



Universidad de Valladolid

PROGRAMA DE DOCTORADO EN FÍSICA

TESIS DOCTORAL:

**Desarrollo de una metodología híbrida basada
en modelos numéricos y algoritmos de
aprendizaje automático para la estimación de
potencia eléctrica generada mediante
tecnología eólica**

Presentada por Álvaro Abad Santiago para optar al
grado de
Doctor/a por la Universidad de Valladolid

Dirigida por:
Dr. Carlos Casanova Mateo
Dra. Julia Sanz Justo

Agradecimientos

Llegar al final de una tesis doctoral es un camino largo y lleno de aprendizajes. Aunque la firma sea solo mía, este trabajo no habría sido posible sin la ayuda, compañía y apoyo de muchas personas, tanto dentro como fuera del ámbito académico. Si me paro a pensar, todo esto comenzó hace ya muchos años, 7 más o menos, cuando entré por casualidad en el LATUV a realizar las “prácticas de empresa” de la carrera.

En primer lugar, quiero agradecer sinceramente a mis directores de tesis, el Dr. Carlos Casanova y la Dra. Julia Sanz, por su guía constante, su paciencia y por confiar en mí desde el principio a pesar de que no nos conocíamos apenas. Su apoyo durante este proceso ha sido esencial, y he podido aprender mucho de ambos tanto en lo académico como en lo personal. También agradecer al Dr. José Luis Casanova, que ha sido también una persona especialmente importante en mi trayectoria académica y con quien he aprendido mucho desde hace, también, mucho tiempo. También me gustaría agradecer a mis compañeros/as de laboratorio, Diego y Pablo, con quienes compartí muchas horas, tanto en el laboratorio como en los cafés. Aunque ya hace algún tiempo que nuestros caminos se separaron, siempre mostraron una gran disposición para ayudar a aquel novato que solo quería aprender.

Agradecer a César y Sancho, de la Universidad de Alcalá de Henares, por su tiempo y ayuda durante este proceso, pues sin ellos todo esto hubiese costado mucho más.

Tampoco quiero olvidarme de mis compañeros y compañeras del Dpto. de Física Aplicada de la Universidad de Valladolid, pues en estos años de doctorado que he compaginado con un puesto como Profesor Asociado, he aprendido mucho de todos ellos viendo el funcionamiento del mundo universitario “desde el otro lado”.

Gracias a mis amigos y amigas por acompañarme en este camino, por soportar mis interminables quejas sobre simulaciones eternas y modelos de Machine Learning que se resistían a funcionar, y por recordarme que hay vida más allá de la tesis. Sus risas, cafés y cañas han sido fundamentales para mantener el equilibrio y la cordura durante todo este proceso. En especial, gracias a Irene, una de las mejores cosas que

me llevo de la carrera, y a Borja, que por alguna razón me sigue aguantando después de 22 años.

A mi familia, por ser siempre mi base. A mis padres, Pilar y Anastasio, por enseñarme el valor del esfuerzo y por apoyarme incluso cuando no entendían del todo en qué consistía esto de 'hacer una tesis'. A mi hermana Alicia, por estar siempre cerca y dejarme su moto.

A todos, gracias por haber formado parte de este viaje. Esta tesis también es un poquito vuestra.

Índice general

Agradecimientos

Índice general

Resumen	1
1. Introducción	3
1.1. La predicción meteorológica en el momento actual	3
1.2. Desafíos en la predicción de generación de energía eólica	5
1.3. La modelización numérica en meteorología	7
1.4. Modelización operativa de la atmósfera	10
1.5. Machine Learning	14
1.6. Introducción a los modelos híbridos	16
2. El modelo WRF (Weather Research and Forecasting)	19
2.1. Desarrollo y colaboración	19
2.2. Ventajas del WRF	25
2.3. El futuro del WRF: potencial y perspectivas de avance en la modelización de la atmósfera	27
3. Configuración del WRF para la realización del estudio	31
3.1. Configuración básica del modelo	31
3.2. Configuración física	36
3.3. Configuración numérica	39
3.4. Preprocesamiento Geográfico (WPS)	41
3.5. Datos de entrada	42
3.6. Postprocesamiento	44
4. Validación de la configuración del modelo numérico WRF	47
4.1. Datos de referencia	49
4.2. Resultados de la validación	51

4.2.1. Valores obtenidos en las métricas	52
4.3. Comparación por niveles de presión y horas	71
4.3.1. Dirección del viento	71
4.3.2. Velocidad del viento	74
5. Metodología híbrida para la predicción de generación de energía	
eólica	78
5.1. Metodología	79
5.1.1. Modelo mesoescalar WRF	79
5.1.2. Modelo de reanálisis ERA5	80
5.1.3. Técnicas de Machine Learning aplicadas	82
5.1.4. Preprocesado y ajuste de hiperparámetros	90
5.1.5. Caso de estudio y variables predictoras consideradas	91
5.1.6. Métricas de evaluación del rendimiento	95
5.2. Resultados obtenidos	96
5.3. Discusión de resultados	101
Conclusiones	103
Bibliografía	105
Anexo	117

Resumen

La creciente tendencia en integrar tecnologías de generación de energía eléctrica a partir de fuentes renovables en los sistemas eléctricos ha derivado en una importante demanda por disponer de sistemas de predicción de los recursos eólico y solar lo más precisos posibles para poder planificar adecuadamente las actividades de producción de energía eléctrica. En el caso concreto de la predicción del recurso eólico, el desafío es ciertamente retador debido a la naturaleza altamente variable y caótica del viento, especialmente en horizontes temporales cortos. En este contexto, los modelos numéricos de predicción meteorológica constituyen herramientas fundamentales para simular el comportamiento atmosférico. Sin embargo, la disruptiva aparición de la inteligencia artificial está transformando la predicción meteorológica tal y como se conoce hasta ahora, pues está demostrando altas capacidades para realizar predicciones con una precisión nunca antes vista.

En esta memoria de tesis doctoral se expone el desarrollo de una metodología híbrida que combina la información física proporcionada por dos modelos numéricos, el WRF (Weather Research and Forecasting) y el reanálisis ERA5, junto con algoritmos de aprendizaje automático, en particular, métodos de ensamble. Esta metodología, basada en integrar, por un lado, la base física que proporcionan los modelos numéricos, y, por otro lado, la capacidad adaptativa y predictiva de los algoritmos de aprendizaje automático, mejora significativamente la precisión en la estimación de la potencia eléctrica generada mediante tecnología eólica a corto plazo.

El estudio de la metodología híbrida se ha llevado a cabo utilizando los datos horarios del año 2022, de la potencia eléctrica generada por un parque eólico ubicado en La Mudarra, municipio de la provincia de Valladolid (Castilla y León, España). En esta memoria se describe la evaluación llevada a cabo para conocer el rendimiento de distintos algoritmos de regresión, observándose una mejora significativa en la precisión cuando se utilizan conjuntamente los datos del WRF y ERA5 frente al uso individual de cada fuente. Entre los algoritmos probados, destacan por su rendimiento Random Forest y LightGBM.

Asimismo, se presentan los resultados de una validación independiente del siste-

ma de predicción WRF que ha sido adaptado y puesto en operación para desarrollar este trabajo. A este fin, se han utilizando datos observacionales del radiosondeo de la estación de Madrid, cercana al aeropuerto Adolfo Suárez Madrid-Barajas. Esta validación se centró en dos franjas horarias (00:00 y 12:00 UTC) durante el mes de julio del año 2015. El análisis permitió comprobar la coherencia de las salidas del modelo WRF con las observaciones reales incluso con datos ausentes, lo que evidenció, además, su capacidad para complementar la información disponible con errores aceptables.

Los resultados obtenidos ponen de manifiesto el potencial las aproximaciones híbridas para generar predicciones más robustas y ajustadas a las condiciones locales. La metodología propuesta resulta versátil y aplicable a diferentes emplazamientos, por lo que puede ayudara contribuir al desarrollo de sistemas de predicción más fiables que favorezcan una mejor integración de la energía eólica en la red eléctrica.

Capítulo 1

Introducción

1.1. La predicción meteorológica en el momento actual

La sociedad moderna demanda, cada vez más, predicciones meteorológicas fiables, ya que su uso está ligado a la planificación de multitud de escenarios estratégicos de relevancia, como pueden ser la gestión de recursos energéticos, la producción agrícola, el transporte o la prevención y gestión de emergencias. En un mundo moderno interconectado y dependiente de los recursos naturales, la capacidad de predecir con la mayor precisión posible las condiciones atmosféricas futuras se ha convertido en un pilar fundamental para el desarrollo sostenible, la transición a fuentes de energía renovables y la eficiencia económica [1].

Dentro de los muchos sectores relevantes, uno de los más destacados es el sector energético, donde el pronóstico del tiempo es vital a la hora de aprovechar recursos naturales como la luz solar o la energía del viento. Ambas fuentes de energía han experimentado una notable expansión en las últimas décadas, convirtiéndose en una fuente clave de generación de energía eléctrica en múltiples mercados energéticos por todo el mundo. Sin embargo, ambas fuentes dependen de factores externos que pueden hacer que se aprovechen en mayor o menor medida. Mientras la radiación solar sigue patrones relativamente predecibles, el comportamiento del viento es ciertamente variable. Esta naturaleza fluctuante de la energía eólica dificulta su integración en la red eléctrica, haciendo que la previsión meteorológica juegue un papel crucial. Una predicción precisa de los recursos eólicos y solares permite optimizar su aprovechamiento, reducir costes y mejorar la eficiencia del sistema eléctrico [2].

Más allá del sector energético, la previsión meteorológica es vital en otras áreas [3] [4], tales como:

- Agricultura: las personas dedicadas a la agricultura utilizan los pronósticos meteorológicos para determinar, entre otros, los tiempos de siembra y cosecha de los cultivos, o los horarios de riego, reduciendo así las pérdidas por sequías naturales o mitigando el impacto de eventos adversos como las inundaciones o las olas de calor.
- Transporte: las predicciones meteorológicas juegan un papel esencial en la aviación y la navegación marítima pues impactan directamente en la seguridad, la regularidad y la eficiencia de las operaciones aéreas y el transporte de mercancías por mares y océanos.
- Gestión de emergencias: ser capaces de pronosticar con antelación eventos extremos como ciclones tropicales, sistemas convectivos de mesoescala u olas de calor permiten tomar medidas de protección de vidas humanas, así como minimizar los daños estructurales.

Por otro lado, el cambio climático es, en el momento actual, uno de los ámbitos de investigación prioritarios. En este sentido, la predicción atmosférica juega un papel muy importante, pues la mejor y más precisa predicción de eventos climáticos extremos se convierte en algo vital. Estos fenómenos aparecen cada vez con más frecuencia y fuerza, lo que requiere de técnicas avanzadas de pronóstico para minimizar, en la medida de lo posible, sus destructivas consecuencias. La transición hacia una economía baja en emisiones de gases de efecto invernadero es uno de los retos más importantes a la hora de contrarrestar los efectos del cambio climático, y esto depende, en gran medida, del desarrollo y la implementación de herramientas que faciliten una efectiva integración de las fuentes de energía renovable en los sistemas eléctricos [5].

Pese a los grandes avances ya existentes en el mundo de la modelización de la atmósfera, el desarrollo de modelos de predicción más precisos es fundamental para seguir mejorando en todos los sectores mencionados anteriormente. Poco a poco, y conforme la dependencia de los recursos naturales aumenta, la necesidad de mejores herramientas que permitan anticipar variaciones atmosféricas se vuelve más evidente. Particularmente, como se ha expuesto en los párrafos anteriores, en el ámbito energético, donde la transición hacia fuentes renovables es una prioridad casi global: contar con pronósticos confiables es clave para garantizar la estabilidad de los sistemas eléctricos. Sin embargo, algunas fuentes de energía renovable, como la eólica, presentan desafíos adicionales debido a su naturaleza variable e intermitente. En este contexto, es importante seguir investigando y desarrollando enfoques innovadores

para mejorar la predicción de la generación eólica, con el fin de avanzar hacia la sostenibilidad y eficiencia del sistema energético moderno.

1.2. Desafíos en la predicción de generación de energía eólica

Toda red eléctrica debe mantener un equilibrio entre la energía eléctrica que genera y la que se consume para funcionar de forma correcta. Este equilibrio es crucial, pues cualquier variación significativa puede provocar graves problemas: fluctuaciones de voltaje, apagones o una sobrecarga de los dispositivos eléctricos conectados a la red. La naturaleza intermitente de la generación eólica complica este equilibrio. En momentos de fuertes vientos, la generación puede superar a la demanda, mientras que en periodos de calma la producción puede reducirse drásticamente, dejando un vacío que debe cubrirse con otros sistemas de generación que puedan entrar en funcionamiento en cualquier momento, principalmente centrales de ciclo combinado.

Además, la falta de estabilidad en la generación de la energía eólica es un problema técnico para las infraestructuras actuales de las redes eléctricas, que no siempre están diseñadas de forma que sean capaces de gestionar fluctuaciones rápidas y significativas de corriente eléctrica. Por esta razón, la dependencia exclusiva de energía eólica puede ser problemática, sobre todo si no se dispone de elementos de almacenamiento de energía, como baterías, o la combinación de la generación eólica con otras fuentes de generación de energía eléctrica más estables como las centrales nucleares.

La integración efectiva de la energía eólica requiere, por tanto, de soluciones a estos problemas principales. El desarrollo de tecnologías de almacenamiento a gran escala, infraestructuras eléctricas más flexible e inteligentes solucionan algunos de estos problemas, pero también es necesario el uso de sistemas de predicción meteorológicos precisos que permitan predecir y anticiparse cada vez mejor a las fluctuaciones del viento [6].

En lo referente a este último punto, existen una serie de problemas bien conocidos por la comunidad técnico-científica:

- **La variabilidad del recurso eólico.** Como ya se ha mencionado, el viento es un elemento que presenta una alta variabilidad espacial y temporal: éste puede cambiar en cuestión de segundos, pasando de una situación de viento constante en dirección e intensidad a presentar ráfagas significativas y dirección variable, o evolucionar desde una situación de varios días de calma a periodos prolongados de vientos fuertes. Además, la geografía puede influir directamente en la

generación de los denominados "vientos locales" que sólo ocurren en una zona concreta; o grandes corrientes que afectan a extensas regiones, dando lugar a escalas espaciales que pueden ir desde los pocos metros a los cientos de kilómetros. Esta variabilidad hace que la predicción del viento sea particularmente compleja, requiriendo de herramientas de predicción capaces de trabajar con fenómenos a múltiples escalas [7].

- **Limitaciones de los modelos numéricos de predicción meteorológica.**

Los modelos numéricos de predicción meteorológica, como el WRF (Weather Research and Forecasting), son potentes herramientas que se usan para simular las condiciones atmosféricas en diferentes escalas tanto temporales como espaciales. Sin embargo, estos modelos tienen limitaciones que afectan a su precisión:

- **Resolución espacial y temporal.** Aunque los modelos de alta resolución son capaces de capturar y predecir fenómenos a escala local, esto puede estar limitado por el hardware, haciendo que los requisitos necesarios para simular determinadas condiciones sean excesivos en tiempo o coste y, en consecuencia, no sean rentables.
- **Incertidumbre en las condiciones iniciales.** Los modelos numéricos dependen de los datos de entrada que se les aporten, datos que pueden contener errores o estar incompletos. Estos errores pueden propagarse a lo largo de la simulación, alterando los resultados finales y, por tanto, afectando a la precisión de los resultados.
- **Fenómenos locales no resueltos.** Determinados fenómenos locales, especialmente aquellos que suceden en escalas de tiempo o de espacio muy concretas, pueden no ser representadas correctamente en los modelos numéricos de predicción. Esto puede ocasionar errores en la integración de determinadas interacciones como, por ejemplo, las que ocurren entre la atmósfera y la superficie, que son muy relevantes para poder obtener unas adecuadas predicciones de viento.

Abordar los desafíos de la predicción eólica requiere, por tanto, de herramientas capaces de capturar la complejidad del viento. Si bien los modelos numéricos de predicción meteorológica son ampliamente usados, estos sistemas, pese a su sofisticación, presentan limitaciones espaciales y temporales, así como incertidumbre en las condiciones iniciales, lo que puede influir significativamente en su capacidad de predicción [8] [1].

1.3. La modelización numérica en meteorología

Un concepto importante en meteorología es la diferenciación entre "tiempo" y "clima". A pesar de que estos términos se emplean indistintamente en el lenguaje cotidiano, poseen significados específicos bien diferenciados:

El tiempo alude a las condiciones atmosféricas en un instante dado y está definido por diversos elementos meteorológicos. Estos elementos meteorológicos son la temperatura, la humedad, la presión atmosférica, la precipitaciones, la velocidad y dirección del viento, etc.

El clima es la síntesis de las condiciones meteorológicas en un lugar determinado, caracterizadas por estadísticas a largo plazo (generalmente al menos 30 años según los estándares de la Organización Meteorológica Mundial [9]) de los elementos meteorológicos en dicho lugar.

Para el análisis del tiempo se utilizan los modelos numéricos operativos, diseñados para predecir su estado a corto plazo, generalmente desde unas pocas horas hasta un máximo de unos 15 días aproximadamente. Por el contrario, los modelos numéricos climáticos son los sistemas utilizados para estudiar el comportamiento del clima a largo plazo, proyectando escenarios que abarcan desde décadas hasta siglos. En cualquier caso, sean del tipo que sean, los modelos numéricos se han convertido en herramientas esenciales para el avance de la ciencia atmosférica contemporánea. Todos ellos tienen una base de funcionamiento común, sustentada en la resolución de las ecuaciones fundamentadas de la Física de la Atmósfera mediante modernas técnicas matemáticas y empleando, como condiciones iniciales de trabajo, grandes volúmenes de datos provenientes de diversas fuentes de información.

Dado que hay una gran variabilidad en las resoluciones requeridas para representar los distintos eventos atmosféricos, los modelos numéricos se clasifican en varias categorías, siendo quizás las dos más relevantes las que se presentan a continuación.

Modelos Numéricos Globales (GCM, Global Circulation Models) Estos modelos son herramientas esenciales para estudiar la evolución de la atmósfera en amplias escalas espaciales. Modelan la atmósfera, los océanos, la criosfera y la biosfera de manera integrada, permitiendo evaluar tendencias globales o estudiar fenómenos como El Niño o la Oscilación del Atlántico Norte (NAO). Su resolución espacial es relativamente baja, ya que buscan capturar procesos de gran escala en lugar de variaciones locales detalladas. Algunos ejemplos de modelos globales son:

- GFS (Global Forecast System): es un modelo numérico de predicción meteorológica desarrollado por la NOAA (Administración Nacional Oceánica y Atmosférica) [10]. El GFS se ejecuta cuatro veces al día (00, 06, 12 y 18 UTC),

y su resolución espacial ha mejorado con el tiempo, y actualmente trabaja con una resolución horizontal de unos 13 km (aunque esto puede variar según la versión y región). Es ampliamente utilizado por meteorólogos de todo el mundo, y también sirve como base para otros modelos regionales de mayor detalle.

- ECMWF (European Centre for Medium-Range Weather Forecasts): es un modelo numérico de predicción meteorológica desarrollado por el Centro Europeo de Predicción Meteorológica a Medio Plazo. Es considerado uno de los modelos más precisos del mundo, sino el mejor, para pronósticos de medio plazo, abarcando hasta 10 a 15 días [11]. Se ejecuta dos veces al día (00 y 12 UTC) y trabaja con una resolución horizontal muy alta (actualmente alrededor de 9 km, aunque esto puede variar según las versiones), lo que le permite representar con gran detalle los sistemas meteorológicos. Es ampliamente utilizado como referencia en la comunidad meteorológica, tanto para previsiones operativas como para investigación. Además del modelo de pronóstico determinista, el ECMWF también cuenta con un sistema de predicción por conjuntos (ensembles), útil para estimar la incertidumbre del pronóstico.
- CESM (Community Earth System Model): es un modelo climático desarrollado principalmente por el NCAR (National Center for Atmospheric Research) en colaboración con otras instituciones. A diferencia del GFS o el ECMWF, el CESM está diseñado para estudiar el comportamiento del sistema climático a largo plazo, no para hacer predicciones del tiempo [12]. CESM es un modelo acoplado, lo que significa que simula múltiples componentes del sistema terrestre de forma integrada: atmósfera, océanos, hielo marino, superficie terrestre y biogeoquímica. Esto le permite representar las interacciones complejas entre estos elementos, lo cual es esencial para estudiar el cambio climático, la variabilidad climática natural y los efectos de diferentes forzamientos externos (como los gases de efecto invernadero o las erupciones volcánicas). Es una herramienta clave en la investigación climática y ha sido ampliamente utilizado en los proyectos de comparación de modelos climáticos globales, como CMIP, que alimentan los informes del IPCC.

Modelos Mesoescalares Para analizar fenómenos atmosféricos regionales con mayor precisión, se emplean modelos numéricos de mesoescala, que proporcionan una resolución superior en comparación con los GCM. Este aumento de la resolución los hace muy útiles para simular las condiciones atmosféricas de una zona concreta, pero son lentos y poco eficientes si se quiere abarcar una región amplia. Uno de los

modelos más empleados en este campo es el WRF, que facilita la simulación de fenómenos atmosféricos en escalas locales y regionales. El WRF se utiliza extensamente en la investigación atmosférica y en aplicaciones operativas. La implementación de modelos mesoescalares como el WRF ha optimizado la capacidad predictiva, aunque persiste un margen de incertidumbre debido a la compleja interacción de factores atmosféricos y geográficos.

