos de la concentración del contaminante, mientras que valores de mayor magnitud de la variable ocurren en simultaneidad con disminuciones de la concentración de PM_{2.5}.

Tabla 5.44: Coeficientes estimados de la relación funcional.

Desfasaje	Con diferenciación	Sin diferenciación
t	7.654e-05	-3.442e-06
t-15	-1.411e-05	-3.438e-05

	Desfasaje	Con diferenciación	Sin diferenciación
t-16	8.115e-06	2.149e-05	
t-18	3.007e-05	6.352e-05	
t-21	1.269e-05	-6.070e-05	

Las correlaciones significativas que perduran luego del procedimiento de filtrado de las series se observaron en los tiempos t, t-15, t-16, t-18 y t-21. Luego de la estimación de la relación funcional se determinó que dos de los cinco coeficientes significativos se vieron influenciados por efectos aleatorios (t y t-21), demostrando que estas relaciones no son significativas en términos estadísticos. En contraste, de los tres restantes demostraron relaciones estadísticamente significativas, dos de ellos directa (t-16 y t-18) y solo uno inversa (t-15).

Referencias bibliográficas

- 10 Achad, M. et al. (2013). Retrieving the relative contribution of aerosol types from single particle analysis and radiation measurements and calculations: A comparison of two independent approaches. *Journal of Aerosol Science*, 64, 11–23.
- Aldas Manzano, J., & Uriel Jimenez, E. (2017). *Análisis multivariante aplicado con r*. Ediciones Paraninfo, SA.
- Al-Janabi, S., Mohammad, M., & Al-Sultan, A. (2020). A new method for prediction of air pollution based on intelligent computation. *Soft Computing*, 24(1), 661–680.
- Alvarado, S. A. et al. (2010). Modeling critical episodes of air pollution by PM10 in santiago, chile. Comparison of the predictive efficiency of parametric and non-parametric statistical models. *Gaceta Sanitaria*, 24(6), 466–472.
- Amarillo, A., Carreras, H. A., Krisna, T., Mignola, M., Busso, I. T., & Wendisch, M. (2021). Exploratory analysis of carbonaceous PM_{2.5} species in urban environments: Relationship with meteorological variables and satellite data. *Atmospheric Environment*, 245, 117987.
- Amegbor, P. M., Sabel, C. E., Mortensen, L. H., Mehta, A. J., & Rosenberg, M. W. (2023). Early-life air pollution and green space exposures as determinants of stunting among children under age five in sub-saharan africa. *Journal of Exposure Science & Environmental Epidemiology*, 1–15.
- Anjos, M., & Lopes, A. (2019). Sea breeze front identification on the northeastern coast of brazil and its implications for meteorological conditions in the sergipe region.

Theoretical and Applied Climatology, 137, 2151–2165.

Apparicio, P. et al. (2016). Cyclists' exposure to air pollution and road traffic noise in central city neighbourhoods of montreal. *Journal of Transport Geography*, 57, 63–69.

Asaly, S., Gottlieb, L.-A., Inbar, N., & Reuveni, Y. (2022). Using support vector machine (SVM) with GPS ionospheric TEC estimations to potentially predict earthquake events. *Remote Sensing*, 14(12), 2822.

Ayturan, Y. A., Ayturan, Z. C., ALTUN, H. O., Kongoli, C., Tuncez, F. D., DURSUN, S., & ÖZTÜRK, A. (2020). *Short-term prediction of PM2. 5 pollution with deep learning methods*.

Badura, M., Batog, P., Drzeniecka-Osiadacz, A., Modzel, P., et al. (2018). Evaluation of low-cost sensors for ambient PM 2.5 monitoring. *Journal of Sensors*, 2018.

Bakheet, S. et al. (2023). Hybrid bag-of-visual-words and FeatureWiz selection for content-based visual information retrieval. *Sensors*, 23(3), 1653.

Bell, B., Hersbach, H., Simmons, A., Berrisford, P., Dahlgren, P., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Radu, R., Schepers, D., et al. (2021). The ERA5 global reanalysis: Preliminary extension to 1950. *Quarterly Journal of the Royal Meteorological Society*, 147(741), 4186–4227.

Bisgaard, S. ren, & Kulahci, M. (2006). Quality quandaries: Studying input-output relationships, part II. *Quality Engineering*, 18(3), 405–410.

Bisgaard, S., & Kulahci, M. (2006). Quality quandaries: Studying input-output relationships, part i. *Quality Engineering*, 18(2), 273–281.

Bishop, C. M. et al. (1995). *Neural networks for pattern recognition*. Oxford university press.

Blanco Becerra, L. C. (2006). Caracterización microbiológica del material particulado como factor de riesgo sobre la salud en la localidad de puente aranda, bogotá DC colombia. *Acta Nova*, 3(2), 257–285.

- Bonilla, J. A., Lopez-Feldman, A., Pereda, P. C., Rivera, N. M., & Ruiz-Tagle, J. C. (2023). Association between long-term air pollution exposure and COVID-19 mortality in latin america. *PLoS One*, 18(1), e0280355.
- Bosilovich, M. G. et al. (2015). *MERRA-2: File specification*.
- Bottou, L., & Bousquet, O. (2007). The tradeoffs of large scale learning. *Advances in Neural Information Processing Systems*, 20.
- Box, G. E., Jenkins, G. M., & Reinsel, G. (1994). *Time series analysis, forecasting and control*. Englewood clifs. NJ: Prentice Hall.
- Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control*. John Wiley & Sons.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140.
- Breiman, L. (2001a). Random forests. *Machine Learning*, 45(1), 5–32.
- Breiman, L. (2001b). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, 16(3), 199–231.
- Busso, I. T. et al. (2021). Modeling air pollution-related hospital admissions employing remote sensing and geographical information systems. *Atmospheric Environment*, 261, 118502.
- Cai, J., Luo, J., Wang, S., & Yang, S. (2018). Feature selection in machine learning: A new perspective. *Neurocomputing*, 300, 70–79.
- Cao, X., Yi, J., Li, Y., Zhao, M., Duan, Y., Zhang, F., & Duan, L. (2024). Characteristics and source apportionment of volatile organic compounds in an industrial area at the zhejiang–shanghai boundary, china. *Atmosphere*, 15(2), 237.
- Carmen Bas, M. del, Ortiz, J., Ballesteros, L., & Martorell, S. (2017). Evaluation of a multiple linear regression model and SARIMA model in forecasting 7Be air concentrations. *Chemosphere*, 177, 326–333.
- Carreras, H. A. et al. (2006). In situ monitoring of urban air in córdoba, argentina using the tradescantia-micronucleus (trad-MCN) bioassay. *Atmospheric Environment*,

40(40), 7824–7830.

Carreras, H. A. et al. (2009). Assessment of human health risk related to metals by the use of biomonitoring in the province of Córdoba, Argentina. *Environmental Pollution*, 157(1), 117–122.

Carreras, H. A., & Pignata, M. (2001). Comparison among air pollutants, meteorological conditions and some chemical parameters in the transplanted lichen Usnea amblyoclada. *Environmental Pollution*, 111(1), 45–52.

