

ESTRATEGIAS DE MODELADO PREDICTIVO DE CONCENTRACIÓN DE MATERIAL PARTICULADO FINO ($PM_{2.5}$) A PARTIR DE SERIES TEMPORALES

por

Rodriguez Nuñez Martin

Presentado ante la Facultad de Ciencias Económicas (FCE), la Facultad de Ciencias Agropecuarias (FCA) y la Facultad de Matemática, Astronomía, Física y

Computación (FAMAF)

para obtener el grado de

Magíster en Estadística Aplicada

de la

Universidad Nacional de Córdoba

Junio 2023

Directora: Monica Balzarini



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

Agradecimientos

Llega el tan ansiado día de defender mi tesis de maestría, luego de muchos años de estudio y trabajo sin duda hay miles de agradecimientos que mencionar. En primer lugar, me siento bendecido y afortunado de haber nacido en la República Argentina. Le agradezco a mi país no solo por haberme brindado educación gratuita de excelencia, si no además por la beca con la cual pude dedicarme exclusivamente a mi investigación, período durante el cual realice mi maestría.

En segundo lugar, creo que el mayor agradecimiento se lo debo a la Dra. Mónica Balzarini, mi directora, quien me acogió como un hijo cuando estaba huérfano en el ámbito académico, brindándome lugar de trabajo y compañía a lo largo de mi investigación de doctorado y tesis de maestría. Y si Mónica fue como mi madre, no puedo olvidarme de mis hermanos, los de la cátedra de estadística y biometría por haberme guiado en este trayecto, con sus comentarios y sugerencias siendo un pilar fundamental para mi investigación -sin contar el gran aporte de amistad, mates, cerveceadas y salidas-.

Obviamente, este trabajo no hubiera sido posible sin momentos de esparcimiento de quien escribe. Aquí es donde debo principalmente agradecer a toda mi familia: Ester, má y Jose, pá, siempre me apoyaron en lo que me hiciera feliz, acompañándome en todo momento y brindándome la posibilidad de estar cómodo en casa estudiando y trabajando. A mi hermano Juan Manuel quien sin dedicarse a la investigación siempre me apoyo acompañándome de cerca y preocupándose por mi salud mental. Y en especial a mi hermano Juan Cruz quien gracias a su notable trayectoria en el

ámbito de la investigación y la ciencia de datos siempre me incentivó a continuar y capacitarme aún más. A mis abuelitos que en paz descansen, y a todos mis familiares que confiaron en mí. Agradezco por que cada miembro me mostró el camino de la perseverancia y de no bajar los brazos, algo que inconscientemente siempre estuvo en mi cabeza a lo largo de los años.

Por otra parte quiero agradecer de igual forma a mis grupos de amigos, la sub 19, los outliers, los del cole y los de la vida, grupos bien distintos, pero que siempre estuvieron allí brindándome toneladas de cariño, siendo también pilares fundamentales a lo largo de los años, ayudándome a solucionar mis problemas y permitiéndome distender con mates, juntadas y giras cuando era necesario. No quiero largarme a nombrarlos de a uno porque no terminaría más, pero cada uno sabe cuánto le agradezco. Además, quería agradecer a una personita muy especial que hoy no está en mi vida pero me acompaña casi todos los años de investigación, motivándome a lograr lo que quería, siendo un gran apoyo anímico y emocional.

Gracias, gracias a todos, les dedico este trabajo, y esta parte de mi vida.

Resumen

La Organización Mundial de la Salud (OMS) estima que nueve de cada diez personas respira aire con altos niveles de contaminante y como resultado siete millones mueren cada año en el mundo por causas relacionadas. La contaminación atmosférica con material particulado fino ($PM_{2.5}$) resulta de relevancia dado sus efectos sobre la salud, directamente vinculados con desórdenes respiratorios, cardiovasculares y circulatorios. La concentración de $PM_{2.5}$ se monitorea temporalmente a frecuencia horaria y la variabilidad de la serie temporal resultante depende de numerosas variables relativas a la condición atmosférica, siendo varias de ellas de público acceso en la nube. Sin embargo, el desarrollo de modelos predictivos tanto lineales como no lineales es limitado. La disponibilidad de un modelo de alto alcance predictivo permitiría construir un sistema de alerta temprana ante eventos nocivos de calidad de aire.

El objeto principal de esta tesis fue comparar metodologías de aprendizaje automático y profundo para predecir series temporales de concentración de $PM_{2.5}$ a partir de variables meteorológicas derivadas de satélites. Se dispusieron sensores de bajo costo para medir esta concentración en ocho sitios con dos usos de suelo diferentes (agrícola/área verde y urbano) ubicados dentro del área metropolitana de la provincia de Córdoba, Argentina. Se obtuvieron series temporales de frecuencia horaria a lo largo de periodos de al menos un año. Se descargaron 164 variables de la plataforma ERA5 y 14, citadas en la literatura relacionada a $PM_{2.5}$, desde los satélites MODIS, MERRA y VIIRS. Se ajustaron, para cada uso de suelo, modelos de regresión lineal, bosques aleatorios y redes neuronales recurrentes, para predecir la concentración de

PM_{2.5}. Además, los modelos predictivos se ajustaron implementando distintas estrategias de selección de variables, base de datos original (completa), reducida por el algoritmo LASSO y por el algoritmo featurewiz. La serie de concentración de PM_{2.5} en el uso de suelo agrícola / área verde tuvo menos variabilidad, destacó menor cantidad de variables de relevancia para la predicción y el menor error cuadrático medio de predicción se obtuvo con el modelo de bosques aleatorios y LASSO. Por el contrario para la serie urbana el algoritmo que demostró mayor capacidad predictiva fue el de redes neuronales recurrentes y la base de datos completa. Para ambas series las predicciones horarias de la concentración de PM_{2.5} esperadas en un horizonte futuro de cinco días mostraron un error de predicción menor al 27%.

Se observó que series temporales con características de gran variabilidad y fluctuaciones entre sus valores son más aptas para ser modeladas a través de redes neuronales recurrentes, antes que bosques aleatorios o modelos lineales. Mientras que el algoritmo de bosques aleatorios mostró un mejor desempeño bajo condiciones de menor variabilidad.

Palabras claves: aprendizaje automático, aprendizaje profundo, selección de características.

Índice general

Capítulo 1: Introducción general	1
1.1. El problema	1
1.2. Motivación	5
1.3. Objetivo general	7
1.4. Objetivos específicos	7
1.5. Resumen gráfico del protocolo analítico	8
Capítulo 2: Variación espacio temporal de la concentración de aerosoles	9
2.1. Introducción	9
2.2. Materiales y métodos	11
2.2.1. Área de estudio	11
2.2.2. Preprocesamiento de datos	18
2.2.3. Análisis estadístico	21
2.3. Resultados y discusión	21
2.4. Conclusión	41
Capítulo 3: Algoritmos para la predicción de la concentración de PM_{2.5}	43
3.1. Introducción	43
3.2. Materiales y métodos	46
3.2.1. Base de datos	46
3.2.2. Protocolo estadístico	48

3.3. Resultados	63
3.3.1. Comparación de algoritmos para el uso del suelo urbano	63
3.3.2. uso del suelo agrícola y área verde	71
3.4. Discusión	78
3.5. Conclusión	82
Comentarios finales	84
Futuras líneas de investigación	86
Anexo 1: Descripción de variables	88
Anexo 2: Bases de datos	194
3.6. Base de datos original	194
3.7. Base de datos reducida	197
3.7.1. Featurewiz para el uso del suelo urbano	197
3.7.2. LASSO para el uso del suelo urbano	199
3.7.3. Featurewiz para el uso del suelo agrícola y área verde	200
3.7.4. LASSO para el uso del suelo agrícola y área verde	202
Anexo 3: Pseudocódigo	204
3.7.5. Análisis exploratorio	204
3.7.6. Modelado predictivo	210
Referencias bibliográficas	221

Capítulo 1

Introducción general

1.1. El problema

La Organización Mundial de la Salud (OMS) estima que nueve de cada diez personas respira aire con altos niveles de contaminantes y como resultado siete millones mueren cada año en el mundo por causas relacionadas (OMS, 2020). La contaminación atmosférica es considerada la forma de polución de mayor relevancia y preocupación dado sus efectos sobre la salud, directamente vinculados con desórdenes respiratorios, cardiovasculares y circulatorios, de acuerdo a la Agencia Ambiental Europea (EEA, 2020) e informes de la OMS que evidencian los impactos de este tipo de contaminación (OMS, 2021a).

Las pautas de calidad del aire publicadas por la OMS (OMS, 2006) establecen los niveles límite de exposición a contaminantes recomendados para preservar la salud pública y garantizar buena calidad de aire. Evitar la exposición a contaminantes atmosféricos por encima de estos límites es importante para preservar la salud de los individuos, en especial para aquellos que son susceptibles dado a la preexistencia de deficiencias cardiovasculares y respiratorias. A pesar de que la calidad del aire ha mejorado desde el Tratado de Aire Limpio desarrollado en 1970 (Rogers, 1990),

sigue siendo sumamente problemática en múltiples ciudades del mundo, lo que lleva a autoridades locales a implementar medidas para reducir la contaminación y educar a la comunidad sobre sus efectos nocivos. Además de su impacto en la salud se destaca el económico asociado, en 2015 la Organización para la Cooperación y el Desarrollo Económico (OCDE) estimó que el impacto económico de los efectos en la salud como consecuencia de la contaminación del aire es de alrededor de 3 trillones de dólares (OECD et al., 2016).

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), conjunto de materia en estado sólido y/o líquido presente en suspensión en la atmósfera, exceptuando 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 su diferencia semántica que demuestra que no son equivalentes, estos términos son utilizados indistintamente. Sus propiedades aerodinámicas engloban las características físicas de mayor relevancia, estas determinan el tipo de transporte y remoción de la atmósfera, además de regular su deposición en el sistema respiratorio y permitir la identificación de la fuente que le dio origen. Su propiedad principal es su diámetro aerodinámico, diámetro de una esfera de densidad unitaria ($1 \frac{kg}{cm^3}$) con las mismas características aerodinámicas que la partícula bajo consideración. Según este pueden clasificarse tres grandes grupos: gruesas (PM_{10} : 2,5-10 μm de diámetro), finas ($PM_{2,5}$: menor a 2,5 μm de diámetro) y ultrafinas ($PM_{0,1}$: menor a 0,1 μm de diámetro) (Gopalakrishnan et al., 2015; Kulkarni et al., 2011).

Según informes de la Organización Mundial de la Salud, el PM es uno de los contaminantes atmosféricos más nocivos para la salud, tanto en países desarrollados como en vías de desarrollo (WHO, 2005). La vía de ingreso de las partículas al cuerpo

humano es por medio del sistema respiratorio. La repercusión en la salud humana depende de su tamaño, componentes adsorbidos, composición química, intensidad de la exposición, duración de la exposición y condición sanitaria del individuo expuesto. A medida que el tamaño disminuye, su potencial impacto sobre la salud 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 (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 el carácter cancerígeno y mutagénico de la partícula, los mayores referentes en cuanto a su capacidad de generar adversidades a la salud como consecuencia de su composición química son los hidrocarburos aromáticos policíclicos (IARC, 2004, 2010). Finalmente, los efectos 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 parte de 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).

Solo países desarrollados tienen la capacidad de generar estándares y normativas de seguimiento y control del PM, un país cuyo actuar es ejemplar 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). A pesar de ser uno de los países que marca tendencia en lo que es el control y gestión de la calidad del aire, este establece protocolos y normativas que responden a la realidad de un país desarrollado económicamente,

además de desarrollar tecnologías cuya accesibilidad requiere elevados presupuestos. Estos estándares representan metas inalcanzables para países en vías de desarrollo, motivo por el cual en los mismos se hace necesario migrar hacia alternativas de bajo costo para la gestión y control de la calidad del aire.

El monitoreo de la calidad del aire es una herramienta fundamental a la hora de identificar y evaluar problemas de contaminación (OMS, 2021b). Junto con modelos predictivos e inventarios de emisión, forman parte integral de los sistemas de gestión y control de la calidad del aire. La predicción efectiva de un fenómeno de contaminación atmosférica es una poderosa herramienta de análisis, con un amplio espectro de aplicación que va desde la evaluación de estrategias de control de emisiones, el análisis de impacto en la calidad del aire, el apoyo en la toma de decisiones de carácter ambiental, hasta la generación de información científica valiosa para comprender mejor la dinámica de la atmósfera y la contaminación en una región. A pesar de esto, su propósito primordial recae en la provisión de información a los científicos, legisladores y planificadores, con el objetivo de fomentar la toma de decisiones para la gestión, mejora del ambiente y preservación de la salud pública.

Para mitigar las nocivas influencias en la salud de la población, es necesario contar con información de variables que permitan describir y predecir la variabilidad del contaminante. Dada la notable influencia de variables meteorológicas y derivadas de información satelital en la variabilidad del PM, es que se recurre a éstas para definir el estado de la atmósfera en un momento determinado (He et al., 2017; Tiwari et al., 2017), con el objetivo de predecir la concentración del contaminante. Además, existen estudios que publican la concentración de PM como más dependiente de la meteorología que de factores antropogénicos (Hooyberghs et al., 2005).

Para contar con estos datos se puede recurrir a la teledetección o sensado remoto, técnica que permite adquirir información mediante instrumentos que no están en contacto físico con los objetos investigados. Los sensores remotos miden la radiación

electromagnética que es reflejada o emitida por éstos. Generalmente, se ubican a bordo de aviones o satélites que orbitan la Tierra (Chuvieco & Hantson, 2010). A partir de la información relevada se logran estimaciones horarias de un gran número de variables climáticas, atmosféricas, terrestres y oceánicas, que llegan a cubrir la totalidad de la superficie terrestre. De esta forma, a partir de los productos satelitales podemos contar con información acerca del estado del sistema Tierra-atmósfera con alta 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, denominado 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). En esta tesis esos datos son usados para la generación de estimaciones futuras (predicciones) de la concentración de PM_{2.5}.

1.2. Motivación

Estudios previos realizados en la ciudad de Córdoba han destacado que se convive con elevadas concentraciones de PM, especialmente en la zona céntrica de la ciudad. La principal fuente de emisión de PM en esta área ha sido vinculada al sector del transporte (Amarillo et al., 2021; Carreras et al., 2006; Carreras & Pignata, 2001). Además, se ha establecido una asociación clara entre los niveles de PM y el incremento de los riesgos de enfermedades respiratorias en la población cordobesa (Busso et al., 2021; Carreras et al., 2009; Mateos et al., 2018). Asimismo, se ha investigado la relación entre los niveles de PM y variables meteorológicas y derivadas de información satelital para el sitio de estudio (Amarillo et al., 2021; Della Ceca et al., 2018).

Sin embargo, ninguno de estos estudios ha abordado la variable respuesta con una frecuencia de variación horaria, ni han seguido un enfoque predictivo que permita anticipar la concentración futura de PM en función de otras covariables. Dado que las condiciones meteorológicas, topográficas y las actividades antropogénicas en una ubicación específica pueden influir en los patrones de emisión de PM, se justifica plenamente la realización de estudios de monitoreo *in situ*. Hasta la fecha, en la ciudad de Córdoba, se han realizado investigaciones que han demostrado los impactos en la salud de este contaminante, así como su relación con covariables y la identificación de su fuente de emisión en función de su composición (Amarillo et al., 2021; Busso et al., 2021; Mateos et al., 2018, 2019). Sin embargo, no se ha llevado a cabo un estudio en el país con el propósito de desarrollar un modelo predictivo de contaminación del aire, con miras a establecer un sistema de alerta temprana ante condiciones adversas de calidad del aire, utilizando técnicas de inteligencia artificial para la concentración horaria de PM_{2.5}. Abarcar este vacío de conocimiento tiene el potencial de dar lugar a un sistema de alerta que permita anticipar condiciones adversas de calidad del aire, lo que a su vez fortalecerá la salud pública y contribuirá a una mejor calidad de vida para la población. Además, la disponibilidad de nuevos datos ayudará a abordar la falta de acceso a información sobre la contaminación del aire en la provincia. La falta de sistemas de monitoreo de referencia en países con recursos limitados ha dado como resultado la ausencia de un control efectivo de la calidad del aire. La creación de redes de vigilancia de la calidad del aire basadas en equipos de referencia certificados es un privilegio de los países económicamente desarrollados debido a los altos costos de adquisición y mantenimiento. Esto representa una barrera significativa para los países en vías de desarrollo, donde la gestión y el control de la calidad del aire se consideran una prioridad secundaria debido a estas limitaciones financieras.

1.3. Objetivo general

Generar modelos predictivos de la concentración de PM_{2.5} sobre sitios con distinto uso del suelo en el área metropolitana de Córdoba.

1.4. Objetivos específicos

1. Estudiar las series temporales de concentración de PM_{2.5} emitido en distintos sitios de muestreo del área metropolitana de la ciudad de Córdoba.
2. Obtener variables satelitales de acceso público potencialmente correlacionadas con la concentración de PM_{2.5}.
3. Construir algoritmos predictivos que aprendan de la variabilidad de las series temporales de concentración de PM_{2.5} según el uso del suelo, que permitan predecir concentraciones en un horizonte de corto plazo.
4. Comparar la capacidad predictiva de algoritmos basados en distintos sub conjuntos de variables satelitales para distintos escenarios respecto al uso del suelo.

1.5. Resumen gráfico del protocolo analítico

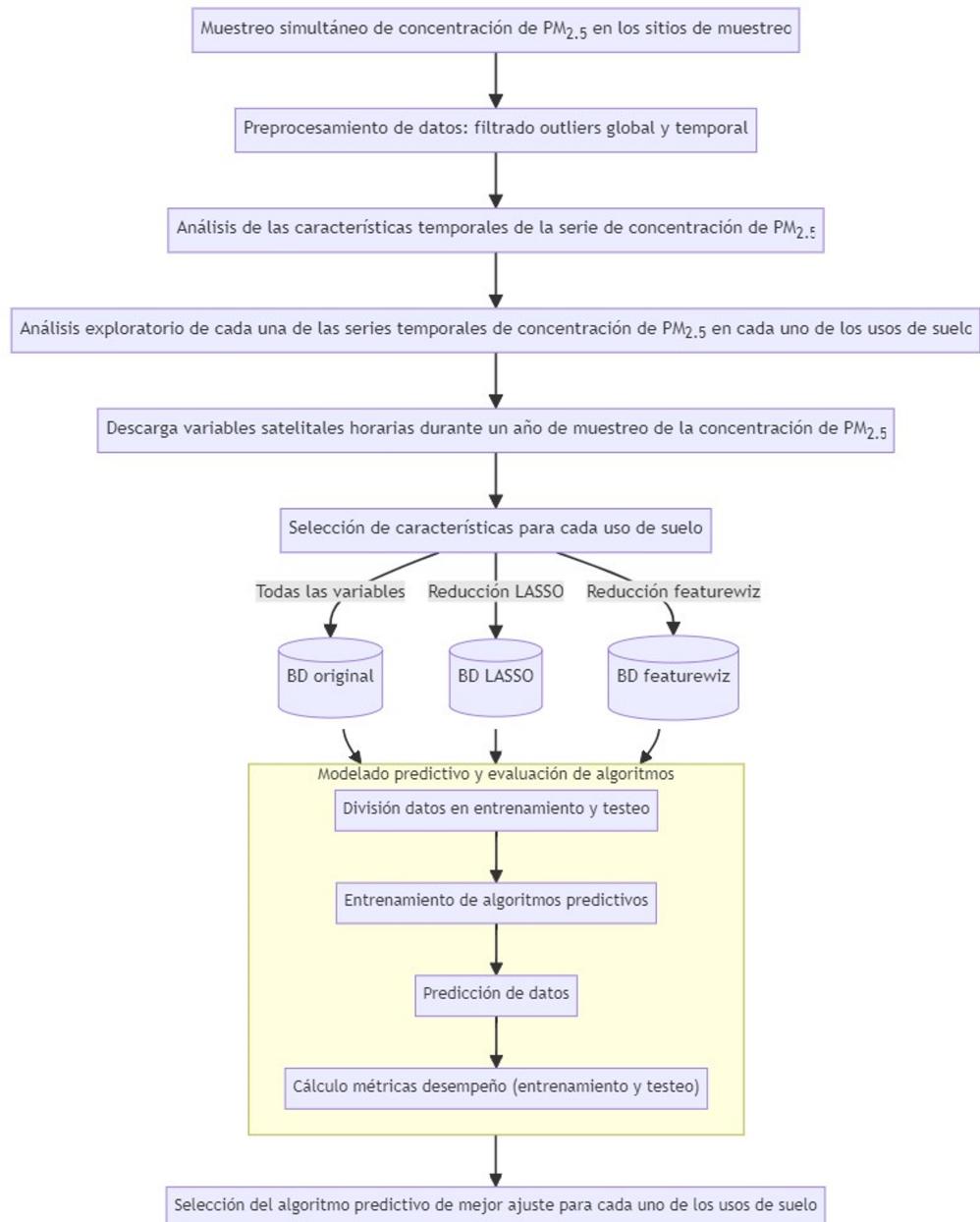


Figura 1.1: Resumen gráfico del protocolo analítico desarrollado desde la descarga de datos hasta el desarrollo y selección de modelos predictivos.

Capítulo 2

Variación espacio temporal de la concentración de aerosoles

2.1. Introducción

Dentro de los contaminantes más preocupantes en entornos urbanos se hallan las partículas en suspensión, las cuales son principalmente producidas por vehículos, industrias y reacciones fotoquímicas (Seinfeld & Pandis, 1998). Esta característica debe su origen a la afección que generan tanto en el medio ambiente, como en la salud humana. A pesar de esto, actualmente la provincia de Córdoba no cuenta con un sistema de monitoreo que informe a la población acerca de las condiciones ambientales con respecto a este contaminante. Sin embargo, gracias a la investigación científica se dispone de numerosos estudios en esta ciudad, los cuales además de informar acerca de las concentraciones, variación y composición, también lo hacen respecto de los impactos en la salud humana (Amarillo et al., 2021; Busso et al., 2021; Mateos et al., 2018, 2019).

En la provincia de Córdoba, los niveles guía de partículas están regulados en la Resolución 105/17, la cual establece el límite de concentración para el PM_{2.5} de 35

$\frac{\mu\text{g}}{\text{m}^3}$ para la media de 24 horas. Este límite es similar al que emplea la Agencia de Protección Ambiental (EPA – EEUU), basándose en la Norma Nacional de Calidad del Aire Ambiente (del inglés, NAAQS) (EPA, 2010). Por otra parte, los estándares que fija la OMS son más restrictivos y es importante tenerlos presentes al analizar concentración de PM_{2.5} (OMS, 2021b).

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 los distintos entornos (Hofman et al., 2016; 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, estos resultados fueron hallados en ciudades de Austria (Gomišček et al., 2004), Grecia (Chaloulakou et al., 2003), entre otras partes del mundo. Similar es lo que ocurre en Córdoba, donde 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). 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 (He et al., 2017; Tiwari et al., 2017). 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).

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 objetivo que se persigue en este capítulo es el de determinar cómo es la variación, tanto espacial como temporal, de la concentración de PM_{2.5} en distintas zonas del área metropolitana de Córdoba. De esta forma, en el presente capítulo se describen las series disponibles y se identifican sus patrones de variación, con el fin de contemplar características propias de las series al momento de desarrollar el modelado predictivo (capítulo 3).