Otros ejemplos de modelos mesoescalares son:

- COSMO (Consortium for Small-scale Modelling) [13]: es un modelo numérico regional de predicción del tiempo, desarrollado por un consorcio de servicios meteorológicos europeos liderado inicialmente por Alemania (DWD). Trabaja en alta resolución (usualmente entre 1 y 7 km, dependiendo de la configuración), lo cual permite captar con mayor detalle las condiciones locales del tiempo. Se alimenta de datos provenientes de modelos globales como el GFS o el ECMWF, que le proporcionan las condiciones en los bordes del dominio regional. El COSMO se ha utilizado ampliamente en Europa y otros países para aplicaciones operativas y de investigación, y ha evolucionado hacia versiones más avanzadas como COSMO-CLM (para estudios climáticos regionales) y ICON, que es el sucesor del COSMO dentro del DWD.
- HARMONIE-AROME [14]: es un modelo numérico desarrollado dentro del consorcio europeo HIRLAM-ALADIN, que reúne varios servicios meteorológicos nacionales, especialmente de Europa del Norte y Occidental. Es un modelo de alta resolución no hidrostático, diseñado para simular fenómenos atmosféricos de escala muy pequeña. Funciona típicamente con una resolución espacial de 2,5 km o menos, lo que le permite captar con gran detalle los procesos atmosféricos locales y orográficos, algo muy útil en regiones montañosas o costeras. Este modelo se centra en pronósticos a muy corto plazo (nowcasting) y corto plazo (hasta 48-60 horas), y suele ejecutarse varias veces al día. HARMONIE-AROME se basa en el núcleo del modelo francés AROME, y está adaptado a las necesidades de los países del consorcio. Se utiliza de forma operativa en países como España (AEMET), Noruega, Países Bajos, Finlandia, entre otros, como parte del sistema conjunto HARMONIE.

En resumen, los modelos globales y regionales desempeñan funciones complementarias en meteorología. Por ello, la integración de ambos enfoques es fundamental para optimizar la comprensión de los fenómenos atmosféricos.

Sin embargo, estos modelos no están exentos de problemas. Cada uno tiene sus propias limitaciones, puntos fuertes y puntos débiles. Especialmente hoy en día, ya

que los continuos avances en las ciencias de la computación permiten no sólo generar cada vez más datos, si no disponer de grandes cantidades de información en enormes bases de datos. Esta ingente cantidad de datos es un problema, pues no puede ser abordado de forma convencional. En este campo, una de las ramas más prometedora es el uso de técnicas de Machine Learning (ML), donde mediante la aplicación de determinados algoritmos se pueden corregir posibles errores sistemáticos en las salidas de los modelos, mejorar la resolución espacial y/o temporal a través de técnicas de downscaling estadístico o incorporar datos de observaciones en tiempo real para ajustar las predicciones.

1.4. Modelización operativa de la atmósfera

La modelización de la atmósfera consiste en la representación matemática y computacional de los procesos físicos y químicos que ocurren en la atmósfera terrestre. Se utilizan ecuaciones matemáticas basadas en las leyes de la Física (como la Termodinámica, la Dinámica de Fluidos y la Radiación) para simular la evolución del tiempo atmosférico.

Ecuaciones físicas en los modelos

Las leyes de la Física se traducen en conjuntos de ecuaciones diferenciales parciales que se resuelven de forma eficiente en los modelos. Las soluciones de estas ecuaciones sustentan nuestra comprensión de cómo se comporta la atmósfera.

Las ecuaciones de Navier-Stokes describen el movimiento de los fluidos, como la atmósfera o los océanos. Se simplifican para su uso en el modelado atmosférico, es decir, se modifican para que los resultados obtenidos sean los más óptimos para las resoluciones espaciales y temporales necesarias en cada caso. Las ecuaciones de Navier-Stokes [15] para un fluido incompresible se escriben como:

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} = -\frac{1}{\rho} \nabla p + \nu \nabla^2 \mathbf{u} + \mathbf{F} \quad (1.1)$$

donde $\mathbf{u}$ es el vector de velocidad, p es la presión, ρ es la densidad, ν es la viscosidad y $\mathbf{F}$ las fuerzas externas.

También se tiene en cuenta la ecuación de continuidad, que asegura la conservación de la masa en el sistema:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0 \quad (1.2)$$

Y la ecuación de la energía, que describe la conservación de la energía en la atmósfera. Incluye términos para la advección, difusión y generación de calor, representada por:

$$\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T = \mathbf{k} \nabla^2 T + \frac{Q}{\rho c_p} \quad (1.3)$$

donde T es la temperatura, $\mathbf{k}$ es la conductividad térmica, Q es la fuente de calor y c_p es la capacidad calorífica a presión constante.

Además, hay que tener en cuenta las diferentes funciones de estado que puede tener el sistema. Para la atmósfera, las ecuaciones de estado relacionan variables presión, temperatura y densidad. Aunque la atmósfera no es un gas ideal, se suele aproximar como tal en muchas aplicaciones de la modelización climática y meteorológica para simplificar los cálculos. La ecuación de estado del gas ideal [16] es:

$$\mathbf{p} = \rho R T \quad (1.4)$$

donde $\mathbf{p}$ es la presión, ρ es la densidad, R es la constante de gas específica y T la temperatura.

El aire es una mezcla de gases (predominantemente nitrógeno y oxígeno, con presencia de otros gases en pequeñas cantidades y vapor de agua) y aerosoles. La ecuación de los gases ideales es una simplificación, pero es bastante adecuada para muchas aplicaciones meteorológicas ya que las desviaciones del comportamiento de los gases ideales suelen ser pequeñas en las condiciones atmosféricas habituales. Si bien las ecuaciones de estado más simples son suficientes en muchas situaciones, en los casos en que se necesita una mayor precisión, como para realizar estudios de química atmosférica, también están disponibles ecuaciones de estado más complejas que incluyen efectos como la humedad y la composición variable del aire.

Limitaciones computacionales y de resolución

Las ecuaciones expuestas anteriormente se resuelven numéricamente en una cuadrícula que discretiza el espacio. El nivel de detalle que puede capturar el modelo está determinado por la resolución de la cuadrícula, pero el costo computacional crece exponencialmente con el aumento de la resolución. Los modelos globales se ejecutan actualmente en resoluciones típicas de 10 a 100 km² para predicciones climáticas y de 1 a 15 km² para predicciones meteorológicas a corto plazo [17] [18].

Incertidumbres en los datos de entrada

La precisión de los modelos depende, en gran medida, de la calidad de los datos de entrada, es decir, de su mayor o menor grado de ajuste a la realidad. Las incertidumbres inherentes a los datos de entrada pueden clasificarse en varias categorías:

- Las observaciones de variables meteorológicas registradas por estaciones de observación de superficie o de altura se utilizan para la primera aproximación y la verificación del modelo. Sin embargo, estos datos pueden contener errores instrumentales (errores debidos a la falta de mantenimiento preventivo, verificaciones o calibraciones) y de muestreo. En relación con este último punto, ha de tenerse en cuenta que existe un sesgo geográfico debido a que las estaciones meteorológicas están distribuidas de manera desigual en el mundo y son relativamente raras en áreas remotas o en el océano.
- Los satélites ofrecen cobertura global, pero habitualmente los datos deben interpretarse utilizando algoritmos inversos que pueden inducir algún error. Por ejemplo, la estimación de la temperatura de la superficie del mar de las mediciones satelitales dependen de la contaminación existente por nubes o por aerosoles [19].
- Las series de datos observacionales suelen presentar inconsistencias, como lagunas de datos en las series temporales o registros erróneos, que requieren ajustes para su corrección y homogenización. Sin embargo, estos procedimientos introducen su propia incertidumbre, ya que distintos métodos de ajuste pueden generar resultados divergentes.

Representación de procesos físicos: parametrizaciones

La parametrización es una técnica útil para representar procesos que ocurren en escalas más pequeñas que la resolución del modelo, como por ejemplo, la formación de nubes o la turbulencia. Sin embargo, estas aproximaciones no siempre son precisas, porque no siempre capturan bien la física subyacente. Por ejemplo, la convección profunda, esencial para la formación de tormentas, es difícil de parametrizar con precisión [20]. Por otro lado, las nubes son un actor esencial dentro del balance energético de la Tierra: reflejan la luz solar y re-irradian radiación infrarroja. La dinámica y el desarrollo de las nubes son procesos complejos y sensibles a la microfísica de las gotas de agua y/o partículas de hielo. Las parametrizaciones actuales de estos procesos son, en ocasiones, básicas y pueden dar lugar a una simulación

inexacta del balance de precipitación y radiación [21]. Además, los océanos y la atmósfera están acoplados. Eventos como El Niño-Oscilación del Sur son consecuencia del acoplamiento entre la atmósfera y la superficie del mar [22]. Para representar con precisión estos procesos es interesante emplear modelos que integren tanto la dinámica oceánica como la atmosférica. Estos modelos deben estar estrechamente acoplados e incorporar enfoques interdisciplinarios, combinando conocimientos de meteorología, oceanografía o química atmosférica. Solo así es posible capturar de manera realista la interacción entre la circulación oceánica y la atmosférica. Tampoco se puede pasar por alto los aerosoles, porque influyen en la formación de nubes y el balance radiativo. La representación de la química atmosférica, que incluye reacciones fotoquímicas y procesos de formación de aerosoles, añade otra capa de complejidad. Los modelos deben incluir estos procesos para predecir con precisión la calidad del aire y los impactos de los aerosoles en las salidas de los modelos [23].

Verificación de los modelos

La segunda cuestión clave en el desarrollo de modelos es la validación de los mismos [24]. Los modelos deben ser validados en relación con las observaciones, pero hay varias razones por las que los modelos difieren de las observaciones, como las limitaciones en la precisión de los registros, la incompletitud de los datos, la resolución inadecuada de los modelos y la representación de los procesos físicos.

Los resultados de los modelos se verifican con observaciones reales de diferentes variables, tales como la temperatura del aire, la cantidad o tipo de precipitación o las concentraciones de gases de efecto invernadero. Sin embargo, las observaciones tienen sus propias incertidumbres y los datos pueden estar limitados espacial y temporalmente. En ocasiones, se utilizan otras fuentes de verificación diferentes a las observaciones reales. Por ejemplo, en el caso de los modelos climáticos se pueden utilizar datos paleoclimáticos, que incluyen información sobre el clima pasado obtenida a partir de proxies de lo más diversos, como los anillos de los árboles, muestras de hielo o sedimentos marinos. En ocasiones, estos datos son útiles para la calibración de dichos modelos que trabajan con escalas de tiempo comparativamente largas, pero también están sujetos a incertidumbres en la interpretación de los indicadores indirectos y la resolución temporal [25].

Por otro lado, se trabaja también en la intercomparación de modelos, esto es, la evaluación cualitativa de la correspondencia y las fluctuaciones de los modelos. Existen programas específicos dedicados a definir los niveles de incertidumbre para la modelización atmosférica, como el Proyecto de Intercomparación de Modelos Acoplados (CMIP) [26].

Por último, añadiendo un nuevo nivel de complejidad al sistema, cabe indicar que también se está trabajando en la combinación de modelos atmosféricos con modelos de impacto, que pueden estimar las consecuencias de, por ejemplo, el cambio climático a nivel sectorial, como la agricultura, la salud o el sector hídrico. La información producida por los modelos se aplica en los modelos de impacto para evaluar los efectos de los cambios en la atmósfera [27]. Las fluctuaciones imprevistas pueden ser tan grandes como los propios modelos a la hora de determinar los impactos, y esto añade más incertidumbre aún.

A la luz de lo anterior, quizás es más fácil comprender por qué la modelización de la atmósfera es muy útil, pero a la vez compleja y con un gran potencial de mejora. Si a los propios factores inherentes al funcionamiento de los modelos le sumamos condiciones ajenas a éstos, como pueden ser eventos imprevistos, naturalezas caóticas o la acción humana, se hace aún más evidente la necesidad de ir mejorando poco a poco este campo. Y de manera específica, la predicción de la energía eólica es un claro ejemplo de desafío, donde se conjungan los problemas inherentes a las limitaciones computacionales de los modelos con la alta variabilidad que presenta el recurso eólico. Debido al alto interés y beneficio que tiene la apuesta por la energía eólica, resolver estos problemas supone una necesidad y, en este contexto, el uso de nuevas técnicas es un campo de gran interés para la investigación.

1.5. Machine Learning

El Machine Learning es una rama de la inteligencia artificial que permite a los sistemas computacionales aprender patrones a partir de datos sin recibir instrucciones de programación directas. Su punto fuerte radica en ser capaz de encontrar patrones y estructuras complejas a partir de conjuntos de datos. Gracias a la combinación de modelos matemáticos y algoritmos de diversa tipología, el Machine Learning se ha convertido en una interesante herramienta para mejorar las predicciones, mejorar la clasificación de información y optimizar procesos que, con otras aproximaciones, requerirían un alto coste computacional [28].

La ciencia moderna está empleando masivamente las técnicas de Machine Learning ya que ha demostrado su capacidad para analizar adecuadamente sistemas de datos no lineales frente al empleo de métodos tradicionales. En contraste con los métodos basados en ecuaciones físicas y estadísticas, el Machine Learning extrae conocimiento directamente de los datos y es ciertamente útil en situaciones que requieren cambios importantes en los modelos o hay implicadas cantidades elevadas de datos. Sin embargo, aunque el Machine Learning reduce la necesidad de modelar

explícitamente todas las relaciones del sistema, su uso no es trivial: la aplicación correcta de los algoritmos de Machine Learning requiere tener, al menos, un conocimiento básico sobre los algoritmos existentes, su funcionamiento y capacidad para aplicar adecuadamente unos criterios mínimos para elegir el algoritmo adecuado para trabajar con un conjunto de datos dado. Una inadecuada aproximación a estas técnicas de Machine Learning da como resultado unas salidas poco útiles o una interpretación incorrecta de los resultados obtenidos.

Entre la amplia selección de algoritmos de Machine Learning desarrollado hasta el momento se pueden encontrar, por ejemplo, los métodos de ensemble como solución óptima para abordar problemas de predicción. Estos métodos de ensemble consisten en combinar varios modelos base que dan como resultado una predicción precisa y robusta. Un ejemplo de este tipo de algoritmos es el conocido Random Forest, un algoritmo que se fundamenta en la construcción de múltiples árboles de decisión. Su funcionamiento consiste en entrenar cada árbol con una muestra aleatoria del conjunto de datos original y, para cada división dentro del árbol, seleccionar aleatoriamente un subconjunto de variables en lugar de considerar todas las variables disponibles. Finalmente, las predicciones individuales de todos los árboles se combinan para obtener un resultado final más estable y, consecuentemente, preciso. En problemas de clasificación, esto suele hacerse mediante un voto mayoritario, mientras que en regresión se promedian las predicciones de todos los árboles. La combinación de múltiples modelos débiles en un solo modelo robusto ha hecho que Random Forest sea una de las herramientas más utilizadas en el análisis de datos científicos.

Las redes neuronales profundas (Deep Learning) han permitido que los modelos de aprendizaje automático identifiquen y extraigan automáticamente patrones y características relevantes a partir de datos en bruto, sin necesidad de una preselección manual de variables. Esto es especialmente útil en contextos donde las relaciones entre los datos son complejas o difíciles de modelar con métodos tradicionales. Gracias a esta capacidad, estos modelos han logrado avances notables en la identificación de dependencias entre variables, el análisis de datos secuenciales y el procesamiento de información en bruto, optimizando la representación y el entendimiento de estructuras subyacentes en los datos [29].

En términos generales, el Machine Learning ha mejorado notablemente el análisis y la predicción en diversos campos científicos. Su habilidad para integrar enfoques fundamentados en datos con modelos físicos ha generado metodologías híbridas que pueden optimizar la precisión y la eficiencia en problemas complejos. No obstante, su utilización exige un conocimiento adecuado sobre las técnicas disponibles y de los

principios subyacentes para asegurar que los modelos sean interpretables, precisos y robustos.

Dadas las limitaciones o complicaciones que presentan los modelos numéricos, especialmente debido al gran volumen de datos que se gestionan, se ha advertido la necesidad de aplicar técnicas complementarias que ayuden a mejorar la precisión de las predicciones realizadas únicamente mediante modelos numéricos. En este campo, una de las ramas más prometedora es el uso de técnicas de Machine Learning, dando lugar a lo que se conoce como modelos híbridos. En estos modelos se busca complementar los datos de salida que nos aportan los modelos meteorológicos con algoritmos de Machine Learning buscando corregir posibles errores sistemáticos en las salidas de los modelos, mejorar la resolución espacial y/o temporal mediante técnicas de downscaling estadístico o incorporar datos de observaciones en tiempo real para ajustar las predicciones [30] [31].

1.6. Introducción a los modelos híbridos

Desde la perspectiva de la Meteorología, los modelos híbridos representan un enfoque diferente e innovador que combina las fortalezas de diferentes modelos numéricos con las capacidades predictivas del Machine Learning. Combinar de manera adecuada estas dos aproximaciones tiene un gran interés en la apuesta por la energía eólica, pues se pueden conseguir predicciones más fiables que, como ya se ha indicado, son imprescindibles para una eficiente predicción de la generación de energía eléctrica mediante tecnologías eólicas [32].

Un modelo híbrido es un sistema que se compone, principalmente, de dos elementos: [33]:

- Modelos numéricos de predicción meteorológica: estos modelos, como el WRF mesoescalar o los modelos globales, se utilizan para resolver las ecuaciones físicas que gobiernan los procesos atmosféricos. Sin embargo, están limitados por los posibles errores en los datos de entrada, en la discretización que realizan en el espacio y en el tiempo para resolver las ecuaciones físicas o en los procesos no resueltos por los modelos y que, por consiguiente, se tienen que parametrizar.
- Técnicas de Machine Learning: tienen una gran utilidad a la hora de corregir errores sistemáticos, mejorar las resoluciones de las predicciones y modificar las salidas de un modelo numérico a las condiciones locales de un lugar.

El Machine Learning se ha establecido en el ámbito científico y tecnológico como una herramienta esencial para el progreso en el análisis de datos, y la Meteorología no es una excepción [34], particularmente cuando se combina con modelos numéricos convencionales. La combinación y utilización de las capacidades de ambos métodos fortalece la habilidad de predecir fenómenos atmosféricos complejos. Esta sinergia puede facilitar la resolución de desafíos específicos o complejos, como la variabilidad del viento y la integración eficiente de la energía eólica en la red eléctrica.

Una de las principales ventajas de los modelos híbridos es su capacidad para aumentar la precisión de las predicciones. A través de técnicas de Machine Learning como las redes neuronales o los modelos de regresión avanzados, es posible identificar errores sistemáticos y corregirlos o, al menos, atenuar sus efectos. Por ejemplo, en la predicción del viento, las técnicas de Machine Learning puede ajustar las salidas de los modelos numéricos para considerar efectos locales, como las brisas o las corrientes de aire forzadas o modificadas por obstáculos.

Mediante técnicas de Machine Learning se puede realizar un downscaling estadístico para optimizar la resolución espacial y temporal de las predicciones producidas por modelos numéricos. Tanto los modelos globales como los mesoescalares [35] ofrecen predicciones de alta resolución; sin embargo, esta resolución puede resultar inadecuada para ciertos estudios. El Machine Learning puede procesar los datos de salida proporcionados por los modelos y optimizar la información para mejorar la resolución. Esto es especialmente importante en zonas complejas donde el viento esté influenciado por factores locales afectando tanto a su dirección como a su velocidad.

El Machine Learning puede adaptarse de varias maneras, ya sea integrando datos de observaciones tras la ejecución de las simulaciones del modelo numérico, incorporando observaciones en tiempo real o ajustándose a modificaciones en el modelo numérico empleado. Esta adaptabilidad es esencial en la predicción a corto plazo, incluso en resoluciones de minutos, donde la precisión es fundamental. Por ejemplo, en el contexto de los parques eólicos, el Machine Learning puede integrar datos de modelos numéricos con mediciones en tiempo real de estaciones meteorológicas o los sensores de los aerogeneradores, para intentar mejorar la predicción a corto plazo. Esto facilita la adaptación a las futuras demandas de la red eléctrica, disminuyendo costos y optimizando la eficiencia del sistema.

Una de las principales fortalezas del Machine Learning es su habilidad para integrar y analizar datos de diversas naturalezas, incluso cuando provienen de fuentes significativamente diferentes. Contrariamente a los modelos tradicionales, que generalmente se restringen a variables meteorológicas, el Machine Learning puede integrar información de diversas categorías, como datos operativos, mediciones

en tiempo real, especificaciones técnicas de equipos o incluso datos históricos de rendimiento. Esta flexibilidad facilita el análisis de problemas complejos desde una perspectiva más amplia, permitiendo la identificación de relaciones no lineales y patrones ocultos que resultarían difíciles de discernir mediante métodos convencionales [36].

En el campo de la energía eólica, el Machine Learning se presenta como un instrumento altamente eficaz. En este contexto, además de optimizar la predicción del viento, permite calcular directamente la generación de energía considerando factores como el desempeño de las turbinas, la orientación de las palas, las condiciones externas y el estado técnico de los equipos. Este enfoque se centra en la optimización de la gestión de los parques eólicos, tanto en el apartado puramente meteorológico como en la operativa de los aerogeneradores. Además, bajo determinadas condiciones, puede disminuir significativamente el costo computacional de los modelos numéricos de alta resolución. Por ejemplo, el uso de técnicas de Machine Learning entrenadas con un histórico de datos permite generar predicciones rápidas sin necesidad de ejecutar el modelo físico completo en cada iteración. De este modo, en lugar de simular directamente un campo de vientos con el WRF a 1 km de resolución, que requiere de una gran capacidad de cómputo para lograr tan alta resolución, se pueden usar técnicas de downscaling basadas en Machine Learning que interpolan y corrigen resultados de modelos con una menor resolución espacial, lo cual conlleva un menor costo computacional, sin comprometer en exceso la precisión de los resultados [37].

Capítulo 2

El modelo WRF (Weather Research and Forecasting)

El modelo de Investigación y Predicción del Tiempo (Weather Research and Forecasting Model (WRF) por sus siglas en inglés [38]) es una herramienta versátil diseñada para la simulación y predicción de la atmósfera. Este modelo es ampliamente utilizado en la investigación meteorológica y en aplicaciones operativas de pronóstico del tiempo. Su desarrollo es el resultado de un esfuerzo colaborativo entre múltiples instituciones, lo que ha permitido la creación de un sistema robusto y altamente adaptable.