Casallas, A., Ferro, C., Celis, N., Guevara-Luna, M. A., Mogollón-Sotelo, C., Guevara-Luna, F. A., & Merchán, M. (2021). Long short-term memory artificial neural network approach to forecast meteorology and PM 2.5 local variables in Bogotá, Colombia. *Modeling Earth Systems and Environment*, 1–14.

Chaloulakou, A. et al. (2003). Measurements of PM10 and PM2.5 particle concentrations in Athens, Greece. *Atmospheric Environment*, 37(5), 649–660.

Chatfield, C. et al. (1994). Time series prediction: Forecasting the future and understanding the past: Neil A. Gershenfeld and Andreas S. Weigend, 1994, 'The future of time series', in: AS Weigend and NA Gershenfeld, eds., (Addison-Wesley, Reading, MA), 1–70. *International Journal of Forecasting*, 10(1), 161–163.

Chatfield, C. (2000). *Time-series forecasting*. Chapman & Hall/CRC.

Chauhan, P. K., Kumar, A., Pratap, V., & Singh, A. K. (2022). Seasonal characteristics of PM1, PM2.5, and PM10 over Varanasi during 2019–2020. *Frontiers in Sustainable Cities*, 4, 909351.

Chen, G., Li, S., Knibbs, L. D., Hamm, N. A., Cao, W., Li, T., Guo, J., Ren, H., Abramson, M. J., & Guo, Y. (2018). A machine learning method to estimate PM2.5 concentrations across China with remote sensing, meteorological and land use information. *Science of the Total Environment*, 636, 52–60.

Chen, M., Dai, F., Yang, B., & Zhu, S. (2019). Effects of neighborhood green space on PM2.5 mitigation: Evidence from five megacities in China. *Building and*

Environment, 156, 33–45.

Chen, M.-H., Chen, Y.-C., Chou, T.-Y., & Ning, F.-S. (2023). PM2. 5 concentration prediction model: A CNN–RF ensemble framework. *International Journal of Environmental Research and Public Health*, 20(5), 4077.

Chen, N., Zhang, X.-Y., Ma, L.-L., Zhao, G.-D., & Yan, Y.-X. (2022). Trends of cancer mortality in xi'an city, china: 2005–2020. *Journal of Cancer Research and Clinical Oncology*, 148(10), 2781–2792.

Chen, R.-C., Dewi, C., Huang, S.-W., & Caraka, R. E. (2020). Selecting critical features for data classification based on machine learning methods. *Journal of Big Data*, 7(1), 52.

Chen, S.-L., Chang, S.-W., Chen, Y.-J., & Chen, H.-L. (2021). Possible warming effect of fine particulate matter in the atmosphere. *Communications Earth & Environment*, 2(1), 1–9.

Chen, Tao et al. (2016). Spatial and temporal variations of PM2. 5 and its relation to meteorological factors in the urban area of nanjing, china. *International Journal of Environmental Research and Public Health*, 13(9), 921.

Chen, Tianqi, & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–794.

Chen, Tianqi, He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., et al. (2015). Xgboost: Extreme gradient boosting. *R Package Version 0.4-2*, 1(4), 1–4.

Cho, K. et al. (2014). On the properties of neural machine translation: Encoder–decoder approaches. *arXiv Preprint arXiv:1409.1259*.

Christopher, S. A., & Gupta, P. (2010). Satellite remote sensing of particulate matter air quality: The cloud-cover problem. *Journal of the Air & Waste Management Association*, 60(5), 596–602.

- Chuvieco, E., & Hantson, S. (2010). Plan nacional de teledetección de media resolución procesamiento estándar de imágenes landsat documento técnico de algoritmos a aplicar. *Universidad de Alalá. España.*
- Cichowicz, R. et al. (2020). Effect of wind speed on the level of particulate matter PM10 concentration in atmospheric air during winter season in vicinity of large combustion plant. *Journal of Atmospheric Chemistry*, 77(1), 35–48.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273–297.
- Costa, L. G., Garrick, J. M., Dao, K., Phillips, A., Marsillach, J., & Cole, T. B. (2022). Traffic-related air pollution and the developing brain. In *Reproductive and developmental toxicology* (pp. 833–843). Elsevier.
- De Oliveira, T. C., Secolin, R., & Lopes-Cendes, I. (2023). A review of ancestry and admixture in latin america and the caribbean focusing on native american and african descendant populations. *Frontiers in Genetics*, 14, 1091269.
- Della Ceca, L. S. et al. (2018). *Análisis espacio-temporal de la distribución de aerosoles atmosféricos en la ciudad de córdoba (argentina) y desarrollo de un modelo predictivo de los niveles de material particulado* [Master's thesis].
- Deng, X., Chen, J., Dai, R., Zhai, Z., He, D., Zhao, L., Jin, X., & Zhang, J. (2023). The effects of planetary boundary layer features on air pollution based on ERA5 data in east china. *Atmosphere*, 14(8), 1273.
- Dhaka, S. K., Longiany, G., Panwar, V., Kumar, V., Malik, S., Singh, N., Dimri, A., Matsumi, Y., Nakayama, T., Hayashida, S., et al. (2022). Trends and variability of PM2. 5 at different time scales over delhi: Long-term analysis 2007-2021. *Aerosol and Air Quality Research*, 22, 220191.
- Dhakal, S. et al. (2021). Exploring a deep LSTM neural network to forecast daily PM 2.5 concentration using meteorological parameters in kathmandu valley, nepal. *Air Quality, Atmosphere & Health*, 14, 83–96.

- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a), 427–431.
- Dimitriou, K. et al. (2015). The dependence of PM size distribution from meteorology and local-regional contributions, in valencia (spain)-a CWT model approach. *Aerosol and Air Quality Research*, 15(5), 1979–1989.
- Ding, W., & Zhu, Y. (2022). Prediction of PM_{2.5} concentration in ningxia hui autonomous region based on PCA-attention-LSTM. *Atmosphere*, 13(9), 1444.
- Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (Vol. 326). John Wiley & Sons.
- Elminir, H. K. (2005). Dependence of urban air pollutants on meteorology. *Science of the Total Environment*, 350(1-3), 225–237.
- EPA. (2010). *Environmental protection agency 40 CFR parts 50 and 58*.
- Feng, F., & Wang, K. (2019). Does the modern-era retrospective analysis for research and applications-2 aerosol reanalysis introduce an improvement in the simulation of surface solar radiation over china? *International Journal of Climatology*, 39(3), 1305–1318.
- Feng, R., Zheng, H., Gao, H., Zhang, A., Huang, C., Zhang, J., Luo, K., & Fan, J. (2019). Recurrent neural network and random forest for analysis and accurate forecast of atmospheric pollutants: A case study in hangzhou, china. *Journal of Cleaner Production*, 231, 1005–1015.
- Finlayson-Pitts, B. J., & Pitts Jr, J. N. (1999). *Chemistry of the upper and lower atmosphere: Theory, experiments, and applications*. Elsevier.
- Fix, E. (1985). *Discriminatory analysis: Nonparametric discrimination, consistency properties* (Vol. 1). USAF school of Aviation Medicine.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 1189–1232.