2.2. Materiales y métodos

2.2.1. 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 re suspensión 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 metalmecánica, alimentarias, metalúrgicas, entre otras. El crecimiento de la flota 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 ocho sitios seleccionados 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.

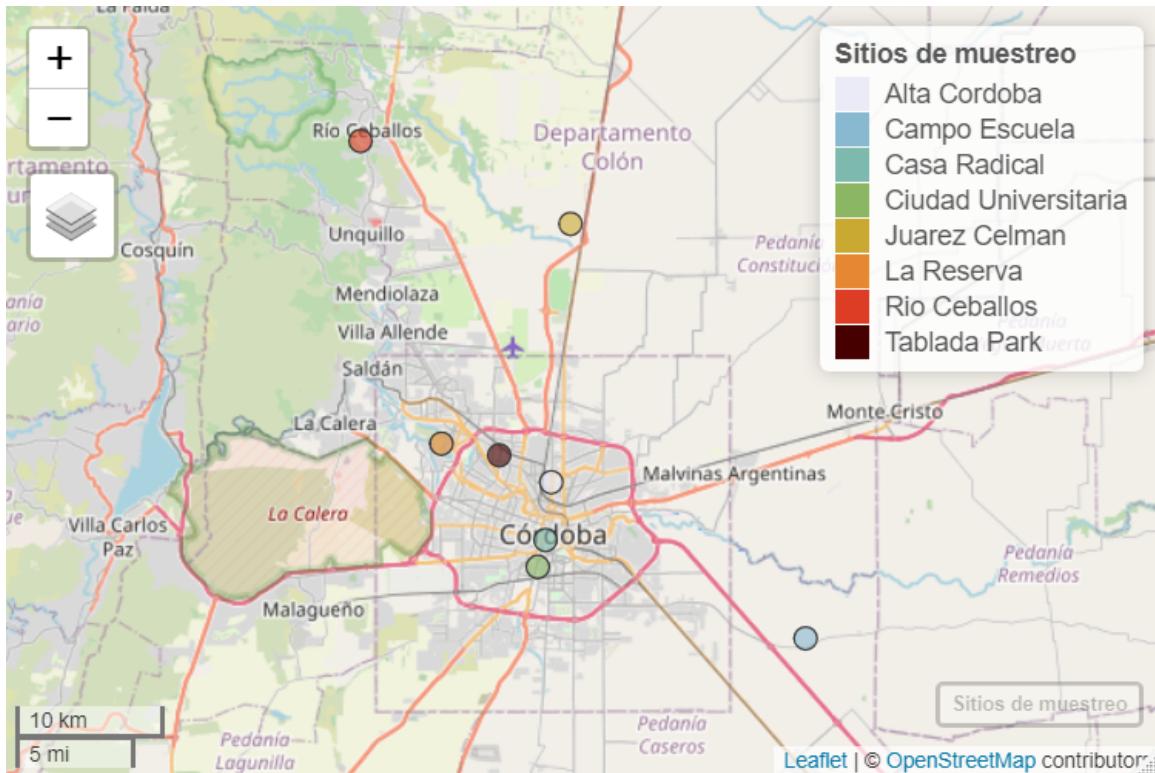


Figura 2.1: Ubicación específica de los sitios de muestreo.

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):

1. Ciudad Universitaria (urbano): 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

de PM_{2.5} 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.

2. Casa Radical (urbano): La Casa Radical, ubicada en barrio Nueva Córdoba, en el centro de la Ciudad, es un edificio donde se llevan a cabo actividades políticas como reuniones y debates del partido Unión Cívica Radical. 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 y menos importante categoría hace referencia a la disposición abierta de edificios que existe.

3. Tablada Park (urbano): Se trata de un hogar familiar ubicado en el barrio

residencial Tablada Park, en la zona noroeste de la Ciudad Capital de la provincia de Córdoba. Este sitio se encuentra muy cercano a la calle 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.

4. Alta Córdoba (urbano): Se trata de un hogar familiar ubicado 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 hogares residenciales 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.

5. La Reserva (urbano): 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 tercera categoría hace referencia a la disposición dispersa de construcciones de pequeño y mediano tamaño.
6. Río Ceballos (agrícola y área verde): 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 re suspensión 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 tercera categoría hace referencia a la disposición dispersa de construcciones de pequeño y mediano tamaño.

7. Campo Escuela (agrícola y área verde): 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 sur este 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.
8. Juarez Celman (agrícola y área verde): 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.

Las características descritas para cada uno de los sitios de muestreo determinan el uso del suelo que los representa en función de su patrón de emisión. El uso del suelo agrícola y área verde agrupa los sitios de Campo Escuela, Juárez Celman y Río Ceballos, mientras que el uso del suelo urbano contempla los sitios de Casa Radical, Ciudad Universitaria, Tablada Park, Alta Córdoba y La Reserva.

2.2.2. 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 refleja 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 satelitales. 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 excepcionales. 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 $\frac{\mu g}{m^3}$. Los valores que superaban esta concentración tendían a describir un salto repentino respecto de los valores en la hora anterior y posterior, además de escapar de la tendencia descripta por la serie temporal en términos generales. 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*. A partir de este resultado fue que se procedió a ajustar un Modelo Lineal Generalizado Mixto que vinculaba la variable respuesta y las distintas variables indicadoras de la escala temporal (hora, día, mes, semana, estación) como variables de efectos aleatorios (Ecuación 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 desvió 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.

$$\log y_i = \beta_1 \text{Hora}_i + \beta_2 \text{Dia}_i + \beta_3 \text{Mes}_i + \beta_4 \text{Semana}_i + \beta_5 \text{Estacion}_i \quad (1)$$

Ecuación 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.

2.2.3. Análisis estadístico

En primera instancia se buscó explorar gráficamente la variabilidad temporal existente de la serie de concentración media horaria de PM_{2.5} en las distintas escalas temporales, con el objetivo de identificar las tendencias, ciclos y patrones que influyen esta serie de tiempo.

En segundo lugar se llevó a cabo un análisis de la concentración de PM_{2.5} en cada uno de los sitios de muestreo, con el objetivo de demostrar la influencia de las condiciones propias de la ubicación espacial en la concentración del contaminante.

Luego se exploró la variabilidad temporal de la serie en cada uno de los usos de suelo, con el objetivo de verificar si las tendencias respecto del análisis promedio realizado en primera instancia se encontraba presente en cada uno de los usos especificados.

Además, este análisis se complementó con uno netamente descriptivo y comparativo entre las series temporales para cada uno de los usos de suelo.

2.3. Resultados y discusión

Filtrado *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 5 datos, cuyos valores de PM_{2.5} se correspondían con eventos anómalos sucedidos en los sitios durante el muestreo, como por ejemplo, podría tratarse de incendios cercanos, caídas del sensor o cualquier otra causa desconocida que influencie de forma repentina el patrón y tendencia descrito por la serie temporal. Estos valores resultaban notablemente diferentes al resto de los datos que pertenecían a la serie.

Filtrado *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 2.1).

Tabla 2.1: Variabilidad explicada por las componentes temporales de 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 2.1, 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 273 valores, lo que representan el 0,4% de la base de datos.

Análisis horario

La figura 2.2 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. 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.

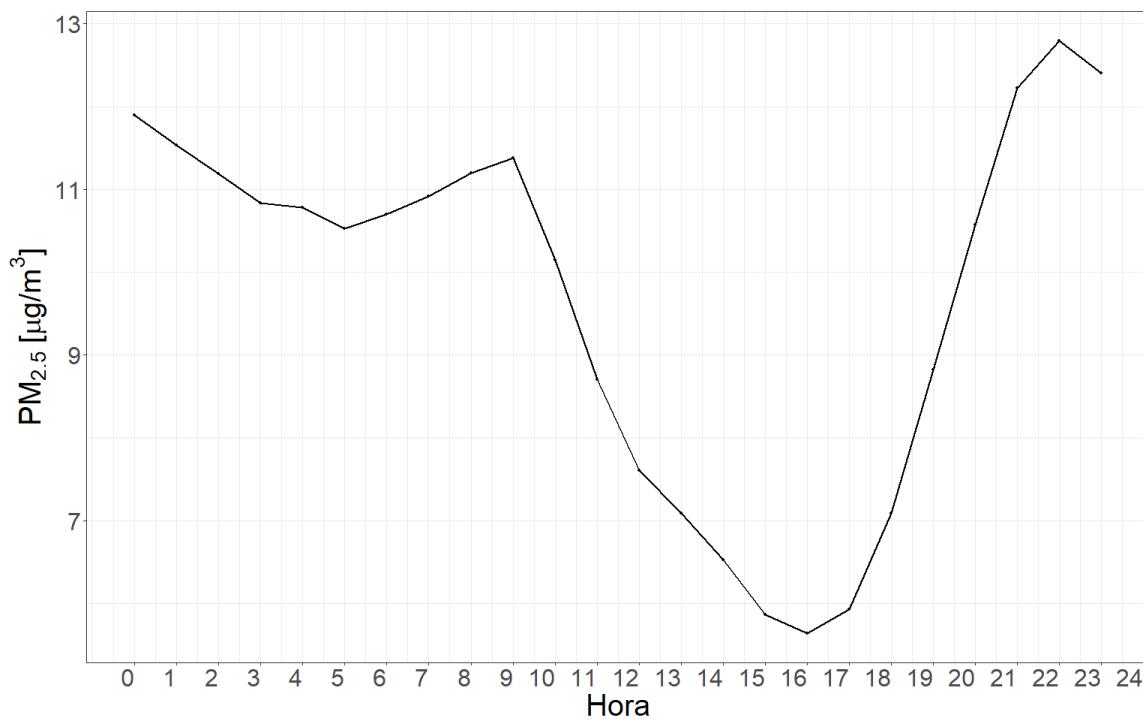


Figura 2.2: Perfil de concentración promedio horario de PM_{2.5} a lo largo de las horas en los sitios en estudio a lo largo de todo el periodo de muestreo.

Análisis diario

En la figura 2.3 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 lunes y domingo se registran las concentraciones más bajas de PM_{2.5}. A partir del lunes las concentraciones 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. 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.

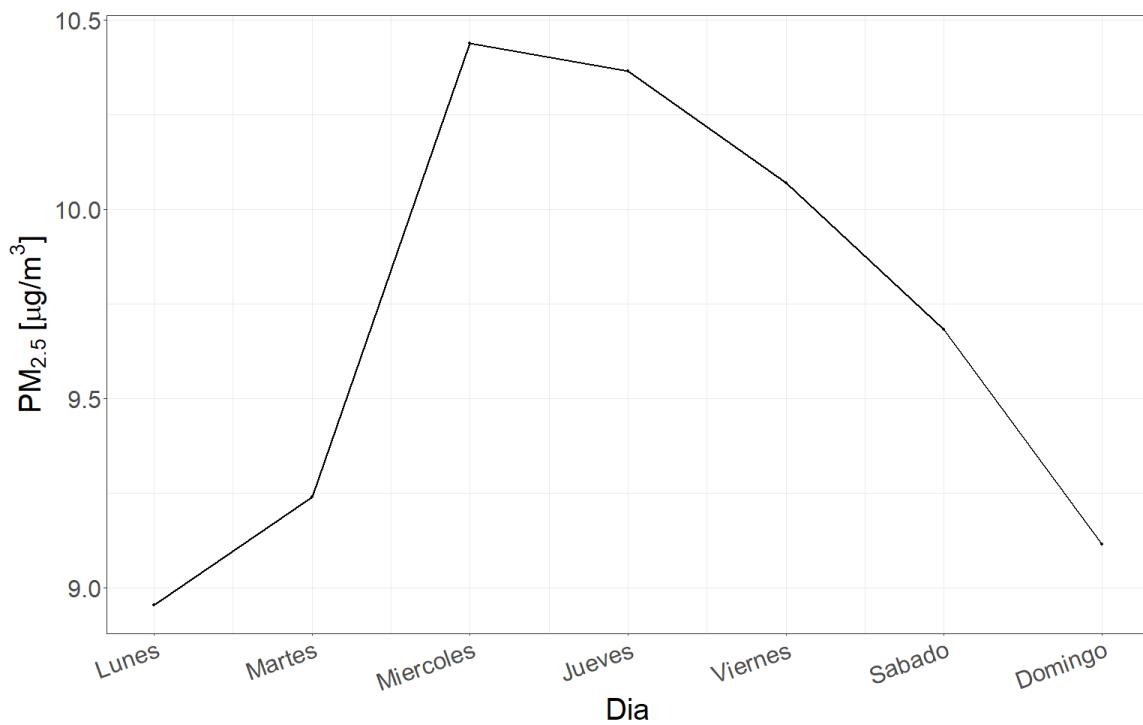


Figura 2.3: Perfil de concentración promedio diario de PM_{2.5} a lo largo de los días de la semana en los sitios en estudio a lo largo de todo el periodo de muestreo.

En la figura 2.4 puede apreciarse el patrón de variabilidad de la concentración de PM_{2.5} a lo largo de las horas del día en los distintos días de la semana. Puede observarse un comportamiento cíclico que evidencia que las concentraciones de 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.

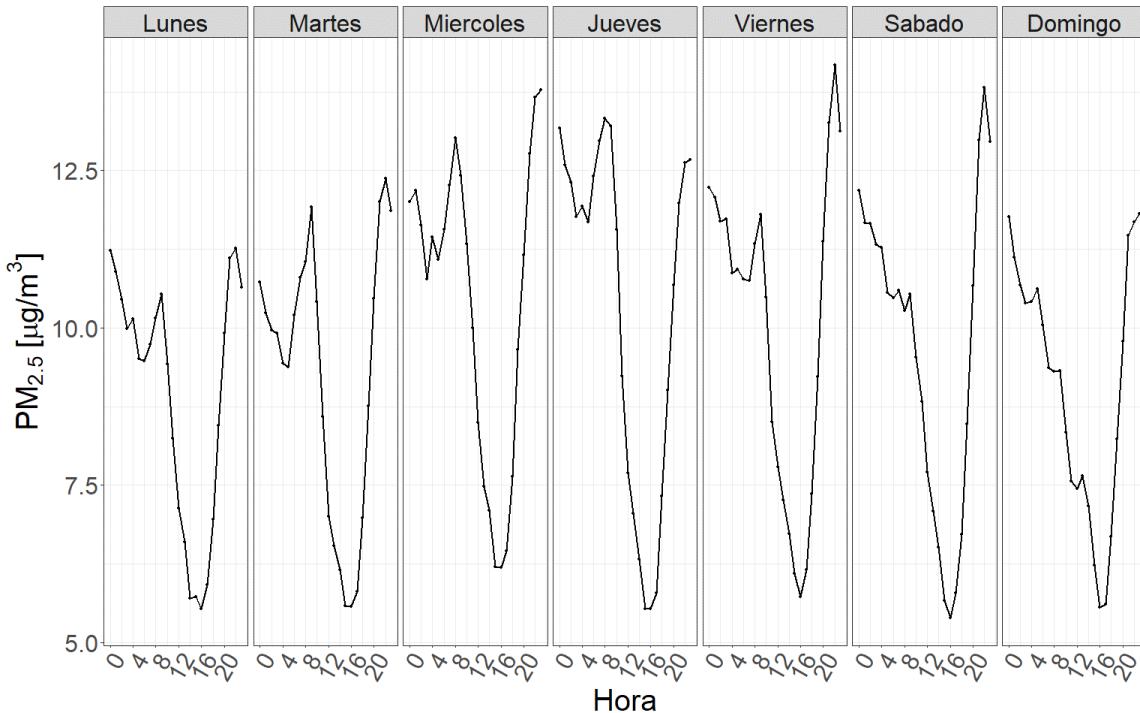


Figura 2.4: Perfil de concentración promedio horario de PM_{2.5} en cada día de la semana a lo largo de la semana en los sitios en estudio a lo largo de todo el periodo de muestreo.

Análisis mensual

En la figura 2.5, 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.

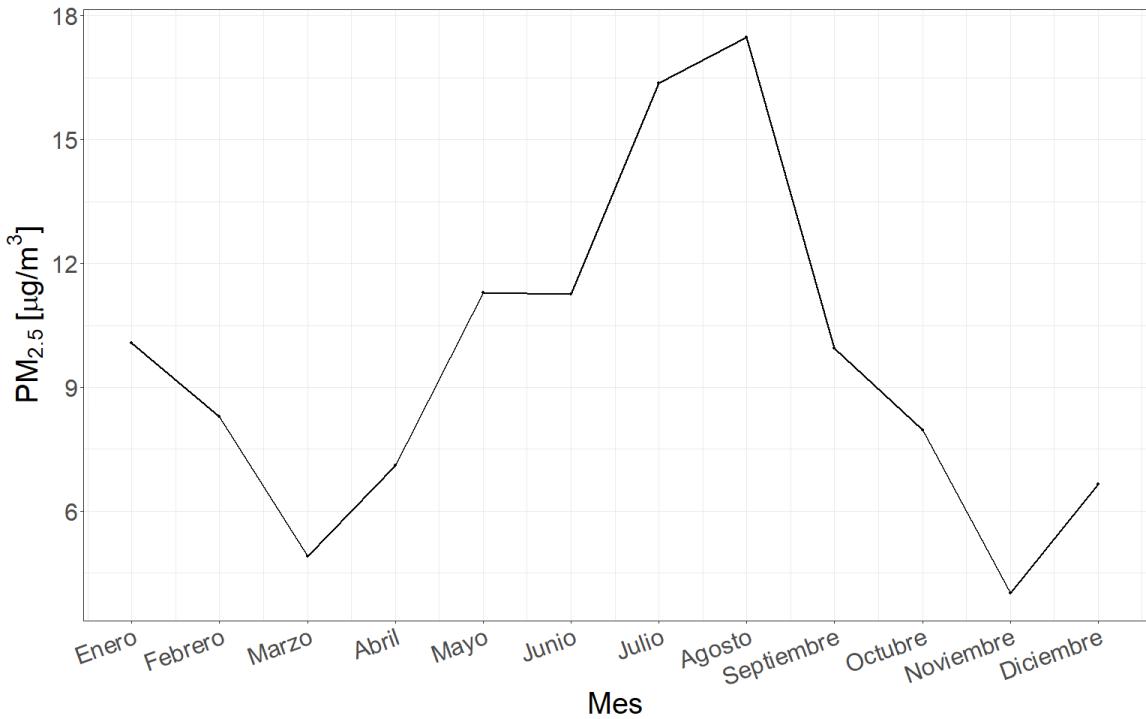


Figura 2.5: Perfil de concentración promedio mensual de PM_{2.5} a lo largo del año en los sitios en estudio a lo largo de todo el periodo de muestreo.

En la figura 2.6 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.

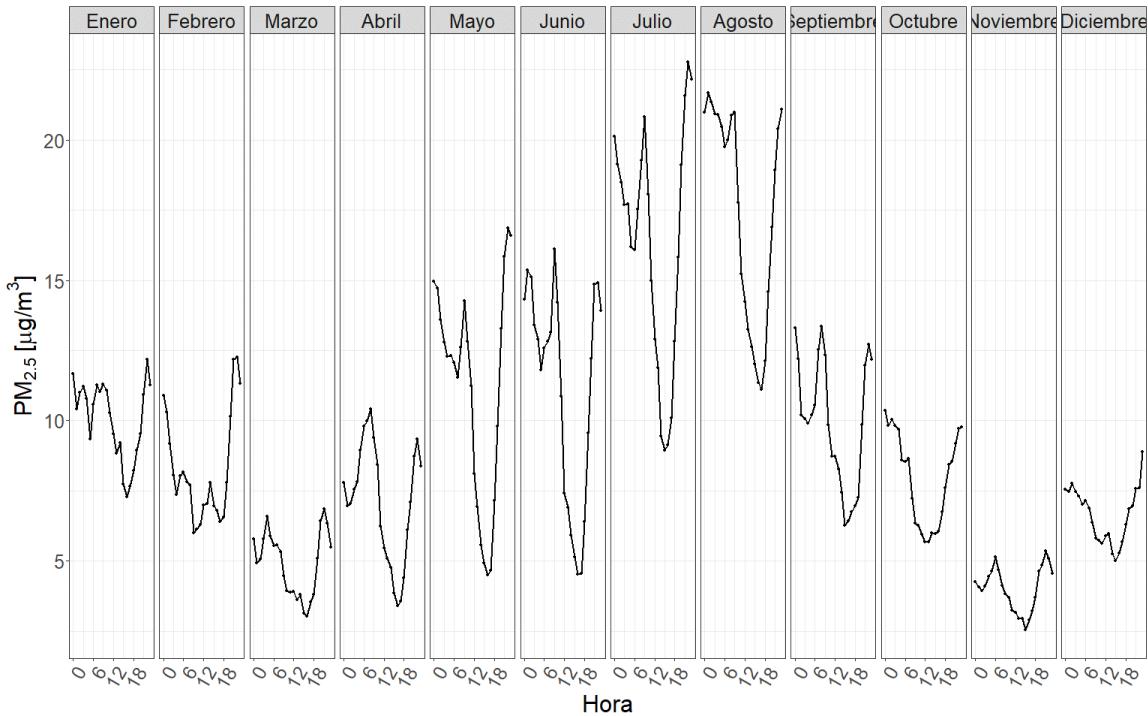


Figura 2.6: Perfil de concentración promedio horario de PM_{2.5} para cada mes a lo largo del año en los sitios en estudio a lo largo de todo el periodo de muestreo.

Análisis por estación del año

La figura 2.7 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.

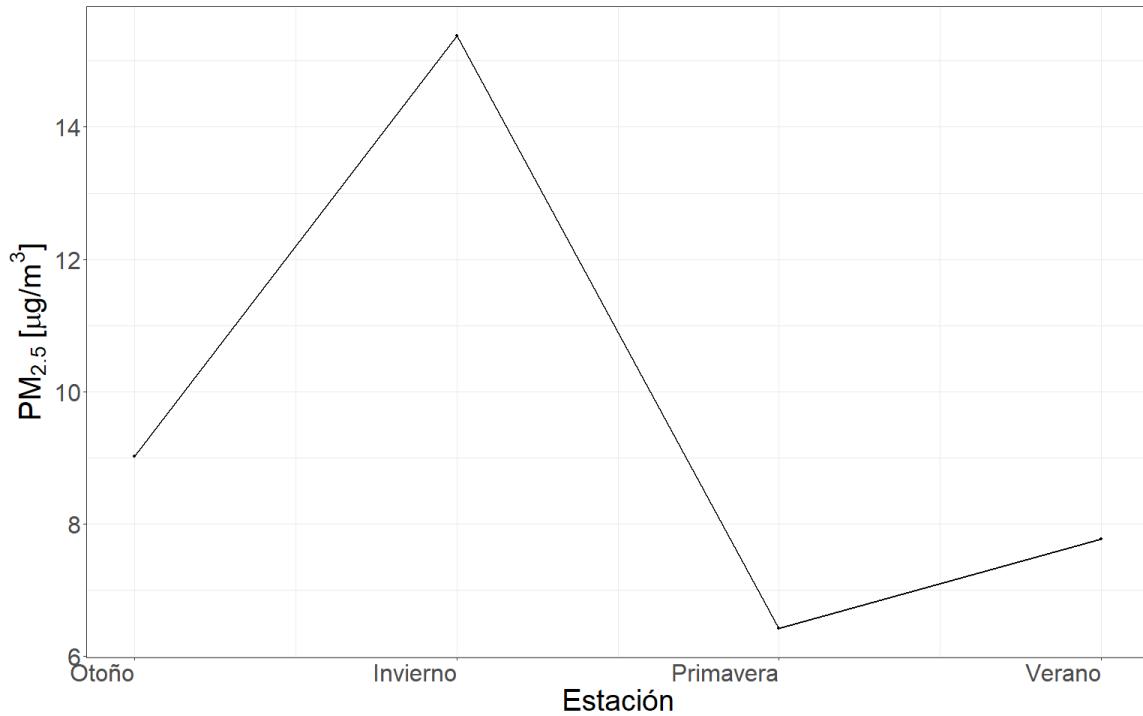


Figura 2.7: Perfil de concentración promedio por estación de PM_{2.5} a lo largo del año en los sitios en estudio a lo largo de todo el periodo de muestreo.