2.1. Desarrollo y colaboración

El WRF es mantenido y desarrollado por un consorcio de instituciones liderado por el National Center for Atmospheric Research (NCAR) [39], la National Oceanic and Atmospheric Administration (NOAA) [40], el National Centers for Environmental Prediction (NCEP) [41] y varias universidades y laboratorios nacionales de los EE.UU. Al estar desarrollado sobre la base de dicha colaboración, el modelo incorpora los avances más actuales en ciencia atmosférica, modelado numérico y técnicas de asimilación de datos.

El NCAR, en Boulder, Colorado, es una de las instituciones de estudios atmosféricos y climáticos más importante a nivel mundial. Se dedica a la investigación de procesos atmosféricos para mejorar las predicciones meteorológicas y climáticas, y es operado por la University Corporation for Atmospheric Research (UCAR) [42]. Su superordenador, Cheyenne, permite simulaciones de alta resolución. Uno de sus objetivos principales es desarrollar herramientas y modelos avanzados en colaboración con otras instituciones a nivel global. La NOAA, creada en 1970, es una agencia del

gobierno de los Estados Unidos encargada de comprender y predecir los cambios en nuestra atmósfera, océanos, costas, la contaminación dañina del aire y el agua y los desastres naturales, como los ciclones tropicales. El National Weather Service (NWS) [43] funciona bajo el paraguas de la NOAA y proporciona pronósticos y advertencias meteorológicas. También controla satélites que son importantes para recopilar datos críticos sobre la atmósfera. Además, la NOAA desempeña un papel fundamental en la gestión de los recursos marinos y costeros y es experta en responder a los desastres naturales, lo que ha permitido salvar muchas vidas y propiedades. Los pronósticos meteorológicos y climáticos avanzados del NCEP son parte de la NOAA. El NCEP tiene nueve centros especializados como el National Hurricane Center (NHC) [44] y el Climate Prediction Center (CPC) [45] que van desde la predicción meteorológica diaria hasta el monitoreo climático a largo plazo. El NCEP utiliza datos de observación y modelos numéricos para producir pronósticos precisos que son utilizados por diversos sectores, tales como la aviación, la agricultura o la gestión de emergencias.

El desarrollo de WRF fue guiado principalmente por dichas organizaciones y comenzó a finales de la década de 1990 con el objetivo de construir una plataforma de modelado común adecuada para su uso en investigación y aplicaciones operativas. En el año 2000 se lanzó la primera versión de WRF y el modelo se ha actualizado y mejorado continuamente desde entonces. La ciencia atmosférica siempre está cambiando, por lo que cada versión contiene nuevas características, mejoras de rendimiento y correcciones de errores con el fin de ir satisfaciendo las necesidades que plantea la comunidad de usuarios.

El WRF es un software de código abierto, lo cual es uno de sus aspectos más relevantes. Está disponible para la comunidad científica global, así como para servicios meteorológicos y otras organizaciones interesadas en la predicción del tiempo y la investigación atmosférica. Además, la licencia abierta del WRF permite a los usuarios modificar y adaptar el código fuente a sus necesidades específicas y no depender únicamente de actualizaciones impulsadas por una única institución, lo que ha fomentado una amplia base de usuarios y una activa comunidad de desarrollo.

La comunidad de usuarios del WRF está formada por personas de universidades y laboratorios de todo el mundo, investigadores y meteorólogos operativos de los servicios meteorológicos nacionales. Esta comunidad desempeña un papel clave en la evolución del modelo: proporciona retroalimentación, informa sobre errores detectados y mejora el modelo con capacidades adicionales. Los foros, talleres y conferencias en línea permiten a los usuarios colaborar entre sí y trabajar juntos para mejorar el software.

Programación y Arquitectura del WRF

El diseño altamente modular y la capacidad de implementación en lenguajes de programación eficientes hacen del WRF una herramienta potente y flexible para muchas aplicaciones. Más adelante se detallará su programación y arquitectura.

WRF está escrito principalmente en Fortran, un lenguaje de programación heredado que históricamente fue favorecido por la comunidad de modelado numérico por su manejo eficiente de cálculos intensivos y el acceso a arquitecturas de cómputo de alto rendimiento. Fortran destaca, además, en el procesamiento de grandes volúmenes de datos y la ejecución de algoritmos complejos en superordenadores.

Además de lo anterior, la elección de Fortran se basa en varias ventajas específicas:

- Rendimiento: Fortran está diseñado para operaciones matemáticas y científicas, permitiendo un alto rendimiento en cálculos numéricos intensivos.
- Optimización: los compiladores de Fortran están altamente optimizados para arquitecturas de alto rendimiento, lo que permite una ejecución más rápida y eficiente.
- Legado: la comunidad de modelado climático y meteorológico tiene una larga historia de desarrollo en Fortran, lo que facilita la integración de códigos y bibliotecas preexistentes.

El WRF tiene una arquitectura modular, por lo que se pueden incorporar o actualizar diversos componentes del modelo sin tener que reescribir todo el código. Esa modularidad lo hace más flexible y adaptable. Los principales componentes del WRF [46] son:

- Núcleo Dinámico (Dynamic Core):
 - Advanced Research WRF (ARW): es el núcleo más utilizado en la comunidad de investigación. El conjunto completo de ecuaciones hidrostáticas y no hidrostáticas se implementa para modelar una amplia gama de fenómenos atmosféricos en varias escalas espaciales y temporales. El ARW es conocido por su flexibilidad y capacidad de configuración.
 - Nonhydrostatic Mesoscale Model (NMM): originalmente fue desarrollado por NCEP para aplicaciones operativas y lo utilizan el Rapid Forecast System (RAP) [47] y el Numerical Weather Forecast System (NAM) [48]. El NMM se ejecuta de manera eficiente y está optimizado para la predicción operativa a corto plazo.

- Físicas atmosféricas (Physics). Las parametrizaciones físicas son componentes esenciales que representan procesos atmosféricos que no se pueden resolver explícitamente dentro del modelo debido a que ocurren a escalas inferiores a la resolución espacial del modelo. Los módulos de física incluyen:
 - Microfísica de las nubes: modela los procesos de formación, crecimiento y precipitación de partículas de agua y hielo. Thompson Microphysics [49] y WSM6 [50] son ejemplos de esquemas microfísicos
 - Convección: transporte de calor y humedad por corrientes ascendentes y descendentes en la atmósfera. El esquema de Kain-Fritsch [51] y el esquema de convección de Tiedtke [52] son algunos ejemplos de esquemas de convección.
 - Radiación: modela la transferencia de radiación solar y terrestre a través de la atmósfera, incluida la absorción, emisión y dispersión por gases y aerosoles. El esquema RRTMG [53] y el esquema Dudhia [54] son ejemplos de esquemas de radiación.
 - Interacciones superficie-atmósfera: modela los intercambios de energía, humedad y momento entre la superficie terrestre/oceánica y la atmósfera. Esto incluye esquemas de capa límite planetaria (PBL) [55] y modelos de superficie terrestre como el modelo de superficie terrestre NOAH [56].
- Asimilación de datos (Data assimilation). la asimilación de datos es un proceso crucial que permite incorporar observaciones atmosféricas en el modelo para mejorar las condiciones iniciales de las simulaciones. El WRF utiliza varios métodos de asimilación de datos, entre ellos:
 - 3DVAR (Three-Dimensional Variational Analysis): un método variacional que minimiza la diferencia entre las observaciones y las simulaciones del modelo.
 - 4DVAR (Four-Dimensional Variational Analysis): similar al 3DVAR, pero incorpora la dimensión temporal, mejorando la consistencia temporal de las asimilaciones.
 - Ensemble Kalman Filter (EnKF) [57]: utiliza un conjunto de simulaciones para representar la incertidumbre en el estado atmosférico y actualizar las condiciones iniciales basadas en las observaciones.
- Preprocesamiento y Postprocesamiento (Pre- and Post-processing):

- WRF Preprocessing System (WPS): este sistema prepara los datos de entrada necesarios para el modelo WRF. Incluye varias utilidades como geogrid (genera los datos geográficos del dominio de simulación), ungrib (procesa los datos meteorológicos en bruto) y metgrid (interpola los datos meteorológicos en la rejilla del modelo).
- Postprocesamiento: incluye herramientas y scripts para analizar y visualizar los resultados de las simulaciones permitiendo generar gráficos, mapas y estadísticas a partir de los datos de salida del WRF. Las herramientas comúnmente utilizadas incluyen NCL (NCAR Command Language) [58], GrADS (Grid Analysis and Display System) [59], y Python con bibliotecas como Matplotlib [60] y PyNGL [61].

El WRF está también diseñado para funcionar de manera eficiente en una amplia gama de plataformas informáticas, desde algunas estaciones de trabajo individuales hasta miles de núcleos en un superordenador. Para lograr esto, incorpora varias técnicas de paralelización y optimización:

- Paralelización por dominio: el dominio de simulación se divide en subdominios más pequeños que pueden procesarse en paralelo por diferentes núcleos de procesamiento. MPI (Message Passing Interface) [62] es una biblioteca estándar para programación paralela que se utiliza para lograr esto.
- Paralelización por hilado: WRF también utiliza OpenMP (Open Multi-Processing) [63] para permitir que varios núcleos en una sola máquina se ejecuten en paralelo mediante la creación de subprocesos de ejecución paralelos en cada subdominio.
- Optimización de código: el código del WRF está diseñado para optimizar los tiempos de cálculo mientras se consume memoria de manera eficiente. Estos incluyen el uso de algoritmos más eficientes, la eliminación de operaciones redundantes y el aprovechamiento de las fortalezas específicas del hardware, como las instrucciones vectoriales.

Por otro lado, el diseño modular del WRF permite a los usuarios personalizar y extender el modelo para satisfacer necesidades específicas. Esto incluye la posibilidad de:

- Agregar nuevos esquemas físicos: los usuarios pueden desarrollar y agregar sus propios esquemas físicos para representar procesos atmosféricos específicos que no están cubiertos por los esquemas existentes.

- Modificar componentes del modelo: dado que el código fuente está disponible, los usuarios pueden modificar cualquier componente del modelo para mejorar su rendimiento o adaptarlo a situaciones particulares.
- Desarrollar nuevas capacidades de asimilación de datos: los usuarios pueden implementar y probar nuevos métodos de asimilación de datos, mejorando la precisión de las condiciones iniciales y, por ende, las predicciones del modelo.

Colaboración y comunidad

Como ya se ha indicado, una de las grandes fortalezas del modelo WRF es su comunidad de usuarios y desarrolladores, muy activa y colaborativa. Son muchos los espacios, tanto oficiales como no oficiales, en los que los usuarios se reúnen para compartir experiencias, resolver dudas y ayudar a que el modelo evolucione. Uno de los principales puntos de encuentro es el foro oficial ^[1], administrado por el NCAR, en el cual se pueden discutir problemas técnicos, obtener ayuda y compartir configuraciones. Este foro, accesible en la página oficial del WRF, es un recurso clave donde tanto expertos del NCAR como de la comunidad general, participan activamente para resolver preguntas.

También existen plataformas creadas por la comunidad fuera de los canales oficiales como wikis y foros no oficiales donde los usuarios de WRF comparten respuestas a diversas cuestiones y problemas que probablemente no estén (aún) cubiertos con documentación elaborada. Además, los usuarios colaboran en redes sociales como LinkedIn, Facebook y Reddit ^[2], dónde existen grupos de intercambio directo de conocimientos que funcionan de manera informal y en los que se resuelven problemas de forma conjunta y en tiempo real. Los recursos oficiales, que se complementan con estos espacios extraoficiales, permiten la conexión y la actividad de la comunidad en diferentes plataformas.

El NCAR y otras de las instituciones involucradas en el desarrollo del modelo brindan soporte técnico oficial del WRF. Este servicio se presta a través de varios canales, entre ellos el foro oficial en el que el personal técnico del NCAR responde a las consultas de los usuarios. El soporte directo a los usuarios ofrece, además de esto, capacitación sobre el uso del modelo ya que el NCAR organiza talleres anuales que incluyen formación práctica sobre el uso del modelo. Algunas instituciones tienen acuerdos más estrechos, sobre un alcance de actividades más específico, que les otorgan un soporte técnico más directo cuando utilizan el WRF para algunas de sus

¹<https://forum.mmm.ucar.edu/>

²<https://www.reddit.com/r/meteorology/>

operaciones. El sistema de soporte en este modelo hace que, ya sea un principiante o un usuario avanzado, no se quede atascado con problemas técnicos y siempre esté al tanto de los desarrollos más recientes del modelo.

La documentación del WRF es extensa y contiene todo lo que se necesita saber. El sitio web del NCAR aloja un manual oficial (Guía del usuario del WRF ³) que abarca desde la instalación más básica hasta la configuración de parámetros avanzados para simulaciones específicas. Además, hay disponible documentación técnica más específica, incluidos los manuales WRF-ARW y WRF-Chem ⁴, que tratan las últimas versiones del modelo utilizado para investigaciones avanzadas o simulaciones químicas. NCAR proporciona, asimismo, enlaces a tutoriales interactivos en línea ⁵ que los usuarios pueden completar mediante ejercicios prácticos. Esto permite que cada cual lleve su propio ritmo de aprendizaje autodidacta y a cualquier persona, con interés o afición por el modelo, aprender sobre él.

2.2. Ventajas del WRF

Existen varias peculiaridades que hacen que el modelo WRF sea único entre los demás modelos meteorológicos. El WRF es extremadamente flexible, permitiendo al usuario elegir entre una gran cantidad de esquemas físicos para representar los diversos procesos en la atmósfera, por ejemplo, esquemas de microfísica de nubes, esquemas convectivos, esquemas de capa límite planetaria y esquemas de radiación. Esta flexibilidad le otorga una ventaja significativa frente a otros modelos que tienen configuraciones más predefinidas y ofrecen menos capacidad de ajuste a las necesidades específicas del usuario. Además, en WRF se pueden definir dominios anidados, es decir, se puede mejorar la resolución en ciertas regiones de una aplicación de modelado más grande. Esto es particularmente útil en aplicaciones donde el usuario necesita información de alta resolución espacial en la ubicación de interés. Otros modelos como COSMO ¹³ o ALADIN ¹⁴ también admiten la anidación, pero WRF es más configurable en la forma de manejar estos dominios anidados.

En términos de datos de entrada, WRF se destaca como un modelo versátil que puede asimilar varios datos de numerosas fuentes de información, tales como estaciones meteorológicas, satélites, radares u otros sensores que se consideren relevantes. Por esta razón, es altamente flexible en función de la escala temporal y espacial. La capacidad del WRF para asimilar datos de alta resolución y utilizar sistemas avanzados de asimilación de datos como el WRF Data Assimilation System (WRF-DAS)

³https://www2.mmm.ucar.edu/wrf/users/docs/user_guide_v4/contents.html

⁴<https://www2.mmm.ucar.edu/wrf/users/docs/>

⁵<https://www2.mmm.ucar.edu/wrf/OnLineTutorial/>

[64], le da una ventaja frente a otros modelos que pueden estar más limitados en cuanto a la tipología de los datos que pueden procesar.

El WRF se utiliza tanto en el ámbito de la investigación como en el operativo, lo que lo convierte en un programa muy flexible. Comparativamente, otros modelos se limitan a un solo tipo de aplicación. Por ejemplo, un modelo similar como es el HARMONIE-AROME [65], está diseñado para la predicción a corto plazo; no se adaptan a la investigación sobre la dinámica atmosférica de mesoescala.

Una debilidad importante de WRF en comparación con otros modelos es su gran demanda de recursos computacionales, debido principalmente a las altas resoluciones con las que trabaja y a los grandes dominios que abarca. Por lo tanto, aunque los modelos como GFS o ECMWF tienen una resolución menor que WRF y son menos adecuados para simulaciones regionales, son más eficientes desde el punto de vista computacional para proporcionar pronósticos globales.

Por otro lado, el requisito de necesitar de una gran capacidad de cómputo puede ser una limitación importante para algunos potenciales usuarios, particularmente en ciertos entornos o instituciones de aprendizaje. Esto a veces conlleva que se prefieran modelos más simples como ALADIN o COSMO en usos operativos rutinarios donde una resolución muy alta puede no ser siempre necesaria.

La gran flexibilidad del WRF es, al mismo tiempo, su debilidad, porque la elección correcta de los esquemas físicos y dinámicos es decisiva para el funcionamiento adecuado del modelo. Modelos como, por ejemplo, el GFS tienen configuraciones bastante establecidas, lo que significa que las configuraciones son más simples de aplicar y hay menos posibilidades de centrarse en configuraciones incorrectas que produzcan resultados más erróneos o incongruentes.

Además, el WRF tiene una curva de aprendizaje más pronunciada que la mayoría de los otros modelos. Esto significa que los usuarios requieren de una formación técnica significativa para manejarlo adecuadamente, mientras que otros modelos operacionales, como el COSMO o el HARMONIE, pueden ser más fáciles de implementar en configuraciones estándar.

La segunda limitación crítica es la alta sensibilidad del modelo WRF con respecto a la configuración del esquema físico. Se ha observado que, dependiendo de las condiciones y la ubicación geográfica, diferentes esquemas de parametrización pueden dar resultados muy diferentes [66] [67] [68]. La flexibilidad de elegir una u otra opción de la lista es una fortaleza, pero se convierte en una debilidad si las personas que lo usan no conocen todas las consecuencias de ciertas configuraciones en los resultados que presenta el modelo. Una implicación directa de esto es la necesidad de validación y calibración, un requisito que quizás sea menos necesario en

el modelado superficial con menos modos posibles de parametrización.

En resumen, los puntos fuertes del WRF frente a otros modelos se pueden resumir en los siguientes puntos:

- Capacidad de personalización: el WRF ha ganado popularidad por el hecho de que permite al usuario ejecutar el modelo con diferentes configuraciones. Esto lo convierte en una herramienta muy útil en la investigación atmosférica avanzada y constituye parte de la base de este trabajo de investigación.
- Resolución espacial: la capacidad de trabajar en cuadrículas de alta resolución lo hacen adecuado para la investigación, como la previsión de eventos extremos localizados o el modelado de fenómenos atmosféricos en entornos urbanos.
- Comunidad activa y soporte continuo: el WRF cuenta con una amplia comunidad de usuarios que colaboran en su mejora continua. Las actualizaciones y revisiones del código son frecuentes, lo que asegura que el modelo esté al día con los avances en la Física de la Atmósfera y la modelización numérica.

Y, por otro lado, sus puntos débiles podrían resumirse en los siguientes puntos:

- Requerimientos computacionales elevados: la capacidad del WRF para funcionar en alta resolución implica un mayor coste computacional, lo que puede ser una limitación para su uso rutinario o en estudios con plazos ajustados.
- Complejidad operativa: debido a su amplia capacidad de personalización, los usuarios deben tener un conocimiento avanzado sobre la ciencia meteorológica y la modelización numérica, lo que puede restringir su adopción en ambientes operacionales que necesitan pronósticos rápidos y confiables.
- Dificultad en la validación: la flexibilidad del WRF en cuanto a los esquemas físicos implica que se requiere un proceso riguroso de validación para garantizar que las simulaciones reflejen correctamente la realidad. Esto puede no ser tan necesario en modelos con parametrizaciones más estandarizadas, donde la validación ha sido más exhaustiva en configuraciones típicas.

2.3. El futuro del WRF: potencial y perspectivas de avance en la modelización de la atmósfera

El WRF se encuentra en una etapa crítica de crecimiento, debido a que las innovaciones que se están realizando tienen el potencial de ampliar la gama de usos

del modelado de la atmósfera. Con las crecientes expectativas en los modelos numéricos, el WRF puede contribuir al comienzo de un nuevo cambio de paradigma en la predicción y la simulación. Otro desarrollo potencial es el fortalecimiento de los aspectos del modelado de los ciclones tropicales, conocidos como huracanes en el Atlántico oeste y el Pacífico este, a través del módulo Hurricane WRF (HWRF) [69]. A medida que la resolución computacional avance hasta el punto en que la circulación secundaria se pueda resolver con mayor detalle, se podrá obtener información sobre características internas como el ojo y la pared del huracán, lo que ayudará a comprender los diversos factores que sustentan alteraciones de corta duración en la intensidad que no se pueden abordar actualmente. Sin embargo, podría abrir posibilidades de utilizar mejores modelos oceánicos que incorporen variaciones espacio-temporales de sus intercambios con la atmósfera necesarias para comprender la energía que impulsa dichos sistemas [70].

Mirando hacia el futuro, se piensa que la inteligencia artificial y el aprendizaje automático serán fundamentales en el desarrollo futuro del WRF. Se cree que estas herramientas son capaces de mejorar la calibración de parámetros, minimizar las incertidumbres y mejorar la adaptabilidad en tiempo real del modelo con respecto a diferentes condiciones meteorológicas.

En cuanto a la resolución multiescala, el WRF tiene un camino prometedor. Las simulaciones globales de alta precisión podrían integrarse dinámicamente con modelos regionales y locales, permitiendo una transición fluida entre diferentes niveles de detalle. Esto sería especialmente relevante en el análisis de algunos fenómenos inusuales que se observan en pequeñas regiones pero que pueden tener un gran impacto, por ejemplo inundaciones locales [71] [72]. Además, la representación mejorada de las cubiertas urbanas hará de WRF la plataforma central para servir al diseño de soluciones de ciudades inteligentes que puedan abordar las próximas olas de calor, episodios de contaminación del aire y otras adversidades en áreas pobladas de alta densidad.

En el contexto de la transformación del clima, WRF tiene un gran potencial para explorar procesos climáticos intrincados y sus implicaciones a escala climática local/regional. Las simulaciones futuras no solo podrían mostrar estados difíciles de pronosticar, como sequías severas o tormentas intensas [73], sino también demostrar su ocurrencia junto con otros estados en el mismo lugar, por ejemplo, ciclones tropicales en áreas que sufren grandes incendios forestales [74]. Si se vinculan a las predicciones climáticas basadas en escenarios del IPCC, estos resultados permitirían evaluaciones más elaboradas de cómo el calentamiento global afecta al clima, lo que resulta muy útil para la planificación y los análisis de riesgos.