- Fuenzalida, H., & Rosenblüth, B. (1990). Prewhitening of climatological time series. *Journal of Climate*, 3(3), 382–393.
- Fuzzi, S., Baltensperger, U., Carslaw, K., Decesari, S., Denier van der Gon, H., Facchini, M. C., Fowler, D., Koren, I., Langford, B., Lohmann, U., et al. (2015). Particulate matter, air quality and climate: Lessons learned and future needs. *Atmospheric Chemistry and Physics*, 15(14), 8217–8299.
- Gama, C., Monteiro, A., Pio, C., Miranda, A. I., Baldasano, J. M., & Tchepel, O. (2018). Temporal patterns and trends of particulate matter over portugal: A long-term analysis of background concentrations. *Air Quality, Atmosphere & Health*, 11, 397–407.
- Ganji, A., Minet, L., Weichenthal, S., & Hatzopoulou, M. (2020). Predicting traffic-related air pollution using feature extraction from built environment images. *Environmental Science & Technology*, 54(17), 10688–10699.
- Gao, X., Ruan, Z., Liu, J., Chen, Q., & Yuan, Y. (2022). Analysis of atmospheric pollutants and meteorological factors on PM_{2.5} concentration and temporal variations in harbin. *Atmosphere*, 13(9), 1426.
- Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., et al. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, 30(14), 5419–5454.
- Gers, F. A. et al. (2000). Learning to forget: Continual prediction with LSTM. *Neural Computation*, 12(10), 2451–2471.
- Gomišček, B., Hauck, H., Stopper, S., & Preining, O. (2004). Spatial and temporal variations of PM₁, PM_{2.5}, PM₁₀ and particle number concentration during the AUPHEP—project. *Atmospheric Environment*, 38(24), 3917–3934.
- Gopalakrishnan, R. et al. (2015). The bipolar diffusion charging of nanoparticles: A review and development of approaches for non-spherical particles. *Aerosol Science and Technology*, 49(12), 1181–1194.

- Grossi, C., Àgueda, A., Vogel, F., Vargas, A., Zimnoch, M., Wach, P., Martíñ, J., López-Coto, I., Bolívar, J., Morgui, J. A., et al. (2016). Analysis of ground-based ^{222}Rn measurements over spain: Filling the gap in southwestern europe. *Journal of Geophysical Research: Atmospheres*, 121(18), 11–021.
- Gu, K., Liu, H., Xia, Z., Qiao, J., Lin, W., & Thalmann, D. (2021). PM_{2.5} monitoring: Use information abundance measurement and wide and deep learning. *IEEE Transactions on Neural Networks and Learning Systems*, 32(10), 4278–4290.
- Guerrero, V. M., & Perera, R. (2004). Variance stabilizing power transformation for time series. *Journal of Modern Applied Statistical Methods*, 3(2), 9.
- Gündoğdu, S., & Elbir, T. (2024). A data-driven approach for PM_{2.5} estimation in a metropolis: Random forest modeling based on ERA5 reanalysis data. *Environmental Research Communications*, 6(3), 035029.
- Guo, D., Wang, R., & Zhao, P. (2020). Spatial distribution and source contributions of PM_{2.5} concentrations in jincheng, china. *Atmospheric Pollution Research*, 11(8), 1281–1289.
- Gupta, A. et al. (2021). Spatio-temporal distribution of INSAT-3D AOD derived particulate matter concentration over india. *Atmospheric Pollution Research*, 12(1), 159–172.
- Gupta, P., Zhan, S., Mishra, V., Aekakkararungroj, A., Markert, A., Paibong, S., Chishtie, F., et al. (2021). Machine learning algorithm for estimating surface PM_{2.5} in thailand. *Aerosol and Air Quality Research*, 21(11), 210105.
- Hahad, O., Daiber, A., & Münzel, T. (2024). Breathing danger: Linking air pollution to cardiovascular disease and increased risk of abdominal aortic aneurysm. In *European Heart Journal* (p. ehae092). Oxford University Press US.
- Hamilton, J. D. (2020). *Time series analysis*. Princeton university press.
- Handschrub, J., Erbertseder, T., & Baier, F. (2023). Systematic evaluation of four satellite AOD datasets for estimating PM_{2.5} using a random forest approach. *Remote*

- Sensing*, 15(8), 2064.
- Haykin, S. (1999). *Neural networks and learning machines*. Pearson Prentice Hall.
- He, Jianhua, Ding, S., & Liu, D. (2019). Exploring the spatiotemporal pattern of PM2. 5 distribution and its determinants in chinese cities based on a multilevel analysis approach. *Science of the Total Environment*, 659, 1513–1525.
- He, Jianjun, Gong, S., Liu, H., An, X., Yu, Y., Zhao, S., Wu, L., Song, C., Zhou, C., Wang, J., et al. (2017). Influences of meteorological conditions on interannual variations of particulate matter pollution during winter in the beijing–tianjin–hebei area. *Journal of Meteorological Research*, 31(6), 1062–1069.
- He, Q., Wang, M., & Yim, S. H. L. (2021). The spatiotemporal relationship between PM 2.5 and aerosol optical depth in china: Influencing factors and implications for satellite PM 2.5 estimations using MAIAC aerosol optical depth. *Atmospheric Chemistry and Physics*, 21(24), 18375–18391.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Hodan, W. M., & Barnard, W. R. (2004). Evaluating the contribution of PM2. 5 precursor gases and re-entrained road emissions to mobile source PM2. 5 particulate matter emissions. *MACTEC Federal Programs, Research Triangle Park, NC*.
- Hodson, T. O. (2022). Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development Discussions*, 2022, 1–10.
- Hofman, J., Staelens, J., Cordell, R., Stroobants, C., Zikova, N., Hama, S., Wyche, K., Kos, G., Van Der Zee, S., Smallbone, K., et al. (2016). Ultrafine particles in four european urban environments: Results from a new continuous long-term monitoring

network. *Atmospheric Environment*, 136, 68–81.

Hooyberghs, J. et al. (2005). A neural network forecast for daily average PM10 concentrations in belgium. *Atmospheric Environment*, 39(18), 3279–3289.

Htun, H. H., Biehl, M., & Petkov, N. (2023). Survey of feature selection and extraction techniques for stock market prediction. *Financial Innovation*, 9(1), 26.

Hu, A., Huang, Y., Chen, Q., Huang, W., Wu, X., Cui, L., Dong, Y., & Liu, J. (2023). Glass transition of amorphous polymeric materials informed by machine learning. *APL Machine Learning*, 1(2).

Hu, W., Zhao, T., Bai, Y., Kong, S., Xiong, J., Sun, X., Yang, Q., Gu, Y., & Lu, H. (2021). Importance of regional PM2. 5 transport and precipitation washout in heavy air pollution in the twain-hu basin over central china: Observational analysis and WRF-chem simulation. *Science of the Total Environment*, 758, 143710.

Huang, K., Xiao, Q., Meng, X., Geng, G., Wang, Y., Lyapustin, A., Gu, D., & Liu, Y. (2018). Predicting monthly high-resolution PM2. 5 concentrations with random forest model in the north china plain. *Environmental Pollution*, 242, 675–683.

Hulley, G., Veraverbeke, S., & Hook, S. (2014). Thermal-based techniques for land cover change detection using a new dynamic MODIS multispectral emissivity product (MOD21). *Remote Sensing of Environment*, 140, 755–765.