La figura 2.8 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.

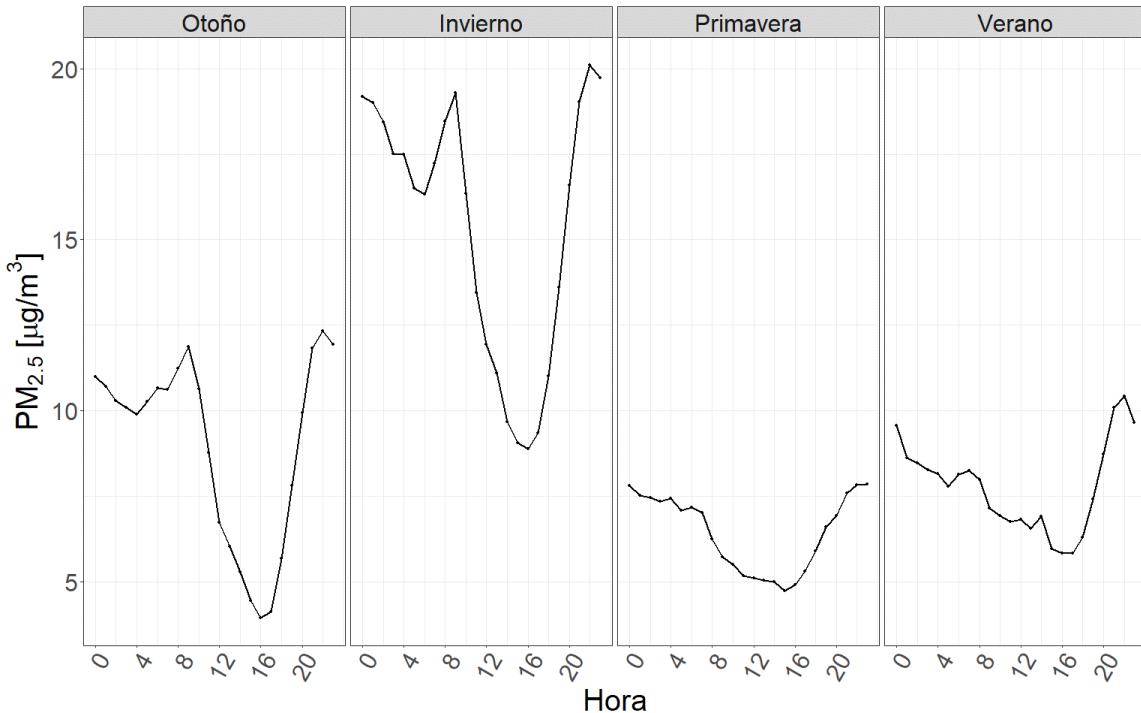


Figura 2.8: Perfil de concentración promedio horario de PM_{2.5} para cada estación a lo largo del año en los sitios en estudio a lo largo de todo el periodo de muestreo.

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 de la concentración de PM_{2.5} a lo largo de escalas temporales de elevada granularidad como lo son los meses y las estaciones podrían indicar que la variación de las condiciones meteorológicas influye en la variabilidad del contaminante. De esta forma, la meteorología característica entre los meses de mayo a agosto favorecerían condiciones adversas de calidad de aire, mientras que el resto del año estas se verían atenuadas. Esto se vincula con previos estudios que destacan inversiones térmicas más frecuentes en el invierno.

Descripción de la concentración promedio de PM_{2.5} en cada uno de los sitios de muestreo

La concentración media anual de PM_{2.5} registrada para cada uno de los sitios de muestreo (Tabla 2.2) exhibe similitud entre sitios del mismo uso del suelo.

Tabla 2.2: Concentración promedio anual de PM_{2.5} en cada sitio de muestreo.

Sitio	Concentración	Desviación estándar
Tablada Park (urbano)	13,89	±12,7
Casa Radical (urbano)	12,1	±10,2
Ciudad Universitaria (urbano)	11,5	±10,7
La Reserva (urbano)	10,3	±10,2
Alta Córdoba (urbano)	9,08	±8,60
Campo Escuela (agrícola/área verde)	8,88	±9,22
Río Ceballos (agrícola/área verde)	6,40	±6,36
Juarez Celman (agrícola/área verde)	6,00	±6,88

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

A partir de la tabla 2.3, pueden identificarse estaciones en las cuales las concentraciones de PM_{2.5} son críticas. En estas estaciones podría suceder que las condiciones atmosféricas sean las responsables de las condiciones adversas de calidad de aire (Amarillo et al., 2021; Mateos et al., 2019; Olcese & Toselli, 1997; Stein & Toselli, 1996). Específicamente, se observan concentraciones máximas durante los meses de mayo, junio, julio y agosto y concentraciones mínimas durante los meses de diciembre, enero, febrero, marzo y abril. 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}. Además, las diferencias entre sitios de muestreo según uso del suelo son mayores en invierno y otoño que en primavera y verano.

Tabla 2.3: Concentración promedio por estación del año de PM_{2.5} en cada sitio de muestreo.

Sitio	Invierno	Otoño	Primavera	Verano
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
Juarez 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.

Análisis descriptivo de las series temporales de concentración de PM_{2.5} para cada uno de los usos de suelo

Análisis horario por uso del suelo

La figura 2.9 permite observar el patrón de variabilidad de la concentración promedio horaria de PM_{2.5} a lo largo de las horas del día, discriminado por los distintos usos de suelo. Como puede apreciarse, el patrón de variabilidad en la concentración de PM_{2.5} a lo largo del día es evidente en los sitios con uso del suelo urbano, aunque se presenta en menor medida en el uso del suelo agrícola y área verde.

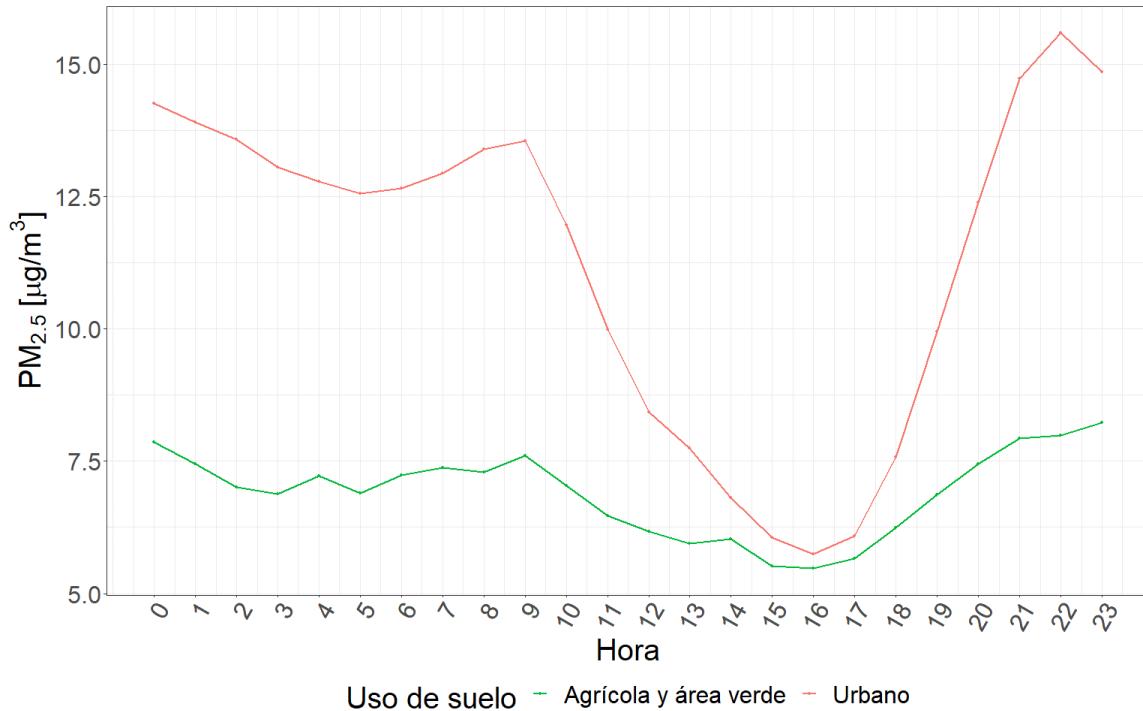


Figura 2.9: Perfil de concentración promedio horario de $\text{PM}_{2.5}$ a lo largo de las horas del día, discriminado por uso del suelo a lo largo de todo el periodo de muestreo.

En el caso de los sitios con uso del suelo urbano, la concentración promedio horaria de $\text{PM}_{2.5}$ presenta un patrón de aumento en la mañana, disminuyendo notablemente por la tarde, para aumentar nuevamente hacia la noche. Por otro lado, en el caso de los sitios con uso del suelo agrícola y área verde, el patrón de variabilidad en la concentración de $\text{PM}_{2.5}$ a lo largo del día se encuentra presente, aunque es menos evidente. La figura 2.9 muestra la presencia del patrón de variabilidad de la concentración promedio horaria de $\text{PM}_{2.5}$ en los distintos usos de suelo.

Análisis diario por uso del suelo

En la figura 2.10 puede apreciarse que para el uso del suelo urbano, las concentraciones de $\text{PM}_{2.5}$ 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. Sin embargo, para

el uso del suelo agrícola y área verde no logra observarse esta variación entre días hábiles y de fin de semana, lo que podría encontrarse vinculado a la variación en las actividades laborales y la notable reducción del tráfico vehicular en estas áreas.

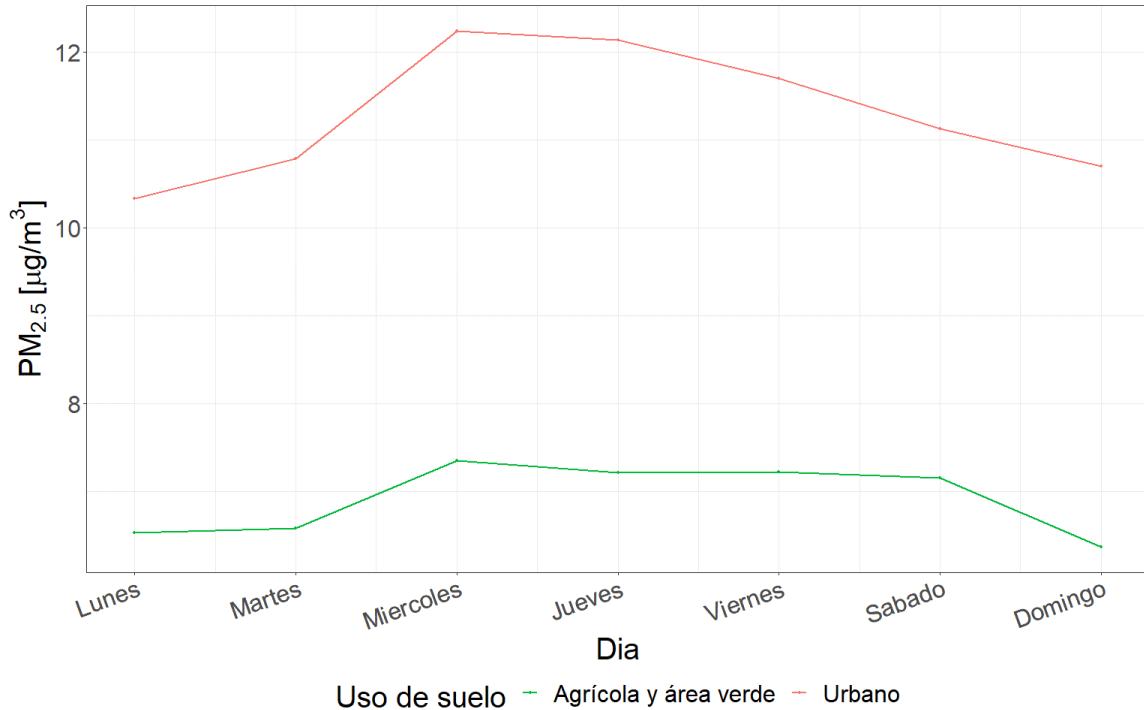


Figura 2.10: Perfil de concentración promedio diario de PM_{2.5} a lo largo de los días de la semana, discriminado por uso del suelo a lo largo de todo el periodo de muestreo.

A partir de la figura 2.11, puede apreciarse que la concentración de PM_{2.5} sigue un patrón de variabilidad similar a lo largo de los días de la semana, independientemente del uso del suelo, a pesar de que existan diferencias en magnitud. Específicamente, se observa que el uso del suelo urbano presenta magnitudes de concentración superiores al uso del suelo agrícola y área verde. Esto podría encontrarse vinculado a la dinámica de movilidad de la población, siendo el tráfico de vehículos una de las fuentes principales que contribuyen a la variabilidad del PM, de manera similar a lo observado por (Pant & Harrison, 2013; Weerakkody et al., 2018). De esta forma en los sitios urbanos, las variaciones en la concentración de PM_{2.5} podrían estar relacionadas con la densidad de personas y vehículos que los frecuentan durante los días laborables, la cual es

notablemente superior a la de los fines de semana, originando una mayor concentración de partículas durante los días laborables. En consecuencia, puede apreciarse que 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 2.10).

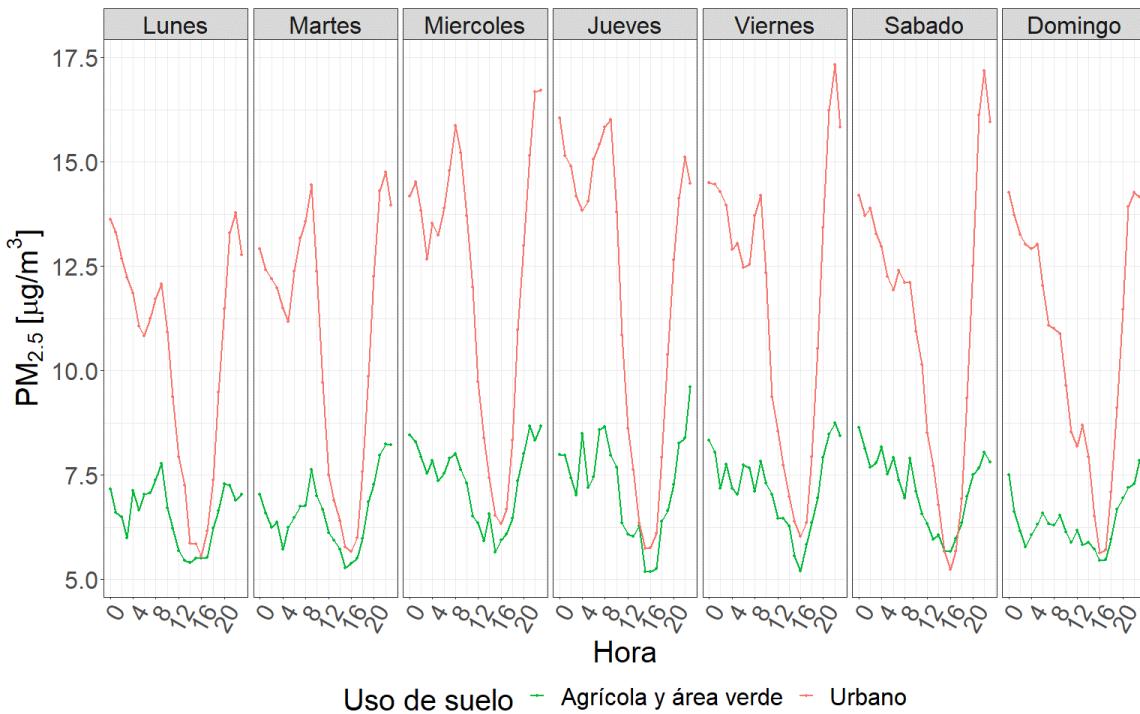


Figura 2.11: Perfil de concentración promedio horario de PM_{2.5} en las horas del día a lo largo de los días de la semana, discriminado por uso del suelo a lo largo de todo el periodo de muestreo.

En contraste, en los sitios cuyo uso del suelo es agrícola y 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.

Análisis mensual por uso del suelo

En la figura 2.12, 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.

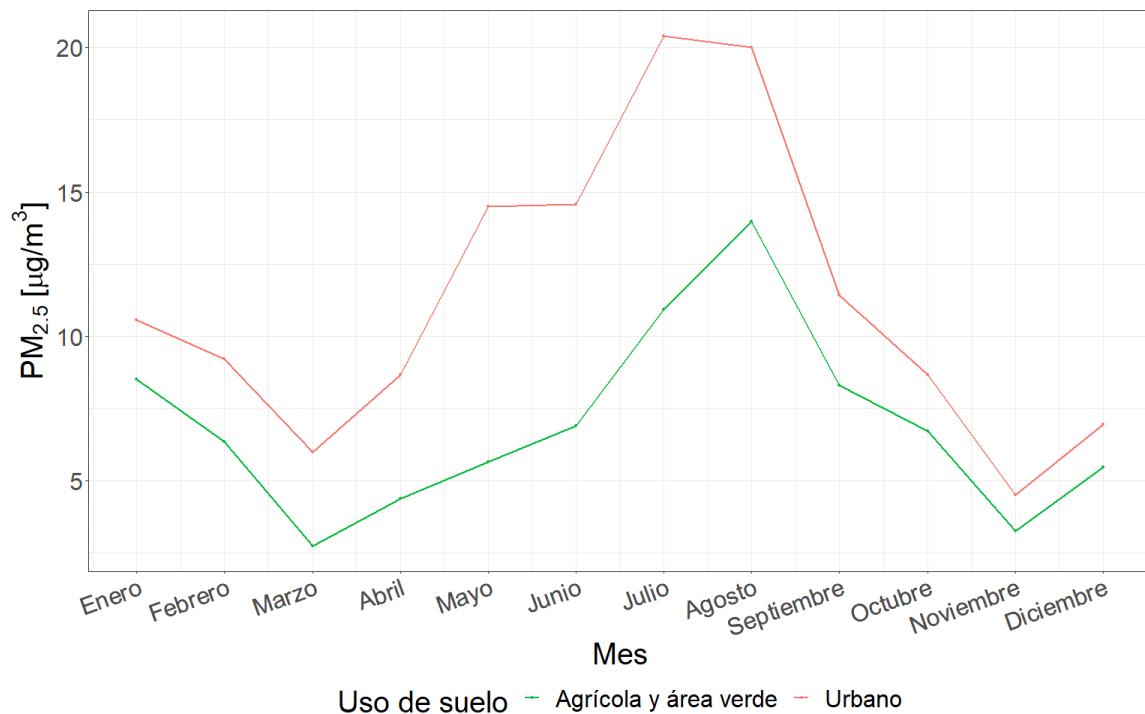


Figura 2.12: Perfil de concentración promedio mensual de PM_{2.5} a lo largo de los meses del año, discriminado por uso del suelo a lo largo de todo el periodo de muestreo.

Análisis según estación del año por uso del suelo

En la figura 2.13, puede apreciarse que el patrón estacional de concentración de PM_{2.5}, previamente observado en la figura 2.7, se mantiene constante en los distintos usos de suelo, a pesar de variar magnitud entre ellos. De esta forma, puede notarse que, al igual que en el análisis general, se registran menores concentraciones en el sitio con uso del suelo agrícola y área verde. Esto sugiere que la estacionalidad en la concentración del PM se encuentra altamente 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 al momento de evaluar los niveles de contaminación atmosférica y cómo estos pueden variar en función de factores ambientales y de uso del suelo.

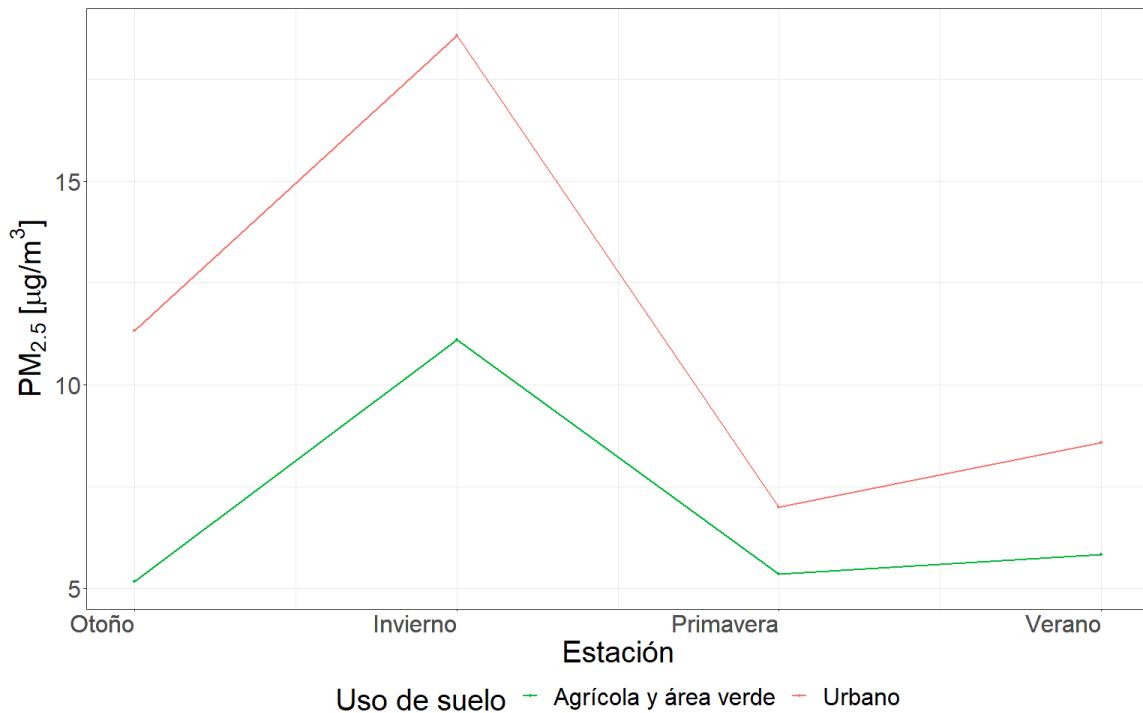


Figura 2.13: Perfil de concentración promedio estacional de PM_{2.5} a lo largo del año, discriminado por uso del suelo a lo largo de todo el periodo de muestreo.

En la figura 2.14 puede observarse el patrón de variabilidad de la concentración promedio horaria de PM_{2.5} a lo largo de las horas del día en las distintas estaciones discriminado en los usos de suelo obtenidos a partir del análisis de clúster realizado. Es interesante destacar que el patrón de variabilidad de la concentración que se observa para el análisis general (figura 2.7) se mantiene constante en los distintos usos de suelo, aunque disminuye en magnitud para el uso del suelo agrícola y área verde.