Otra aplicación interesante se refiere a su capacidad para predecir estados atmosféricos más antiguos, un proceso denominado asimilación inversa de datos o retrospección. Sin embargo, cabe señalar que debido a las propiedades caóticas y no lineales de las ecuaciones fundamentales del modelo, no se puede demostrar que el flujo dinámico retroceda en el tiempo; sin embargo, incluso en este caso, el uso de ciertas herramientas de alta tecnología permite realizar tales estimaciones. En la actualidad, mediante el acoplamiento con técnicas de asimilación de datos, el WRF puede imitar una atmósfera anterior incorporando observaciones previas. Este enfoque es crucial para estudiar eventos históricos severos y desastrosos y comprender la sinóptica compleja de los eventos meteorológicos que han ocurrido. En segundo lugar, los procedimientos de optimización inversa permiten realizar cambios en los datos iniciales para el cálculo secuencial de forma progresiva, de modo que los resultados se ajusten a los estados finales dados; por lo tanto, estos enfoques proporcionan medios bastante eficaces para rastrear fenómenos atmosféricos de forma retrospectiva. Este tipo de simulación puede ser fundamental en muchos casos de trazabilidad, por ejemplo, el seguimiento de los orígenes de la contaminación o la reconstrucción del movimiento de partículas en el aire. En el futuro, estas capacidades podrían beneficiarse de las técnicas de inteligencia artificial, ya que los hechos y los historiales son amplios y podrían entrenarse para inferir resultados más precisos sobre las alteraciones retrospectivas [75] [76] [77] [78] [79].

WRF-Chem [80] tiene un gran futuro para el modelado de la calidad del aire. Poco a poco, a medida que las resoluciones mejoren junto con la integración de fuentes de datos en vivo basadas en eventos, se estima que se proporcionarán simulaciones más realistas de eventos como los incendios forestales [81] y los efectos de las emisiones industriales [82] en la salud pública. Estos niveles de precisión serán muy útiles, en particular, al diseñar métodos prácticos para operar medidas de resiliencia en tiempo real, especialmente en vecindarios congestionados.

En el área de la energía y las ventas en particular, las futuras capacidades de WRF pueden mejorar la planificación y la gestión de recursos relacionados con la energía renovable y las ventas en la agricultura. En concreto, la mejora en la predicción de las condiciones de viento, radiación solar y precipitaciones a escala local podrán redundar en una mejora en la generación de energía eólica, hidráulica y solar. Por otro lado, mediante la vinculación de modelos meteorológicos con modelos de crecimiento de cultivos, se conseguirá un apoyo a la toma de decisiones en sectores agrícolas sensibles al cambio climático para garantizar la producción sostenible de cultivos en climas variables para regiones de alta volatilidad climática [83].

Por último, pero no por ello menos importante, la llegada de la computación a

escala de exaescala supondrá un punto de inflexión para el WRF [84]. El sistema combinado de datos de observación en tiempo real y la tecnología de asimilación de datos ha experimentado un enorme progreso y la capacidad para ejecutar simulaciones globales de alta resolución en tiempo real revolucionará no sólo las aplicaciones meteorológicas operativas de los satélites geoestacionarios de EUMETSAT, sino también los estudios climáticos y ambientales con productos y mejoras de datos únicos. Debido a la capacidad de abordar la turbulencia a escala de vórtice y simular procesos de alta complejidad que ocurren en todo el mundo, el WRF será una herramienta principal para trabajar con los problemas del siglo XXI en las interconexiones entre el clima, la sociedad y los ecosistemas [85].

En resumen, el futuro del WRF se perfila como una etapa de transformación radical en su capacidad para abordar fenómenos atmosféricos y climáticos cada vez más complejos. Desde la mejora de la previsión de la ocurrencia de fenómenos meteorológicos adversos hasta su acoplamiento con modelos climáticos, químicos y urbanos, el WRF está bien posicionado para convertirse en un elemento central en los análisis y pronósticos de un clima que empeora debido al cambio climático. Este avance continuo no sólo será importante para la meteorología, sino que afectará a diversos campos, como la gestión de riesgos, la salud pública, la energía renovable o la agricultura sostenible, y ofrecerá nuevas cotas para la ciencia atmosférica. Además, las sugerencias de que puede reconstruir estados atmosféricos pasados y proporcionar una mejor visión retrospectiva pueden contribuir a nuevas posibilidades para la investigación histórica y la gestión de los ecosistemas.

Capítulo 3

Configuración del WRF para la realización del estudio

3.1. Configuración básica del modelo

Dominio Espacial

El modelo WRF ha sido configurado con un esquema multinivel que incluye un dominio principal (D01) que cubre todo el globo y dos dominios anidados (D02 y D03) con una resolución espacial progresivamente más alta para estudiar un área de interés específica con más detalle.

Dominio principal (D01) **Figura 3.1:** este dominio abarca la práctica totalidad de la superficie terrestre entre las latitudes $-88,657^\circ$ a $88,657^\circ$ y longitudes de $-178,667^\circ$ a $178,667^\circ$, proporcionando así una cobertura global que es capaz de simular fenómenos planetarios a gran escala, como los patrones de circulación atmosférica, las interacciones entre la atmósfera y el océano y los transportes de masas de aire. Con una resolución horizontal de 294,3 km, tiene suficientemente resolución como para resolver las características globales de la dinámica atmosférica. Este dominio tiene un tamaño de malla de 136×68 puntos de red, lo que le permite un adecuado balance entre detalle espacial y eficiencia computacional.

Dominio 2 (D02) **Figura 3.1:** este dominio se encuentra anidado dentro del dominio global. Se ha centrado en una amplia región, que incluye toda Europa, el Medio Oriente y el norte de África. Las coordenadas de cobertura de este dominio abarcan desde $15,224^\circ$ a $65,373^\circ$ en latitud y de $-46,222^\circ$ a $62,222^\circ$ en longitud. Este dominio tiene una resolución horizontal de aproximadamente 98,1 km y puede

capturar fenómenos de mesoescala, como sistemas de tormentas regionales o patrones de viento locales. Tiene un tamaño de cuadrícula de 124×58 puntos, lo que asegura una representación más detallada de los procesos en esta región.

Dominio 3 (D03) **Figura 3.1:** el dominio anidado de mayor resolución se ha centrado en la Península Ibérica. Las latitudes van de $30,985^\circ$ a $48,716^\circ$, y las longitudes van de $-11,911^\circ$ a $3,911^\circ$. Este dominio tiene una resolución horizontal de aproximadamente 19,6 km, que permiten simular procesos de pequeña escala, como precipitaciones locales intensa, interacciones topográficas o la evolución de los sistemas convectivos (su malla, compuesta por 91×101 puntos, permite capturar detalles de pequeña escala de los procesos atmosféricos).

Proporciones de anidamiento: se emplean razones de escala de **3:1** (Dominio 2 a Dominio 1) y **5:1** (Dominio 3 a Dominio 2) para la configuración de dominios anidados, debido a que la transición entre resoluciones debe ser gradual. Esto facilita la resolución de fenómenos a menor escala, manteniendo simultáneamente la dinámica del dominio global.

La posición relativa de los dominios secundarios dentro del dominio global se establece mediante puntos de red:

- **Dominio 2:** la ubicación inicial corresponde a (51, 40), que corresponde a la columna 51 y la fila 40 en la malla del Dominio 1.
- **Dominio 3:** la posición inicial en la que se inicia es (40, 19), que corresponde a la columna 40 y la fila 19 de la cuadrícula correspondiente al Dominio 2.

Estas ubicaciones representan los puntos de la malla en los que la esquina inferior izquierda del dominio anidado se encuentra contenida en la malla del dominio global en WRF. Estos índices son siempre relativos al dominio que contiene al anidado.

Cálculo de coordenadas geográficas: el punto de inicio del dominio principal sirve como referencia para convertir dichas posiciones de la cuadrícula en coordenadas geográficas (latitud y longitud) correspondientes. Bajo esta perspectiva, las posiciones (i, j) se transforman en desplazamientos horizontales y verticales mediante la resolución espacial del dominio global:

- dx : la distancia entre puntos en longitud se denomina distancia espacial (resolución horizontal).
- dy : la distribución de la latitud entre puntos (resolución vertical).

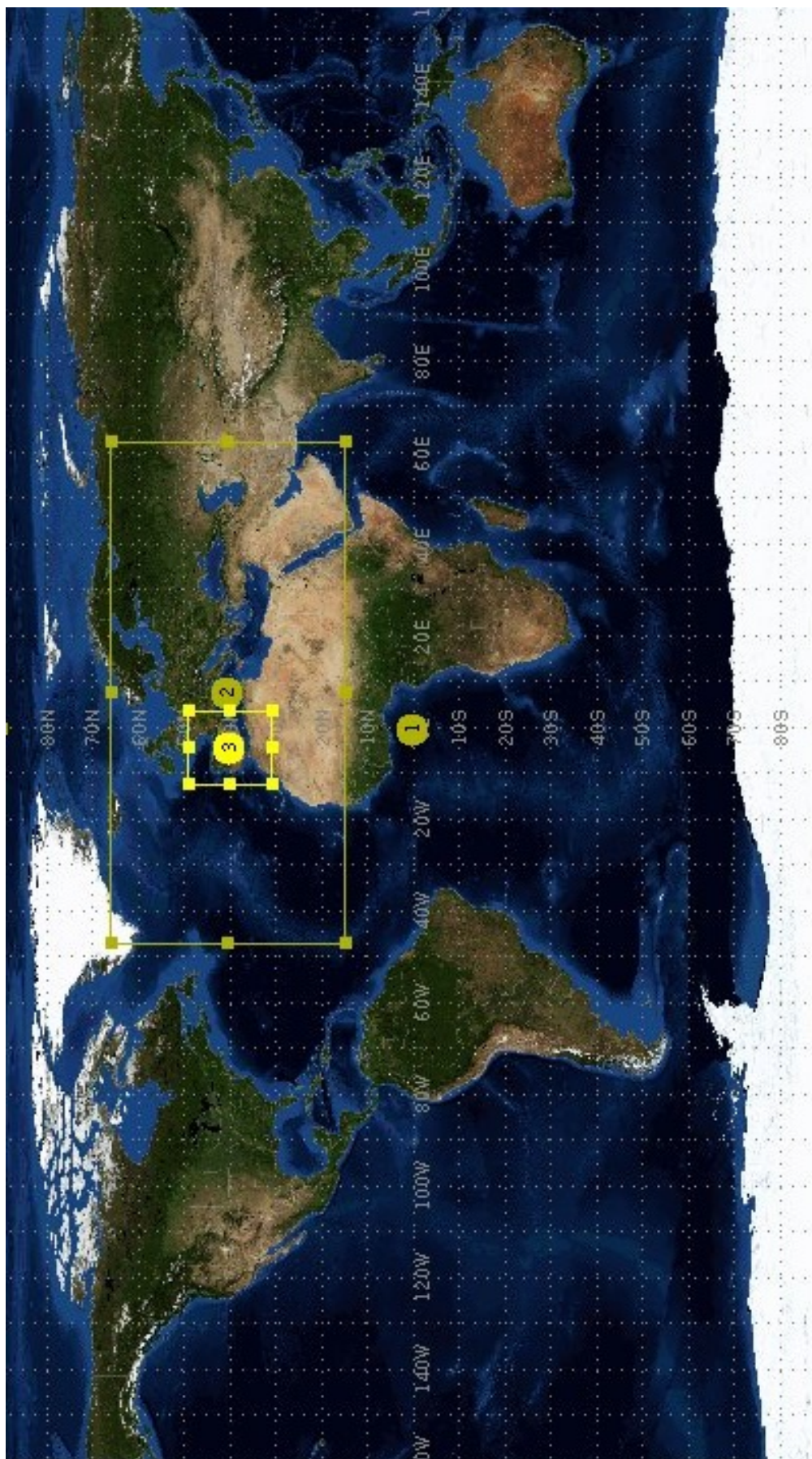


Figura 3.1: Representación gráfica de los dominios $D01$ (Globo entero), $D02$ (Centrado en Europa) $D03$ (Península Ibérica)

Por ejemplo:

- la implementación de este proceso con 51 columnas (dx) resulta en un desplazamiento de $51 \times 294,3$ km, que corresponde a la resolución correspondiente al Dominio 1;
- esto implica que la transmisión de 40 filas (dy) representa un desplazamiento de $40 \times 294,3$ km.

Esta disposición sitúa el Dominio 2 en una localización específica dentro del Dominio 1. El mismo principio se aplica al Dominio 3 pero, en este caso, con las resoluciones y posiciones relativas al Dominio 2. Por ejemplo, en el Dominio 2, si la resolución del Dominio 2 es de 98,1 km, entonces las 40 columnas y 19 filas corresponden a desplazamientos de $40 \times 98,1$ km y $19 \times 98,1$ km respectivamente.

Relación entre dominios y coordenadas absolutas: cada dominio inicia su sistema de cuadrícula ubicado en la posición (1, 1), situada en la esquina inferior izquierda del mismo. El sistema de referencia del Dominio 1 se emplea para determinar las coordenadas absolutas (latitud y longitud) de los puntos de la cuadrícula en los dominios anidados, garantizando así la consistencia geográfica entre una escala y la otra.

Esta metodología permite posicionar adecuadamente los dominios anidados y las áreas de interés pueden ser examinadas meticulosamente manteniendo la conexión con los patrones globales del dominio en su totalidad.

Proyección cartográfica: el modelo se configura mediante una proyección geográfica denominada proyección cartográfica de tipo Latitud-Longitud (Lat-Lon). La Tierra se representa como una estructura cuadrangular compuesta por líneas de latitud y longitud. Las líneas de latitud, que se encuentran paralelas entre sí, representan círculos de igual distancia angular desde el ecuador hasta los polos; las líneas de longitud, que convergen en los polos, constituyen divisiones equivalentes de la Tierra. La superficie terrestre se caracteriza por cada punto ubicado en la superficie, que se establece mediante un par de coordenadas angulares, latitud (ϕ) y longitud (λ), expresadas en grados angulares. La proyección simple Lat-Lon es ampliamente empleada en simulaciones globales debido a su sencillez.

En el modelo WRF, la proyección Lat-Lon se caracteriza por los parámetros del polo de latitud establecido en 90° (eje de rotación terrestre), del polo de longitud establecido en 0 (Meridiano de Greenwich) y del meridiano central establecido en 180° para determinar la simetría del dominio global en la proyección. Mediante estas

modificaciones, es posible preservar la coherencia geográfica en las simulaciones a gran escala.

La proyección Lat-Lon no está exenta de restricciones. A pesar de ser adecuada para investigaciones a escala global, la distorsión resultante de las áreas distantes del ecuador es considerable, particularmente en las regiones polares, donde las distancias y áreas representadas pueden diverger significativamente de las distancias y áreas reales. Esta propiedad también la hace menos apropiada para investigaciones locales o regionales que requieren mediciones espaciales precisas de puntos distantes del ecuador.

Además de la proyección Lat-Lon, otras proyecciones cartográficas se emplean con frecuencia en la modelización atmosférica, dependiendo del dominio de interés y la característica geográfica que se pretende representar. La proyección cónica conforme a la metodología de Lambert (LCC) representa una de las opciones más comunes. En este caso, se proyecta la superficie terrestre sobre un cono que sea tangente o secante a la esfera terrestre, lo que permite preservar las formas locales (conformidad). En esta proyección, las líneas de latitud se definen como arcos de círculos concéntricos, mientras que las líneas de longitud se definen como líneas rectas que convergen en un punto fuera del mapa. La LCC resulta particularmente atractiva para dominios de tamaño medio con una orientación este-oeste, tales como continentes o naciones elongadas horizontalmente, con la necesidad de un equilibrio óptimo entre precisión y facilidad de implementación. Sin embargo, lejos de los paralelos estándar, definidos por el cono de proyección, las distorsiones en áreas y distancias aumentan.

Otra proyección significativa es la Mercator, ampliamente reconocida por su aplicación en mapas de navegación. Esta representación gráfica proyecta la superficie terrestre sobre un cilindro tangente al ecuador, y las líneas de latitud y longitud se ubican en ángulos rectos. Esta proyección conserva los ángulos y direcciones locales, pero genera una distorsión significativa en las áreas y distancias, exagerando considerablemente el tamaño relativo de las regiones de alta latitud conforme se aproximan a los polos. En consecuencia, su aplicación se circunscribe a estudios ecuatoriales o representaciones donde las distorsiones en las latitudes elevadas no tienen relevancia.

Otra alternativa empleada en la modelización atmosférica, particularmente en investigaciones de regiones polares, es la proyección estereográfica polar. Esta proyección representa la superficie terrestre desde un punto diametralmente opuesto al polo hasta un plano tangente al polo contrario. Se trata de una proyección conforme, dado que conserva las formas locales, y resulta particularmente provechosa para investigaciones detalladas en el Ártico o la Antártida. No obstante, experimenta una

distorsión significativa lejos del polo proyectado, lo cual limita su aplicación más allá de las regiones polares.

En resumen, la aplicación de la proyección Lat-Lon se encuentra justificada, dado que se trata de una simulación de carácter global. Este tipo de proyección es el más apropiado para investigaciones de gran envergadura, dado que ofrece una representación directa y sencilla de las coordenadas geográficas sin la necesidad de transformaciones complejas. Adicionalmente, su capacidad para interactuar con la mayoría de los conjuntos de datos globales lo establece como una opción óptima en términos de computación, dado que no conllevará la sobrecarga computacional que tendrían proyecciones más precisas.

Configuración del período de simulación

La configuración del período de simulación seleccionado para este estudio comprende un total de un año, desde las 00:00 UTC del 1 de enero de 2021 hasta las 00:00 UTC del 1 de enero de 2022. El intervalo temporal fue seleccionado con el objetivo de examinar tanto los patrones atmosféricos a gran escala como los detalles regionales y locales de sucesos meteorológicos particulares que tuvieron lugar durante este período.

Se estableció un intervalo temporal para la asimilación de datos y las condiciones de frontera de 21.600 segundos (6 horas), mientras que los intervalos de salida varían entre distintos dominios con el objetivo de optimizar la resolución temporal:

- **Dominio 1:** Cada 120 minutos.
- **Dominio 2:** Cada 60 minutos.
- **Dominio 3:** Cada 15 minutos.

Esta configuración asegura la representación detallada y a escala global de los procesos atmosféricos.

3.2. Configuración física

A continuación se exponen los esquemas físicos seleccionados para implementar la parametrización de procesos subgrid que no pueden ser resueltos de manera explícita por el modelo debido a las restricciones de resolución espacial. Se procederá a detallar los esquemas predominantes seleccionados para este estudio, junto con las justificaciones técnicas pertinentes.

Esquema de microfísica de nubes: se ha empleado el *WRF Single-Moment 3-class* (`mp_physics = 3`) para representar los procesos que participan en la formación y evolución de las nubes. Este representa un esquema de microfísica de un solo momento, en el que se calcula exclusivamente la masa de los hidrometeoros, sin considerar su distribución numérica. Incorpora tres categorías de hidrometeoros: agua en estado líquido, nieve y lluvia.

El esquema *WRF Single-Moment 3-class* es eficaz dado que encapsula los procesos esenciales de condensación, evaporación, congelación y deposición, eliminando el elevado grado de detalle exigido por esquemas de mayor complejidad como los multimomento.

Adicionalmente se caracteriza por su sencillez y eficiencia computacional, factores que resultan particularmente apropiados para simulaciones en un amplio espacio geográfico o durante un periodo extenso. La restricción de este modelo radica en que no representa de manera explícita los procesos de granizo ni la distribución espectral de las partículas, aspectos que podrían resultar cruciales en simulaciones de alta resolución o en eventos convectivos severos.

Radiación de onda larga: el modelo *RRTM (Rapid Radiative Transfer Model)* (`ra_lw_physics = 1`) ha sido seleccionado para simular la transmisión de radiación de onda larga a través de la atmósfera. Este modelo estima la absorción y la emisión de radiación infrarroja por los gases atmosféricos tales como el dióxido de carbono, el metano o el vapor de agua, además de las nubes.

Debido a su robustez y exactitud, RRTM se emplea extensivamente en simulaciones a nivel regional y global. Se aplica un modelo de bandas correlacionadas que disminuye de manera significativa los tiempos de cálculo mediante la división del espectro radiativo en bandas discretas. Específicamente, este enfoque proporciona una representación adecuada del efecto invernadero y otros procesos radiativos cruciales en la dinámica atmosférica.

Este esquema resulta esencial en simulaciones tanto a nivel regional como a escala global, en las que los flujos radiales de onda larga constituyen un componente esencial en la distribución de energía atmosférica. Aunque es muy preciso, puede volverse computacionalmente costoso en configuraciones con una alta densidad de capas verticales.

Radiación de onda corta: para la parametrización que simula la transferencia de radiación de onda corta desde la atmósfera hasta la superficie terrestre se ha seleccionado el esquema *Dudhia* (`ra_sw_physics = 1`). Este esquema, caracterizado

por su simplicidad y eficiencia, es habitualmente empleado para la realización de simulaciones a mediana y gran escala.

El modelo Dudhia incorpora los efectos de la reflexión y absorción de la radiación provenientes de la superficie terrestre y las nubes. Sin embargo, no incorpora un análisis exhaustivo de la dispersión atmosférica, lo cual podría comprometer su exactitud en simulaciones de alta resolución, donde su relevancia es más significativa. Su elevada eficiencia lo establece como una alternativa habitual para investigaciones climáticas y meteorológicas, donde la observación de la dinámica global de la radiación prioritaria.

Esquema de capa límite: el esquema *YSU (Yonsei University)* (`bl_pbl_physics` = 1) ha sido el seleccionado para la modelización de la capa límite planetaria, esto es, la zona en la que la atmósfera y la superficie terrestre interactúan. El esquema representa la mezcla turbulenta en dicha capa límite planetaria y tiene como objetivo considerar los procesos de transmisión de calor, humedad y momento dentro de dicha capa.

El esquema YSU emplea un enfoque de primer orden para la parametrización turbulenta, en el que los flujos verticales se determinan a partir de los gradientes de las variables meteorológicas. Una particularidad de este esquema radica en su desarrollo basado en un perfil de mezcla inversa, lo que optimiza la representación de la estratificación estable y los flujos convectivos dentro de la capa límite planetaria.