Huot, F., Hu, R. L., Goyal, N., Sankar, T., Ihme, M., & Chen, Y.-F. (2022). Next day wildfire spread: A machine learning dataset to predict wildfire spreading from remote-sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–13.

IARC. (2004). IARC monographs on the evaluation of carcinogenic risks to humans. *IARC, Lyon*, 1, 1–1452.

IARC. (2010). Some non-heterocyclic polycyclic aromatic hydrocarbons and some related exposures. Working group on the evaluation of carcinogenic risks to humans and others. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*, 92, 1.

- Imane, S., Oumaima, B., Kenza, K., Laila, I., Youssef, E. M., Zineb, S., & Mohamed, E. J. (2022). A review on climate, air pollution, and health in north africa. *Current Environmental Health Reports*, 9(2), 276–298.
- INDEC, A. (2010). Censo nacional de población. *Hogares y Viviendas*.
- Jain, S., Sharma, S., Vijayan, N., & Mandal, T. (2020). Seasonal characteristics of aerosols (PM_{2.5} and PM₁₀) and their source apportionment using PMF: A four year study over delhi, india. *Environmental Pollution*, 262, 114337.
- Jamei, M., Ali, M., Karimi, B., Karbasi, M., Farooque, A. A., & Yaseen, Z. M. (2023). Surface water electrical conductivity and bicarbonate ion determination using a smart hybridization of optimal boruta package with elman recurrent neural network. *Process Safety and Environmental Protection*, 174, 115–134.
- Jayaratne, R., Liu, X., Thai, P., Dunbabin, M., & Morawska, L. (2018). The influence of humidity on the performance of a low-cost air particle mass sensor and the effect of atmospheric fog. *Atmospheric Measurement Techniques*, 11(8), 4883–4890.
- Jena, C., Ghude, S. D., Kumar, R., Debnath, S., Govardhan, G., Soni, V. K., Kulkarni, S. H., Beig, G., Nanjundiah, R. S., & Rajeevan, M. (2021). Performance of high resolution (400 m) PM_{2.5} forecast over delhi. *Scientific Reports*, 11(1), 4104.
- Jiang, N., Dong, Z., Xu, Y., Yu, F., Yin, S., Zhang, R., Tang, X., et al. (2018). Characterization of PM₁₀ and PM_{2.5} source profiles of fugitive dust in zhengzhou, china. *Aerosol and Air Quality Research*, 18(2), 314–329.
- Jiang, T., Chen, B., Nie, Z., Ren, Z., Xu, B., & Tang, S. (2021). Estimation of hourly full-coverage PM_{2.5} concentrations at 1-km resolution in china using a two-stage random forest model. *Atmospheric Research*, 248, 105146.
- Jiménez-Guerrero, P., Jorba, O., Baldasano, J. M., & Gassó, S. (2008). The use of a modelling system as a tool for air quality management: Annual high-resolution simulations and evaluation. *Science of the Total Environment*, 390(2-3), 323–340.
- Justice, C., Giglio, L., Korontzi, S., Owens, J., Morisette, J., Roy, D., Descloitres,

- J., Alleaume, S., Petitcolin, F., & Kaufman, Y. (2002). The MODIS fire products. *Remote Sensing of Environment*, 83(1-2), 244–262.
- Kalashnikov, D. A., Schnell, J. L., Abatzoglou, J. T., Swain, D. L., & Singh, D. (2022). Increasing co-occurrence of fine particulate matter and ground-level ozone extremes in the western united states. *Science Advances*, 8(1), eabi9386.
- Kang, G. K., Gao, J. Z., Chiao, S., Lu, S., & Xie, G. (2018). Air quality prediction: Big data and machine learning approaches. *Int. J. Environ. Sci. Dev*, 9(1), 8–16.
- Khalid, S., Khalil, T., & Nasreen, S. (2014). A survey of feature selection and feature extraction techniques in machine learning. *2014 Science and Information Conference*, 372–378.
- Khan, M. F. et al. (2010). Characterization of PM_{2.5}, PM_{2.5–10} and PM_{>10} in ambient air, yokohama, japan. *Atmospheric Research*, 96(1), 159–172.
- Kim, K.-H. et al. (2015). A review on the human health impact of airborne particulate matter. *Environment International*, 74, 136–143.
- Kim, Y., Park, S.-B., Lee, S., & Park, Y.-K. (2023). Comparison of PM_{2.5} prediction performance of the three deep learning models: A case study of seoul, daejeon, and busan. *Journal of Industrial and Engineering Chemistry*, 120, 159–169.
- Kleine Deters, J., Zalakeviciute, R., Gonzalez, M., Rybarczyk, Y., et al. (2017). Modeling PM 2.5 urban pollution using machine learning and selected meteorological parameters. *Journal of Electrical and Computer Engineering*, 2017.
- Kulkarni, P. et al. (2011). *Aerosol measurement: Principles, techniques, and applications*. John Wiley & Sons.
- Kumar, A., Ambade, B., Sankar, T. K., Sethi, S. S., & Kurwadkar, S. (2020). Source identification and health risk assessment of atmospheric PM_{2.5}-bound polycyclic aromatic hydrocarbons in jamshedpur, india. *Sustainable Cities and Society*, 52, 101801.
- Kumar, P. et al. (2011). Dynamics and dispersion modelling of nanoparticles from

road traffic in the urban atmospheric environment—a review. *Journal of Aerosol Science*, 42(9), 580–603.

Kursa, M. B., Jankowski, A., & Rudnicki, W. R. (2010). Boruta—a system for feature selection. *Fundamenta Informaticae*, 101(4), 271–285.

Kurtz, M., Lezon, C., Boyer, P., & Tasat, D. (2022). *Malnutrition and air pollution in latin america: Impact of two stressors on children's health*.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1-3), 159–178.

Lalchandani, V., Srivastava, D., Dave, J., Mishra, S., Tripathi, N., Shukla, A., Sahu, R., Thamban, N., Gaddamidi, S., Dixit, K., et al. (2022). Effect of biomass burning on PM_{2.5} composition and secondary aerosol formation during post-monsoon and winter haze episodes in delhi. *Journal of Geophysical Research: Atmospheres*, 127(1), e2021JD035232.

Lehmann, E. L., & Casella, G. (2006). *Theory of point estimation*. Springer Science & Business Media.

Li, D. et al. (2020). Correlation study of planetary-boundary-layer-height retrievals from cl51 and chm15k ceilometers with application to pm_{2.5} dynamics in new york city. *EPJ Web of Conferences*, 237, 03010.

Li, H., Guo, B., Han, M., Tian, M., Zhang, J., et al. (2015). Particulate matters pollution characteristic and the correlation between PM (PM 2.5, PM 10) and meteorological factors during the summer in shijiazhuang. *Journal of Environmental Protection*, 6(05), 457.

Li, Y., Li, C., Li, M., & Liu, Z. (2019). Influence of variable selection and forest type on forest aboveground biomass estimation using machine learning algorithms. *Forests*, 10(12), 1073.