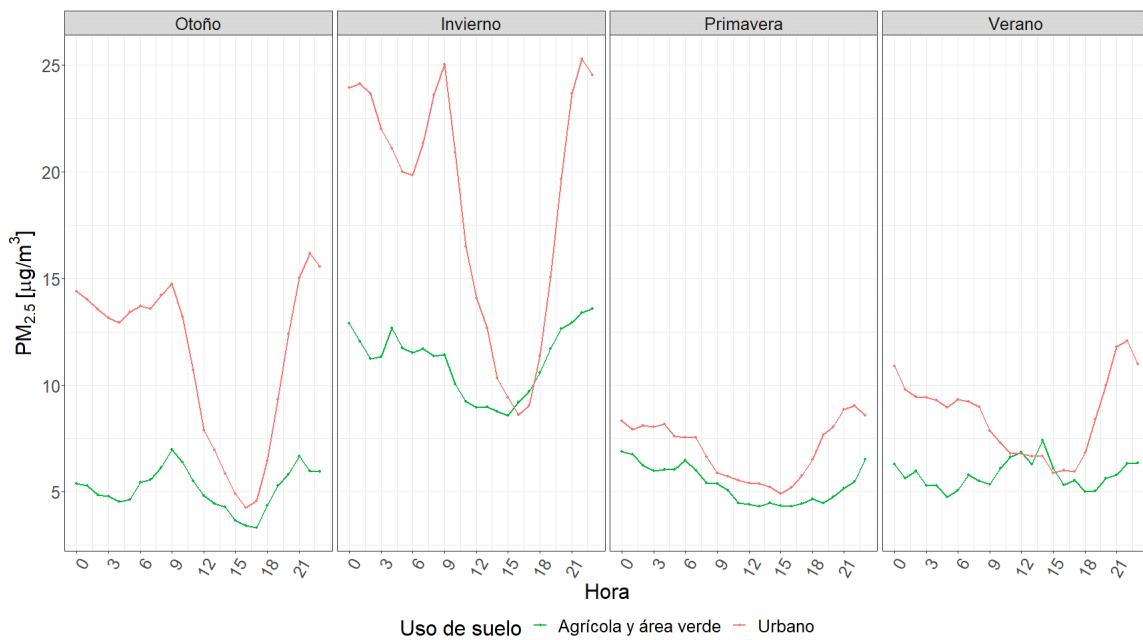


Figura 2.14: Perfil de concentración promedio horario de PM_{2.5} a lo largo de las horas del día en las distintas estaciones del año, discriminado por uso del suelo a lo largo de todo el periodo de muestreo.

Los resultados sugieren que las condiciones meteorológicas características de la época fría generan condiciones adversas de calidad del aire, independientemente del uso del suelo (Amarillo et al., 2021; Mateos et al., 2019; Olcese & Toselli, 1997; Stein & Toselli, 1996). Además, es importante mencionar que 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.

Medidas descriptivas

A partir de los datos de la tabla 2.4 puede apreciarse que la media para la serie de datos urbanos es superior que para la serie agrícola y área verde. Este hecho podría llevarnos a interpretar que a primera vista la concentración de PM_{2,5} es superior en entornos urbanos. Sin embargo, es importante destacar que esta medida de posición puede verse influenciada por valores extremos.

Tabla 2.4: Medidas descriptivas de series temporales promedio para los distintos usos de suelo.

Medida	Urbano	Agrícola y área verde
Media	11,54	6,78
Desvió estándar	9,90	6,10
Máximo	81,80	47,02
Mínimo	0	0
Percentil 50	8,41	4,92
Percentil 75	15,72	9,30
Percentil 90	25,49	15,16
Percentil 95	32,11	19,57
RI	7,31	4,38

Rango intercuartílico (RI).

Por lo tanto, para realizar la comparación también se ha empleado la mediana o percentil 50. Puede apreciarse que la serie proveniente de entornos urbanos posee un percentil 50 de 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 la serie agrícola y área verde es de 4,92 $\frac{\mu g}{m^3}$, reforzando la idea de que las series de datos provenientes de entornos urbanos muestran mayores concentraciones que la proveniente de entornos agrícolas y áreas verdes.

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. Este hecho evidencia en términos generales que un mayor porcentaje de datos se encuentra por debajo del mismo límite numérico, indicando que la serie de datos urbana toma valores de mayor magnitud que la serie de datos agrícola y 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 la serie urbana. De esta forma queda en evidencia que la serie de tiempo urbana toma valores de mayores magnitudes que la serie de tiempo agrícola y área verde, encontrándose sus registros más dispersos respecto a su media. 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.

A partir de la figura 2.15 se puede notar en primer lugar una diferencia en el tamaño de la caja para los distintos usos de suelo, esta diferencia es representativa de la mayor dispersión observada para la serie urbana y la mayor concentración de datos en torno a su media para la serie agrícola y área verde. Además, puede apreciarse la influencia de valores extremos en ambas series generando un corrimiento de la media respecto de la mediana, este corrimiento es superior para la serie de datos urbana, indicando un mayor porcentaje de valores extremos.

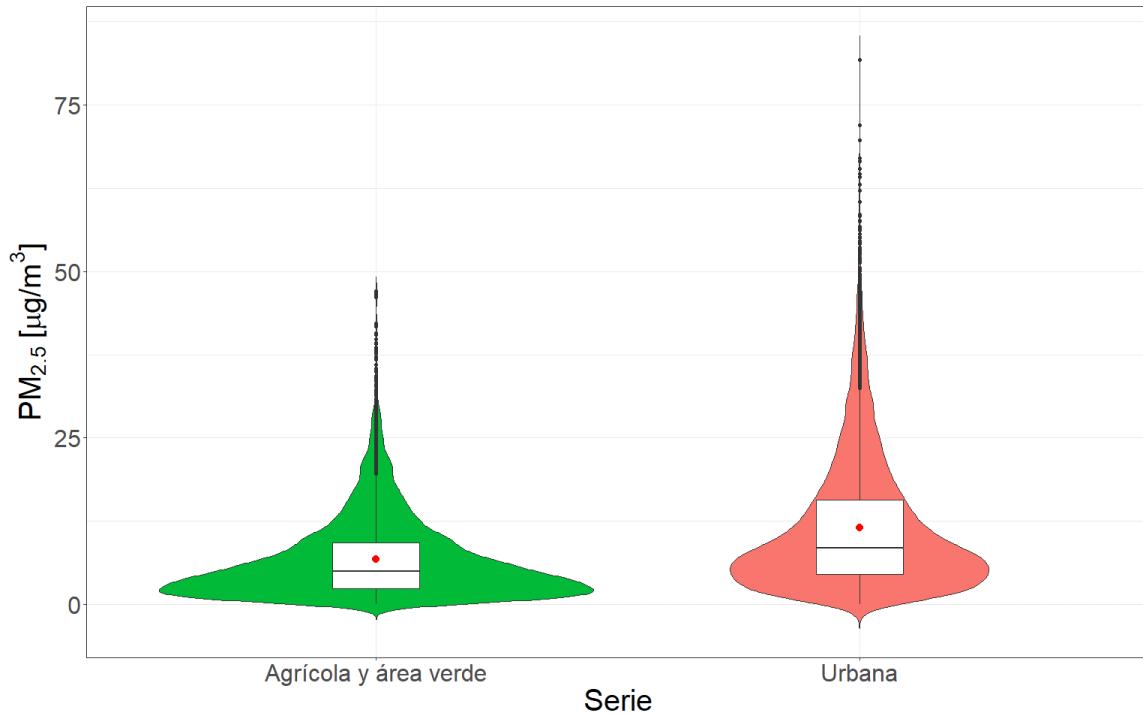


Figura 2.15: Diagrama de violín que compara la concentración promedio horaria de PM_{2.5} en las series con distintos usos de suelo a lo largo de todo el periodo de muestreo. La media de las serie se encuentra indicada por el punto rojo dentro de la caja.

A partir de la figura 2.16 puede notarse que se trata de series temporales que logran diferenciarse en función de sus comportamientos a lo largo del tiempo. Es importante destacar que la serie agrícola y de área verde muestra un rango de variación significativamente inferior en comparación con la serie urbana. Puede observarse que los valores de la serie urbana oscilan en un rango de magnitud mucho mayor en comparación con la serie agrícola y de área verde. La serie de datos urbana demuestra ser más inestable, pasando desde picos de elevada magnitud a valores mínimos de forma sucesiva y repetitiva a lo largo del tiempo.

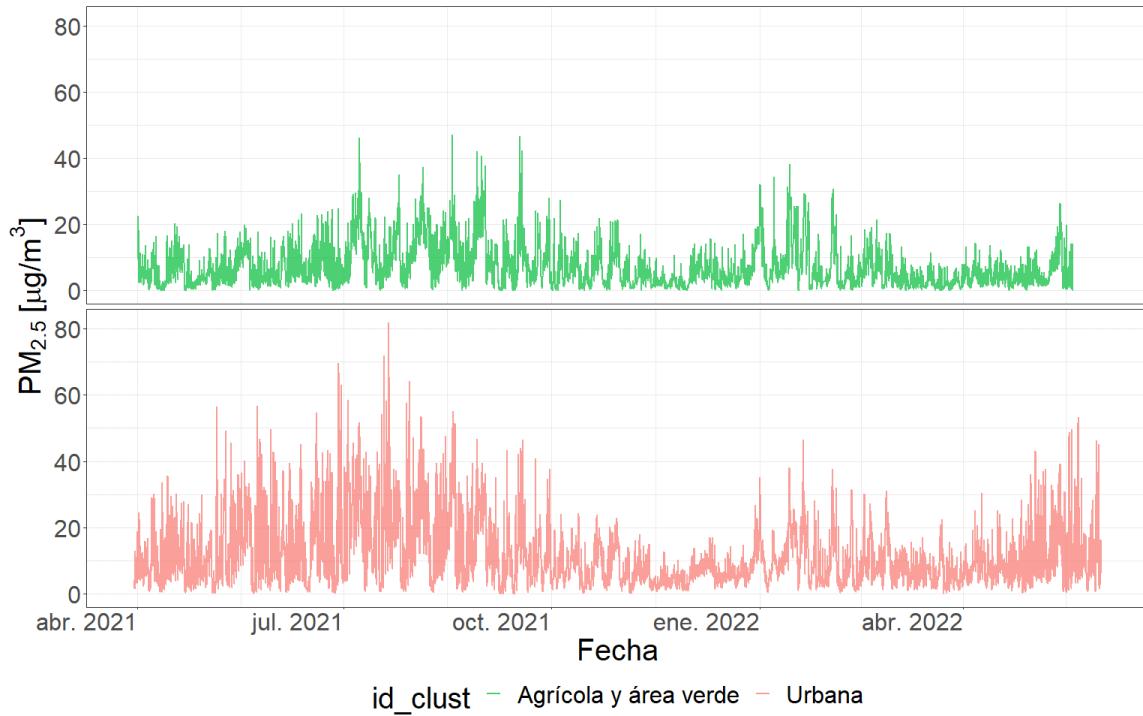


Figura 2.16: Serie de tiempo de la concentración promedio horaria de $\text{PM}_{2.5}$ en entornos de distinto uso del suelo a lo largo de todo el periodo de muestreo.

2.4. Conclusión

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 $\text{PM}_{2.5}$ fue la distribución *gamma*.

Las características temporales de la serie de concentración de $\text{PM}_{2.5}$ permiten explicar parte de la variabilidad de la misma. A su vez las características espaciales, referidas a la ubicación y la influencia diferencial de la microescala en cada uno de los usos de suelo, permite explicar la discrepancia en la magnitud de concentraciones de $\text{PM}_{2.5}$ existentes.

Las características temporales de la serie promedio de concentración de $\text{PM}_{2.5}$ 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 mostrar un patrón de variación similar, puede destacarse que el mismo se ve atenuado en cuanto a la magnitud de las concentraciones en la serie agrícola y área verde en comparación con la serie urbana. Este hecho refleja que el patrón de variación de 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.

El análisis exploratorio confirmó la diferencia existente entre las series temporales provenientes de distintos usos de suelo. La serie de datos proveniente de entornos agrícolas y área verde demostró concentraciones de menor magnitud, mayor concentración de datos en torno de su media, menores fluctuaciones entre valores de elevada y baja magnitud que la caracterizaron como una serie temporal de menor variabilidad en comparación con la urbana.

La variación de las escalas temporales de mayor magnitud, como los meses y estaciones, podría vincularse a la variación de las condiciones meteorológicas. Esta variación potencial para explicar la variabilidad de la concentración de PM_{2.5} podría ser independientemente del uso del suelo que se esté considerando.

Capítulo 3

Algoritmos para la predicción de la concentración de PM_{2.5}

3.1. Introducción

El modelado predictivo de contaminantes atmosféricos ha tomado gran importancia como base para la implementación de mecanismos de alerta y reducción de este tipo de contaminación. La medición directa del contaminante a través de un sistema de monitoreo no garantiza una cobertura completa del territorio, sin mencionar sus elevados costes de instalación y mantenimiento. Por este motivo, los modelos predictivos ganaron terreno como una forma efectiva de proveer alertas tempranas, con el objetivo de proteger la salud pública (Titus, 1990). Estas alertas permiten dar aviso a las autoridades con la finalidad de iniciar el plan de mitigación correspondiente para reducir sus impactos. La gestión de la contaminación puede ser abordada a través de la implementación de mecanismos de atenuación de emisiones como lo es la regulación del tráfico (Cesaroni et al., 2012; Font et al., 2019), como también de reducción de concentración por medio del lavado a través de la pulverización de agua, además es común aplicar medidas más exigentes en sitios donde poblaciones más vulnerables

son afectadas. Esta es la razón por la cual puede hallarse una extensiva literatura relacionada al modelado de contaminantes atmosféricos.

Los métodos de modelado que mejor desempeño han alcanzado en términos de la exactitud de sus predicciones provienen de la rama de la inteligencia artificial (He et al., 2016; Navares & Aznarte, 2020a, 2020b; Shang et al., 2019). El aprendizaje automático es una subárea de la inteligencia artificial, cuyo propósito recae en la creación de algoritmos que aprendan las interrelaciones existentes entre las señales de entrada y las de salida. Mediante el aprendizaje automático se posibilita la creación de programas capaces de derivar el comportamiento de una variable respuesta (variable de salida) a partir de la información de las características (variables de entrada) en un momento dado. Cabe destacar que el aprendizaje profundo es un conjunto de algoritmos de aprendizaje automático que busca modelar abstracciones de alto nivel en los datos usando arquitecturas computacionales que incluyen transformaciones no lineales de forma sucesiva sobre los datos de base (Bengio et al., 2013; Tong et al., 2019; Yeo et al., 2021).

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 material particulado fino ($PM_{2.5}$) (He et al., 2017; Tiwari et al., 2017). Numerosos estudios han evidenciado estas asociaciones, en su mayoría encontrando un vínculo inverso con respecto al material particulado (PM), las variables citadas en la literatura como explicativas de este patrón son: la velocidad del viento (Cichowicz et al., 2020; Usman et al., 2022; XiaoYang et al., 2019), la precipitación (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) 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 como la profundidad óptica de aerosoles (AOD) (Gupta et al., 2021; Soni et al., 2018) y el tráfico de vehículos (Pant

& Harrison, 2013; Weerakkody et al., 2018). Tanto para la presión atmosférica, como para la humedad relativa, existen resultados que vinculan las variables tanto de forma directa (H. Li et al., 2015; Nogarotto & Pozza, 2020) como inversa (Tao Chen et al., 2016; Tian, Qiao, & Xu, 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 de vinculación 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).

Enmarcado dentro del Servicio de Cambio Climático Copernicus (Copernicus Climate Change Service - C3S) de la Agencia Espacial Europea (ESA), el Centro Europeo para Previsiones Meteorológicas a Medio Plazo (European Centre for Medium-Range Weather Forecasts - ECMWF) genera grandes conjuntos de datos climáticos a escala global. Para ello se combinan modelos climáticos y datos observacionales de todo el mundo. Este producto es la quinta generación de Reanálisis Europeo (European ReAnalysis - ERA5), sucesor de ERA-Interim (Muñoz-Sabater et al., 2021). Estos productos, ponen a disposición información para todo el mundo 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. Sin embargo, aun cuando existen modelos predictivos mecanicista 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.

En esta tesis se compararon distintos enfoques de modelado predictivo basados en el aprendizaje automático y aprendizaje profundo implementando como datos de entrada variables meteorológicas, ambientales y temporales, obtenidas principalmente de satélites. El objetivo de este trabajo es ajustar un modelo predictivo capaz de ser empleado como sistema de alerta ante condiciones adversas de concentración de PM_{2.5}.

3.2. Materiales y métodos

3.2.1. 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 variables generadas a partir de información satelital (Tabla 3.1). Entre estas se destacan las variables generadas por el modelo de 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. ERA5 permitió disponer de información de 164 variables atmosféricas, oceánicas y terrestres, que fueron incorporadas a la base de datos para desarrollar la tesis. Estas estimaciones cuentan con una resolución temporal de una hora y una espacial de 0.25 grados. Además, se incluyeron otras covariables, citadas en la literatura como relevantes para el PM_{2.5}, provenientes de las mediciones del sensor MODIS a bordo de los satélites Terra y Aqua de la Administración Nacional de Aeronáutica y el Espacio (NASA)

(Hulley et al., 2014; Lyapustin et al., 2018, 2019; Savtchenko et al., 2004), datos obtenidos a partir del análisis retrospectivo de la era moderna para investigación y aplicaciones, versión 2 (MERRA-2) (Bosilovich et al., 2015) y la base de datos de Google Traffic (Y.-J. Wu et al., 2007). Una descripción completa de cada una de las variables involucradas se encuentra en el Anexo 1.

Tabla 3.1: 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 μm) 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*).

Las variables temporales (estación, mes, día del mes, día de la semana y hora de la lectura del contaminante) fueron incluidas en su forma categórica original y luego de ser codificadas de manera cíclica a través de la inclusión de sus senos y cosenos. Es importante destacar que todas las variables numéricas se sometieron a un proceso de estandarización o normalización. Esto implicó restar la media y dividir por la desviación estándar de cada variable, con el propósito de uniformar sus escalas y

establecer una desviación estándar de 1. Esta metodología se implementó para evitar problemas de escala y facilitar la comparación entre las variables en el análisis.

La inclusión de variables categóricas 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.

3.2.2. Protocolo estadístico

Métodos de inteligencia artificial comparados

La **regresión lineal múltiple** (*multiple linear regression*) es una técnica estadística que se utiliza para modelar una variable dependiente como combinación lineal de múltiples variables independientes. Este tipo de análisis es aplicable tanto a datos de sección transversal como a series temporales (Aldas Manzano & Uriel Jimenez, 2017). El supuesto subyacente en el modelo es que existe una relación lineal entre la variable dependiente y las variables independientes. El aprendizaje de la interrelación entre las variables se logra mediante el ajuste de los coeficientes de regresión, estimados a partir del conjunto de datos de entrenamiento y validados con los de testeo. La regresión lineal busca determinar el hiperplano que mejor ajusta a los datos de entrenamiento, lo cual implica encontrar los coeficientes que minimicen la diferencia entre los valores observados y predichos por el modelo. El ajuste fue llevado a cabo por medio del método de mínimos cuadrados ordinarios, este ajusta un modelo lineal cuyos coeficientes resultantes buscan minimizar la sumatoria cuadrada de residuos entre los valores observados en el conjunto de datos de entrenamiento, y los valores

predichos por la aproximación lineal. Luego los resultados son validados a través de las predicciones en el conjunto de datos de testeo.

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). Estos métodos combinan los resultados de múltiples modelos independientes para mejorar el rendimiento predictivo del modelo ensamblado (Figura 3.1). Uno de los métodos de ensamblado más reconocido es el empaquetado (*bagging* o *bootstrap aggregating*) (Breiman, 1996). El empaquetado crea múltiples conjuntos de datos mediante remuestreo con reemplazo de los datos de entrenamiento. Al combinar los resultados de los modelos estimados en cada uno de los conjuntos de datos obtenidos por remuestreo, se aumenta la estabilidad y precisión de las predicciones, porque se reduce la varianza.

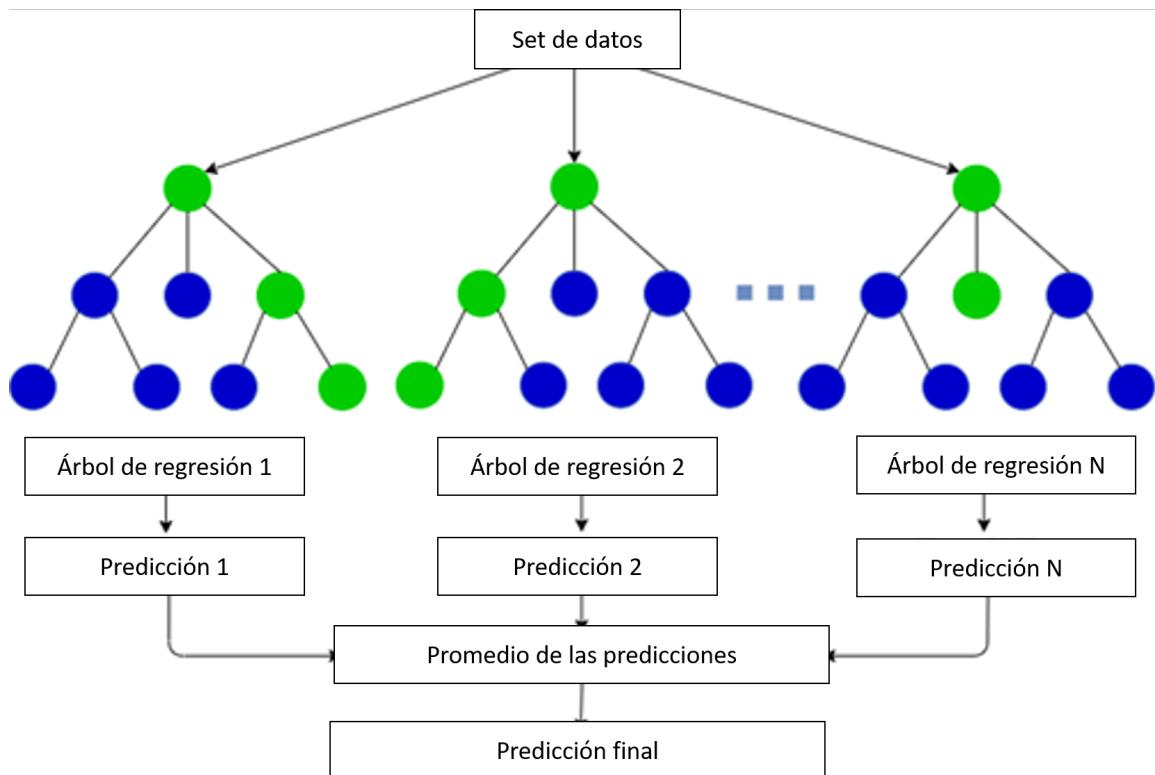


Figura 3.1: Diagrama de funcionamiento del algoritmo de bosques aleatorios.

Posteriormente, Breiman (Breiman, 2001) 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 los distintos subconjuntos generados por el remuestreo.

En resumen, el algoritmo de bosques aleatorios combina los resultados de múltiples árboles de regresión independientes, cada uno generado sobre distintos subconjuntos de datos creados a partir de los datos originales y agrega las predicciones resultantes de los distintos árboles a través de un promedio.

Las **redes neuronales artificiales** (*artificial neural networks (ANN)*) son algoritmos que permiten modelar procesos no lineales de manera eficiente, a partir de la información de entrada recolectada a través de un vector denominado capa de entrada. Una vez ingresada la información 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 3.2). Cada neurona genera una respuesta excitatoria a las señales recibidas por medio de una función de activación. Existen distintas funciones, 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

regresión 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. 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 tradicionales, incluyen bucles o conexiones entre unidades que permiten que la información persista en los sucesivos capas de la red (Figura 3.2). 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, es decir reciente, 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 desfasajes de tiempo.