Este esquema ha sido concebido para la realización de simulaciones a nivel regional y global en las que se busca un equilibrio entre exactitud y eficiencia. No obstante, puede no ser adecuado para simulaciones de alta resolución en las que la capa límite presenta mayor complejidad, como en el caso de flujos turbulentos a muy pequeña escala.

Esquema de superficie terrestre: para modelar los intercambios de energía y humedad entre la superficie terrestre y la atmósfera, se ha utilizado el esquema *Noah Land Surface Model* (`sf_surface_physics` = 1). Este modelo calcula flujos de calor sensible, calor latente y energía radiante, además de considerar variables como la humedad del suelo, la evapotranspiración y la cobertura vegetal.

El esquema Noah es ampliamente utilizado debido a su capacidad para representar procesos terrestres complejos de manera eficiente. Incluye varias capas de suelo y parametriza la interacción entre la radiación solar y la superficie terrestre, lo que le hace ideal para simulaciones que involucran procesos de superficie significativos, como el ciclo hidrológico o los cambios en la cobertura vegetal.

Convección: la representación de la convección profunda se ha realizado mediante el esquema *Kain-Fritsch* (`cu_physics = 1`). Este esquema parametriza el proceso de convección profunda considerando la formación, crecimiento y disipación de las nubes de origen convectivo.

El esquema Kain-Fritsch se activa en dominios con resoluciones moderadas (mayores a 10 km), donde los procesos convectivos no pueden resolverse explícitamente debido a las limitaciones espaciales del modelo. Este enfoque asegura que los efectos de la convección, como el transporte de calor y humedad, se representen adecuadamente en el modelo. Sin embargo, en dominios de alta resolución, como el Dominio 3, la convección se resuelve directamente, por lo que este esquema no es utilizado en ese escenario.

3.3. Configuración numérica

La disposición numérica del modelo WRF constituye un elemento crucial, dado que dicta la resolución de las ecuaciones dinámicas y termodinámicas que caracterizan el comportamiento de la atmósfera. A continuación se detallan las configuraciones numéricas empleadas en este trabajo junto con sus correspondientes justificaciones.

Tamaño del paso temporal: Considerando la elección habitual en casos de estudio similares, se seleccionó un intervalo temporal de 90 segundos para mantener una estabilidad numérica del modelo. La elección de este valor se fundamenta en la resolución espacial del dominio de mayor resolución (Dominio 3), donde el espaciado horizontal es de aproximadamente 19,6 km. Esta elección tiene su explicación en que el paso temporal de los modelos numéricos de predicción, particularmente el WRF, está regida por el criterio de estabilidad de *Courant-Friedrichs-Lewy* (CFL), que establece que la distancia recorrida por una onda durante un paso de tiempo no debe exceder el espaciado entre puntos de la malla. En otras palabras, $\Delta t \leq \Delta x / c$, donde Δt es el paso temporal, Δx es la resolución espacial, y c es la velocidad característica de las ondas en el modelo.

Las velocidades intrínsecas de las ondas y los gradientes espaciales presentan una mayor intensidad en dominios de alta resolución, como el Dominio 3, por lo que resulta crucial mantener un intervalo temporal reducido para prevenir inestabilidades numéricas. A pesar de que el costo computacional incrementa con una duración de tiempo reducida, se asegura una adecuada estabilidad de la simulación.

Dominio no hidrostático: se ha empleado la configuración no hidrostática del modelo `Non_hydrostatic = .true.` al objeto de facilitar la resolución de fenómenos atmosféricos que conllevan movimientos verticales relevantes. En un modelo hidrostático, para fenómenos de gran magnitud donde los movimientos verticales son reducidos en comparación con los horizontales, se presupone que las fuerzas de presión y gravedad mantienen un equilibrio. No obstante, esta metodología resulta insuficiente para simulaciones de alta resolución en las que la convección profunda, las ondas de montaña y los sistemas orográficos producen movimientos verticales intensos.

El modelo adopta una perspectiva no hidrostática y resuelve de manera explícita todas las ecuaciones del movimiento, incorporando los términos de aceleración vertical. La exactitud de los fenómenos a pequeña escala se logra a expensas de incrementar el coste en términos de computación. En este trabajo se ha utilizado el modo no hidrostático para representar los procesos locales presentes en el Dominio 3.

Opciones de diferencias finitas: Para la realización de las simulaciones no se implementó difusión horizontal o vertical explícitamente (`diff_opt = 0`), sino que se recurrió a los esquemas implícitos de mezcla turbulenta y advección del modelo para representar estos procesos. La difusión explícita incorpora términos adicionales a las ecuaciones del modelo con el objetivo de suavizar los gradientes y evitar inestabilidades numéricas. Sin embargo, también puede disminuir las propiedades reales de la dinámica atmosférica.

La simulación se optimiza al obviar la difusión explícita, dado que conserva los pormenores de los fenómenos atmosféricos, tales como las corrientes en chorro, las líneas de inestabilidad y las ondas de montaña que podrían ser atenuadas mediante una difusión excesiva. No obstante, esta configuración presenta problemas particulares de estabilidad que requieren un control meticuloso con el transcurso del tiempo y las condiciones de contorno.

Amortiguación vertical: se implementó un esquema de amortiguación vertical (`zdamp = 5000`, coeficiente = 0.2) con el objetivo de prevenir fluctuaciones numéricas en las capas superiores del aire sin alterar de manera significativa los procesos dinámicos de interés en las capas inferiores. Este esquema atenúa las ondas acústicas y gravitacionales amplificadas a elevadas altitudes, atribuibles a la reducción de la densidad del aire y al incremento de los gradientes verticales.

Otras consideraciones numéricas: la configuración numérica incluye también esquemas avanzados para la resolución de las ecuaciones de Navier-Stokes en la atmósfera:

- **Esquemas de advección:** se emplean esquemas de advección de alto nivel que tienen la capacidad de capturar intensos gradientes sin introducir oscilaciones numéricas de relevancia.
- **Metodologías de filtrado:** el modelo emplea filtros implícitos para regular las oscilaciones no físicas, especialmente en las bandas superiores del espectro energético.
- **Resolución vertical:** la disposición comprende 41 niveles verticales entre la superficie y la estratosfera.

Impacto de la configuración numérica: La configuración numérica adoptada en este trabajo permite encontrar un equilibrio entre precisión, estabilidad y eficiencia computacional. Las opciones seleccionadas facilitan la definición de los fenómenos en cada dominio del modelo, con el objetivo de abarcar los procesos tanto a gran escala como a nivel de microescala. En concreto, la configuración no hidrostática, la variación temporal y la falta de difusión explícita aseguran una simulación adecuada de los procesos atmosféricos locales.

3.4. Preprocesamiento Geográfico (WPS)

Se ha elegido el sistema de preprocesamiento WPS para la generación de los datos de entrada requeridos durante el proceso de simulación. Esto es, en lo referente a la interpolación de información meteorológica inicial y de frontera, así como a la generación de campos geográficos.

Datos Geográficos: la información geográfica comprende cuantificaciones más exactas del albedo debido a partículas, el uso del suelo, la altitud y los factores del suelo. Estos datos se derivaron de fuentes bien conocidas, entre las que se incluyen:

- **Albedo y Leaf Area Index (LAI):** Datos obtenidos del sensor MODIS, con resoluciones de 10 m y 30 s.
- **Topografía:** GMTED2010, con una resolución de 30 segundos de arco.

- **Uso del suelo y cobertura vegetal:** datos del sensor MODIS, clasificado en categorías estándar que permiten modelar las interacciones entre la superficie y la atmósfera.
- **Orografía (Gravity Wave Drag - GWD):** datos disponibles en múltiples resoluciones, desde 1° hasta 30s.

Presión vertical: Se seleccionaron intervalos de presión que oscilan entre 1000 hPa (101325 Pa), correspondiente a la proximidad del nivel del mar, y 50 hPa (5000 Pa), correspondiente al nivel de la estratosfera. Estos niveles se dispersaron a lo largo de las cuarenta y una capas verticales seleccionadas, facilitando de este modo una simulación precisa de los procesos atmosféricos desde la capa superficial inferior hasta la estratosfera.

3.5. Datos de entrada

Datos meteorológicos

Los datos meteorológicos iniciales y de frontera se obtuvieron a partir del conjunto de datos *NCEP GFS 0,25 Degree Global Forecast Grids Historical Archive* [86], disponible en el *Research Data Archive (RDA)* del *National Center for Atmospheric Research (NCAR)*. Las características más relevantes de este conjunto de datos son las siguientes:

- **Fuente de los datos:** Sistema Global de Pronóstico (GFS) del NCEP.
- **Cobertura espacial:** global, con latitudes de -90° a 90° y longitudes de -180° a 180° , en una cuadrícula de resolución de $0,25^\circ \times 0,25^\circ$ (1440 $\times$ 721 puntos).
- **Cobertura temporal:** desde el 15 de enero de 2015 hasta el día de hoy. En este estudio, se utilizaron datos de dos períodos diferentes. El primero correspondiendo al mes de julio de 2015. El segundo al año 2022 en su totalidad.
- **Frecuencia temporal:** los datos están disponibles con una frecuencia de 6 horas.
- **Variables incluidas:** :
 - Altura geopotencial.
 - Temperatura del aire.

- Viento (componentes U y V) en varios niveles de presión.
 - Humedad relativa y específica.
 - Presión atmosférica a nivel del mar.
 - Temperatura de la superficie del mar (SST).
 - Radiación de onda corta y larga.
 - Precipitación acumulada.
- **Formato de los datos:** GRIB2, formato estandarizado para el manejo de grandes volúmenes de datos meteorológicos.

Datos geográficos

Además de los datos meteorológicos, se incorporaron datos geográficos detallados para definir la topografía, el uso del suelo y otras características de la superficie terrestre. Los principales conjuntos de datos utilizados han sido:

- **Topografía:** GMTED2010, con una resolución de 30 segundos de arco.
- **Uso del suelo y cobertura vegetal:** datos del sensor MODIS, clasificados en categorías estándar.
- **Leaf Area Index (LAI) y albedo:** proporcionados por el sensor MODIS, con resoluciones de 10m y 30s respectivamente.
- **Orografía:** datos de Gravity Wave Drag (GWD), disponibles en múltiples resoluciones (1°, 2°, 10 m, 20 m, y 30 s).

Estos datos fueron procesados mediante el módulo **geogrid** del sistema WPS, que generó los archivos `geo_em.d01.nc`, `geo_em.d02.nc`, y `geo_em.d03.nc` para cada dominio.

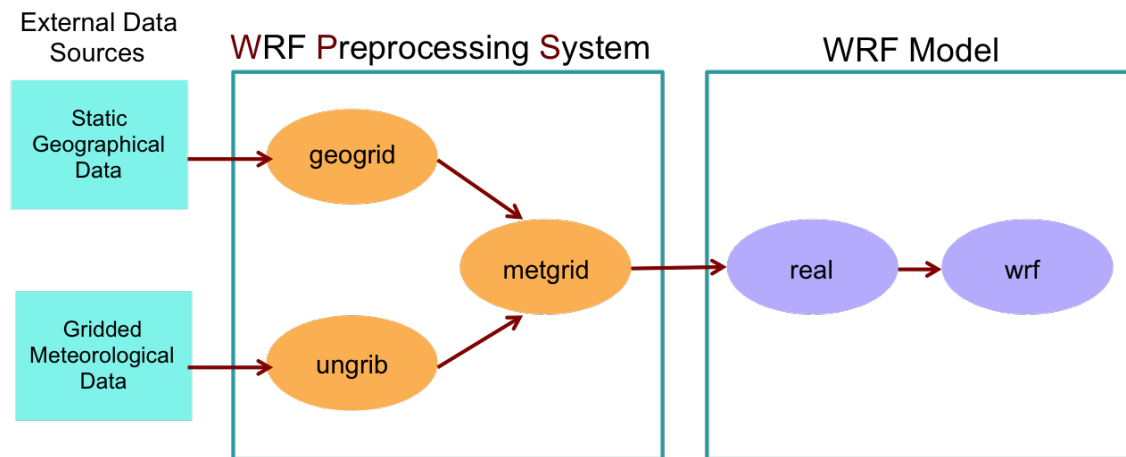


Figura 3.2: Esquema de funcionamiento de los diferentes procesos del WRF

3.6. Postprocesamiento

La evaluación de los resultados de la simulación se llevó a cabo mediante una combinación de diversas herramientas, las cuales abarcan la interpretación y el análisis de datos y métodos estadísticos, además de la representación visual de los escenarios meteorológicos simulados. La selección de estas herramientas se fundamentó en su aptitud para el análisis de datos científicos en formatos como NetCDF y en la representación gráfica gráfica.

Python

Prácticamente la totalidad del post-procesamiento se llevó a cabo en Python, lenguaje de programación que ha adquirido popularidad en la comunidad científica debido a su versatilidad y su abundante colección de paquetes especializados. Las principales tareas que se han ejecutado han sido las siguientes:

Representación Gráfica: para la creación de mapas, gráficos y figuras, se utilizaron librerías como:

- **Matplotlib:** para generar gráficos básicos, como series temporales o histogramas.
- **Cartopy:** esencial para crear mapas detallados que representen variables meteorológicas, como la distribución de la temperatura, la velocidad del viento o las precipitaciones.

- **Seaborn:** utilizada para visualizaciones más complejas, como gráficos de densidad y correlación.

Análisis estadístico: se realizaron cálculos estadísticos avanzados con el soporte de las siguientes librerías:

- **NumPy:** para realizar operaciones numéricas rápidas y eficientes.
- **Pandas:** utilizada para manipular y analizar conjuntos de datos tabulares.
- **SciPy:** aplicada en el cálculo de métricas como la correlación, la desviación estándar y otras estadísticas descriptivas.

Aplicación de Machine Learning: Con el fin de identificar patrones y realizar predicciones basadas en los resultados del modelo, se emplearon técnicas de aprendizaje automático utilizando la librería Scikit-learn. Esto incluyó:

- Clasificación y regresión para analizar las relaciones entre variables.
- Clustering para identificar agrupaciones naturales en los datos simulados.
- Evaluación de métricas de rendimiento para validar los modelos ajustados.

Panoply

El software Panoply, desarrollado por *NASA/GISS*, se empleó como un instrumento auxiliar en el análisis visual acelerado de los datos generados por el modelo. Este software proporciona una herramienta gráfica de fácil manejo para la visualización de archivos en formato *NetCDF*, facilitando así la validación de las salidas del modelo.

Principales funcionalidades utilizadas:

- Visualización de campos meteorológicos: se analizaron variables como temperatura, viento y presión en diferentes niveles atmosféricos.
- Inspección rápida: Panoply permitió verificar la coherencia espacial y temporal de los datos simulados, asegurando que no hubiera anomalías evidentes en las salidas.

Especificaciones del sistema: el software fue ejecutado en un sistema operativo *Windows 11*, utilizando *Java 11.0.20*, con una asignación de memoria de 10880 MiB. Estas configuraciones permitieron manejar eficientemente los archivos de gran tamaño generados por las simulaciones.

Sinergia entre herramientas

Python, junto con Panoply, permitió realizar el análisis y la representación visual de datos. Python también se empleó en el análisis cuantitativo, así como en la creación de representaciones gráficas específicas, mientras que Panoply facilitó la evaluación de la calidad general de los datos y la realización de verificaciones visuales iniciales. Esto contribuyó a garantizar la precisión y la capacidad de análisis gráfico de los datos obtenidos.

Capítulo 4

Validación de la configuración del modelo numérico WRF

En el campo de la meteorología, los modelos numéricos computacionales constituyen el fundamento para la simulación y predicción de los fenómenos atmosféricos. Como ya se ha expuesto anteriormente, estos modelos se construyen sobre la base de una serie de ecuaciones cuya resolución, en diferentes escalas espaciales y temporales, permite simular el comportamiento y la evolución de la atmósfera. No obstante, para conocer el grado de destreza que tiene un modelo concreto para generar escenarios realistas es imprescindible evaluar la adecuación del modelo a la realidad observada, esto es, validarlo.

¿Qué es la validación de un modelo? La validación del modelo implica establecer con qué grado de exactitud los resultados que produce un modelo numérico emulan los datos meteorológicos observados. El procedimiento consiste, básicamente, en comparar salidas cualitativas generadas por el modelo, tales como la temperatura, las presiones atmosféricas, las velocidades del viento o la precipitación estimada con datos observacionales reales provenientes de fuentes confiables, tales como estaciones meteorológicas o boyas oceánicas.

La validación contribuye no solo a corroborar el grado de exactitud del modelo numérico, sino también a conocer sus limitaciones. Este proceso permite establecer los entornos o las aplicaciones en los que el modelo numérico puede ser de utilidad, puede aportar un valor añadido, especialmente en contextos en los que existen tomas de decisiones basadas en las salidas de los modelos numéricos, tales como la agricultura, la aviación, la gestión de desastres o la adaptación al cambio climático. Así, es posible determinar si las proyecciones que genera el modelo numérico son adecuadas para representar fenómenos meteorológicos dentro de un margen de error

tolerable, un aspecto esencial para delimitar conveniente sus prestaciones en cada caso concreto. De forma análoga, la validación posibilita la identificación de tendencias en los errores o sesgos que pueden comprometer la calidad de un modelo y suministra, al mismo tiempo, información crucial para la modificación de parámetros, la mejora de procesos físicos o el incremento de la resolución. Además, facilita la comparación entre un modelo y otro con el objetivo de discernir cuál se adapta mejor a una área geográfica o fenómeno atmosférico específico. Finalmente, la validación permite presentar los resultados de forma clara y estructurada, lo que facilita que otros científicos puedan replicarlos en sus propios estudios y/o utilizarlos para comparaciones.

En qué consiste la validación de un modelo: Habitualmente se suele seleccionar una serie de variables meteorológicas que conformarán el conjunto de aquellas que serán intercomparadas entre las obtenidas por el modelo y las registradas mediante elementos de observación solventes y con la debida trazabilidad en sus errores de medida. Estas variables, tales como la temperatura, la precipitación, la humedad relativa, la velocidad del viento o la presión atmosférica pueden ser estimadas en diversas escalas espaciales (yendo desde lo más local hacia ámbitos más globales) y temporales (horario, diario, semanal, mensual, etc.).

La validación de un modelo numérico implica la utilización de métricas estandarizadas y comúnmente aceptadas por la comunidad científica para establecer su skill o grado de pericia. Entre las métricas más habituales se encuentra el error absoluto medio, que refleja la suma total de las diferencias absolutas entre los valores pronosticados y observados, el error cuadrático medio, que toma en cuenta más los errores grandes que los pequeños; el sesgo, que señala si el modelo tiende a sobrestimar o subestimar, y el coeficiente de correlación, que muestra la correlación entre los valores pronosticados y observados. Adicionalmente se pueden emplear elementos de estudios gráficos tales como diagramas de dispersión, mapas de diferencias o series temporales con el objetivo de facilitar la identificación de errores o patrones coincidentes.

Como se ha indicado anteriormente, la validación de un modelo conlleva el uso de un conjunto de datos observados de alta fiabilidad. No obstante, con carácter previo a ser utilizados en el proceso de validación, estos datos deben ser sometidos a controles de calidad para asegurar que se trabaja con series de datos homogéneas (aquella que es representativa del clima de una determinada región en los alrededores del punto de observación). En ocasiones, cuando conjunto de datos de observación de que se dispone es insuficiente o tiene lagunas (falta de datos), se pueden aplicar

métodos de interpolación para el relleno de las lagunas existentes, tales como el inverso de la distancia o la regresión múltiple.

En este trabajo, la validación de las salidas generadas por el modelo numérico frente a observaciones reales se ha realizado utilizando las métricas previamente mencionadas, que serán definidas en detalle posteriormente, así como mediante instrumentos gráficos. Esta aproximación facilita la comprensión de los resultados de validación y permite establecer con detalle sus fortalezas y debilidades.

4.1. Datos de referencia

Para llevar a cabo la validación de la configuración del modelo numérico WRF se han adoptado una metodología consistente en comparar las salidas del WRF con medidas obtenidas por una estación de observación en altura (sondeo aerológico). Concretamente se han empleado las medidas de la estación ubicada en las proximidades -del aeropuerto de Madrid-Barajas cuyo código de identificación de la Organización Meteorológica Mundial (OMM) es el 08221. Los sondeos aerológicos son mediciones que se realizan en la atmósfera con el fin de conocer su estado en diferentes niveles de altitud. Generalmente, se utilizan radiosondas que se elevan con globos meteorológicos y registran medidas de la temperatura, la humedad, la presión y, especialmente interesante para nuestro estudio, el viento, resultan fundamentales a la hora de evaluar la capacidad predictiva de los modelos numéricos. El radiosondeo ofrece información precisa en la vertical de la atmósfera, lo que permite una comparación directa con las salidas del modelo en distintos niveles de presión atmosférica.

La elección de la estación de observación en altura de Madrid se fundamenta en su alta disponibilidad, y fiabilidad de los datos, ya que es gestionada por la Agencia Estatal de Meteorología que forma parte de la red mundial de observación de la OMM. Es importante destacar que la estación 08221 se encuentra ubicada en una región con una orografía de cierta complejidad, significativamente influenciada por la proximidad de la Sierra de Guadarrama (Sistema Central). Esta cadena montañosa actúa como una barrera natural a las masas de aire, afectando de manera significativa la dinámica atmosférica tanto a nivel local como regional. En particular, la interacción entre la topografía y las condiciones sinópticas puede dar lugar a fenómenos como flujos canalizados, inversión térmica, cizalladura vertical del viento o formación de ondas de montaña, todos ellos de gran interés en el estudio y predicción del viento.