Lim, C.-H., Ryu, J., Choi, Y., Jeon, S. W., & Lee, W.-K. (2020). Understan-

ding global PM_{2.5} concentrations and their drivers in recent decades (1998–2016). *Environment International*, 144, 106011.

Lima, A. N., Philot, E. A., Trossini, G. H. G., Scott, L. P. B., Maltarollo, V. G., & Honorio, K. M. (2016). Use of machine learning approaches for novel drug discovery. *Expert Opinion on Drug Discovery*, 11(3), 225–239.

Liu, Q., Wu, R., Zhang, W., Li, W., & Wang, S. (2020). The varying driving forces of PM_{2.5} concentrations in chinese cities: Insights from a geographically and temporally weighted regression model. *Environment International*, 145, 106168.

Liu, Z., Hu, B., Wang, L., Wu, F., Gao, W., & Wang, Y. (2015). Seasonal and diurnal variation in particulate matter (PM 10 and PM 2.5) at an urban site of beijing: Analyses from a 9-year study. *Environmental Science and Pollution Research*, 22, 627–642.

López, M. L., Ceppi, S., Palancar, G. G., Olcese, L. E., Tirao, G., & Toselli, B. M. (2011). Elemental concentration and source identification of PM10 and PM_{2.5} by SR-XRF in córdoba city, argentina. *Atmospheric Environment*, 45(31), 5450–5457.

Lubo-Robles, D., Devegowda, D., Jayaram, V., Bedle, H., Marfurt, K. J., & Pranter, M. J. (2020). Machine learning model interpretability using SHAP values: Application to a seismic facies classification task. *SEG International Exposition and Annual Meeting*.

Lyapustin, Alexei et al. (2018). MODIS collection 6 MAIAC algorithm. *Atmospheric Measurement Techniques*, 11(10), 5741–5765.

Lyapustin, Alexei et al. (2019). MAIAC thermal technique for smoke injection height from MODIS. *IEEE Geoscience and Remote Sensing Letters*, 17(5), 730–734.

Lyapustin, A., Korkin, S., Wang, Y., Quayle, B., & Laszlo, I. (2012). Discrimination of biomass burning smoke and clouds in MAIAC algorithm. *Atmospheric Chemistry and Physics*, 12(20), 9679–9686.

Lyapustin, Alexei, Wang, Y., Laszlo, I., Kahn, R., Korkin, S., Remer, L., Levy, R.,

& Reid, J. (2011). Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm. *Journal of Geophysical Research: Atmospheres*, 116(D3).

Ma, H., Cao, X., Ma, X., Su, H., Jing, Y., & Zhu, K. (2022). Improving the wind power density forecast in the middle-and high-latitude regions of china by selecting the relatively optimal planetary boundary layer schemes. *Atmosphere*, 13(12), 2034.

Ma, Z., Luo, W., Jiang, J., Wang, B., Ma, Z., Lin, J., & Liu, D. (2023). Spatial and temporal characteristics analysis and prediction model of PM2. 5 concentration based on SpatioTemporal-informer model. *Plos One*, 18(6), e0287423.

Maior, C. S., & Silva, T. (2024). Time-series failure prediction on small datasets using machine learning. *IEEE Latin America Transactions*, 22(5), 362–371.

Martin, P. B. (2005). *Contaminación del aire por material particulado en la ciudad de buenos aires* [PhD thesis]. Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales.

Mateos, A. C. et al. (2018). Land use and air quality in urban environments: Human health risk assessment due to inhalation of airborne particles. *Environmental Research*, 161, 370–380.

Mateos, A. C. et al. (2019). Influence of meteorological variables and forest fires events on air quality in an urban area (córdoba, argentina). *Archives of Environmental Contamination and Toxicology*, 77(2), 171–179.

McFarlane, C., Isevulambire, P. K., Lumbuenamo, R. S., Ndinga, A. M. E., Dhammapala, R., Jin, X., McNeill, V. F., Malings, C., Subramanian, R., Westervelt, D. M., et al. (2021). First measurements of ambient PM2. 5 in kinshasa, democratic republic of congo and brazzaville, republic of congo using field-calibrated low-cost sensors. *Aerosol and Air Quality Research*, 21(7), 200619.

Meszaros, E. et al. (1999). *Fundamentals of atmospheric aerosol chemistry*. Akadémiai kiado.

Mhawish, A., Banerjee, T., Sorek-Hamer, M., Lyapustin, A., Broday, D. M., &

- Chatfield, R. (2019). Comparison and evaluation of MODIS multi-angle implementation of atmospheric correction (MAIAC) aerosol product over south asia. *Remote Sensing of Environment*, 224, 12–28.
- Miranda, R. M. de, Fatima Andrade, M. de, Fornaro, A., Astolfo, R., Andre, P. A. de, & Saldiva, P. (2012). Urban air pollution: A representative survey of PM2. 5 mass concentrations in six brazilian cities. *Air Quality, Atmosphere & Health*, 5(1), 63–77.
- Mlambo, C., Ngonisa, P., Ntshangase, B., Ndlovu, N., & Mvuyana, B. (2023). Air pollution and health in africa: The burden falls on children. *Economies*, 11(7), 196.
- Moran, M. D., Lupu, A., Zhang, J., Savic-Jovcic, V., & Gravel, S. (2018). A comprehensive performance evaluation of the next generation of the canadian operational regional air quality deterministic prediction system. *Air Pollution Modeling and Its Application XXV* 35, 75–81.
- Mostafi, S., & Elgazzar, K. (2021). An open source tool to extract traffic data from google maps: Limitations and challenges. *2021 International Symposium on Networks, Computers and Communications (ISNCC)*, 1–8.
- Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., et al. (2021). ERA5-land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13(9), 4349–4383.
- Murtagh, F., & Legendre, P. (2014). Ward’s hierarchical agglomerative clustering method: Which algorithms implement ward’s criterion? *Journal of Classification*, 31(3), 274–295.
- Muthukumar, P., Nagrecha, K., Comer, D., Calvert, C. F., Amini, N., Holm, J., & Pourhomayoun, M. (2022). PM2. 5 air pollution prediction through deep learning using multisource meteorological, wildfire, and heat data. *Atmosphere*, 13(5), 822.
- Noble, C. A., Vanderpool, R. W., Peters, T. M., McElroy, F. F., Gemmill, D. B., & Wiener, R. W. (2001). Federal reference and equivalent methods for measuring fine

particulate matter. *Aerosol Science & Technology*, 34(5), 457–464.

Nogarotto, D. C., & Pozza, S. A. (2020). A review of multivariate analysis: Is there a relationship between airborne particulate matter and meteorological variables? *Environmental Monitoring and Assessment*, 192(9), 1–14.

Nordberg, G. F. et al. (2014). *Handbook on the toxicology of metals*. Academic press.

Novaes Neto, J. A. (2022). *Modelo preditivo de capacidade de pagamento para prospecção PF: Atraindo e fidelizando clientes no cenário de open finance*.

Olcese, L. E., & Toselli, B. M. (1997). Effects of meteorology and land use on ambient measurements of primary pollutants in cordoba city, argentina. *Meteorology and Atmospheric Physics*, 62(3), 241–248.