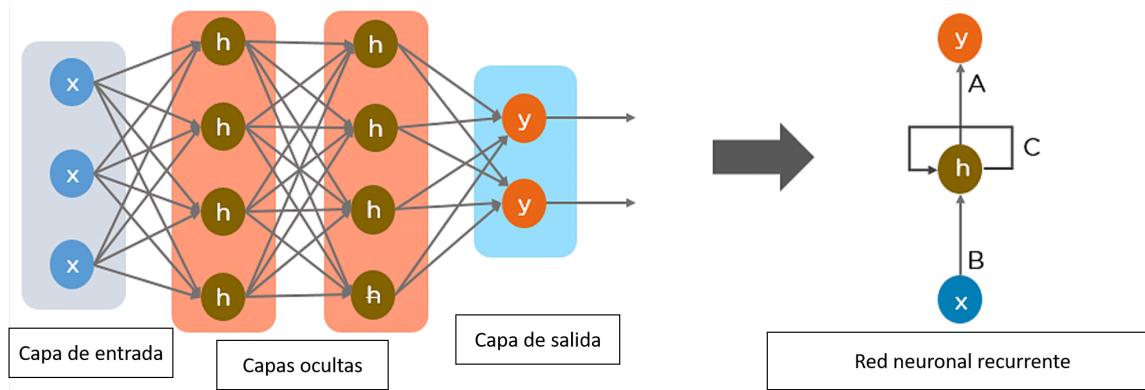


Figura 3.2: Diagrama de funcionamiento del algoritmo de redes neuronales recurrentes.

Para solventar esta falencia, en 1997 fueron introducidas las **redes de memoria a corto 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), las cuales mantienen las características propias de las *LSTM* con una estructura más simple, reduciendo la cantidad de parámetros a ajustar y permitiendo una ejecución de mayor rapidez y eficiencia.

Metodología de modelado

Se evaluaron múltiples alternativas para desarrollar el ajuste del modelo predictivo de concentración de PM_{2.5}, entre ellas distintas formas de condicionarlo, con el objetivo de aumentar el desempeño en la obtención de predicciones. Se generaron modelos para cada estación del año (4 modelos predictivos), día de la semana (7 modelos predictivos), mes del año (12 modelos predictivos), hora del día (24 modelos predictivos) y uso del

suelo (2 modelos predictivos). Mediante estadística descriptiva se determinó que el uso del suelo es una fuente de variación importante de las series temporales de PM_{2.5}. Y por ello el desarrollo del algoritmo predictivo fue llevado a cabo condicionado por esta variable, de forma que se entrenaron dos algoritmos predictivos, uno para cada uno de los usos de suelo presente en la base de datos.

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, con el objetivo de obtener un modelo predictivo que permita disponer de predicciones de concentración de contaminantes atmosféricos con cinco días de anticipación. Este desfasaje se haya fundamentado en que la disponibilidad de los datos de ERA5 de forma gratuita se obtiene con una latencia de 120 horas. De esta forma, en el momento presente se dispone de datos que ocurrieron hace 120 horas. De allí surge la necesidad de realizar el entrenamiento del modelo con las variables desfasadas para que el modelo, en función de lo que ocurrió hace 120 horas, logre predecir lo que sucederá en la hora siguiente.

Selección de variables

Fueron analizadas tres alternativas respecto de la cantidad de variables predictoras a incluir en el modelado predictivo. De esta forma pudo determinarse si los algoritmos resultantes de mejor ajuste se mantenían constantes bajo distintas versiones de la base de datos o si la selección de variables estaba jugando un rol fundamental en los resultados obtenidos. Las alternativas evaluadas contemplaron el ajuste bajo una base de datos que involucrase todas las variables regresoras (base de datos original), otra que involucra las seleccionadas por el algoritmo LASSO y otra base de datos que involucra las variables seleccionadas por el algoritmo featurewiz. Todos los análisis se realizaron con la versión 3.10 de Python, empleando diferentes módulos, entre ellos *scikit-learn*, un módulo de aprendizaje automático creado a partir del paquete *scipy*.

para la selección por LASSO, y *featurewiz*, una biblioteca de Python para crear y seleccionar las características más importantes de una base de datos, para aplicar la selección del algoritmo de featurewiz.

El algoritmo LASSO es una técnica de selección de variables que utiliza una penalización (L1) para reducir la complejidad del modelo y mejorar su capacidad de generalización. Su objetivo es encontrar el subconjunto óptimo de variables más relevantes para desarrollar un modelo predictivo. Para lograr esto, LASSO minimiza una función de costo regularizada, que introduce un término de penalización L1 (Ecuación 2). La penalización actúa como un mecanismo de selección de variables, ya que algunos coeficientes de regresión se reducirán a cero, lo que indica que las variables correspondientes no contribuyen significativamente al modelo. Una de las características útiles de LASSO es que permite reducir la dimensionalidad de los datos al eliminar las variables menos importantes (Robert, 1994).

$$\text{Pérdida} = \text{Error}(Y - \hat{Y}) + \lambda \sum_{i=1}^n |w_i| \quad (2)$$

Ecuación 2: Función de costo regularizada que implementa el algoritmo de LASSO que introduce un término de penalización L1.

Componentes:

- **Y:** valores reales u observados en un conjunto de datos. Estos son los valores de la variable respuesta que se intentan predecir en un problema de regresión.
- **\hat{Y} :** predicciones del modelo. Representa los valores estimados por el modelo para la variable de respuesta.
- **λ :** hiperparámetro de ajuste que controla la fuerza de la regularización L1 en el método Lasso. Un valor más alto de λ aumenta la penalización de los coeficientes de las características, lo que puede llevar a la selección automática de características al forzar algunos coeficientes a cero.
- **w_i :** coeficientes asociados con cada una de las características en un modelo de regresión. Estos coeficientes representan la contribución de cada característica a la predicción de la variable de respuesta. En el contexto de Lasso, se aplicará la penalización L1 a estos coeficientes para promover la selección automática de características.

El algoritmo *featurewiz* (Bakheet et al., 2023; 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 *MRRM* (*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*, el cual es un método de ensamblado (*ensamble*) perteneciente a la familia de los algoritmos de *boosting*, cuya característica principal es que el entrenamiento del modelo predictivo se genera de forma secuencial (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*. De esta forma, el algoritmo logra encontrar el conjunto óptimo de características, que permite entrenar un modelo de aprendizaje automático robusto.

Las variables incluidas en cada una de las bases de datos original y reducidas por selección de características, para cada uno de los usos de suelo se encuentran descriptas en el Anexo 2.

Metodología de entrenamiento

La metodología implementada para llevar a cabo el ajuste y evaluación de los modelos predictivos, constó en un procedimiento iterativo de división de datos, entrenamiento del modelo y evaluación por medio de las métricas de desempeño. La base de datos fue dividida en subconjuntos de entrenamiento, validación y testeo 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 respectivas proporciones de los datos para cada uno de los subconjuntos (entrenamiento, validación y testeo), fue de 70 %, 15 % y 15 %, respectivamente. Posteriormente se procedió a ajustar cada uno de los algoritmos de modelado predictivo (regresión lineal múltiple, bosques aleatorios y redes neuronales recurrentes) en la base de datos de entrenamiento y validación. Una vez entrenados los modelos se prosiguió con el cálculo de las 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 ambas particiones de los datos. Este procedimiento de entrenamiento y evaluación del modelo predictivo fue realizado de forma iterativa un total de diez veces por medio de la variación de las semillas aleatorias que determinaron las particiones que dieron origen a cada uno de los subconjuntos de datos. Finalizada la iteración, se llevó a cabo el cálculo promedio de las métricas de desempeño para las diez corridas del procedimiento tanto para la base de datos de entrenamiento como la de testeo. Finalmente se compararon los resultados promedio de las métricas de desempeño a partir de los cuales se determinó cuál de los algoritmos predictivos fue el que generó el ajuste de mayor capacidad predictiva.

Optimización modelo predictivo

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 obtener los valores de los mismos que garantizaron el ajuste de mayor desempeño predictivo. Además, para el caso de las redes neuronales se profundizó en el estudio de diferentes arquitecturas para llevar a cabo el modelado predictivo.

Criterio para interpretar los resultados del modelo

Las contribuciones de cada una de las variables a las predicciones se cuantificaron con el método de los valores de las explicaciones aditivas de Shapley (SHAP) (Lubo-Robles et al., 2020). Los valores SHAP son una medida de atribución a las variables que permite interpretar modelos complejos de aprendizaje automático. Cada valor SHAP es la contribución de cada 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 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 (Figura 3.3). 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.

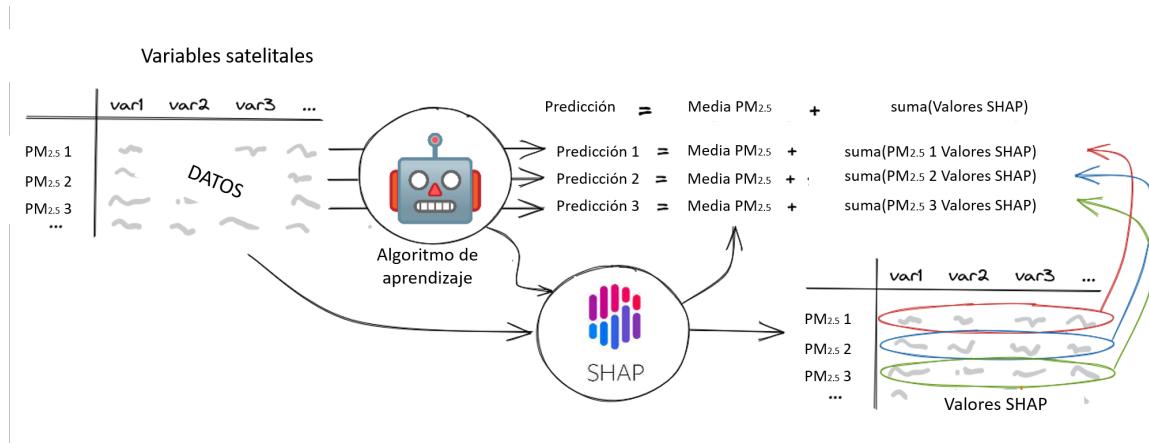


Figura 3.3: Diagrama de funcionamiento del algoritmo de valores SHAP.

Evaluación del modelo predictivo

Para comparar el desempeño predictivo de los distintos algoritmos ajustados tanto en el conjunto de datos de entrenamiento como en el de testeo, se implementaron cinco métricas de rendimiento: la raíz del error cuadrático medio de predicción (*RMSE*), el error absoluto medio (*MAE*), el coeficiente de determinación (R^2), el coeficiente de correlación al cuadrado (r^2) y el coeficiente de concordancia de Lin (*CCC*).

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 3). 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; Zhang et al., 2014).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (3)$$

Ecuación 3: 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.
- **ŷ_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.

Error absoluto medio (*mean absolute error (MAE)*): mide la media de las diferencias absolutas entre los valores predichos por el modelo y los valores observados, siendo menos sensible a valores atípicos a diferencia del *RMSE* (Ecuación 4). Una ventaja

del MAE es que es fácil de interpretar, ya que se mide en las mismas unidades que la variable de respuesta. Un valor de MAE más bajo indica que el modelo es más preciso.

$$MAE = \sum_{i=1}^N |y_i - \hat{y}_i| \quad (4)$$

Ecuación 4: Error absoluto medio (*mean absolute error (MAE)*).

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.

Otra métrica de evaluación de modelos predictivos es el coeficiente de determinación (*coefficient of determination (R^2)*) (Ecuación 5). Calculada bajo las predicciones en la base de datos de entrenamiento estima la proporción de la varianza de la variable respuesta que puede ser explicada por el modelo, brindando una medida de bondad de ajuste del modelo. Por otra parte, bajo una base de datos de testeo representa la capacidad predictiva del modelo ajustado. El R^2 toma valores entre 0 y 1, donde 0 indica que el modelo no explica nada de la variabilidad de la variable respuesta (entrenamiento), o bien que la capacidad predictiva del modelo es deficiente (testeo), mientras que 1 indica que el modelo explica la totalidad de la variabilidad (entrenamiento), o bien que estamos frente a un modelo perfecto (testeo).

$$R^2 = \frac{SSR}{SST} \quad (5)$$

Ecuación 5: coeficiente de determinación (*coefficient of determination (R^2)*).

Componentes:

- **SSR (suma de cuadrados de la regresión o *Sum of Squares due to Regression*):** representa la variación explicada por el modelo de regresión. Mide cuánto de la variación en la variable dependiente puede ser explicada por las variables independientes o predictoras del modelo. En otras palabras, la SSR representa la contribución del modelo a la explicación de la variación en los datos.
- **SST (suma de cuadrados totales o *Total Sum of Squares*):** representa la variación total en los valores observados de la variable dependiente.

Donde SSR es la suma de cuadrados de la regresión también conocida como variación explicada por el modelo (Ecuación 6) y SST es la variación total en los datos también conocida como suma de cuadrados totales (Ecuación 7).

$$SSR = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2 \quad (6)$$

Ecuación 6: Suma de cuadrados de la regresión o *Sum of Squares due to Regression*.

Componentes:

- \hat{y}_i : predicción del modelo para el i-ésimo punto de datos.
- \bar{y} : es la media de los valores observados y_i .

$$SST = \sum_{i=1}^N (y_i - \bar{y})^2 \quad (7)$$

Ecuación 7: Suma de cuadrados totales o *Total Sum of Squares*.

Componentes:

- y_i : valores observados.
- \bar{y} : es la media de los valores observados y_i .

Donde y_i es el valor y para la observación i, \bar{y} es la media del valor y y \hat{y}_i es el valor predicho de y para la observación i.

El coeficiente de correlación de Pearson (*Pearson's correlation coefficient (r)*) es una medida estadística que indica el grado de covariación lineal entre dos variables. Este oscila entre valores de 1 y -1, un valor de 1 indica una correlación lineal positiva perfecta, lo que significa que cuando una variable aumenta, la otra variable también aumenta en la misma proporción (Ecuación 8). Por otra parte un valor de -1 indica una correlación lineal negativa perfecta, lo que significa que cuando una variable aumenta, la otra variable disminuye proporcionalmente. Finalmente un valor de 0 indica que no hay correlación entre las variables involucradas. Es importante destacar que el coeficiente de correlación de Pearson solo mide la asociación lineal entre dos variables, por lo que puede no ser adecuado para evaluar la relación entre variables cuando esta no es lineal. Cabe destacar que en esta tesis la presente métrica se implementa para evaluar tanto la bondad de ajuste como la capacidad predictiva del modelo ajustado, por este motivo las variables comparadas son los valores observados contra

los predichos por el modelo ajustado. Su cálculo bajo los resultados de una base de datos de entrenamiento permiten obtener una medida de bondad de ajuste, mientras que bajo los de una de testeo reflejan la capacidad predictiva del modelo ajustado. Se espera que mientras mejor sea el ajuste del modelo, más cercano a 1 sea el valor de esta métrica.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2(y_i - \bar{y})^2}} \quad (8)$$

Ecuación 8: Coeficiente de correlación de Pearson (*Pearson's correlation coefficient*).

Componentes:

- x_i : valor de la variable x en el i-ésimo punto de datos. Son los valores observados de la primera variable.
- \bar{x} : es la media (promedio) de los valores de la variable x, es decir, la suma de todos los valores de x dividida por n.
- y_i : valor de la variable y en el i-ésimo punto de datos. Son los valores observados de la segunda variable.
- \bar{y} : es la media (promedio) de los valores de la variable y, es decir, la suma de todos los valores de y dividida por n.

Otra métrica útil para evaluar el ajuste de un modelo es el coeficiente de concordancia de Lin (*Lin's Concordance Correlation Coefficient (CCC)*), este evalúa la concordancia entre los valores observados y predichos por el modelo ajustado (Ecuación 9). A diferencia del coeficiente de correlación de Pearson, el *CCC* mide la concordancia o acuerdo entre las variables comparadas. Los valores que toma el *CCC* oscilan entre 1 y -1, donde un valor de 1 indica una concordancia perfecta, mientras que un valor de -1 indica una discordancia perfecta. Un valor de 0 indica que no hay concordancia entre ambas variables. Es importante destacar que el *CCC* es una métrica más robusta que el coeficiente de correlación de Pearson, ya que no se ve afectado por valores atípicos o distribuciones no normales. Es importante destacar que los resultados obtenidos bajo una base de datos de entrenamiento permiten evaluar la bondad de ajuste del modelo predictivo, mientras que bajo los de una de testeo refleja la capacidad predictiva del modelo ajustado.

$$\frac{2s_{12}}{(\bar{x}_1 - \bar{x}_2)^2 + s_1^2 + s_2^2} \quad (9)$$

Ecuación 9: Coeficiente de concordancia de Lin (*Lin's Concordance Correlation Coefficient*). Componentes:

- s_{12} : covarianza entre las dos mediciones o poblaciones. Mide cómo las dos mediciones varían juntas. Una covarianza positiva indica una relación positiva, mientras que una covarianza negativa indica una relación negativa.
- \bar{x} : media de las dos mediciones o poblaciones. Estas medias representan el promedio de las mediciones en cada conjunto de datos.
- s^2_i : varianza de las dos mediciones o poblaciones. Estas varianzas miden la dispersión o variabilidad de las mediciones en cada conjunto de datos.

Donde \bar{x}_1 y \bar{x}_2 son la media de cada una de las variables, s_1^2 y s_2^2 son la varianza de cada una de las variables y s_{12} es su covarianza.

Según las métricas calculadas sobre los datos de testeo, valores más bajos de RMSE y MAE y más altos de R^2 , r y CCC indican un rendimiento del modelo más deseable y predicciones más acertadas, o bien con menos errores.

Por otra parte, estos resultados de RMSE, MAE y R^2 evaluadas sobre los datos de entrenamiento, proveen medidas de bondad de ajuste, indicativas de la capacidad del modelo ajustado para explicar la variabilidad en los datos.

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 la hora que no se tiene dato como la media de la concentración en la hora anterior y posterior. Cabe destacar 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 superior o similar al de este modelo de referencia refleja una elevada capacidad predictiva del modelo ajustado.

3.3. Resultados

3.3.1. Comparación de algoritmos para el uso del suelo urbano

Base de datos original

En esta sección se desarrolló el modelado predictivo a partir de la base de datos que involucró todas las variables obtenidas a partir de información satelital, una descripción detallada de cada una de ellas puede hallarse en el Anexo 1. La base de datos original empleada para llevar a cabo el modelado predictivo involucró un total de 193 variables predictoras. Los resultados del modelo predictivo desarrollado en esta sección se presentan en las tablas 3.2 y 3.3.

Tabla 3.2: Métricas de ajuste de modelos predictivos aplicados en el contexto de aprendizaje automático al set de datos de entrenamiento para el uso del suelo urbano. Los valores de *RMSE* y *MAE* se encuentran en las mismas unidades que la concentración de PM_{2.5}, es decir [$\frac{\mu\text{g}}{\text{m}^3}$].

Modelo	<i>RMSE</i>	<i>MAE</i>	<i>R</i> ²	<i>r</i>	<i>CCC</i>
Redes neuronales recurrentes	1,657	1,174	0,972	0,987	0,986
Bosques Aleatorios	6,823	4,919	0,529	0,754	0,639
Regresión lineal múltiple	7,840	5,788	0,378	0,615	0,548
<i>Modelo de referencia</i>	3,740	2,316	0,859	0,929	0,929

Raíz del error cuadrático medio (*RMSE*), error absoluto medio (*MAE*), coeficiente de determinación (*R*²), coeficiente de correlación de Pearson (*r*) y coeficiente de concordancia de Lin (*CCC*).

Tabla 3.3: Métricas de ajuste de los algoritmos predictivos de aprendizaje automático en el set de datos de testeo para el uso del suelo urbano. Los valores de *RMSE* y *MAE* se encuentran en las mismas unidades que la variable respuesta, es decir [$\frac{\mu\text{g}}{\text{m}^3}$].

Modelo	<i>RMSE</i>	<i>MAE</i>	<i>R</i> ²	<i>r</i>	<i>CCC</i>
Redes neuronales recurrentes	3,747	2,472	0,857	0,927	0,924
Bosques Aleatorios	7,135	5,168	0,482	0,715	0,601
Regresión lineal múltiple	7,929	5,888	0,360	0,601	0,540
<i>Modelo de referencia</i>	3,740	2,326	0,858	0,930	0,930

Raíz del error cuadrático medio (*RMSE*), error absoluto medio (*MAE*), coeficiente de determinación (*R*²), coeficiente de correlación de Pearson (*r*) y coeficiente de concordancia de Lin (*CCC*).

Los resultados observados en las tablas 3.2 y 3.3 muestran que las redes neuronales recurrentes demostraron ser el algoritmo predictivo de mejor desempeño para el uso del suelo urbano bajo la base de datos original tanto en el conjunto de datos de entrenamiento como en el de testeo. El desempeño predictivo alcanzado demuestra cómo las características propias de aprendizaje del algoritmo logran captar las fuentes de variabilidad y realizar predicciones acertadas de la variable respuesta. Además, puede apreciarse que los bosques aleatorios y la regresión lineal múltiple no lograron superar el modelo de referencia propuesto, demostrando una menor capacidad predictiva que las redes neuronales recurrentes para la serie de PM_{2.5} para el uso del suelo urbano.

Interpretación del modelo

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

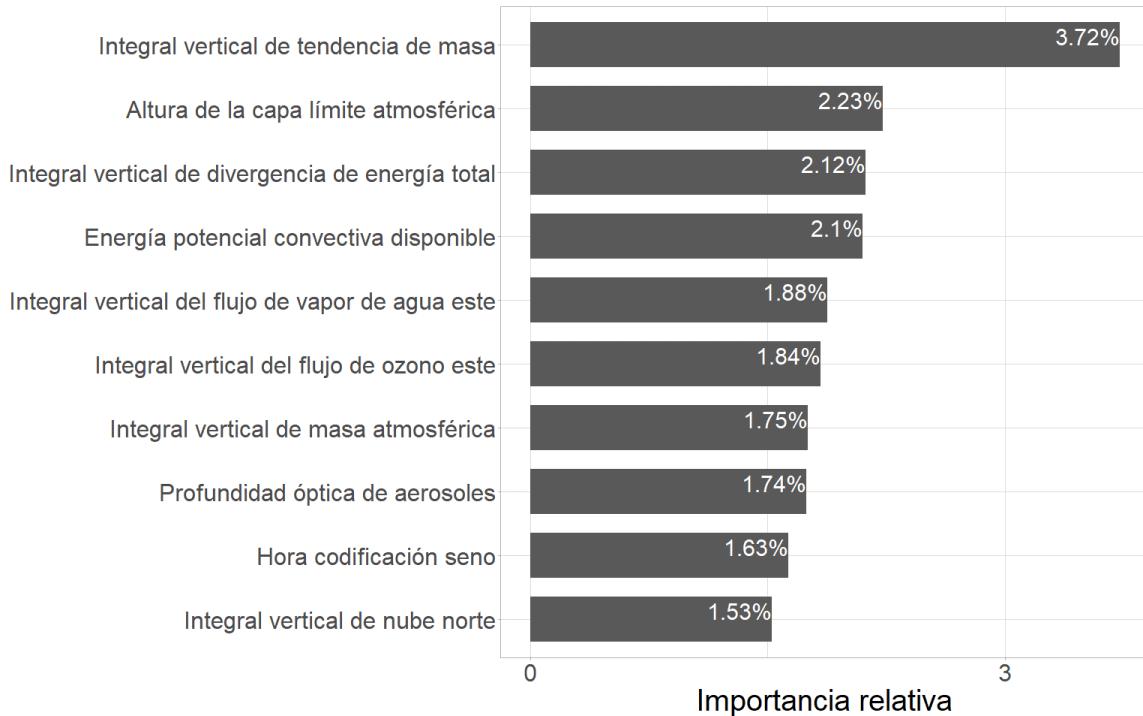


Figura 3.4: Importancia relativa de las variables en el modelo predictivo para la serie de datos urbanos bajo el enfoque de la base de datos original.