Estas características hacen de los radiosondeos de la estación 08221 una herra-

mienta especialmente útil para la validación de las salidas del modelo WRF, ya que permiten analizar el comportamiento de las predicciones en un entorno con condiciones atmosféricas diversas y complejas. En este contexto, la evaluación se centra en comparar los perfiles verticales previstos por el modelo numérico con los perfiles observados mediante radiosondeo, en una serie de niveles de presión previamente seleccionados. Esta metodología ha sido aplicada en otros estudios similares ([87], [88], [89]) en los que los radiosondeos son considerados como una fuente fiable para ser utilizados como datos de referencia para procesos de verificación. Por tanto, la validación llevada a cabo no solo permite comprobar el grado de ajuste del modelo numérico en condiciones reales, sino también identificar posibles sesgos o limitaciones estructurales del sistema predictivo. Al comparar directamente las predicciones del modelo con observaciones reales obtenidas a través de radiosondeo, se obtiene una medida objetiva y detallada del rendimiento del sistema, tanto en términos generales como en lo que respecta específicamente a la variable del viento, que constituye el eje central de este estudio.

Dado que la estación proporciona valores cada 12 horas, concretamente a las 00:00 horas y a las 12:00 horas UTC, se han evaluado los resultados ofrecidos por el modelo numérico WRF en distintos niveles de presión atmosférica en estas dos horas, pudiendo así evaluar su comportamiento en cada una de las horas de medida. Además, se ha realizado un análisis global de todo el periodo de validación, que va desde el 9 de julio de 2015 a las 00:00 horas UTC hasta el 31 de julio de 2015 a las 12:00 horas UTC. Este enfoque permite una evaluación detallada del desempeño del modelo WRF en distintos momentos y condiciones atmosféricas, así como para tener una visión global del mismo.

Respecto a la cobertura temporal, aunque el periodo de validación es relativamente corto, la intención de esta validación no es demostrar la funcionalidad general del modelo WRF. En esta validación se busca comprobar que la configuración específica utilizada para este estudio es fiable y que el modelo está funcionando correctamente. El WRF ya ha sido ampliamente validado en multitud de escenarios y configuraciones por otros autores [90] [91] [92] [93] [94] [95], por lo que este análisis se centra exclusivamente en verificar que el rendimiento de nuestra configuración es adecuado. Teniendo en cuenta que las salidas del modelo WRF van a ser utilizadas posteriormente para desarrollar una metodología híbrida de predicción en una zona llana, sin ningún tipo de sistema montañoso a su alrededor, parece razonable concluir que, si el proceso de validación en una zona de orografía más compleja, como Madrid, es exitoso, la configuración que se ha seleccionado e implementado para poner operativo el modelo numérico WRF es adecuada y, por consiguiente, será capaz de generar

salidas confiables en una zona llana, donde el flujo de aire no es condicionado por accidentes geográficos.

4.2. Resultados de la validación

Para evaluar el rendimiento y la confianza en los resultados aportados por el WRF, se ha llevado a cabo un análisis comparativo entre las predicciones del modelo y los valores observados en la estación de observación en altura de Madrid. A este fin, se han comparado los datos del radiosondeo de la estación de Madrid con los datos del punto más cercano proporcionado por el modelo numérico WRF. El objetivo de esta validación ha sido comprobar que el WRF, dentro de los errores e incertidumbres inherentes al modelo, es capaz de proporcionar valores de velocidad y dirección del viento similares a los registros reales y, en consecuencia, representativos de la realidad observada en el entorno del aeropuerto [96] [90] [91] [97].

Para cuantificar el error en las predicciones, se han utilizado las tres métricas principales ya comentadas antes [98]:

- Error medio absoluto (MAE por sus siglas en inglés): mide el promedio de los errores absolutos entre los valores pronosticados y los observados, proporcionando una medida de la precisión general del modelo numérico.

$$MAE = \frac{1}{N} \sum_{i=1}^N |P_i - O_i| \quad (4.1)$$

- Raíz del error cuadrático medio (RMSE por sus siglas en inglés): es similar al MAE, pero al elevar los errores al cuadrado antes de promediarlos se penalizan más los valores atípicos o especialmente dispares, lo que permite identificar la presencia de errores grandes en la predicción.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2} \quad (4.2)$$

- Sesgo (Bias, en inglés): indica la tendencia sistemática del modelo a sobreestimar o subestimar los valores observados. Un Bias positivo sugiere sobreestimación, mientras que uno negativo indica subestimación.

$$Bias = \frac{1}{N} \sum_{i=1}^N (P_i - O_i) \quad (4.3)$$

Donde, en las ecuaciones anteriores, N es el número total de predicciones realizadas, P_i el valor predicho por el modelo en la observación i , O_i el valor observado (real) en la observación i y $|P_i - O_i|$ el valor absoluto del error para cada observación individual.

- Desviación estándar (σ): es una medida de dispersión que indica cuánto se alejan, en promedio, los datos de un conjunto respecto a su media.

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (4.4)$$

Donde N el número total de elementos del conjunto de datos, x_i el valor de cada observación individual y μ la media poblacional, dada por:

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (4.5)$$

En los siguientes apartados se presentan los resultados obtenidos durante el proceso de validación acompañados de ejemplos ilustrativos. No obstante, todas las gráficas necesarias para el análisis detallado de los resultados se encuentran en el anexo a esta memoria de tesis doctoral, donde se presentan las comparaciones visuales de las predicciones con los datos reales, así como la evolución temporal de los errores a lo largo del periodo de validación.

4.2.1. Valores obtenidos en las métricas

En la Tabla [4.1](#) se presentan los valores medios globales de las métricas de error utilizadas para evaluar el rendimiento del modelo numérico WRF en la predicción de la velocidad y la dirección del viento:

Métrica	Valor medio	Desviación estándar
Velocidad (m/s)		
MAE	4.40	1.30
RMSE	5.34	1.66
Bias	0.17	2.88
Dirección ($^{\circ}$)		
MAE	29.21	11.57
RMSE	43.02	14.05
Bias	13.54	15.29

Tabla 4.1: Valores medios y desviación estándar de las métricas para la velocidad y dirección del viento para el tiempo total de validación.

Estos valores ofrecen una visión general del desempeño del modelo WRF en ambas variables. Para entender mejor el significado de estos errores y sus implicaciones, se tratarán por separado, junto con sus correspondientes gráficas y datos adicionales, cada una de los elementos del vector viento: velocidad y dirección.

Errores en la predicción de la velocidad del viento: con carácter general, los valores obtenidos para las métricas de error de la Tabla 4.1 muestran que el modelo WRF arroja unos resultados que podrían calificarse como consistentes en lo que a predicción de la velocidad del viento respecta ya que el MAE presenta un valor similar al que se ha obtenido en estudios similares de verificación de predicción de viento a corto plazo, lo que es indicativo de que la configuración seleccionada e implementada puede considerarse como adecuada [99].

En las siguientes gráficas se presentan los errores obtenidos en cada uno de los dos momentos en los que se lanzan las sondas aerológicas: las 00:00 y las 12:00 horas UTC. Este análisis permite identificar si hay diferencias cuando se consideran momentos horarios en los que ha presencia (12:00 horas UTC) o ausencia (00:00 horas UTC) de radiación solar.

Si se observa la Figura 4.1, se puede apreciar que el RMSE es algo superior al MAE, aunque son del mismo orden de magnitud. Este resultado es indicativo de la existencia de algunos días en los que los errores son significativamente mayores que la media, lo que eleva esta métrica debido a su sensibilidad a los valores extremos. En términos generales podría decirse que la precisión del modelo es elevada, pero

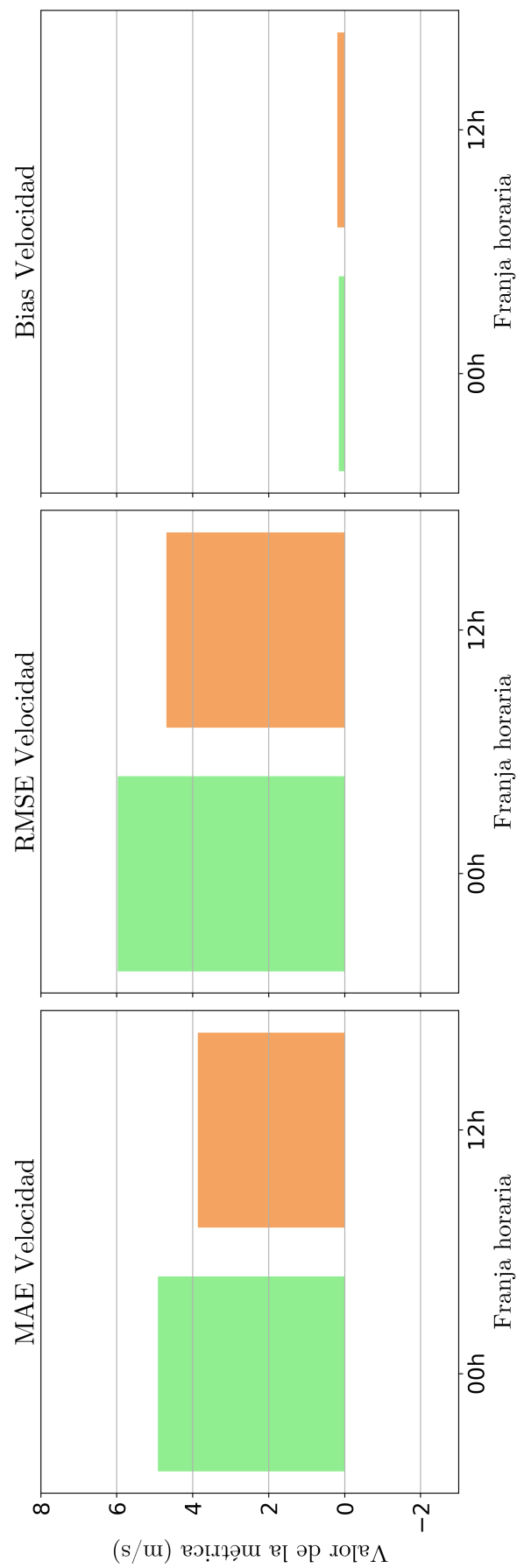


Figura 4.1: Métricas de error para las velocidades del viento generadas por el modelo WRF.

hay ciertos días en los que el modelo se desvía sensiblemente del valor real de esta variable. Si se estudia más detenidamente en este suceso puede advertirse que la principal fuente de este error está relacionada en la forma en la que el modelo numérico WRF representa los cambios de velocidad del viento a diferentes niveles de presión. Mientras que las observaciones del radiosondeo meteorológico muestran, en ciertos días y horas, unas variaciones bruscas de la velocidad, el modelo WRF tiene a generar transiciones más suaves y graduales, dando lugar a la diferencia significativas entre los valores medidos por el sondeo aerológico y el modelo WRF. Sin embargo, aunque el modelo no parece haber sido capaz de representar siempre con exactitud estos cambios bruscos, si que sigue correctamente la tendencia general de la evolución de la velocidad del viento. Por ejemplo, en un hipotético caso en el que la velocidad del viento pasa de 5 a 15 m/s de un nivel de presión al inmediatamente siguiente, y suponiendo que la velocidad de viento se mantenga en 15 m/s durante los siguientes niveles, el WRF no emula ese cambio de una manera tan inmediata, pero sí modela una progresión ascendente en la velocidad del viento. Esta progresión es más suave y escalonada en comparación con los datos observados, pero resulta oportuno indicar que el WRF si que simula correctamente la dirección del cambio del viento. Dicho de otra manera, aunque el modelo no refleja la magnitud exacta del salto en la velocidad del viento de un nivel de presión al inmediatamente superior, si que es capaz de capturar la variación y tendencia de este cambio, aunque de manera más gradual.

Un ejemplo de esta situación se puede observar en las Figuras [4.2](#) y [4.3](#).

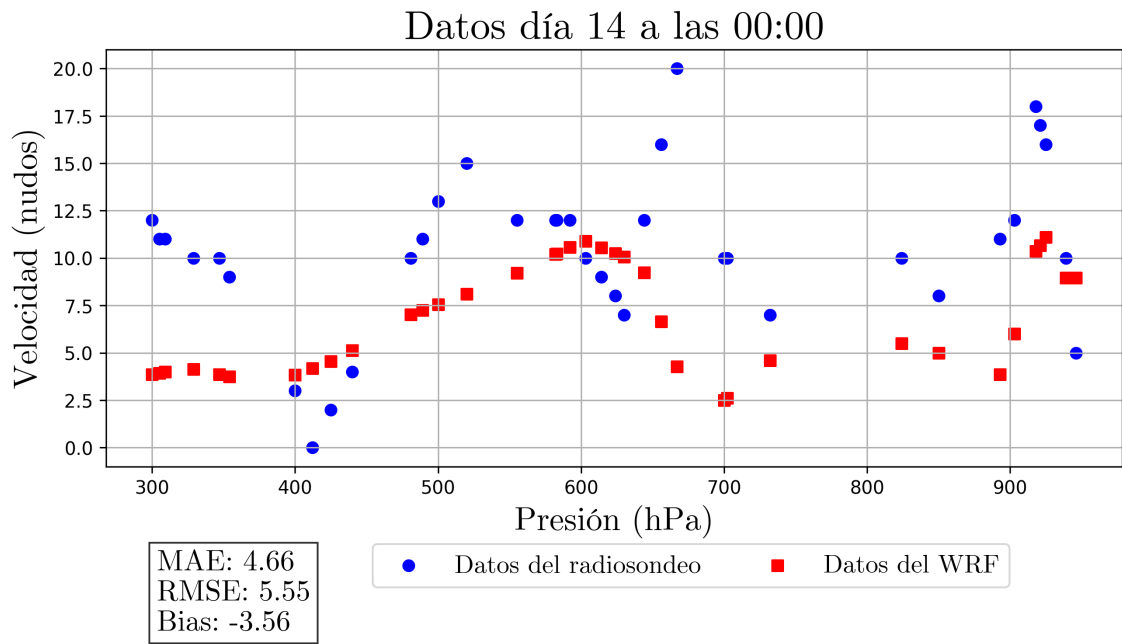


Figura 4.2: Ejemplo del día 14 de julio de 2015 a las 00:00 UTC donde se observa una alta variabilidad en la velocidad del viento en diferentes niveles de presión y cómo el WRF tiende a seguir esa evolución.

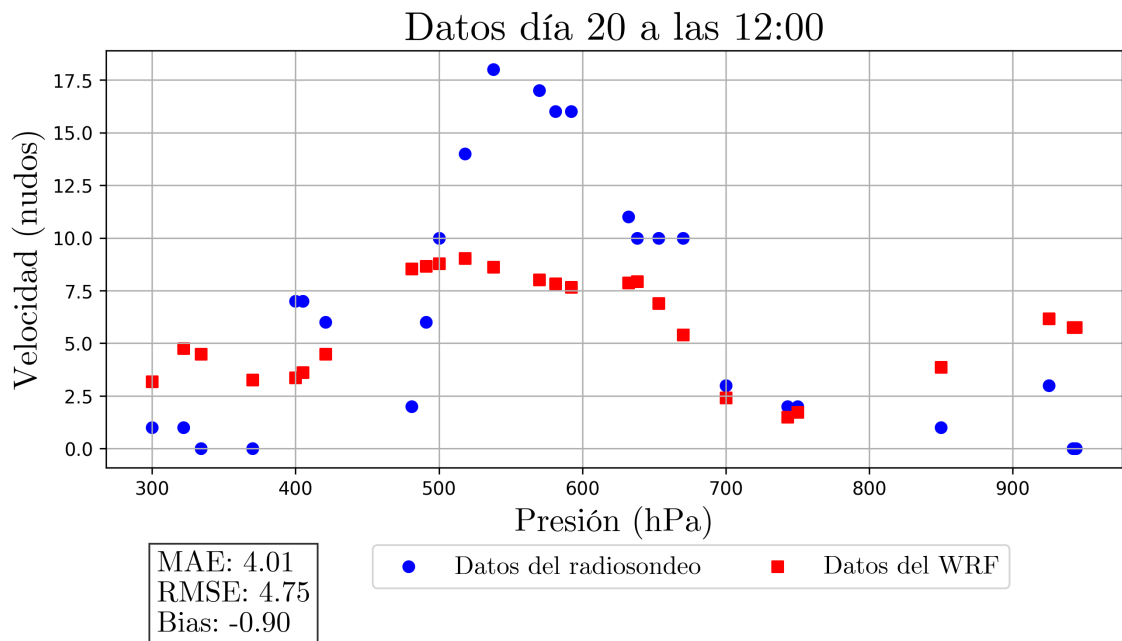


Figura 4.3: Ejemplo del 20 de julio de 2015 a las 12:00 UTC donde se observa una alta variabilidad en la velocidad del viento en diferentes niveles de presión y cómo el WRF tiende a seguir esa evolución.

Dado que el RMSE penaliza los errores más abultados al elevarlos al cuadrado, esta diferencia en la dinámica de representación de la velocidad del viento del modelo

Presión (hPa)	Radiosondeo (m/s)	WRF (m/s)
946.00	5.00	8.95
939.00	10.00	8.95
925.00	16.00	11.10
921.00	17.00	10.67
918.00	18.00	10.35
903.00	12.00	6.00
893.00	11.00	3.86
850.00	8.00	4.98
824.00	10.00	5.51
732.00	7.00	4.60
702.00	10.00	2.62
700.00	10.00	2.49
667.00	20.00	4.28
656.00	16.00	6.65
644.00	12.00	9.24
630.00	7.00	10.06
624.00	8.00	10.24
614.00	9.00	10.55
603.00	10.00	10.90
592.00	12.00	10.57
583.00	12.00	10.23
582.00	12.00	10.20
555.00	12.00	9.22
520.00	15.00	8.10
500.00	13.00	7.54
489.00	11.00	7.24
481.00	10.00	7.02
440.00	4.00	5.12
425.00	2.00	4.56
412.00	0.00	4.19
400.00	3.00	3.85
354.00	9.00	3.74
347.00	10.00	3.85
329.00	10.00	4.14
309.00	11.00	3.99
305.00	11.00	3.93
300.00	12.00	3.86

Tabla 4.2: Velocidades del viento del radiosondeo de Madrid y del modelo WRF por nivel de presión del día 14 de julio de 2015 a las 00:00 horas UTC.

Presión (hPa)	Radiosondeo (m/s)	WRF (m/s)
944.00	0.00	5.76
942.00	0.00	5.76
925.00	3.00	6.17
850.00	1.00	3.87
750.00	2.00	1.73
743.00	2.00	1.50
700.00	3.00	2.41
670.00	10.00	5.40
653.00	10.00	6.90
638.00	10.00	7.94
632.00	11.00	7.88
592.00	16.00	7.66
581.00	16.00	7.84
570.00	17.00	8.01
538.00	18.00	8.62
518.00	14.00	9.03
500.00	10.00	8.79
491.00	6.00	8.67
481.00	2.00	8.54
421.00	6.00	4.48
405.00	7.00	3.63
400.00	7.00	3.36
370.00	0.00	3.26
334.00	0.00	4.49
322.00	1.00	4.75
300.00	1.00	3.19

Tabla 4.3: Velocidades del viento del radiosondeo de Madrid y del modelo WRF por nivel de presión del día 20 de julio de 2015 a las 12:00 horas UTC.

WRF contribuye a un valor más elevado del RMSE frente al MAE. Aunque el modelo WRF parece arrojar predicciones razonables en condiciones normales, cuando se presentan episodios de alta variabilidad, la diferencia entre los valores observados y los modelizados se amplifica. No obstante, es importante recalcar que, aunque existen días en los que los errores son más elevados, el modelo WRF en su conjunto se comporta de acuerdo con lo esperado. En la mayoría de los días, las predicciones generadas por el WRF presentan un buen ajuste con los datos observados, como se pueden ver en los ejemplos de las Figuras 4.4 y 4.5 donde es posible evidenciar que el WRF es capaz de representar de manera fiable la evolución de la velocidad del viento en el punto de estudio. En lo que respecta al Bias, muestra un valor relativamente pequeño en comparación con el MAE y el RMSE, siendo éste un valor negativo, lo que indica que, en promedio, el modelo WRF tiende a subestimar ligeramente la velocidad del viento. Si bien esta diferencia no es significativa en términos absolutos, sugiere que el WRF introduce un pequeño sesgo hacia valores ligeramente inferiores a los observados.

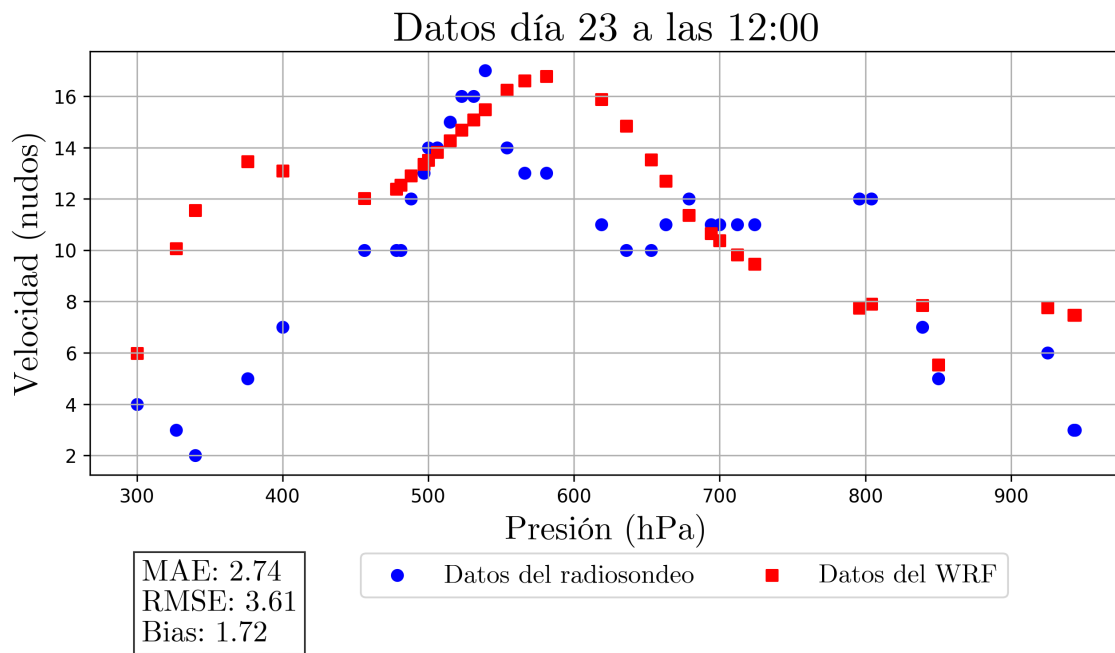


Figura 4.4: Ejemplo del día 23 de julio de 2015 a las 12:00 UTC donde se observa que el WRF genera una predicción muy ajustada la observación real.