Olcese, L. E., & Toselli, B. M. (2002). Some aspects of air pollution in córdoba, argentina. *Atmospheric Environment*, 36(2), 299–306.

Omar, N., & Syed-Fadzil, S. (2011). Assessment of passive thermal performance for a penang heritage shop house. *Procedia Engineering*, 20, 203–212.

OMS. (2021). *WHO global air quality guidelines: Particulate matter (PM_{2.5} and PM₁₀), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. World Health Organization.

Palmer, T., & Hagedorn, R. (2006). *Predictability of weather and climate*. Cambridge University Press.

Pant, P., & Harrison, R. M. (2013). Estimation of the contribution of road traffic emissions to particulate matter concentrations from field measurements: A review. *Atmospheric Environment*, 77, 78–97.

Paparrizos, J., & Gravano, L. (2015). K-shape: Efficient and accurate clustering of time series. *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, 1855–1870.

Paul, S., Coops, N., Johnson, M., Krzic, M., Chandna, A., & Smukler, S. (2020).

Mapping soil organic carbon and clay using remote sensing to predict soil workability for enhanced climate change adaptation. *Geoderma*, 363, 114177.

Pearce, J. L., Beringer, J., Nicholls, N., Hyndman, R. J., & Tapper, N. J. (2011). Quantifying the influence of local meteorology on air quality using generalized additive models. *Atmospheric Environment*, 45(6), 1328–1336.

Peña, D. (2005). *Análisis de series temporales*. Alianza.

Pengcheng, L. et al. (2020). Prediction of compressive strength of high-performance concrete by random forest algorithm. *IOP Conference Series: Earth and Environmental Science*, 552, 012020.

Peters, J., Van den Bossche, J., Reggente, M., Van Poppel, M., De Baets, B., & Theunis, J. (2014). Cyclist exposure to UFP and BC on urban routes in antwerp, belgium. *Atmospheric Environment*, 92, 31–43.

Pey Betrán, J. (2008). *Caracterización fisico-química de los aerosoles atmosféricos en el mediterraneo occidental*. Universitat Politècnica de Catalunya.

Pontius, R. G. et al. (2008). Components of information for multiple resolution comparison between maps that share a real variable. *Environmental and Ecological Statistics*, 15(2), 111–142.

Probst, W. N., Stelzenmüller, V., & Fock, H. O. (2012). Using cross-correlations to assess the relationship between time-lagged pressure and state indicators: An exemplary analysis of north sea fish population indicators. *ICES Journal of Marine Science*, 69(4), 670–681.

Puliafito, E. (2009). Gestión de la calidad del aire en argentina. *Contaminación Atmosférica En Argentina Contribuciones de La II Reunión Anual PROIMCA*, 67.

Qu, Y., Zhao, M., Wang, T., Li, S., Li, M., Xie, M., & Zhuang, B. (2022). Lidar-and UAV-based vertical observation of spring ozone and particulate matter in nanjing, china. *Remote Sensing*, 14(13), 3051.

R Core Team. (2021). *R: A language and environment for statistical computing*. R

Foundation for Statistical Computing. <https://www.R-project.org/>

Radovic, M. et al. (2017). Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. *BMC Bioinformatics*, 18(1), 1–14.

Rai, P. K. (2016). Impacts of particulate matter pollution on plants: Implications for environmental biomonitoring. *Ecotoxicology and Environmental Safety*, 129, 120–136.

Rai, P. K., & Panda, L. L. (2014). Leaf dust deposition and its impact on biochemical aspect of some roadside plants of aizawl, mizoram, north east india. *International Research Journal of Environment Sciences*, 3(11), 14–19.

Rasheed, A., Aneja, V. P., Aiyyer, A., Rafique, U., et al. (2015). Measurement and analysis of fine particulate matter (PM2. 5) in urban areas of pakistan. *Aerosol and Air Quality Research*, 15(2), 426–439.

Razavi, S., & Vogel, R. (2018). Prewhitening of hydroclimatic time series? Implications for inferred change and variability across time scales. *Journal of Hydrology*, 557, 109–115.

Rocchini, D., Andreo, V., Förster, M., Garzon-Lopez, C. X., Gutierrez, A. P., Gillespie, T. W., Hauffe, H. C., He, K. S., Kleinschmit, B., Mairota, P., et al. (2015). Potential of remote sensing to predict species invasions: A modelling perspective. *Progress in Physical Geography*, 39(3), 283–309.

Rojano, R. E., Angulo, L. C., & Restrepo, G. (2013). Niveles de partículas suspendidas totales (PST), PM10 y PM2. 5 y su relación en lugares públicos de la ciudad riohacha, caribe colombiano. *Información Tecnológica*, 24(2), 37–46.

Rumao, S. (2019). *Exploration of variable importance and variable selection techniques in presence of correlated variables*. Rochester Institute of Technology.

Rumelhart, D. E. et al. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536.

Saaroni, H., Pearlmutter, D., & Hatuka, T. (2015). Human-biometeorological

conditions and thermal perception in a mediterranean coastal park. *International Journal of Biometeorology*, 59, 1347–1362.

Sacks, J. D., Stanek, L. W., Luben, T. J., Johns, D. O., Buckley, B. J., Brown, J. S., & Ross, M. (2011). Particulate matter-induced health effects: Who is susceptible? *Environmental Health Perspectives*, 119(4), 446–454.

Salameh, D., Detournay, A., Pey, J., Pérez, N., Liguori, F., Saraga, D., Bove, M. C., Brotto, P., Cassola, F., Massabò, D., et al. (2015). PM_{2.5} chemical composition in five european mediterranean cities: A 1-year study. *Atmospheric Research*, 155, 102–117.

Samek, L., Stegowski, Z., Styszko, K., Furman, L., Zimnoch, M., Skiba, A., Kistler, M., Kasper-Giebl, A., Rozanski, K., & Konduracka, E. (2020). Seasonal variations of chemical composition of PM_{2.5} fraction in the urban area of krakow, poland: PMF source attribution. *Air Quality, Atmosphere & Health*, 13(1), 89–96.

Santosa, F., & Symes, W. W. (1986). Linear inversion of band-limited reflection seismograms. *SIAM Journal on Scientific and Statistical Computing*, 7(4), 1307–1330.

Sarda-Espinosa, A., Sarda, M. A., & LazyData, T. (2018). Package “dtwclust.” *Pobrane z: Http://Cran. Ma. Imperial. Ac. Uk/Web/Packages/Dtwclust/Dtwclust. Pdf*.

Saunders, C., Gammerman, A., & Vovk, V. (1998). *Ridge regression learning algorithm in dual variables*.

Savtchenko, A., Ouzounov, D., Ahmad, S., Acker, J., Leptoukh, G., Koziana, J., & Nickless, D. (2004). Terra and aqua MODIS products available from NASA GES DAAC. *Advances in Space Research*, 34(4), 710–714.

Sax, S. N., Koutrakis, P., Ruiz Rudolph, P. A., Cereceda-Balic, F., Gramsch, E., & Oyola, P. (2007). Trends in the elemental composition of fine particulate matter in santiago, chile, from 1998 to 2003. *Journal of the Air & Waste Management Association*, 57(7), 845–855.