A partir de la figura 3.4 pueden apreciarse las diez variables de mayor importancia relativa para el modelo predictivo en el uso del suelo urbano bajo la base de datos original. Entre estas se destaca que nueve de las diez son variables derivadas de información satelital. Por otra parte, la variable hora codificación seno que se encuentra vinculada a la escala temporal, sugiere estructura de variación periódica en la concentración de PM_{2.5}. Los resultados muestran que estas diez variables reúnen el 20 % de la importancia relativa, a su vez las cinco primeras reúnen el 12 % y las dos primeras menos del 6 %.

Base de datos reducida por el algoritmo LASSO

En esta sección se desarrolló el modelado predictivo a partir de la base de datos reducida por medio del algoritmo de selección de variables de *LASSO*. La base de datos original quedó reducida a un total de 45 variables predictoras. Los resultados

del modelo predictivo desarrollado en esta sección se presentan en las tablas 3.4 y 3.5.

Tabla 3.4: Métricas de ajuste de modelos predictivos aplicados en el contexto de aprendizaje automático al set de datos de entrenamiento para el uso del suelo urbano. Los valores de $RMSE$ y MAE se encuentran en las mismas unidades que la concentración de $PM_{2.5}$, es decir $[\frac{\mu g}{m^3}]$.

Modelo	$RMSE$	MAE	R^2	r	CCC
Redes neuronales recurrentes	2,091	1,513	0,955	0,978	0,977
Bosques Aleatorios	6,756	4,846	0,538	0,755	0,653
Regresión lineal múltiple	8,175	6,052	0,323	0,569	0,489
<i>Modelo de referencia</i>	3,740	2,316	0,859	0,929	0,929

Raíz del error cuadrático medio ($RMSE$), error absoluto medio (MAE), coeficiente de determinación (R^2), coeficiente de correlación de Pearson (r) y coeficiente de concordancia de Lin (CCC).

Tabla 3.5: Métricas de ajuste de los algoritmos predictivos de aprendizaje automático en el set de datos de testeo para el uso del suelo urbano. Los valores de $RMSE$ y MAE se encuentran en las mismas unidades que la variable respuesta, es decir $[\frac{\mu g}{m^3}]$.

Modelo	$RMSE$	MAE	R^2	r	CCC
Redes neuronales recurrentes	3,804	2,506	0,853	0,925	0,921
Bosques Aleatorios	7,049	5,068	0,494	0,719	0,619
Regresión lineal múltiple	8,180	6,062	0,319	0,566	0,488
<i>Modelo de referencia</i>	3,740	2,326	0,858	0,930	0,930

Raíz del error cuadrático medio ($RMSE$), error absoluto medio (MAE), coeficiente de determinación (R^2), coeficiente de correlación de Pearson (r) y coeficiente de concordancia de Lin (CCC).

Los resultados observados en la tabla 3.4 y 3.5 muestran que para el caso de una base de datos reducida bajo el algoritmo de LASSO las redes neuronales recurrentes demuestran ser el algoritmo predictivo de mayor desempeño para el uso del suelo urbano tanto en el conjunto de datos de entrenamiento como en el de testeo. Estos resultados son similares a los obtenidos bajo la base de datos original (Tabla 3.2 y 3.3). Los resultados obtenidos muestran una capacidad predictiva superior para el caso de las redes neuronales recurrentes en comparación con los demás algoritmos.

Interpretación del modelo

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

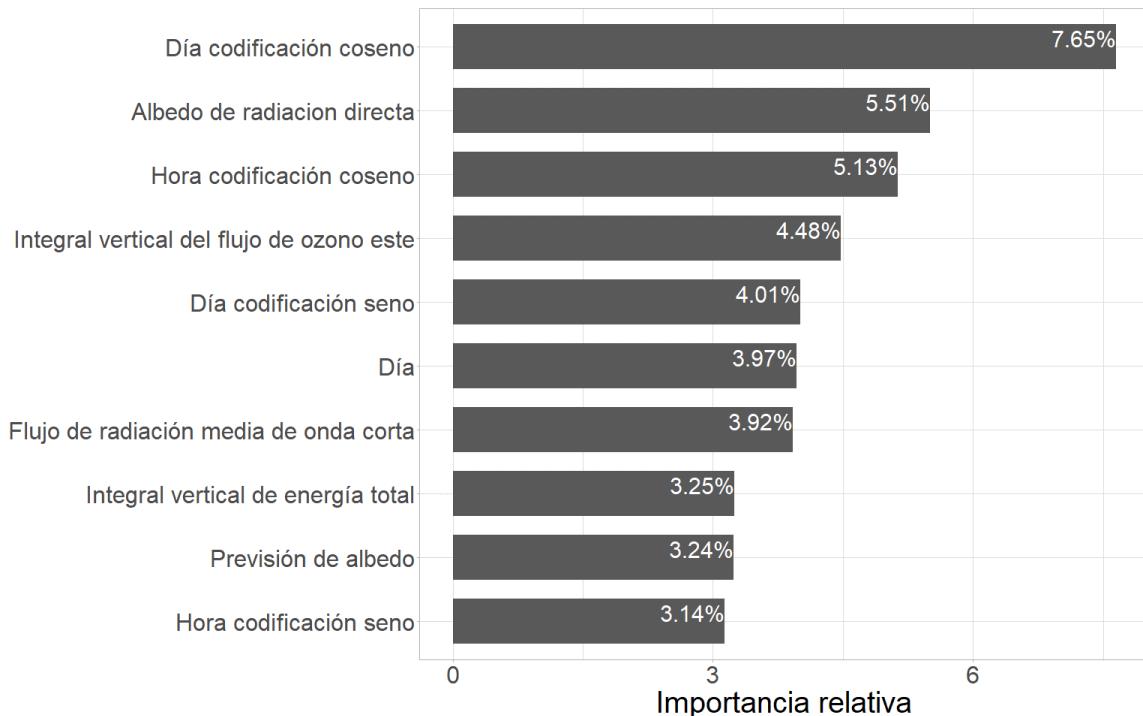


Figura 3.5: Importancia relativa de las variables en el modelo para la serie de datos urbanos, bajo una base de datos reducida por medio del algoritmo de LASSO.

A partir de la figura 3.5 pueden apreciarse las diez variables de mayor importancia relativa para el modelo predictivo en el uso del suelo urbano bajo una base de datos

reducida por el algoritmo de LASSO. Entre estas se destaca que cinco de las diez son variables obtenidas a partir de información satelital, mientras que las restantes cinco son variables temporales, sugiriendo una estructura de variación periódica en la concentración de PM_{2.5}. Los resultados muestran que estas diez variables reúnen el 44 % de la importancia relativa, a su vez las cinco primeras reúnen el 26 % y las dos primeras menos del 15 %.

Base de datos reducida por el algoritmo featurewiz

En esta sección se desarrolló el modelado predictivo a partir de una base de datos reducida por medio del algoritmo de selección de variables de *featurewiz*. La base de datos original quedó reducida a un total de 66 variables. Los resultados del modelo predictivo desarrollado en esta sección se presentan en las tablas 3.6 y 3.7.

Tabla 3.6: Métricas de ajuste de modelos predictivos aplicados en el contexto de aprendizaje automático al set de datos de entrenamiento para el uso del suelo urbano. Los valores de *RMSE* y *MAE* se encuentran en las mismas unidades que la concentración de PM_{2.5}, es decir [$\frac{\mu g}{m^3}$]

Modelo	<i>RMSE</i>	<i>MAE</i>	<i>R</i> ²	<i>r</i>	<i>CCC</i>
Redes neuronales recurrentes	1,943	1,390	0,960	0,981	0,980
Bosques Aleatorios	6,833	4,922	0,527	0,753	0,638
Regresión lineal múltiple	8,234	6,121	0,313	0,560	0,477
<i>Modelo de referencia</i>	3,740	2,316	0,859	0,929	0,929

Raíz del error cuadrático medio (*RMSE*), error absoluto medio (*MAE*), coeficiente de determinación (*R*²), coeficiente de correlación de Pearson (*r*) y coeficiente de concordancia de Lin (*CCC*).

Tabla 3.7: Métricas de ajuste de los algoritmos predictivos de aprendizaje automático en el set de datos de testeo para el uso del suelo urbano. Los valores de *RMSE* y *MAE* se encuentran en las mismas unidades que la variable respuesta, es decir [$\frac{\mu g}{m^3}$].

Modelo	<i>RMSE</i>	<i>MAE</i>	<i>R</i> ²	<i>r</i>	<i>CCC</i>
Redes neuronales recurrentes	3,897	2,587	0,845	0,921	0,918
Bosques Aleatorios	7,143	5,173	0,481	0,714	0,600
Regresión lineal múltiple	8,265	6,173	0,305	0,553	0,473
<i>Modelo de referencia</i>	3,740	2,326	0,858	0,930	0,930

Raíz del error cuadrático medio (*RMSE*), error absoluto medio (*MAE*), coeficiente de determinación (*R*²), coeficiente de correlación de Pearson (*r*) y coeficiente de concordancia de Lin (*CCC*).

Los resultados observados en la tabla 3.6 y 3.7 muestran que para el caso de una base de datos reducida bajo el algoritmo de *featurewiz* las redes neuronales recurrentes demuestran ser el algoritmo predictivo de mayor desempeño para el uso del suelo urbano tanto en el conjunto de datos de entrenamiento como en el de testeo. Estos resultados son similares a los obtenidos para los anteriores enfoques evaluados (Tabla 3.2, 3.3, 3.4 y 3.5).

Interpretación del modelo

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

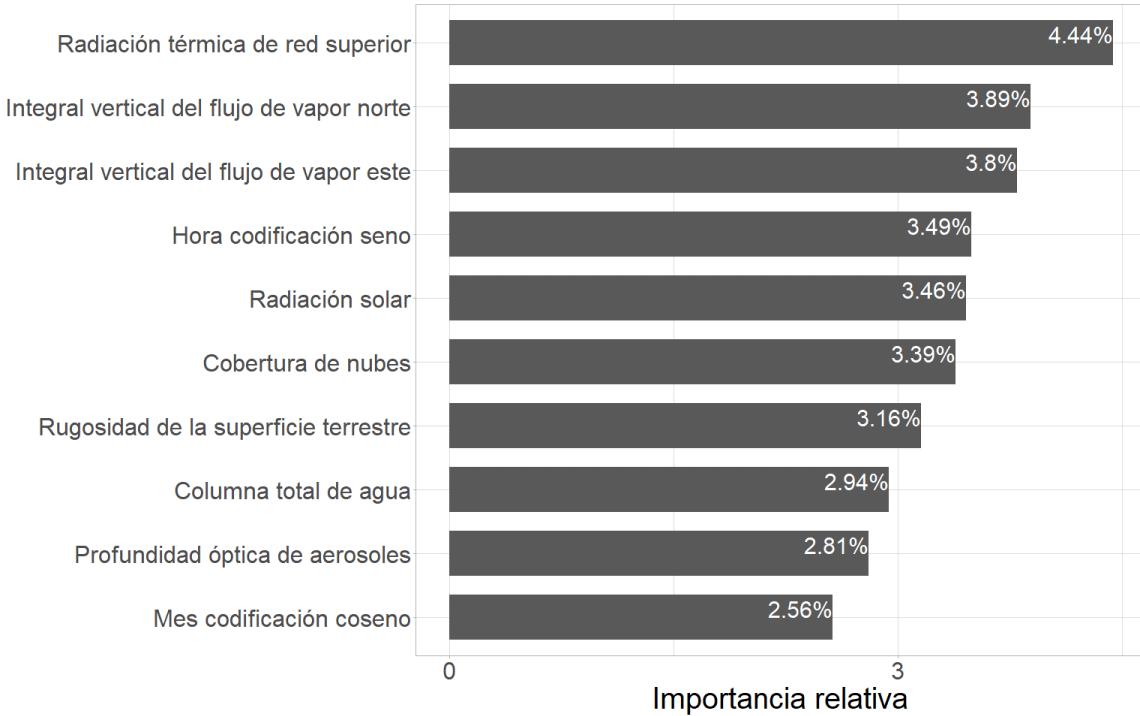


Figura 3.6: Importancia relativa de las variables en el modelo predictivo para la serie de datos urbanos bajo una base de datos reducida por medio del algoritmo de featurewiz.

A partir de la figura 3.6 pueden apreciarse las diez variables de mayor importancia relativa para el modelo predictivo en el uso del suelo urbano bajo una base de datos reducida por medio del algoritmo de featurewiz. Entre estas se destaca que ocho de las diez son variables obtenidas a partir de información satelital, demostrando la dependencia existente de la variable respuesta respecto de estas. Por otra parte, dos de las variables se hallan vinculadas a la escala temporal, exhibiendo la dependencia de la variable respuesta frente a sus patrones de variación temporal. Los resultados muestran que estas diez variables reúnen el 34 % de la importancia relativa, a su vez las cinco primeras reúnen el 20 % y las dos primeras menos del 8 %.

3.3.2. uso del suelo agrícola y área verde

Base de datos original

En esta sección se desarrolló el modelado predictivo a partir de la base de datos original que cuenta con un total de 193 variables predictoras, las cuales se encuentran descriptas en el Anexo 1. Los resultados del modelo predictivo desarrollado en esta sección se presentan en las tablas 3.8 y 3.9.

Tabla 3.8: Métricas de ajuste de modelos predictivos aplicados en el contexto de aprendizaje automático al set de datos de entrenamiento para el uso de agrícola y área verde. Los valores de *RMSE* y *MAE* se encuentran en las mismas unidades que la concentración de PM_{2.5}, es decir [$\frac{\mu\text{g}}{\text{m}^3}$].

Modelo	<i>RMSE</i>	<i>MAE</i>	<i>R</i> ²	<i>r</i>	<i>CCC</i>
Bosques Aleatorios	1,016	0,657	0,973	0,989	0,985
Redes neuronales recurrentes	1,363	0,940	0,950	0,975	0,973
Regresión lineal múltiple	4,941	3,681	0,356	0,597	0,525
<i>Modelo de referencia</i>	2,813	1,569	0,792	0,896	0,896

Raíz del error cuadrático medio (*RMSE*), error absoluto medio (*MAE*), coeficiente de determinación (*R*²), coeficiente de correlación de Pearson (*r*) y coeficiente de concordancia de Lin (*CCC*).

Tabla 3.9: Métricas de ajuste de los algoritmos predictivos de aprendizaje automático en el set de datos de testeo para el uso de agrícola y área verde. Los valores de *RMSE* y *MAE* se encuentran en las mismas unidades que la variable respuesta, es decir [$\frac{\mu\text{g}}{\text{m}^3}$].

Modelo	<i>RMSE</i>	<i>MAE</i>	<i>R</i> ²	<i>r</i>	<i>CCC</i>
Bosques Aleatorios	2,673	1,749	0,810	0,903	0,888
Redes neuronales recurrentes	2,846	1,843	0,784	0,887	0,881

Modelo	<i>RMSE</i>	<i>MAE</i>	<i>R</i> ²	<i>r</i>	<i>CCC</i>
Regresión lineal múltiple	5,013	3,741	0,334	0,579	0,511
<i>Modelo de referencia</i>	2,718	1,554	0,804	0,903	0,903

Raíz del error cuadrático medio (*RMSE*), error absoluto medio (*MAE*), coeficiente de determinación (*R*²), coeficiente de correlación de Pearson (*r*) y coeficiente de concordancia de Lin (*CCC*).

Los resultados observados en las tablas 3.8 y 3.9 muestran que el algoritmo predictivo de mayor capacidad son los bosques aleatorios para el uso del suelo agrícola y área verde bajo la base de datos original. Tanto las redes neuronales como los bosques aleatorios superan el modelo de referencia propuesto en la base de datos de entrenamiento, a pesar de esto el algoritmo de bosques aleatorios muestra un mayor exactitud en sus predicciones. Por otra parte, las métricas obtenidas para la base de datos de testeo evidencian que solo el algoritmo de bosques aleatorios supera el modelo de referencia propuesto.

Interpretación del modelo

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

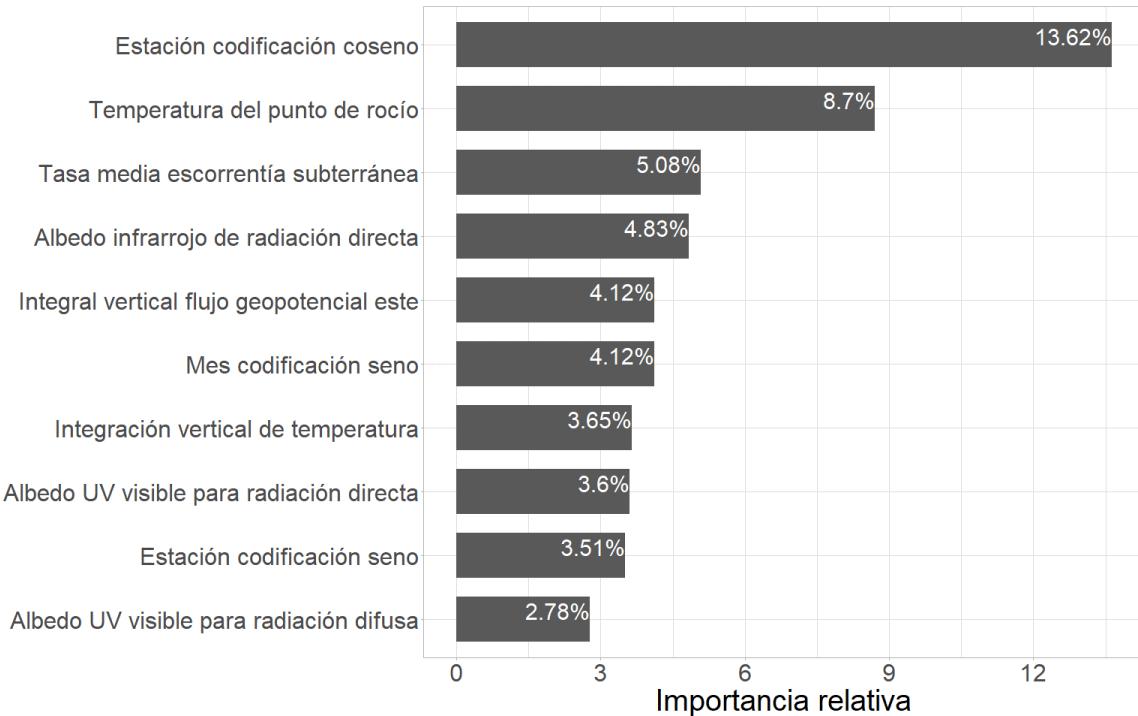


Figura 3.7: Importancia relativa de las variables en el modelo predictivo para la serie de datos agrícola y área verde en la base de datos original.

A partir de la figura 3.7 pueden apreciarse las diez variables de mayor importancia relativa para el modelo predictivo en el uso del suelo agrícola y área verde bajo la base de datos original. Se destaca que tres de las diez son variables temporales, entre ellas se encuentra la variable estación codificación coseno la cual es la de mayor importancia relativa. Por otra parte las restantes son variables derivadas a partir de información satelital demostrando la capacidad de las mismas para explicar la variabilidad de la concentración de PM_{2.5}. Los resultados muestran que estas diez variables reúnen el 55 % de la importancia relativa, a su vez las cinco primeras reúnen el 36 % y las dos primeras menos del 22 %.

Base de datos reducida por el algoritmo de LASSO

En esta sección se desarrolló el modelado predictivo a partir de la base de datos reducida por medio del algoritmo de selección de variables de *LASSO*. La base de

datos original quedó reducida a un total de 40 variables predictoras. Los resultados del modelo predictivo desarrollado en esta sección se presentan en las tablas 3.10 y 3.11.

Tabla 3.10: Métricas de ajuste de modelos predictivos aplicados en el contexto de aprendizaje automático al set de datos de entrenamiento para el uso de agrícola y área verde. Los valores de $RMSE$ y MAE se encuentran en las mismas unidades que la concentración de $PM_{2.5}$, es decir $[\frac{\mu g}{m^3}]$.

Modelo	$RMSE$	MAE	R^2	r	CCC
Bosques Aleatorios	0,975	0,625	0,975	0,990	0,986
Redes neuronales recurrentes	1,530	1,051	0,937	0,970	0,967
Regresión lineal múltiple	5,250	3,879	0,273	0,523	0,429
<i>Modelo de referencia</i>	2,813	1,569	0,792	0,896	0,896

Raíz del error cuadrático medio ($RMSE$), error absoluto medio (MAE), coeficiente de determinación (R^2), coeficiente de correlación de Pearson (r) y coeficiente de concordancia de Lin (CCC).

Tabla 3.11: Métricas de ajuste de los algoritmos predictivos de aprendizaje automático en el set de datos de testeo para el uso de agrícola y área verde. Los valores de $RMSE$ y MAE se encuentran en las mismas unidades que la variable respuesta, es decir $[\frac{\mu g}{m^3}]$.

Modelo	$RMSE$	MAE	R^2	r	CCC
Bosques Aleatorios	2,550	1,650	0,827	0,912	0,899
Redes neuronales recurrentes	2,784	1,796	0,794	0,893	0,888
Regresión lineal múltiple	5,241	3,900	0,272	0,523	0,430
<i>Modelo de referencia</i>	2,718	1,554	0,804	0,903	0,903

Raíz del error cuadrático medio ($RMSE$), error absoluto medio (MAE), coeficiente de determinación (R^2), coeficiente de correlación de Pearson (r) y coeficiente de concordancia de Lin (CCC).

Los resultados observados en las tablas 3.10 y 3.11, muestran que el algoritmo predictivo de mayor capacidad son los bosques aleatorios para el uso del suelo agrícola y área verde bajo la base de datos reducida por medio del algoritmo de *LASSO*. Estos resultados fueron similares a los exhibidos para el caso de la base de datos original (Tabla 3.8 y 3.9).