Presión (hPa)	Radiosondeo (m/s)	WRF (m/s)
944.00	3.00	7.48
943.00	3.00	7.48
925.00	6.00	7.77
850.00	5.00	5.53
839.00	7.00	7.86
804.00	12.00	7.90
796.00	12.00	7.75
724.00	11.00	9.46
712.00	11.00	9.82
700.00	11.00	10.37
694.00	11.00	10.65
679.00	12.00	11.36
663.00	11.00	12.69
653.00	10.00	13.52
636.00	10.00	14.83
619.00	11.00	15.87
581.00	13.00	16.77
566.00	13.00	16.60
554.00	14.00	16.25
539.00	17.00	15.49
531.00	16.00	15.08
523.00	16.00	14.68
515.00	15.00	14.27
506.00	14.00	13.82
500.00	14.00	13.51
497.00	13.00	13.36
488.00	12.00	12.90
481.00	10.00	12.54
478.00	10.00	12.39
456.00	10.00	12.01
400.00	7.00	13.09
376.00	5.00	13.45
340.00	2.00	11.55
327.00	3.00	10.07
300.00	4.00	5.99

Tabla 4.4: Velocidades del viento del radiosondeo de Madrid y del modelo WRF por nivel de presión del día 23 de julio de 2015 a las 12:00 horas UTC.

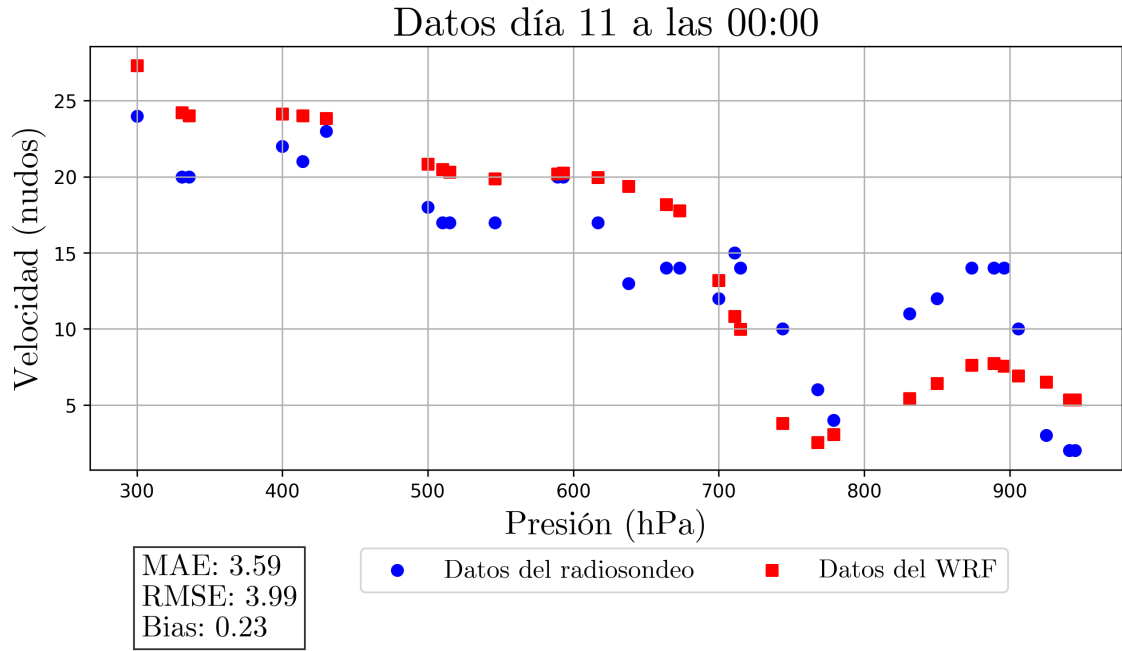


Figura 4.5: Ejemplo del día 11 de julio de 2015 a las 00:00 UTC donde se observa que el WRF genera una predicción muy ajustada la observación real.

La consistencia temporal de las predicciones de la velocidad del viento generadas por el modelo WRF puede analizarse a través de la evolución temporal de las métricas de error. Esta evolución muestra un comportamiento altamente correlacionado entre las predicciones del modelo y las observaciones reales Figura 4.6. A lo largo de todo el periodo de validación se puede observar que, cuando una métrica aumenta, las demás tienden a comportarse de igual manera, destacando el comportamiento casi simétrico del MAE y el RMSE. Es decir, todas siguen una misma tendencia. Este patrón sugiere que los errores en la predicción del viento responden a un conjunto de factores comunes que afectan de forma simultánea a la precisión del modelo.

Además, en la Figura 4.6 se puede advertir un diferente comportamiento del valor absoluto en los dos momentos de observación considerados: los errores en la observación de las 00:00 horas UTC tienden a ser mayores que los de la observación de las 12:00 horas UTC. Esta diferencia podría explicarse si se tiene en cuenta el procedimiento de simulación que tiene implementado el WRF: el modelo toma como condiciones iniciales para la simulación las correspondientes a las 00:00 horas UTC de un día determinado y con estas realiza las simulaciones correspondientes a la salida de las 00:00 horas UTC del día siguiente, es decir, entre las condiciones iniciales y la salida del modelo hay un periodo de 24 horas. Estas mismas condiciones iniciales de las 00:00 horas UTC son las que utiliza para generar las salidas de las 12:00 horas UTC. En consecuencia, el tiempo transcurrido desde la asimilación inicial de

Presión (hPa)	Radiosondeo (m/s)	WRF (m/s)
945.00	2.00	5.34
941.00	2.00	5.34
925.00	3.00	6.51
906.00	10.00	6.92
896.00	14.00	7.56
889.00	14.00	7.73
874.00	14.00	7.63
850.00	12.00	6.41
831.00	11.00	5.43
779.00	4.00	3.05
768.00	6.00	2.55
744.00	10.00	3.79
715.00	14.00	9.98
711.00	15.00	10.84
700.00	12.00	13.20
673.00	14.00	17.76
664.00	14.00	18.19
638.00	13.00	19.37
617.00	17.00	19.95
593.00	20.00	20.25
589.00	20.00	20.19
546.00	17.00	19.88
515.00	17.00	20.30
510.00	17.00	20.47
500.00	18.00	20.82
430.00	23.00	23.85
414.00	21.00	24.01
400.00	22.00	24.15
336.00	20.00	24.01
331.00	20.00	24.24
300.00	24.00	27.32

Tabla 4.5: Velocidades del viento por nivel de presión para cada fuente de datos el día 11 de julio de 2015 a las 00:00 horas UTC.

datos hasta la generación de la predicción es mayor en la franja horaria de las 00:00 horas que en la de las 12:00 horas UTC, por lo que parece razonable concluir que la capacidad de predecibilidad del modelo es mayor a las 12:00 horas que a las 00:00 horas UTC.

Errores en la predicción de la velocidad del viento: en cuanto al análisis de los errores en la dirección del viento, se pueden observar grandes similitudes con el caso de la velocidad. La estructura de los errores sigue un patrón similar, con el MAE siendo menor que el RMSE, lo que reafirma la idea de que, en términos generales, el modelo mantiene una precisión adecuada en la estimación de la variable, como se observa en la Figura [4.7](#).

Como en el caso de la velocidad, el RMSE es superior al MAE, lo que vuelve a sugerir la presencia de algunos días con errores más elevados que afectan a esta métrica. Esto implica que, aunque la tónica general de las predicciones de la dirección del viento muestran una desviación moderada respecto a los valores observados, existen situaciones puntuales en las que la diferencia entre la salida del WRF y el sondeo aerológico es grande. Sin embargo, un aspecto fundamental en la evaluación de los posibles errores en la estimación de la dirección del viento es la relación entre el Bias y las demás métricas. A diferencia de la velocidad, en este caso el Bias es casi siempre positivo y, además, mayor que el RMSE y, por tanto, que el MAE. Este comportamiento indica que el modelo tiende a sobrestimar sistemáticamente la dirección del viento en comparación con los valores medidos en la estación. Algunos ejemplos de este comportamiento se pueden observar en las figuras [4.8](#) y [4.9](#).

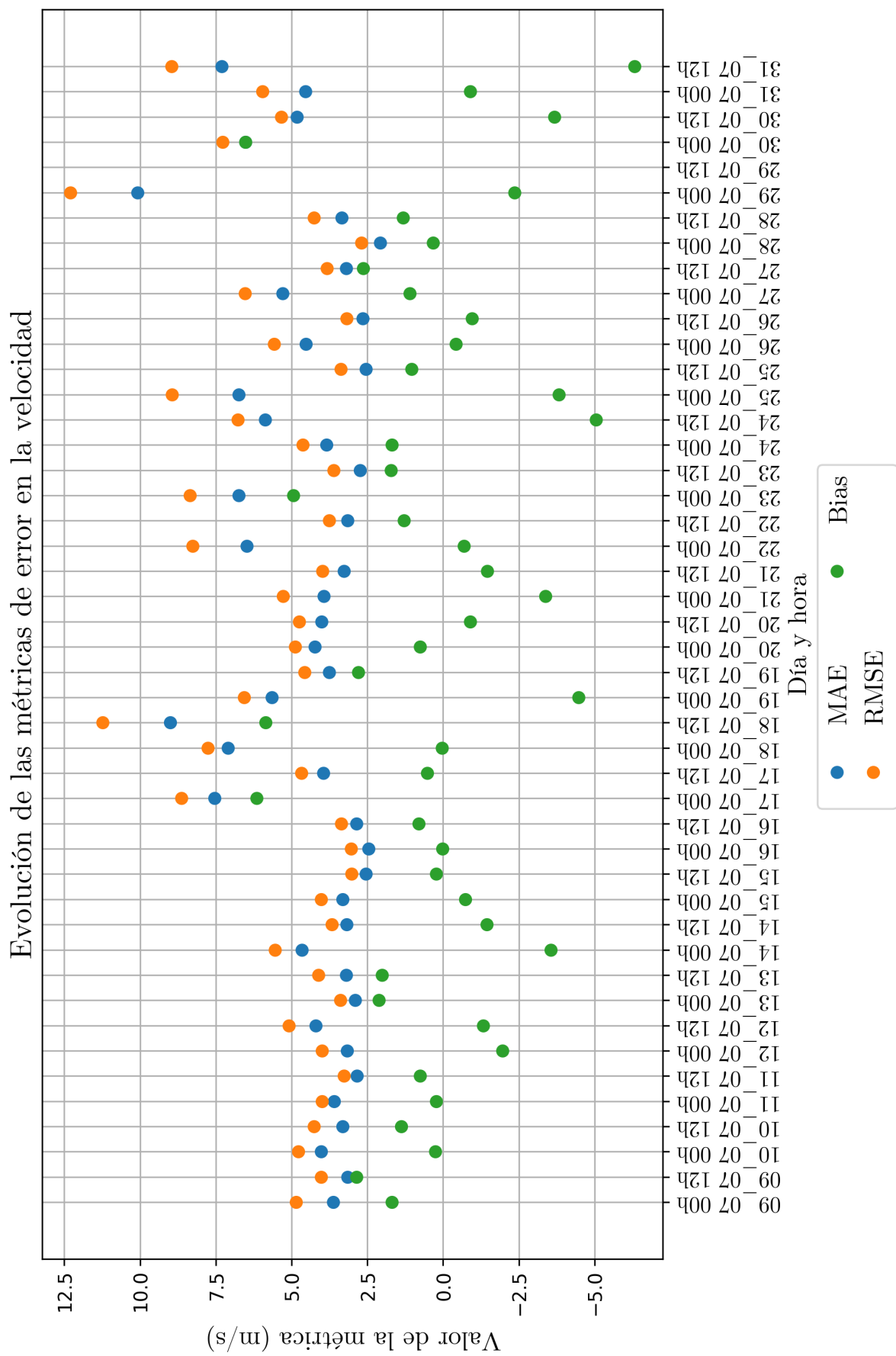


Figura 4.6: Evolución temporal en todo el periodo de estudio de las métricas de error asociadas a la velocidad del viento.

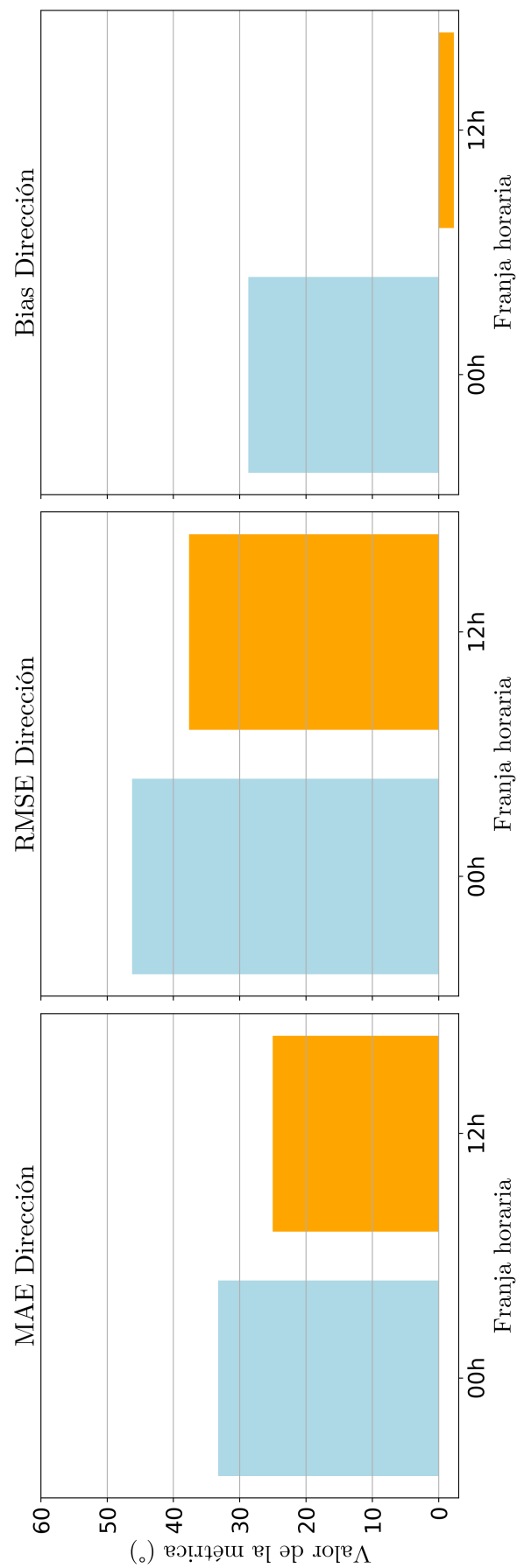


Figura 4.7: Métricas de error para las direcciones del viento generadas por el modelo WRF.

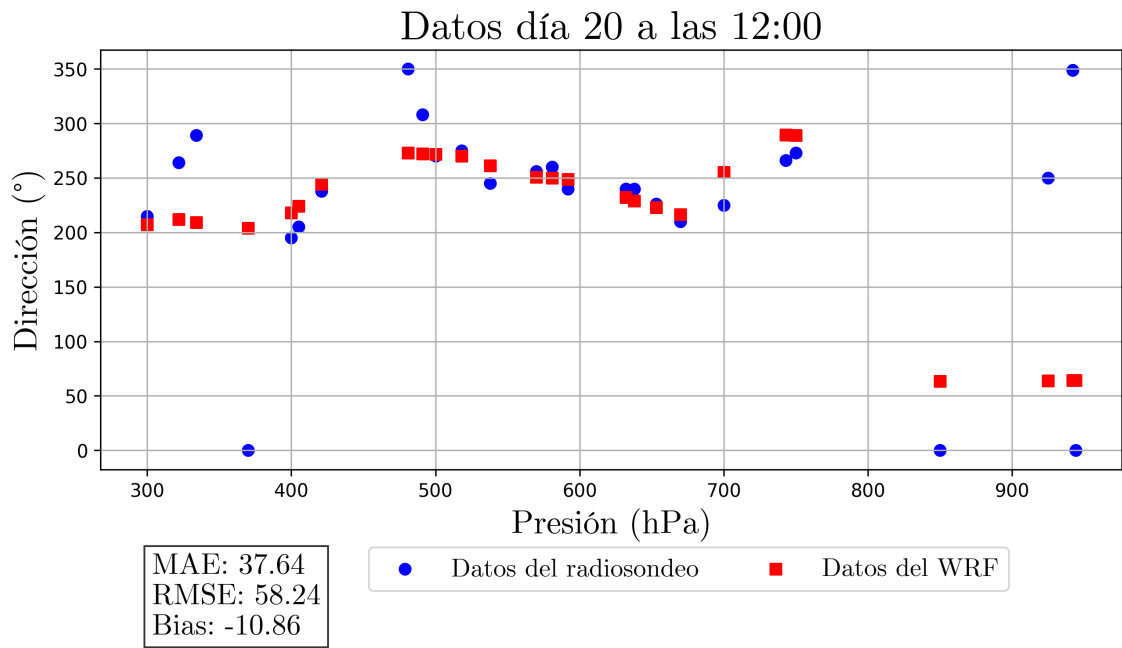


Figura 4.8: Ejemplo del día 20 de julio de 2015 a las 12:00 UTC donde se observa que el modelo WRF reproduce de forma razonable el perfil vertical de la dirección del viento.

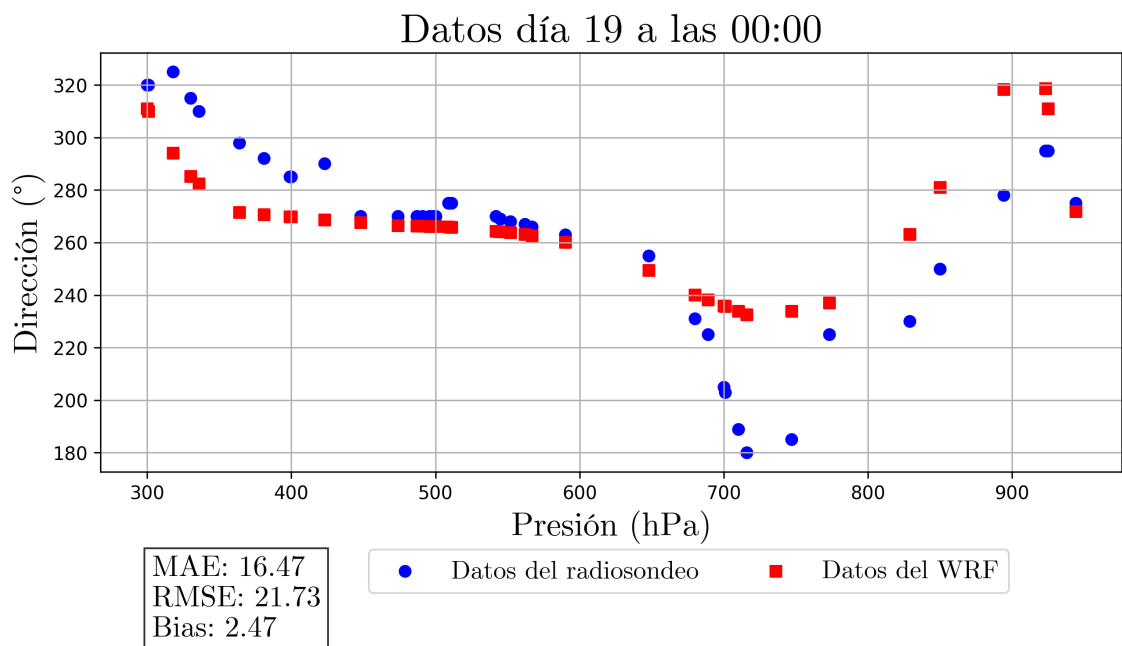


Figura 4.9: Ejemplo del día 19 de julio de 2015 a las 00:00 UTC donde se observa una muy buena concordancia entre la dirección del viento estimada por el modelo WRF y los datos del radiosondeo, con errores bajos y escasa desviación sistemática.

El hecho de que el Bias sea superior al RMSE y al MAE sugiere que el error sistemático tiene un peso considerable en la evaluación global del modelo. Esto sig-

Presión (hPa)	Radiosondeo (°)	WRF (°)
944.00	0.00	64.31
942.00	349.00	64.31
925.00	250.00	64.00
850.00	0.00	63.38
750.00	273.00	289.01
743.00	266.00	289.38
700.00	225.00	255.14
670.00	210.00	216.57
653.00	226.00	222.98
638.00	240.00	229.02
632.00	240.00	232.08
592.00	240.00	248.80
581.00	260.00	249.83
570.00	256.00	250.86
538.00	245.00	261.32
518.00	275.00	270.04
500.00	270.00	271.57
491.00	308.00	272.29
481.00	350.00	273.08
421.00	238.00	244.13
405.00	205.00	224.27
400.00	195.00	218.06
370.00	0.00	203.79
334.00	289.00	209.24
322.00	264.00	212.20
300.00	215.00	207.10

Tabla 4.6: Direcciones del viento por nivel de presión para cada fuente de datos el día 20 de julio de 2015 a las 12:00 horas UTC.