- Saxena, M., Sharma, A., Sen, A., Saxena, P., Mandal, T., Sharma, S., Sharma, C., et al. (2017). Water soluble inorganic species of PM10 and PM2. 5 at an urban site of delhi, india: Seasonal variability and sources. *Atmospheric Research*, 184, 112–125.
- Sayahi, T., Butterfield, A., & Kelly, K. (2019). Long-term field evaluation of the plantower PMS low-cost particulate matter sensors. *Environmental Pollution*, 245, 932–940.
- Schlesinger, R. B. (2007). The health impact of common inorganic components of fine particulate matter (PM2. 5) in ambient air: A critical review. *Inhalation Toxicology*, 19(10), 811–832.
- Scotto, M. G. (1998). Introduction to time series and forecasting. *Journal of Applied Statistics*, 25(5), 717.
- Seinfeld, J. H., & Pandis, S. N. (1998). From air pollution to climate change. *Atmospheric Chemistry and Physics*, 1326.
- Serio, L., Puccetti, C., & Oderigo, J. (2020). Revisión de los monitoreos de calidad del aire en la ciudad de buenos aires. *Agronomía & Ambiente*, 40(1).
- Sfetsos, A., & Vlachogiannis, D. (2010). A new approach to discovering the causal relationship between meteorological patterns and PM10 exceedances. *Atmospheric Research*, 98(2-4), 500–511.
- Solomon, S., Qin, D., Manning, M., Averyt, K., & Marquis, M. (2007). *Climate change 2007-the physical science basis: Working group i contribution to the fourth assessment report of the IPCC* (Vol. 4). Cambridge university press.
- Soni, M. et al. (2018). Particulate matter estimation over a semi arid region jaipur, india using satellite AOD and meteorological parameters. *Atmospheric Pollution Research*, 9(5), 949–958.
- Stanojević, G. B., Miljanović, D. N., Doljak, D. L., Ćurčić, N. B., Radovanović, M. M., Malinović-Milićević, S. B., & Hauriak, O. (2019). Spatio-temporal variability of annual PM2. 5 concentrations and population exposure assessment in serbia for the

period 2001–2016. *Journal of the Geographical Institute “Jovan Cvijić” SASA*, 69(3), 197–211.

Stein, A. F., & Toselli, B. M. (1996). Street level air pollution in Córdoba city, Argentina. *Atmospheric Environment*, 30(20), 3491–3495.

Stewart, I. D., & Oke, T. R. (2012). Local climate zones for urban temperature studies. *Bulletin of the American Meteorological Society*, 93(12), 1879–1900.

Storey, M. A., & Price, O. F. (2022). Prediction of air quality in Sydney, Australia as a function of forest fire load and weather using Bayesian statistics. *Plos One*, 17(8), e0272774.

Suleiman, A. et al. (2020). A comparative study of using random forests (RF), extreme learning machine (ELM) and deep learning (DL) algorithms in modelling roadside particulate matter (PM10 & PM2.5). *IOP Conference Series: Earth and Environmental Science*, 476, 012126.

Sun, X., Zhao, T., Liu, D., Gong, S., Xu, J., & Ma, X. (2020). Quantifying the influences of PM2.5 and relative humidity on change of atmospheric visibility over recent winters in an urban area of East China. *Atmosphere*, 11(5), 461.

Suparwito, H., Polina, A. M., Budiraharjo, M., et al. (2021). Student perceptions analysis of online learning: A machine learning approach. *Indonesian Journal of Information Systems*, 4(1), 64–75.

Tai, A. P. et al. (2010). Correlations between fine particulate matter (PM2.5) and meteorological variables in the United States: Implications for the sensitivity of PM2.5 to climate change. *Atmospheric Environment*, 44(32), 3976–3984.

Tao, M., Wang, J., Li, R., Wang, L., Wang, L., Wang, Z., Tao, J., Che, H., & Chen, L. (2019). Performance of MODIS high-resolution MAIAC aerosol algorithm in China: Characterization and limitation. *Atmospheric Environment*, 213, 159–169.

Thao, N. N. L., Pimonsree, S., Prueksakorn, K., Thao, P. T. B., & Vongruang, P. (2022). Public health and economic impact assessment of PM2.5 from open

biomass burning over countries in mainland southeast asia during the smog episode. *Atmospheric Pollution Research*, 13(6), 101418.

Thompson, W. H., & Fransson, P. (2016). On stabilizing the variance of dynamic functional brain connectivity time series. *Brain Connectivity*, 6(10), 735–746.

Thunis, P. et al. (2011). A procedure for air quality models benchmarking. *Joint Research Centre (JRC)*. Ispra, Italy.

Tian, G., Qiao, Z., & Xu, X. (2014). Characteristics of particulate matter (PM10) and its relationship with meteorological factors during 2001–2012 in beijing. *Environmental Pollution*, 192, 266–274.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.

Tiwari, S., Dumka, U., Gautam, A., Kaskaoutis, D., Srivastava, A., Bisht, D., Chakrabarty, R., Sumlin, B., & Solmon, F. (2017). Assessment of PM2. 5 and PM10 over guwahati in brahmaputra river valley: Temporal evolution, source apportionment and meteorological dependence. *Atmospheric Pollution Research*, 8(1), 13–28.

Tiwari, S., Srivastava, A. K., Bisht, D. S., Parmita, P., Srivastava, M. K., & Attri, S. (2013). Diurnal and seasonal variations of black carbon and PM2. 5 over new delhi, india: Influence of meteorology. *Atmospheric Research*, 125, 50–62.

Unkašević, M., Vukmirović, Z., Tošić, I., & Lazić, L. (2003). Effects of uncontrolled particulate matter release on precipitation under warfare conditions. *Environmental Science and Pollution Research*, 10(2), 89–97.

Usman, F., Zeb, B., Alam, K., Valipour, M., Ditta, A., Sorooshian, A., Roy, R., Ahmad, I., & Iqbal, R. (2022). Exploring the mass concentration of particulate matter and its relationship with meteorological parameters in the hindu-kush range. *Atmosphere*, 13(10), 1628.

Valappil, N. K. M., Hamza, V., & Oliveira Júnior, J. F. de. (2023). Evaluation of daily average temperature trends in kerala, india, using MERRA-2 reanalysis data: A

climate change perspective. *Environmental Science and Pollution Research*, 30(10), 26663–26686.

Van den Bossche, J., Peters, J., Verwaeren, J., Botteldooren, D., Theunis, J., & De Baets, B. (2015). Mobile monitoring for mapping spatial variation in urban air quality: Development and validation of a methodology based on an extensive dataset. *Atmospheric Environment*, 105, 148–161.

Van Poppel, M. et al. (2013). Methodology for setup and data processing of mobile air quality measurements to assess the spatial variability of concentrations in urban environments. *Environmental Pollution*, 183, 224–233.

VandenBoer, T. C., Brown, S. S., Murphy, J. G., Keene, W. C., Young, C. J., Pszenny, A., Kim, S., Warneke, C., Gouw, J. A. de, Maben, J. R., et al. (2013). Understanding the role of the ground surface in HONO vertical structure: High resolution vertical profiles during NACHTT-11. *Journal of Geophysical Research: Atmospheres*, 118(17), 10–155.