Interpretación del modelo

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

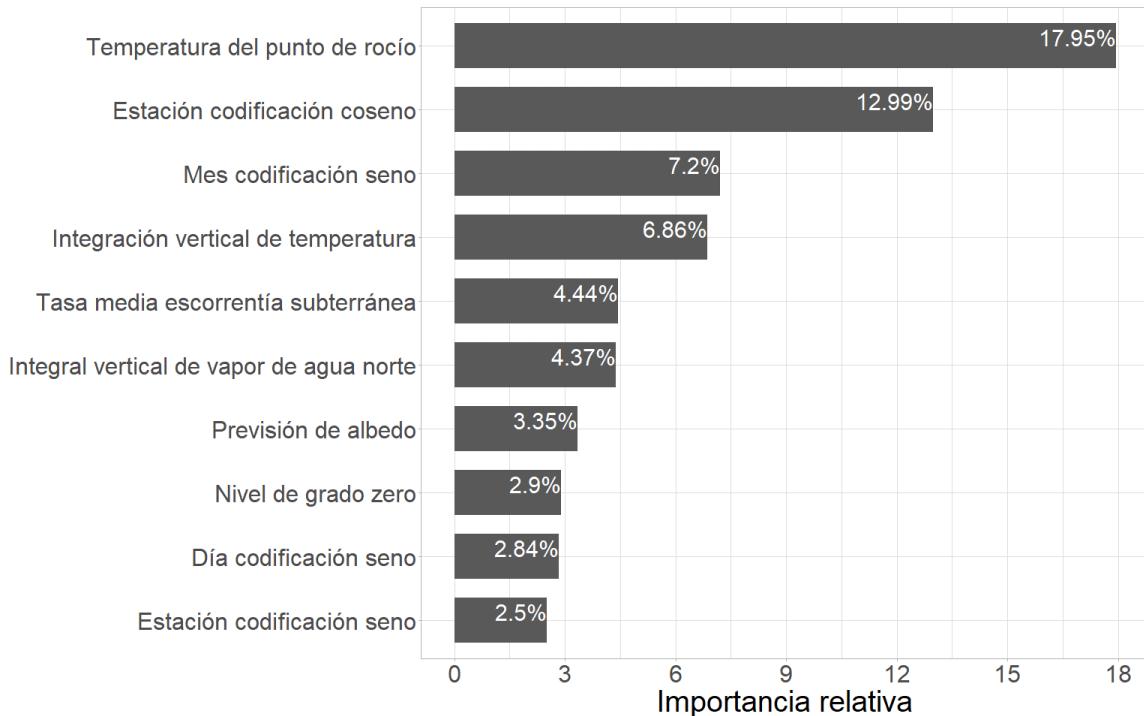


Figura 3.8: Importancia relativa de las variables en el modelo para la serie de datos agrícola y área verde para una base de datos reducida por medio del algoritmo de LASSO.

A partir de la figura 3.8 pueden apreciarse las diez variables de mayor importancia relativa para el modelo predictivo en el uso del suelo agrícola y área verde bajo una base de datos reducida por medio del algoritmo de LASSO. Entre estas se destaca que seis de las diez, son variables obtenidas a partir de información satelital, demostrando la

dependencia de la variable repuesta respecto de estas. Por otra parte las restantes son variables vinculada a la escala temporal, lo cual exhibe la importancia de contemplar los patrones de variación temporales a la hora de predecir la concentración de PM_{2.5}. Los resultados muestran que estas diez variables reúnen el 65 % de la importancia relativa, a su vez las cinco primeras reúnen el 50 % y las dos primeras menos del 30 %.

Base de datos reducida por el algoritmo de featurewiz

En esta sección se desarrolló el modelado predictivo a partir de una base de datos reducida por medio del algoritmo de selección de variables de *featurewiz*. La base de datos original quedó reducida a un total de 54 variables. Los resultados del modelo predictivo desarrollado en esta sección se presentan en las tablas 3.12 y 3.13.

Tabla 3.12: Métricas de ajuste de modelos predictivos aplicados en el contexto de aprendizaje automático al set de datos de entrenamiento para el uso de agrícola y área verde. Los valores de *RMSE* y *MAE* se encuentran en las mismas unidades que la concentración de PM_{2.5}, es decir [$\frac{\mu g}{m^3}$].

Modelo	RMSE	MAE	<i>R</i> ²	<i>r</i>	CCC
Bosques Aleatorios	1,020	0,662	0,973	0,989	0,985
Redes neuronales recurrentes	1,432	0,992	0,944	0,973	0,971
Regresión lineal múltiple	5,176	3,850	0,293	0,542	0,454
<i>Modelo de referencia</i>	2,813	1,569	0,792	0,896	0,896

Raíz del error cuadrático medio (*RMSE*), error absoluto medio (*MAE*), coeficiente de determinación (*R*²), coeficiente de correlación de Pearson (*r*) y coeficiente de concordancia de Lin (CCC).

Tabla 3.13: Métricas de ajuste de los algoritmos predictivos de aprendizaje automático en el set de datos de testeo para el uso de agrícola y área verde. Los valores de $RMSE$ y MAE se encuentran en las mismas unidades que la variable respuesta, es decir [$\frac{\mu g}{m^3}$].

Modelo	$RMSE$	MAE	R^2	r	CCC
Bosques Aleatorios	2,702	1,769	0,806	0,901	0,885
Redes neuronales recurrentes	2,819	1,820	0,788	0,890	0,886
Regresión lineal múltiple	5,197	3,878	0,284	0,534	0,448
<i>Modelo de referencia</i>	2,718	1,554	0,804	0,903	0,903

Raíz del error cuadrático medio ($RMSE$), error absoluto medio (MAE), coeficiente de determinación (R^2), coeficiente de correlación de Pearson (r) y coeficiente de concordancia de Lin (CCC).

Los resultados observados en las tablas 3.12 y 3.13, muestran que el algoritmo predictivo de mayor capacidad son los bosques aleatorios para el uso del suelo agrícola y área verde bajo una base reducida por medio del algoritmo de *featurewiz*. Estos resultados fueron similares a los exhibidos para el caso de la base de datos original y la reducida por medio del algoritmo de *LASSO* (Tabla 3.8, 3.9, 3.10 y 3.11).

Interpretación del modelo

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

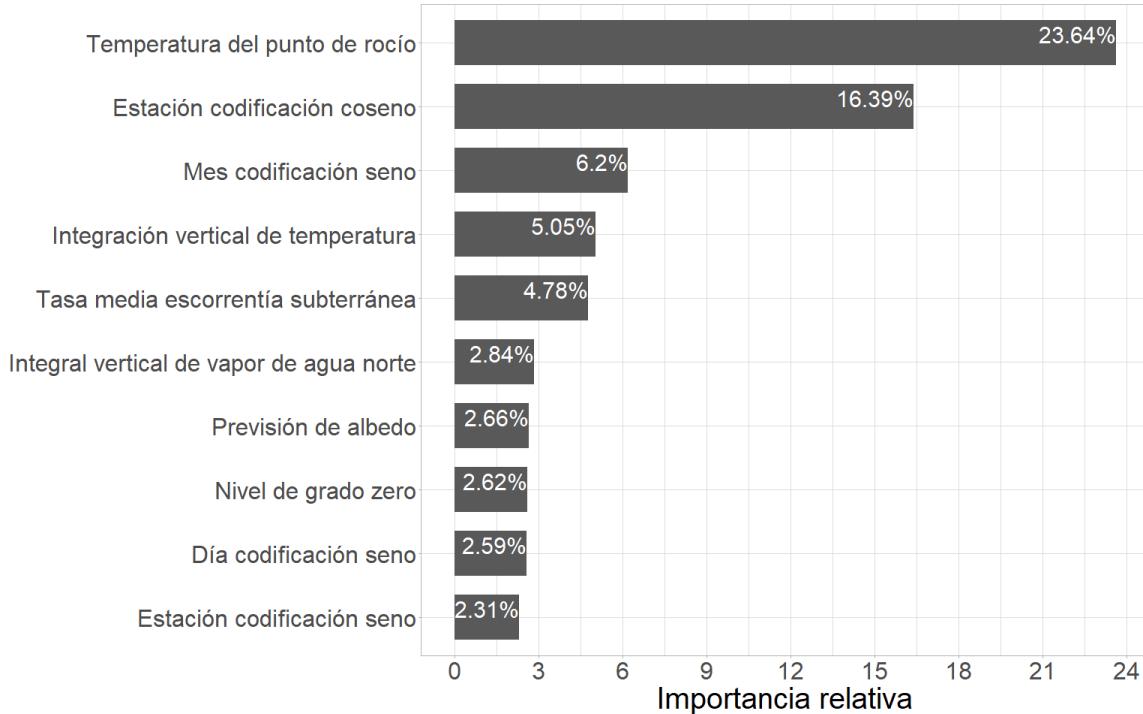


Figura 3.9: Importancia relativa de las variables en el modelo para la serie de datos agrícola y área verde para una base de datos reducida bajo el algoritmo de *featurewiz*.

A partir de la figura 3.9 pueden apreciarse las diez variables de mayor importancia relativa para el modelo predictivo en el uso del suelo agrícola y área verde bajo una base de datos reducida por medio del algoritmo de *featurewiz*. Entre estas se destaca que seis de las diez son variables obtenidas a partir de información satelital, mientras que las restantes son variables vinculadas a la escala temporal. Los resultados muestran que estas diez variables reúnen el 69 % de la importancia relativa, a su vez las cinco primeras reúnen el 56 % y las dos primeras menos del 40 %.

3.4. Discusión

Es interesante destacar que para el uso del suelo urbano los distintos enfoques de bases de datos exhiben resultados similares en cuanto al algoritmo predictivo de mayor capacidad. El algoritmo de redes neuronales recurrentes obtuvo los resultados

de mayor exactitud no solo bajo todos los enfoques de bases de datos, si no también para cada una de sus particiones (entrenamiento y testeo). Resultados similares fueron hallados por Wongrin et al. (2023), demostrando que las redes neuronales recurrentes, en especial las *LSTM* destacan de entre múltiples algoritmos predictivos como las de mayor precisión (Wongrin et al., 2023). Por otra parte, otros estudios también 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; Feng et al., 2019). En esta tesis la mayor capacidad predictiva para el uso del suelo urbano fue alcanzada por las redes neuronales recurrentes entrenadas bajo la base de datos original, alcanzando una *RMSE* de 1,657 $\frac{\mu\text{g}}{\text{m}^3}$ en el set de datos de entrenamiento y una de 3,747 $\frac{\mu\text{g}}{\text{m}^3}$ en el de testeo.

Por otra parte, en el uso del suelo agrícola y área verde, el algoritmo predictivo que demostró mayor capacidad fue el de bosques aleatorios, manteniendo constantes sus resultados bajo los distintos enfoques de bases de datos evaluados. 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 fue alcanzada por los bosques aleatorios entrenados bajo la base de datos reducida por medio del algoritmo de *LASSO*, alcanzando una *RMSE* de 1,943 $\frac{\mu\text{g}}{\text{m}^3}$ en el set de datos de entrenamiento y una de 3,897 $\frac{\mu\text{g}}{\text{m}^3}$ en el de testeo.

A partir de los resultados obtenidos para los distintos tipos de modelos predictivos ajustados en el contexto de aprendizaje automático en ambos usos de suelo, puede notarse que el algoritmo predictivo de mejor ajuste varía en función del uso del suelo analizado. Este resultado podría encontrarse vinculado no solo a características propias de cada una de las series temporales, si no también a las de aprendizaje de los algoritmos empleados en el modelado predictivo.

En primera instancia, el cambio podría encontrarse vinculado a las características propias de la serie temporal en cada uno de los usos de suelo. En el caso de la serie de tiempo para el 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 (Figuras 3.7, 3.8 y 3.9). Por el contrario la variabilidad fue mayor en la serie de concentración de PM_{2.5} correspondiente al uso del suelo urbano.

Es destacable cómo en las primeras cinco variables se reúne un porcentaje considerable de la importancia relativa para explicar la variable respuesta en el uso del suelo agrícola y área verde. Además, puede destacarse que dentro de las diez variables de mayor importancia relativa, bajo los tres métodos de selección de variables para el modelado predictivo siempre se mantienen constantes seis variables predictoras, difiriendo sutilmente en el orden de importancia que reciben. Estas son: temperatura del punto de rocío, estación codificación coseno, integral vertical de temperatura, mes codificación seno, tasa media de escorrentía subterránea y estación codificación seno. El resultado refleja cómo en un número bajo de variables se concentra la información necesaria para modelar la variabilidad de la serie, independientemente del enfoque de modelado que se esté considerando. La concentración de información se debe a la baja variabilidad de la serie la cual requiere menor información para describir sus fluctuaciones 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 al 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 fueron retenidas en múltiples árboles individuales

del bosque, generando una influencia de mayor peso en las predicciones. Dado el gran porcentaje de información contenida en una cantidad reducida de variables, los bosques aleatorios lograron internalizar de forma efectiva las interrelaciones 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 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. La serie urbana muestra 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 este caso se requiere un número de variables superior para reunir el mismo porcentaje de importancia relativa (Figuras 3.4, 3.5 y 3.6). Es importante destacar que entre los distintos enfoques de modelado predictivo, ninguna de las variables predictoras se repite dentro de las diez de mayor importancia relativa. Esto ocurre como consecuencia de la gran variabilidad existente, lo cual genera que las variables retenidas bajo los distintos métodos de selección de variables difieran notablemente. Es importante destacar que la importancia relativa otorgada a cada una de las variables participantes en el modelo varía en un rango acotado. A diferencia del uso del suelo agrícola y área verde, no hay variables en el uso del suelo urbano que demuestren supremacía en cuanto a su importancia relativa.

Los algoritmos de redes neuronales ajustados no identificaron variables predictoras de gran 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 del uso del suelo urbano, 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 3.4, 3.5 y 3.6, se presentaron los porcentaje de importancia relativa para el uso del suelo urbano, los cuales evidenciaron ser inferiores a los del uso del suelo agrícola y área verde. Resultados similares fueron reportados por Ghassen et. al., quienes demostraron que mejores resultados eran alcanzados mientras un mayor número de variables era involucrado al implementar redes neuronales recurrentes, particularmente LSTM (Chniti et al., 2017).

Dadas las características de la serie temporal urbana el modelado predictivo fue más apto 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, puede 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 una serie 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 resultó en un mayor desempeño predictivo. De esta forma en cada predicción puede verse reflejada la influencia de cada una de las variables predictoras.

Además, las redes neuronales se desempeñaron mejor cuando se trabajó con la base de datos original en comparación con las reducidas; resultado que posiblemente podría encontrarse asociado con la homogeneidad de las influencias relativas de las variables y las características de aprendizaje del algoritmo de redes neuronales recurrentes.

3.5. Conclusión

Los resultados obtenidos indican 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.

La similitud existente entre las métricas obtenidas para el modelo de referencia y las de los modelos predictivos desarrollados demuestran 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.

Las características propias de cada una de las series temporales bajo los distintos usos de suelo y las de aprendizaje de cada uno de los algoritmos predictivos evaluados, fueron determinantes en los resultados de mejor ajuste para cada uno de los usos de suelo.

La serie de tiempo de concentración de PM_{2.5} en el uso del suelo agrícola y área verde exhibe una menor variabilidad, la cual impacta en la cantidad de información necesaria para explicar sus variaciones. Este hecho es sustentado en que bajo los distintos enfoques de bases de datos las variables: temperatura del punto de rocío, estación codificación coseno, integral vertical de temperatura, mes codificación seno, tasa media de escorrentía subterránea y estación codificación seno permanecen siempre contenidas dentro de las diez de mayor importancia relativa. Por otra parte, para la serie de tiempo proveniente de entornos urbanos, al demostrar tener una mayor variabilidad, se necesita mayor cantidad de información para explicar su variación, motivo por el cual los mejores resultados en el modelado predictivo se hallan bajo el enfoque de la base de datos original.

La consistencia de los resultados bajo los distintos enfoques de bases de datos permiten concluir que series temporales con características de mayor variabilidad son más aptas para ser modeladas a través de redes neuronales recurrentes. Mientras que series temporales de menor variabilidad son más aptas para ser modeladas bajo el algoritmo de bosques aleatorios.

Comentarios finales

Los resultados del estudio destacan que la distribución *gamma* fue la que mejor ajustó a los datos de concentración de PM_{2.5} recopilados en diversos sitios de muestreo en el área metropolitana de la ciudad de Córdoba. 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, lo que resalta 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 derivadas a partir de información satelital para explicar la variabilidad de la concentración de PM_{2.5}. Persiguiendo el objetivo del ajuste de modelos predictivos se recopilaron numerosas variables que permitiesen representar las características de la atmósfera en un momento dado. La base de datos final contó con un total de 193 variables, a partir de esta se exploraron distintas técnicas de modelado predictivo. Cabe destacar que el enfoque de modelado se encontró enmarcado en uno meramente computacional y no mecanicista. La diferencia reside que en este caso el objetivo principal es obtener un modelo de elevada capacidad predictiva, a diferencia del otro enfoque que busca inferir la relación de vinculación entre las variables involucradas.

Los resultados obtenidos indicaron 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, obteniendo resultados comparables e incluso superiores al modelo base propuesto. Cabe destacar que la predicción con cinco días de anticipación posee potencial para alertar a la población ante condiciones adversas de calidad de aire. Además, es importante notar que las características de la series temporales bajo cada uno de los usos de suelo y las de aprendizaje de cada algoritmo fueron determinantes en los resultados de mejor ajuste. La serie de tiempo de concentración de PM_{2.5} en el uso de suelo agrícola y área verde exhibió una menor variabilidad, requiriendo menor cantidad de información para explicar sus variaciones. Finalmente los algoritmos de redes neuronales recurrentes resultaron más aptos para modelar la serie de mayor variabilidad, mientras que la serie de menor variabilidad obtuvo resultados de mayor capacidad bajo el algoritmo de bosques aleatorios.

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 dos series una que demuestre un alto grado de variabilidad y otra con uno 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.

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: 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*) (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 $8.55 \mu\text{m}$, $11.03 \mu\text{m}$ y $12.02 \mu\text{m}$, 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 2: Bases de datos

3.6. Base de datos original

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

3.7. Base de datos reducida

3.7.1. Featurewiz para el uso del suelo urbano

La base de datos reducida por medio del algoritmo de featurewiz para el uso del suelo urbano contó con un total de 77 variables estas son: hora del día, día de la semana, mes del año, día del mes, estación del año, mes del año codificación seno, mes del año codificación coseno, día de la semana codificación seno, día de la semana codificación coseno, día del mes codificación seno, día del mes codificación coseno, hora

del día codificación seno, hora del día codificación coseno, estación del año codificación seno, estación del año codificación coseno, 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, 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).

3.7.2. LASSO para el uso del suelo urbano

La base de datos reducida por medio del algoritmo de LASSO para el uso del suelo urbano contó con un total de 55 variables estas son: hora del día, día de la semana, mes del año, día del mes, estación del año, mes del año codificación seno, mes del año codificación coseno, día de la semana codificación seno, día de la semana codificación coseno, día del mes codificación seno, día del mes codificación coseno, hora del día codificación seno, hora del día codificación coseno, estación del año codificación seno, estación del año codificación coseno, componente v del viento a 100 m, componente u del viento neutro a 10 m, componente v del viento neutro a 10 m, temperatura a 2 m, cielo despejado radiación solar directa en superficie, tensión superficial de la onda de gravedad hacia el este, albedo previsto, rugosidad superficial prevista, velocidad de fricción, índice k, índice de área foliar vegetación alta, temperatura máxima en 2m desde el postprocesado anterior, disipación media de la capa límite, tasa media de evaporación, tasa media de evaporación potencial, presión media, flujo medio de radiación directa de onda corta en superficie, flujo de radiación de onda larga descendente en superficie media, flujo medio superficial descendente de radiación de onda corta, flujo medio superficial neto de radiación de onda larga, flujo de calor sensible medio en superficie, flujo medio superficial descendente de radiación de onda corta, albedo cercano al ir para radiación directa, radiación solar neta superficial, radiación térmica neta superficial, presión superficial, flujo de calor

sensible superficial, radiación solar superficial descendente, ozono total en columna, albedo visible uv para radiación directa, integral vertical de divergencia de flujo de masa, integral vertical del flujo de masa hacia el este, integral vertical del flujo de ozono hacia el este, integral vertical de la masa de la atmósfera, integral vertical de energía potencial e interna, integral vertical de energía total, nivel de cero grados, velocidad del viento a 100 m, velocidad del viento a 10 m e incremento de la profundidad óptica de aerosoles (merra).

3.7.3. Featurewiz para el uso del suelo agrícola y área verde

La base de datos reducida por medio del algoritmo de featurewiz para el uso del suelo agrícola y área verde contó con un total de 84 variables estas son: hora del día, día de la semana, mes del año, día del mes, estación del año, mes del año codificación seno, mes del año codificación coseno, día de la semana codificación seno, día de la semana codificación coseno, día del mes codificación seno, día del mes codificación coseno, hora del día codificación seno, hora del día codificación coseno, estación del año codificación seno, estación del año codificación coseno, 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, 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).

3.7.4. LASSO para el uso del suelo agrícola y área verde

La base de datos reducida por medio del algoritmo de LASSO para el uso del suelo agrícola y área verde contó con un total de 59 variables estas son: hora del día, día de la semana, mes del año, día del mes, estación del año, mes del año codificación seno, mes del año codificación coseno, día de la semana codificación seno, día de la semana codificación coseno, día del mes codificación seno, día del mes codificación coseno, hora del día codificación seno, hora del día codificación coseno, estación del año codificación seno, estación del año codificación coseno, componente u del viento a 100 m, componente v del viento a 100 m, energía potencial convectiva disponible, albedo previsto, disipación de ondas de gravedad, flujo de calor sensible instantáneo en superficie, fracción de precipitación a gran escala, disipación media de la onda de gravedad, fracción media de precipitación a gran escala, tensión superficial turbulenta media hacia el norte, presión media, tasa media de escorrentía subsuperficial, flujo medio de radiación de onda larga neta superior, gradiente vertical medio de refractividad dentro de la capa de captura, nubosidad media, albedo cercano al ir para la radiación directa, tensión superficial de ondas gravitatorias hacia el norte, tensión superficial turbulenta hacia el norte, flujo de calor latente superficial, radiación térmica neta superficial, radiación térmica superficial hacia abajo, ozono total en columna, agua de lluvia total en columna, evaporación total, índice de totales totales, altura superior de la capa de captura, tipo de vegetación baja, albedo visible uv para radiación directa, integral vertical de divergencia del flujo de energía cinética, integral vertical de divergencia del flujo de energía térmica, integral vertical del flujo de energía cinética hacia el este, integral vertical del flujo de ozono hacia el este, integral vertical del flujo de agua líquida de las nubes hacia el norte, integral vertical del flujo térmico 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, integral vertical de energía potencial e interna, velocidad del viento

neutro a 10 m, profundidad óptica de aerosoles (merra), incremento de la profundidad óptica de aerosoles (merra), temperatura de la superficie terrestre (modis) y tráfico de vehículos (google).