Venkatesh, B., & Anuradha, J. (2019). A review of feature selection and its methods. *Cybernetics and Information Technologies*, 19(1), 3–26.

Vinoj, V., & Pandey, S. K. (2022). Role of meteorology in atmospheric aerosols and air pollution over south asia. In *Asian atmospheric pollution* (pp. 97–110). Elsevier.

Wang, C., Jia, M., Xia, H., Wu, Y., Wei, T., Shang, X., Yang, C., Xue, X., & Dou, X. (2019). Relationship analysis of PM 2.5 and boundary layer height using an aerosol and turbulence detection lidar. *Atmospheric Measurement Techniques*, 12(6), 3303–3315.

Wang, M., & Wang, H. (2021). Spatial distribution patterns and influencing factors of pm 2.5 pollution in the yangtze river delta: Empirical analysis based on a gwr model. *Asia-Pacific Journal of Atmospheric Sciences*, 57, 63–75.

Wang, Y. et al. (2021). Using recursive feature selection with random forest to improve protein structural class prediction for low-similarity sequences. *Computational*

and Mathematical Methods in Medicine, 2021.

Watmough, G. R., Marcinko, C. L., Sullivan, C., Tschirhart, K., Mutuo, P. K., Palm, C. A., & Svenning, J.-C. (2019). Socioecologically informed use of remote sensing data to predict rural household poverty. *Proceedings of the National Academy of Sciences*, 116(4), 1213–1218.

Weerakkody, U. et al. (2018). Quantification of the traffic-generated particulate matter capture by plant species in a living wall and evaluation of the important leaf characteristics. *Science of the Total Environment*, 635, 1012–1024.

Westergaard, G., Erden, U., Mateo, O. A., Lampo, S. M., Akinci, T. C., & Topsakal, O. (2024). Time series forecasting utilizing automated machine learning (AutoML): A comparative analysis study on diverse datasets. *Information*, 15(1), 39.

WHO et al. (2021). *WHO global air quality guidelines: Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. World Health Organization.

Willmott, C. J., & Matsuura, K. (2006). On the use of dimensioned measures of error to evaluate the performance of spatial interpolators. *International Journal of Geographical Information Science*, 20(1), 89–102.

Wright, R., Flynn, L., Garbeil, H., Harris, A., & Pilger, E. (2002). Automated volcanic eruption detection using MODIS. *Remote Sensing of Environment*, 82(1), 135–155.

Wu, B., Zhu, W., Yan, N., Xing, Q., Xu, J., Ma, Z., & Wang, L. (2020). Regional actual evapotranspiration estimation with land and meteorological variables derived from multi-source satellite data. *Remote Sensing*, 12(2), 332.

Wu, H., Cao, Y., Wang, T., You, J., Yang, T., Qu, Y., Yan, S., Yang, H., Mu, X., Gao, L., et al. (2024). How does PM2. 5 impact the urban vertical temperature structure? A case study in nanjing. *Aerosol and Air Quality Research*, 24(4), 230214.

Wu, J., Yao, F., Li, W., & Si, M. (2016). VIIRS-based remote sensing estimation

of ground-level PM_{2.5} concentrations in beijing–tianjin–hebei: A spatiotemporal statistical model. *Remote Sensing of Environment*, 184, 316–328.

Wu, Y.-J. et al. (2007). A google-map-based arterial traffic information system. *2007 IEEE Intelligent Transportation Systems Conference*, 968–973.

XiaoYang, L., HongQuan, S., ShiYan, Z., SiQi, L., YunFeng, K., HaoMing, X., HaiPeng, Z., et al. (2019). Particulate matter pollution in chinese cities: Areal-temporal variations and their relationships with meteorological conditions (2015–2017). *Environmental Pollution*, 246, 11–18.

Xing, Y., Xu, Y., & Shi, M. (2016). LY (2016). The impact of PM_{2.5} on the human respiratory system. *Journal of Thoracic Disease*, 8(1), E69–74.

Xu, Y. et al. (2021). Influence of AOD remotely sensed products, meteorological parameters, and AOD–PM 2.5 models on the PM 2.5 estimation. *Stochastic Environmental Research and Risk Assessment*, 35, 893–908.

Yan, J.-W., Tao, F., Zhang, S.-Q., Lin, S., & Zhou, T. (2021). Spatiotemporal distribution characteristics and driving forces of PM_{2.5} in three urban agglomerations of the yangtze river economic belt. *International Journal of Environmental Research and Public Health*, 18(5), 2222.

Yang, Q., Yuan, Q., Li, T., Shen, H., & Zhang, L. (2017). The relationships between PM_{2.5} and meteorological factors in china: Seasonal and regional variations. *International Journal of Environmental Research and Public Health*, 14(12), 1510.

Yang, X., Zhao, C., Guo, J., & Wang, Y. (2016). Intensification of aerosol pollution associated with its feedback with surface solar radiation and winds in beijing. *Journal of Geophysical Research: Atmospheres*, 121(8), 4093–4099.

Yang, Z., Mahendran, R., Yu, P., Xu, R., Yu, W., Godellawattage, S., Li, S., & Guo, Y. (2022). Health effects of long-term exposure to ambient PM_{2.5} in asia-pacific: A systematic review of cohort studies. *Current Environmental Health Reports*, 9(2), 130–151.

- Zhang, H., Chen, G., Hu, J., Chen, S.-H., Wiedinmyer, C., Kleeman, M., & Ying, Q. (2014). Evaluation of a seven-year air quality simulation using the weather research and forecasting (WRF)/community multiscale air quality (CMAQ) models in the eastern united states. *Science of the Total Environment*, 473, 275–285.
- Zhang, T., Zang, L., Mao, F., Wan, Y., & Zhu, Y. (2020). Evaluation of himawari-8/AHI, MERRA-2, and CAMS aerosol products over china. *Remote Sensing*, 12(10), 1684.
- Zhao, H., Liu, Y., Gu, T., Zheng, H., Wang, Z., & Yang, D. (2022). Identifying spatiotemporal heterogeneity of PM_{2.5} concentrations and the key influencing factors in the middle and lower reaches of the yellow river. *Remote Sensing*, 14(11), 2643.
- Zhao, Yujie et al. (2022). Analysis on the change law and influencing factors of atmospheric particulate matter concentration in xiamen city. *International Conference on Statistics, Applied Mathematics, and Computing Science (CSAMCS 2021)*, 12163, 232–237.
- Zhao, Yan, Cai, J., Zhu, X., Donkelaar, A. van, Martin, R. V., Hua, J., & Kan, H. (2020). Fine particulate matter exposure and renal function: A population-based study among pregnant women in china. *Environment International*, 141, 105805.
- Zhong, J., Zhang, X., Gui, K., Wang, Y., Che, H., Shen, X., Zhang, L., Zhang, Y., Sun, J., & Zhang, W. (2021). Robust prediction of hourly PM_{2.5} from meteorological data using LightGBM. *National Science Review*, 8(10), nwaa307.
- Zhou, H., Dai, Z., Wu, C., Ma, X., Zhu, L., & Wu, P. (2024). Comparison of different impact factors and spatial scales in PM_{2.5} variation. *Atmosphere*, 15(3), 307.
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320.