Anexo 3: Pseudocódigo

3.7.5. Análisis exploratorio

- Código R

```
# Importar librerías necesarias:
```

```
library(tidyverse)
library(univariateML)
library(lubridate)
library(xts)
library(reshape2)
library(TSstudio)
library(forecast)
library(flextable)
library(openair)
library(methods)
library(ggpubr)
library(urca)
library(future)
library(RTransferEntropy)
library(tseries)
```

```
library(urroot)

# Lectura de datos y preprocesamiento de datos:

datos<- read.table("base_de_datos.txt",sep=";", header=T, dec=".")  
  
datos$date <- as.POSIXct(datos$date, format = "%Y-%m-%d %H:%M:%S", tz = "UTC")  
datos$fecha <- as.POSIXct(datos$fecha, format = "%Y-%m-%d", tz = "UTC")  
datos$id <- as.factor(datos$id)  
datos$hora <- as.numeric(datos$hora)  
datos$mes <- as.numeric(datos$mes)  
datos$dia <- as.numeric(datos$dia)  
datos$semana <- as.numeric(datos$semana)  
  
datos$dayofyear<-yday(datos$date)  
  
datos$dia_2 <- recode(datos$dia,  
                      "1"="Domingo",  
                      "2"="Lunes",  
                      "3"="Martes",  
                      "4"="Miercoles",  
                      "5"="Jueves",  
                      "6"="Viernes",  
                      "7"="Sabado")  
  
datos$mes_2 <- recode(datos$mes,  
                      "1"="Enero",
```

```
"2"="Febrero",
"3"="Marzo",
"4"="Abril",
"5"="Mayo",
"6"="Junio",
"7"="Julio",
"8"="Agosto",
"9"="Septiembre",
"10"="Octubre",
"11"="Noviembre",
"12"="Diciembre")

datos$estacion <- as.numeric(datos$estacion)

datos$estacion_2 <- recode(datos$estacion,
  "1"="Otoño",
  "2"="Invierno",
  "3"="Primavera",
  "4"="Verano")

estacion_name <- c("1"="Otoño",
  "2"="Invierno",
  "3"="Primavera",
  "4"="Verano")

mes_name <- c("1"="Enero",
  "2"="Febrero",
```

```
"3"="Marzo",
"4"="Abril",
"5"="Mayo",
"6"="Junio",
"7"="Julio",
"8"="Agosto",
"9"="Septiembre",
"10"="Octubre",
"11"="Noviembre",
"12"="Diciembre")

dia_name <- c("8"="Domingo",
"2"="Lunes",
"3"="Martes",
"4"="Miercoles",
"5"="Jueves",
"6"="Viernes",
"7"="Sabado")

datos<-datos %>%
  mutate(id_comp = if_else(id == "CES", "Campo Escuela",
                           if_else(id == "CGA", "Alta Cordoba",
                                   if_else(id == "CGI", "Rio Ceballos",
                                         if_else(id == "CRA", "Casa Radica",
                                               if_else(id == "CUN", "Ciudad Universitaria",
                                                     if_else(id == "JO", "Jardines del Oeste",
                                                       if_else(id == "PES", "Parque Ecológico San Jerónimo"))))))
```

```
    id_comp = as.factor(id_comp)
)

datos$id_clust[datos$id_clust == "Agricola/AreaVerde"] <- "Agrícola y área verde"

# Cálculo promedios horarios:

prom <- datos %>%
  group_by(date) %>%
  summarise_if(is.numeric, mean, na.rm = TRUE)

pm_ts<-prom$PM25 %>% ts(freq= 24)

# Análisis exploratorio y visualizaciones:

## Ejemplo:

### Análisis horario

datos3 <- datos %>%
  group_by(hora) %>%
  summarise(mean_pm25=mean(PM25), .groups = "drop")
names(datos3)<-c("hora","PM25")

datos3$hora <- as.numeric(datos3$hora)
```

```
datos3 %>%

  ggplot(aes(x=hora, y=PM25)) +
  geom_point(size=1) +
  geom_line(size=1) +
  theme_bw() +
  ylab(expression(PM[2.5] ~ paste("[", mu, "g/", m^3, "]"))) +
  theme(text = element_text( size=30),
        axis.text.x=element_text(angle=0, hjust=1),
        legend.position="bottom") +
  scale_x_continuous("Hora",breaks=seq(0, 24, 1))

### Análisis horario por uso del suelo.

datos$hora <- as.numeric(datos$hora)

# Diario cada lugar

datos3 <- datos %>%
  group_by(id_clust,hora) %>%
  summarise(mean_pm25=mean(PM25), .groups = "drop")
names(datos3)<-c("id_us","hora","PM25")

datos3$hora<-as.numeric(datos3$hora)

datos3 %>%
  ggplot(aes(x=hora, y=PM25, color=id_us)) +
  geom_point(size=1) +
  geom_line(size=1) +
  theme_bw() +
```

```
ylab(expression(PM[2.5]~paste("[" , mu , "g/", m^3, "]")))+  
theme(text = element_text( size=30),  
      axis.text.x=element_text(angle=60, hjust=1),  
      legend.position="bottom") +  
guides(colour=guide_legend(title="uso del suelo")) +  
scale_x_continuous("Hora", labels = as.character(datos3$hora), breaks = datos3$
```

3.7.6. Modelado predictivo

- Código python

```
# Importar librerías necesarias:  
  
from numpy.random import seed  
seed(1)  
import tensorflow  
tensorflow.random.set_seed(2)  
import numpy as np  
import os  
import pandas as pd  
import matplotlib.pyplot as plt  
import random  
  
# Lectura y preprocesamiento de los datos:  
  
os.chdir('Proyecto')  
datos = pd.read_csv("base_de_datos.txt",sep=";")  
datos = datos.drop(["ano"],axis=1)
```

```
## Encoding de variables temporales:  
  
def Enc_temp(df,temporales):  
  
    # Ordenar valores  
  
    df = df.sort_index(ascending=True)  
  
  
    # Hacer el encoding de las temporales  
  
    for i,col in enumerate(temporales):  
  
        df[col+'sin'] = np.sin(2 * np.pi * df[col]/df[col].max())  
        df[col+'cos'] = np.cos(2 * np.pi * df[col]/df[col].max())  
  
    return(df)  
  
  
datos = Enc_temp(datos,['mes','dia','hora','dia_mes','estacion'])  
print("Shape datos encoding temporales:",datos.shape)  
  
  
## División en X e y  
  
X = datos.drop(["PM25"],axis=1)  
y = datos[["PM25"]]  
  
print("Shape X:",X.shape)  
print("Shape y:",y.shape)  
  
  
## Desfasaje de la variable respuesta  
  
N = 120  
  
y = y.iloc[N-1:].copy()  
print("Shape y_lag:",y.shape)  
X = X.iloc[:-(N-1)].copy()  
print("Shape X_lag:",X.shape)
```

```
## Crear semillas aleatorias
randoms = random.sample(range(1, 5000), 10)
randoms = [2474, 2905, 4850, 1194, 77, 3684, 3544, 520, 4365, 3769]

## Data frame de resultados

results = pd.DataFrame(columns = ['Modelo', 'mean_absolute_error','mean_squared_error','mean_squared_log_error','r2_score'])

# Modelado predictivo:

## Ejemplo:

### Random Forest:

from sklearn.metrics import mean_absolute_percentage_error, mean_absolute_error, mean_squared_error, r2_score
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestRegressor
model = RandomForestRegressor()

mae = []
mse = []
rmse = []
me = []

for i in randoms:
```

```
# Divido entrenamiento y testeo estratificado
X['mes_dia'] = X['mes'].astype(str) + X['dia'].astype(str)

# División entrenamiento y testeo
X_training, X_test_red, y_training, y_test_red = train_test_split(X, y, test_size=0.2, random_state=42)
X_train_red, X_val_red, y_train_red, y_val_red = train_test_split(X_training, y_training, test_size=0.5, random_state=42)

X_train_red = X_train_red.drop(['mes_dia'], axis=1)
X_test_red = X_test_red.drop(['mes_dia'], axis=1)
X_val_red = X_val_red.drop(['mes_dia'], axis=1)

X_train_red = X_train_red.sort_index(ascending=True)
X_test_red = X_test_red.sort_index(ascending=True)
y_train_red = y_train_red.sort_index(ascending=True)
y_test_red = y_test_red.sort_index(ascending=True)
X_val_red = X_val_red.sort_index(ascending=True)
y_val_red = y_val_red.sort_index(ascending=True)

X_train = pd.concat([X_train_red, X_val_red], axis=0)
y_train = pd.concat([y_train_red, y_val_red], axis=0)
X_train = X_train.sort_index(ascending=True)

y_train = y_train.sort_index(ascending=True)

X_test = X_test_red.copy(deep=True)
y_test = y_test_red.copy(deep=True)

# Estandarizo
```

```
columns_x = list(X_train.columns)

scaler_x = StandardScaler()

X_train_std = X_train.copy()

X_train_std[columns_x] = scaler_x.fit_transform(X_train[columns_x])

X_test_std = X_test.copy()

X_test_std[columns_x] = scaler_x.transform(X_test[columns_x])


columns_y = list(y_train.columns)

scaler_y = StandardScaler()

y_train_std = y_train.copy()

y_train_std[columns_y] = scaler_y.fit_transform(y_train[columns_y])

y_test_std = y_test.copy()

y_test_std[columns_y] = scaler_y.transform(y_test[columns_y])


# Modelo

model.fit(
    X_train_std,
    y_train_std,
    # eval_set=[(X_test_std, y_test_std)],
    # eval_metric="mean_squared_error",
    # early_stopping_rounds=100
)
```

Predicciones

```
preds = model.predict(X_test_std)

inv_yhat = scaler_y.inverse_transform(preds.reshape(-1,1))

inv_y = scaler_y.inverse_transform(y_test_std)
```

```
# Recopilo métricas

mae.append(mean_absolute_error(inv_y,inv_yhat))

mse.append(mean_squared_error(inv_y,inv_yhat, squared=True))

rmse.append(mean_squared_error(inv_y,inv_yhat, squared=False))

me.append(max_error(inv_y,inv_yhat))

print(np.mean(mae))

print(np.mean(mse))

print(np.mean(rmse))

print(np.mean(me))

resultado_parcial = {"Modelo":"RanfomForest_Long_AA",

"mean_absolute_error":np.mean(mae), 

"mean_squared_error":np.mean(mse), 

"root_mean_squared_error":np.mean(rmse), 

"max_error":np.mean(me)

}

results=results.append(resultado_parcial, 

ignore_index = True)

results.head(25)

### Gated Recurrent Units (GRU):


from tensorflow.keras.layers import GRU
```

```
mae = []
mse = []
rmse = []
me = []

for i in randoms:

    X['mes_dia'] = X['mes'].astype(str) + X['dia'].astype(str)

    X_training, X_test_red, y_training, y_test_red = train_test_split(X, y, test_size=0.2, random_state=i)
    X_train_red, X_val_red, y_train_red, y_val_red = train_test_split(X_training, y_training, test_size=0.2, random_state=i)

    X_train_red = X_train_red.drop(['mes_dia'], axis=1)
    X_test_red = X_test_red.drop(['mes_dia'], axis=1)
    X_val_red = X_val_red.drop(['mes_dia'], axis=1)

    X_train_red = X_train_red.sort_index(ascending=True)
    X_test_red = X_test_red.sort_index(ascending=True)
    y_train_red = y_train_red.sort_index(ascending=True)
    y_test_red = y_test_red.sort_index(ascending=True)
    X_val_red = X_val_red.sort_index(ascending=True)
    y_val_red = y_val_red.sort_index(ascending=True)

columns_x = list(X_train_red.columns)
scaler_x = StandardScaler()

X_train_red[columns_x] = scaler_x.fit_transform(X_train_red[columns_x])
X_test_red[columns_x] = scaler_x.transform(X_test_red[columns_x])
X_val_red[columns_x] = scaler_x.transform(X_val_red[columns_x])

columns_y = list(y_train_red.columns)
scaler_y = StandardScaler()
```

```
y_train_red[columns_y] = scaler_y.fit_transform(y_train_red[columns_y])
y_test_red[columns_y] = scaler_y.transform(y_test_red[columns_y])
y_val_red[columns_y] = scaler_y.transform(y_val_red[columns_y])

X_train_red = X_train_red.to_numpy().reshape((X_train_red.shape[0], 1, X_train_red.shape[1]))
X_test_red = X_test_red.to_numpy().reshape((X_test_red.shape[0], 1, X_test_red.shape[1]))
X_val_red = X_val_red.to_numpy().reshape((X_val_red.shape[0], 1, X_val_red.shape[1]))

y_train_red=y_train_red.to_numpy()
y_test_red = y_test_red.to_numpy()
y_val_red = y_val_red.to_numpy()

columns_x = list(X_train_red.columns)
scaler_x = StandardScaler()
X_train_red[columns_x] = scaler_x.fit_transform(X_train_red[columns_x])
X_test_red[columns_x] = scaler_x.transform(X_test_red[columns_x])
X_val_red[columns_x] = scaler_x.transform(X_val_red[columns_x])

columns_y = list(y_train_red.columns)
scaler_y = StandardScaler()
y_train_red[columns_y] = scaler_y.fit_transform(y_train_red[columns_y])
y_test_red[columns_y] = scaler_y.transform(y_test_red[columns_y])
y_val_red[columns_y] = scaler_y.transform(y_val_red[columns_y])

X_train_red = X_train_red.to_numpy().reshape((X_train_red.shape[0], 1, X_train_red.shape[1]))
X_test_red = X_test_red.to_numpy().reshape((X_test_red.shape[0], 1, X_test_red.shape[1]))
X_val_red = X_val_red.to_numpy().reshape((X_val_red.shape[0], 1, X_val_red.shape[1]))
```

```
y_train_red=y_train_red.to_numpy()
y_test_red = y_test_red.to_numpy()
y_val_red = y_val_red.to_numpy()

# Modelado

es = EarlyStopping(monitor='val_loss', mode='min', verbose=1, patience=30, restore_best_weights=True)

regressor = Sequential()
regressor.add(GRU(units = 200,return_sequences=True, input_shape=(X_train_red.shape[1], X_train_red.shape[2])))
regressor.add(Dropout(0.3))

regressor.add(GRU(units = 100,return_sequences=True))
regressor.add(Dropout(0.3))

regressor.add(GRU(units = 100))
regressor.add(Dropout(0.3))

regressor.add(Dense(units = 1, activation = 'linear'))

regressor.compile(optimizer = keras.optimizers.Adam(1e-3), loss = 'mean_squared_error',
                   metrics = ['mean_squared_error'])
regressor.fit(X_train_red, y_train_red, epochs=200, batch_size=64, validation_data=(X_val_red, y_val_red))

# Predicciones

yhat = regressor.predict(X_test_red)
inv_yhat = scaler_y.inverse_transform(yhat)
inv_yhat = inv_yhat[:,0]
```

```
test_y = y_test_red.reshape((len(y_test_red), 1))

inv_y = scaler_y.inverse_transform(test_y)

inv_y = inv_y[:,0]

mae.append(mean_absolute_error(inv_y,inv_yhat))

mse.append(mean_squared_error(inv_y,inv_yhat, squared=True))

rmse.append(mean_squared_error(inv_y,inv_yhat, squared=False))

me.append(max_error(inv_y,inv_yhat))

print(np.mean(mae))

print(np.mean(mse))

print(np.mean(rmse))

print(np.mean(me))

resultado_parcial = {"Modelo":"GRU_3",

"median_absolute_error":np.mean(mae),

"mean_squared_error":np.mean(mse),

"root_mean_squared_error":np.mean(rmse),

"max_error":np.mean(me)

}

results=results.append(resultado_parcial,

ignore_index = True)

results.head(15)

# Luego de correr todos los modelos, continuamos con la comparación
```

```
results.head(50)
```

Comparar

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.
- 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 PM2. 5 species in urban environments: Relationship with meteorological variables and satellite data. *Atmospheric Environment*, 245, 117987.
- 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.
- 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., et al.others. (2021). The ERA5 global reanalysis: Preliminary extension to 1950. *Quarterly Journal of the Royal Meteorological Society*, 147(741), 4186–4227.

- Bengio, Y. et al. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), 1798–1828.
- 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.
- Bosilovich, M. G. et al. (2015). *MERRA-2: File specification*.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Busso, I. T. et al. (2021). Modeling air pollution-related hospital admissions employing remote sensing and geographical information systems. *Atmospheric Environment*, 261, 118502.
- 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.
- Cesaroni, G., Boogaard, H., Jonkers, S., Porta, D., Badaloni, C., Cattani, G., ...

- Hoek, G. (2012). Health benefits of traffic-related air pollution reduction in different socioeconomic groups: The effect of low-emission zoning in rome. *Occupational and Environmental Medicine*, 69(2), 133–139.
- 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.
- Chen, G., Li, S., Knibbs, L. D., Hamm, N. A., Cao, W., Li, T., ... 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, 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. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794).
- Chen, Tianqi, He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., et al.others. (2015). Xgboost: Extreme gradient boosting. *R Package Version 0.4-2*, 1(4), 1–4.
- Chniti, G. et al. (2017). E-commerce time series forecasting using LSTM neural network and support vector regression. In *Proceedings of the international conference on big data and internet of thing* (pp. 80–84).
- Cho, K. et al. (2014). On the properties of neural machine translation: Encoder-decoder approaches. *arXiv Preprint arXiv:1409.1259*.

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.

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).

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.

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.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (Vol. 326). John Wiley & Sons.

EEA. (2020). Air pollution: how it affects our health. Retrieved November 10, 2020, from <https://www.eea.europa.eu/themes/air/health-impacts-of-air-pollution>

EPA. (2010). Environmental protection agency 40 CFR parts 50 and 58.

Feng, R., Zheng, H., Gao, H., Zhang, A., Huang, C., Zhang, J., ... 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.

- Font, A. et al. (2019). A tale of two cities: Is air pollution improving in paris and london? *Environmental Pollution*, 249, 1–12.
- Gers, F. A. et al. (2000). Learning to forget: Continual prediction with LSTM. *Neural Computation*, 12(10), 2451–2471.
- Gomišček, B. et al. (2004). Spatial and temporal variations of PM1, PM2. 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.
- 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.
- Hamilton, J. D. (2020). *Time series analysis*. Princeton university press.
- Haykin, S. (1999). *Neural networks and learning machines*. Pearson Prentice Hall.
- He, J., Gong, S., Liu, H., An, X., Yu, Y., Zhao, S., et al.others. (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, J., Yu, Y., Xie, Y., Mao, H., Wu, L., Liu, N., & Zhao, S. (2016). Numerical model-based artificial neural network model and its application for quantifying impact factors of urban air quality. *Water, Air, & Soil Pollution*, 227(7), 1–16.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al.others. (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.

- Hofman, J., Staelens, J., Cordell, R., Stroobants, C., Zikova, N., Hama, S., et al.others. (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.
- Hu, A., Huang, Y., Chen, Q., Huang, W., Wu, X., Cui, L., ... Liu, J. (2023). Glass transition of amorphous polymeric materials informed by machine learning. *APL Machine Learning*, 1(2).
- Huang, K., Xiao, Q., Meng, X., Geng, G., Wang, Y., Lyapustin, A., ... 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. et al. (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.
- 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.
- INDEC, A. (2010). Censo nacional de población. *Hogares y Viviendas*.
- Khan, M. F. et al. (2010). Characterization of PM2. 5, PM2. 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.
- Kulkarni, P. et al. (2011). *Aerosol measurement: Principles, techniques, and applications*.

cations. John Wiley & Sons.

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.

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 pm2. 5 dynamics in new york city. In *EPJ web of conferences* (Vol. 237, p. 03010). EDP Sciences.

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.

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 PM2. 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. In *SEG international exposition and annual meeting*. OnePetro.

Lyapustin, A. et al. (2018). MODIS collection 6 MAIAC algorithm. *Atmospheric Measurement Techniques*, 11(10), 5741–5765.

Lyapustin, A. et al. (2019). MAIAC thermal technique for smoke injection height from MODIS. *IEEE Geoscience and Remote Sensing Letters*, 17(5), 730–734.

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.
- Meszaros, E. et al. (1999). *Fundamentals of atmospheric aerosol chemistry*. Akadémiai kiado.
- Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., et al.others. (2021). ERA5-land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13(9), 4349–4383.
- Navares, R., & Aznarte, J. L. (2020a). Deep learning architecture to predict daily hospital admissions. *Neural Computing and Applications*, 32(20), 16235–16244.
- Navares, R., & Aznarte, J. L. (2020b). Predicting air quality with deep learning LSTM: Towards comprehensive models. *Ecological Informatics*, 55, 101019.
- 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.
- OECD, I. et al. (2016). Energy and air pollution: World energy outlook special report 2016.

- 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.
- OMS. (2006). *Air quality guidelines: Global update 2005: Particulate matter, ozone, nitrogen dioxide, and sulfur dioxide*. World Health Organization.
- OMS. (2020). Air pollution. Retrieved November 10, 2020, from <https://www.who.int/health-topics/air-pollution>
- OMS. (2021a). *Review of evidence on health aspects of air pollution: REVIHAAP project: Technical report*. World Health Organization. Regional Office for Europe.
- OMS. (2021b). *WHO global air quality guidelines: Particulate matter (PM_{2.5} and PM₁₀), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. World Health Organization.
- 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.
- 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. In *IOP conference series: Earth and environmental science* (Vol. 552, p. 012020). IOP Publishing.
- 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.
- 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.

- Puliafito, E. (2009). Gestión de la calidad del aire en argentina. *Contaminación Atmosférica En Argentina Contribuciones de La II Reunión Anual PROIMCA*, 67.
- Radovic, M. et al. (2017). Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. *BMC Bioinformatics*, 18(1), 1–14.
- Robert, T. (1994). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B*, 58, 267.
- Rogers, P. G. (1990). The clean air act of 1970. *EPA J.*, 16, 21.
- Rumelhart, D. E. et al. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536.
- 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., et al.others. (2015). PM2. 5 chemical composition in five european mediterranean cities: A 1-year study. *Atmospheric Research*, 155, 102–117.
- 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.
- 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. et al. (2020). Revisión de los monitoreos de calidad del aire en la ciudad

- de buenos aires. *Agronomía & Ambiente*, 40(1).
- Shang, Z. et al. (2019). A novel model for hourly PM_{2.5} concentration prediction based on CART and EELM. *Science of The Total Environment*, 651, 3043–3052.
- 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.
- 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.
- 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 & PM_{2.5}). In *IOP conference series: Earth and environmental science* (Vol. 476, p. 012126). IOP Publishing.
- Tai, A. P. et al. (2010). Correlations between fine particulate matter (PM_{2.5}) and meteorological variables in the united states: Implications for the sensitivity of PM_{2.5} to climate change. *Atmospheric Environment*, 44(32), 3976–3984.
- 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.
- Titus, J. G. (1990). Greenhouse effect, sea level rise, and barrier islands: Case study of long beach island, new jersey.
- Tiwari, S., Dumka, U., Gautam, A., Kaskaoutis, D., Srivastava, A., Bisht, D., ... Solmon, F. (2017). Assessment of PM_{2.5} and PM₁₀ over guwahati in brahmaputra river valley: Temporal evolution, source apportionment and meteorological dependence.

- Atmospheric Pollution Research*, 8(1), 13–28.
- Tong, W. et al. (2019). Deep learning PM2. 5 concentrations with bidirectional LSTM RNN. *Air Quality, Atmosphere & Health*, 12(4), 411–423.
- Usman, F., Zeb, B., Alam, K., Valipour, M., Ditta, A., Sorooshian, A., ... 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.
- 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., et al.others. (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.
- 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., ... 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, 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.
- 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.

WHO, E. (2005). WHORO: Air quality guidelines global update. In *Report on a working group meeting. In. Bonn, germany*.

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.

Wongrin, W. et al. (2023). Comparison of statistical and deep learning methods for forecasting PM 2.5 concentration in northern thailand. *Polish Journal of Environmental Studies*, 32(2).

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, Y.-J. et al. (2007). A google-map-based arterial traffic information system. In *2007 IEEE intelligent transportation systems conference* (pp. 968–973). IEEE.

XiaoYang, L., HongQuan, S., ShiYan, Z., SiQi, L., YunFeng, K., HaoMing, X., et al.others. (2019). Particulate matter pollution in chinese cities: Areal-temporal variations and their relationships with meteorological conditions (2015-2017). *Environmental Pollution*, 246, 11–18.

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.

Yeo, I. et al. (2021). Efficient PM2. 5 forecasting using geographical correlation based on integrated deep learning algorithms. *Neural Computing and Applications*, 33(22), 15073–15089.

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.

Zhao, Yujie et al. (2022). Analysis on the change law and influencing factors of atmospheric particulate matter concentration in xiamen city. In *International conference on statistics, applied mathematics, and computing science (CSAMCS 2021)* (Vol. 12163, pp. 232–237). SPIE.

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.