Presión (hPa)	Radiosondeo (°)	WRF (°)
944.00	275.00	271.92
925.00	295.00	310.94
923.00	295.00	318.68
894.00	278.00	318.27
850.00	250.00	280.97
829.00	230.00	263.17
773.00	225.00	237.00
747.00	185.00	233.83
716.00	180.00	232.63
710.00	189.00	233.87
701.00	203.00	235.73
700.00	205.00	235.94
689.00	225.00	238.21
680.00	231.00	240.06
648.00	255.00	249.39
590.00	263.00	260.07
567.00	266.00	262.58
562.00	267.00	263.13
552.00	268.00	263.81
545.00	269.00	264.21

Presión (hPa)	Radiosondeo (°)	WRF (°)
542.00	270.00	264.39
511.00	275.00	265.90
509.00	275.00	265.93
500.00	270.00	266.09
499.00	270.00	266.11
497.00	270.00	266.15
496.00	270.00	266.16
491.00	270.00	266.25
487.00	270.00	266.32
474.00	270.00	266.58
448.00	270.00	267.59
423.00	290.00	268.69
400.00	285.00	269.85
399.00	285.00	269.91
381.00	292.00	270.73
364.00	298.00	271.48
336.00	310.00	282.33
330.00	315.00	285.25
318.00	325.00	294.06
301.00	320.00	310.02
300.00	320.00	310.96

Tabla 4.7: Direcciones del viento por nivel de presión para cada fuente de datos el día 19 a las 00h (dividida en dos columnas))

nifica que, más allá de errores puntuales en días concretos, existe una tendencia a sobreestimar la dirección del viento. Este sesgo positivo podría implicar una desviación estructural en la representación de la dirección del viento, algo que podría estar relacionado directamente con el diseño del modelo WRF. Sin embargo, aunque esta desviación es un factor a tener en cuenta, no supone un problema crítico para el objeto de este trabajo, pues el modelo híbrido con el que se pretende trabajar dispone de herramientas variadas en Machine Learning que pueden corregir este tipo de errores sistemáticos. Al incorporar en posterior fases técnicas de aprendizaje automático, este sesgo puede compensarse de manera efectiva, mejorando la precisión final de la predicción, a pesar de que los datos en bruto puedan no parecer adecuados. En general, la modelización de la dirección del viento es más compleja que la de la velocidad, ya que presenta una mayor tasa de variabilidad en función del tiempo, y los errores obtenidos son consistentes con los resultados obtenidos en estudios similares citados anteriormente.

En lo que respecta al análisis en función de las dos horas de observación, en las figuras [4.7](#) y [4.10](#), se puede observar un comportamiento similar al del caso de la velocidad del viento: mayor error en la franja horaria de las 00:00 horas UTC frente al de las 12:00 horas UTC y una evolución simétrica para el MAE, el RMSE y el Bias, aunque en este caso todos los valores son positivos.

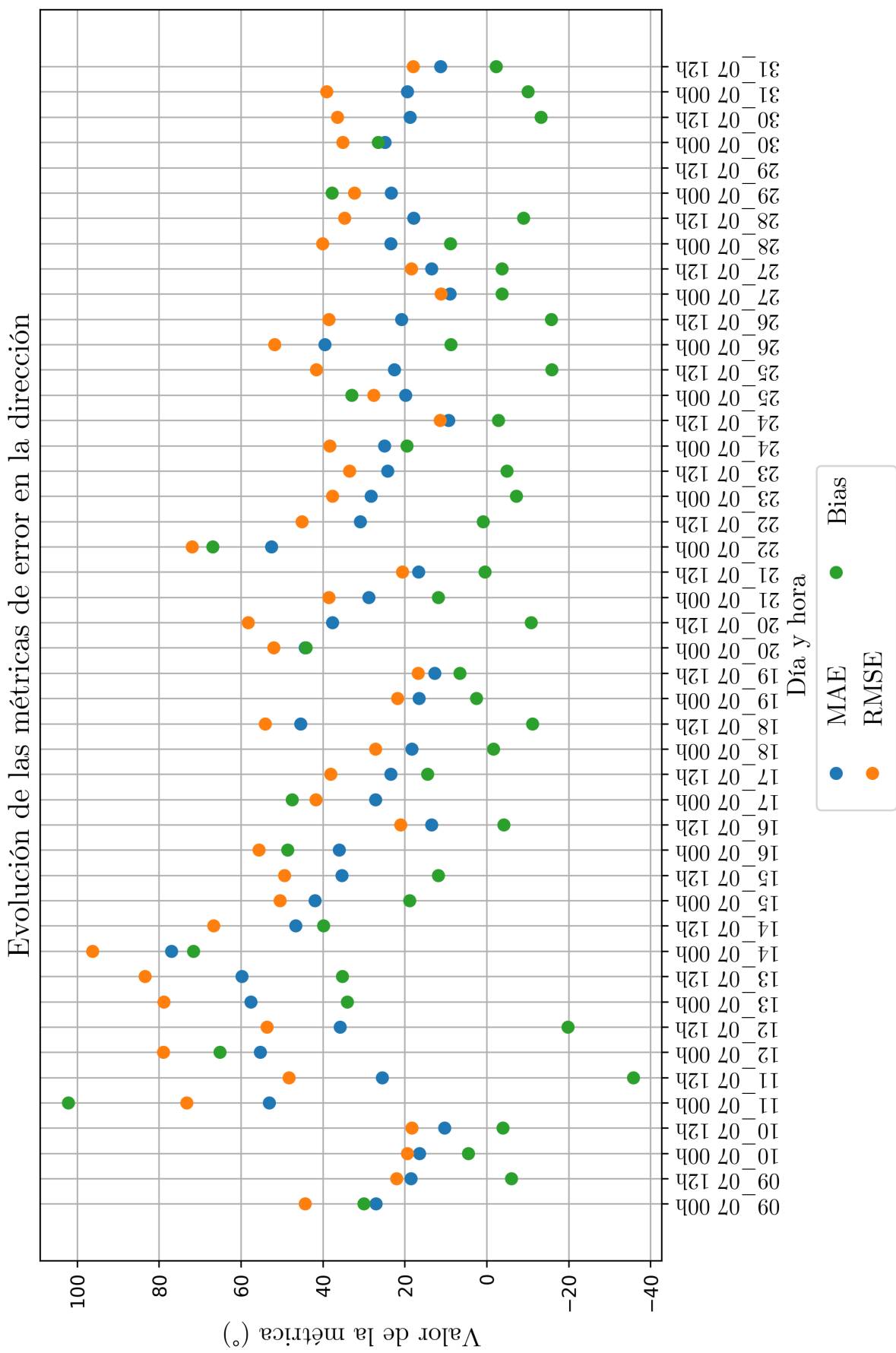


Figura 4.10: Evolución temporal de las métricas de error asociadas a la dirección del viento.

4.3. Comparación por niveles de presión y horas

Adicionalmente al análisis mostrado en los puntos anteriores, es interesante estudiar de manera más detallada el comportamiento del modelo WRF en los niveles de presión más utilizados en los estudios meteorológicos. De esta manera es posible evaluar de una forma más detallada el comportamiento del modelo en aquellos niveles más comúnmente empleados para así conocer si existen influencias debidas a procesos dinámicos y termodinámicos existentes en los diferentes niveles de presión.

Además, este análisis vertical puede revelar, por ejemplo, si el modelo WRF tiende a generar resultados con una mayor o menor dispersión en unos niveles u otros (por ejemplo, en los niveles bajos donde el relieve del terreno y la fricción son factores muy influyentes, o en los niveles más altos, donde predominan procesos como las corrientes de chorro o la subsidencia).

Para este estudio se han considerado los niveles de presión atmosférica considerados como estándar por la OMM: 300, 400, 500, 700, 850 y 925 hPa [100]. Asimismo, se han mantenido las dos horas de análisis utilizadas en el estudio temporal, 00:00 y 12:00 horas UTC, con el fin de detectar posibles variaciones en el comportamiento del modelo asociadas al momento del día que se estudia.

4.3.1. Dirección del viento

En las siguientes figuras se presentan las gráficas comparativas para la dirección del viento. Para facilitar la comparativa en cada figura se ha trazado la correspondiente bisectriz de forma que cuanto más se acerquen los puntos representados a la misma ello indicará una mejor correlación entre las medidas obtenidas por los sondeo aerológicos y las salidas del modelo numérico WRF.

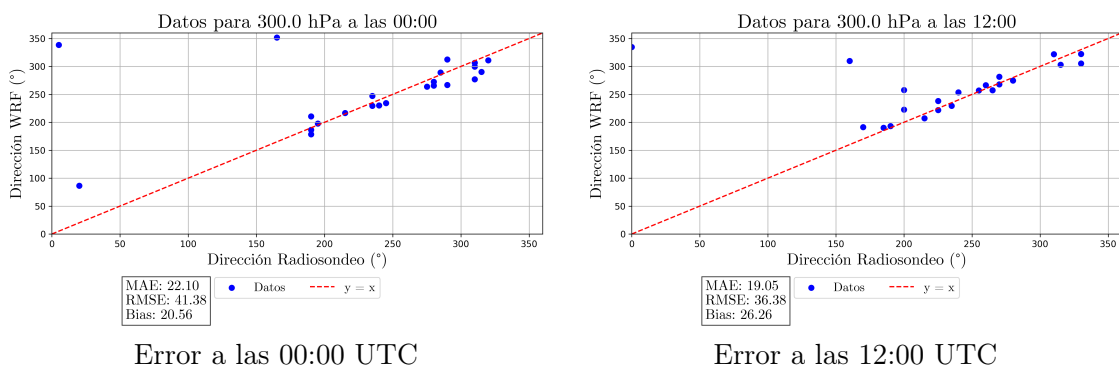
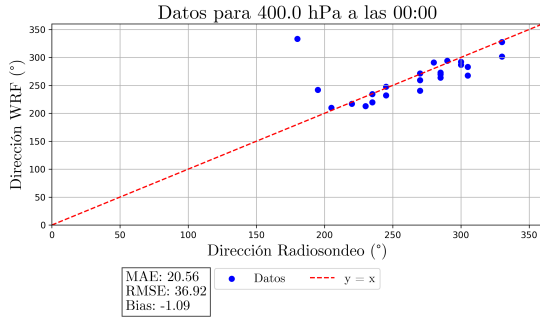
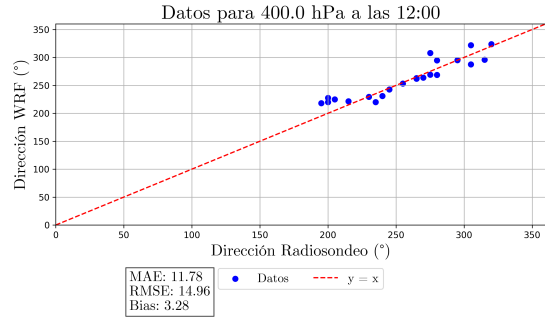


Figura 4.11: Comparación de los valores obtenidos por el modelo y los del radiosondeo para dirección a 300 hPa.

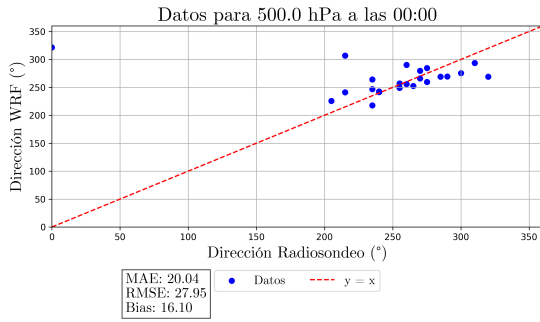


Error a las 00:00 UTC

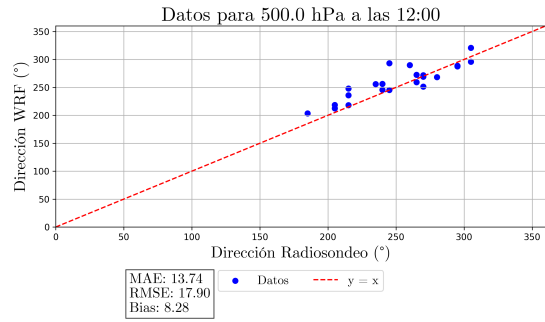


Error a las 12:00 UTC

Figura 4.12: Comparación de los valores obtenidos por el modelo y los del radiosondeo para dirección a 400 hPa.

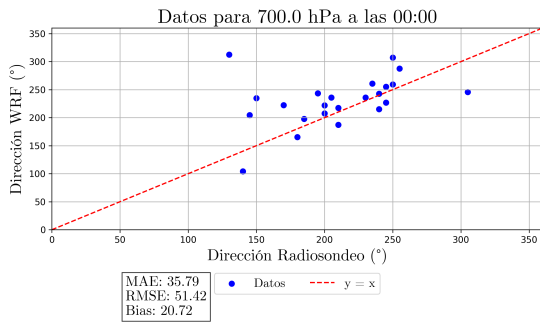


Error a las 00:00 UTC

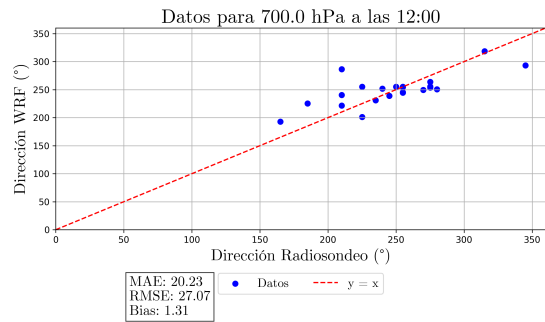


Error a las 12:00 UTC

Figura 4.13: Comparación de los valores obtenidos por el modelo y los del radiosondeo para dirección a 500 hPa.

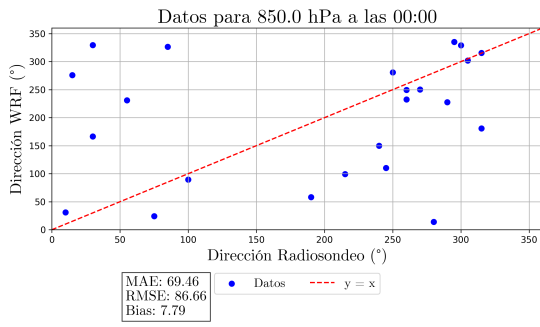


Error a las 00:00 UTC

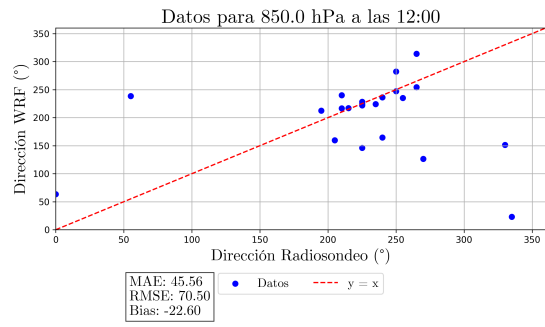


Error a las 12:00 UTC

Figura 4.14: Comparación de los valores obtenidos por el modelo y los del radiosondeo para dirección a 700 hPa.

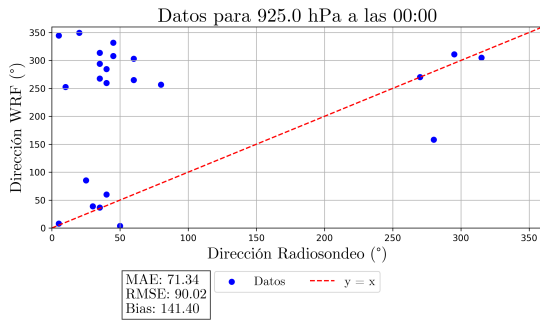


Error a las 00:00 UTC

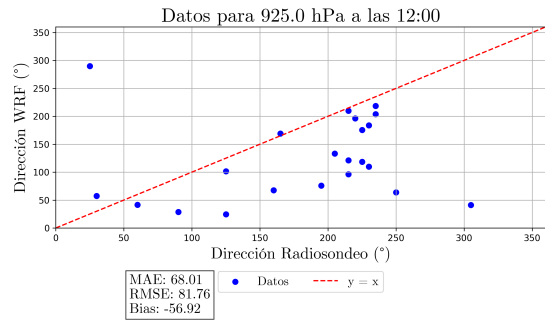


Error a las 12:00 UTC

Figura 4.15: Comparación de los valores obtenidos por el modelo y los del radiosondeo para dirección a 850 hPa.



Error a las 00:00 UTC



Error a las 12:00 UTC

Figura 4.16: Comparación de los valores obtenidos por el modelo y los del radiosondeo para dirección a 925 hPa.

Niveles de presión (hPa)	00:00 horas			12:00 horas		
	MAE (hPa)	RMSE (hPa)	Bias (hPa)	MAE (hPa)	RMSE (hPa)	Bias (hPa)
300	22.19	41.38	20.56	19.05	36.38	26.26
400	20.56	36.92	-1.09	11.78	14.96	3.28
500	20.04	27.95	16.10	13.74	17.90	8.28
700	35.79	51.42	20.72	20.23	27.07	1.31
850	69.46	86.66	7.79	45.56	70.50	-22.60
925	71.34	90.02	141.40	68.01	81.76	-56.92

Tabla 4.8: Valores de MAE, RMSE y Bias para las direcciones en cada nivel de presión.

Es posible observar que los valores proporcionados por el WRF presentan un mayor ajuste a la realidad, es decir, menor error, en los niveles más altos de la tro-

posfera, correspondientes a presiones más bajas. Esto se debe a que, en estos niveles, las condiciones atmosféricas tienden a ser más estables y están menos influenciadas por factores locales como la orografía o las interacciones con la superficie terrestre. Además, los modelos numéricos suelen representar mejor la dinámica de la atmósfera libre, donde predominan los procesos a gran escala y el efecto de la fricción es prácticamente despreciable.

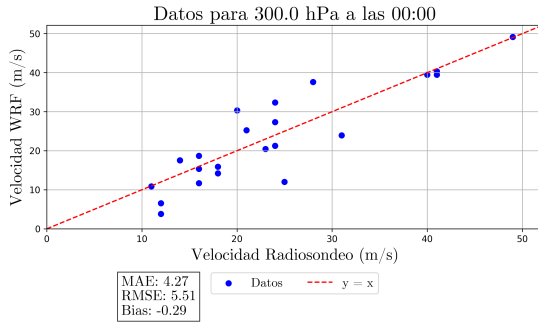
En contraposición, los errores más significativos se concentran en los niveles de 850 y 925 hPa. Estos niveles corresponden a capas bajas de la atmósfera, muy próximas a la superficie, donde la orografía, el tipo de cobertura del suelo, y otros efectos superficiales como los intercambios de energía y momento afectan de manera notable a la dinámica del viento. La complejidad de estos procesos, junto con la dificultad para representarlos con precisión en los modelos, contribuye a un aumento del error en estas capas (téngase en cuenta que estos efectos no son resueltos explícitamente por los modelos sino que son considerados a través de las parametrizaciones físicas). Esta situación es esperable y coherente la literatura habitual de referencia sobre validación de modelos meteorológicos [101] [102].

En cuanto a los dos momentos de observación considerados (00:00 y 12:00 UTC), se observa que, aunque el modelo mantiene un comportamiento general consistente, existe un patrón sistemático en el sesgo: los errores tienden a ser mayores en las simulaciones correspondientes a las 00:00 UTC que en las de las 12:00 UTC. Esta diferencia sugiere que el modelo proporciona predicciones más precisas cuando se simulan las condiciones para las 12:00 UTC, lo que podría explicarse por el hecho de que, en la práctica operativa, las simulaciones de las 12:00 UTC se generan con más antelación que las de las 00:00 UTC, disponiendo así de condiciones iniciales potencialmente más consolidadas, lo cual es consistente con el estudio de las direcciones y velocidades del apartado anterior.

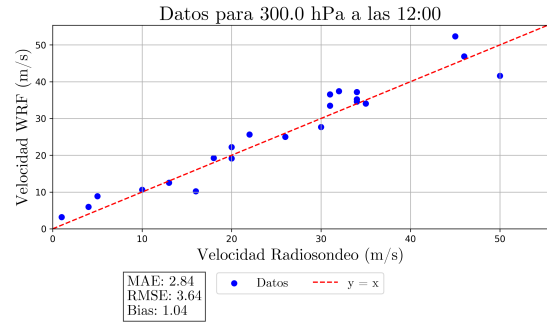
El resto de niveles de presión, especialmente los situados en la atmósfera media y alta, muestran un muy buen ajuste entre las predicciones del modelo y los datos de referencia. Este comportamiento confirma la fiabilidad del modelo en capas donde los fenómenos atmosféricos están mejor representados y menos condicionados por las irregularidades del terreno o la variabilidad local.

4.3.2. Velocidad del viento

En las siguientes figuras se presentan las gráficas comparativas para la velocidad del viento. Al igual que en la comparativa de la dirección, en cada figura se ha trazado la correspondiente bisectriz para facilitar su interpretación.

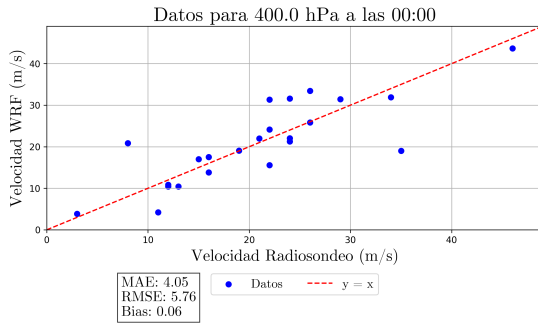


Error a las 00:00 UTC

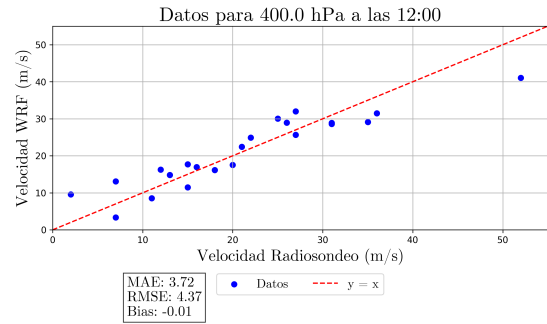


Error a las 12:00 UTC

Figura 4.17: Comparación de los valores obtenidos por el modelo y los del radiosondeo para velocidad a 300 hPa.

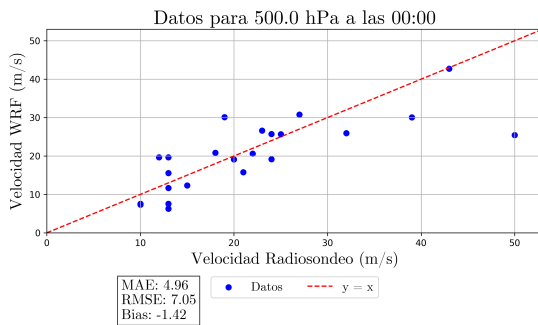


Error a las 00:00 UTC

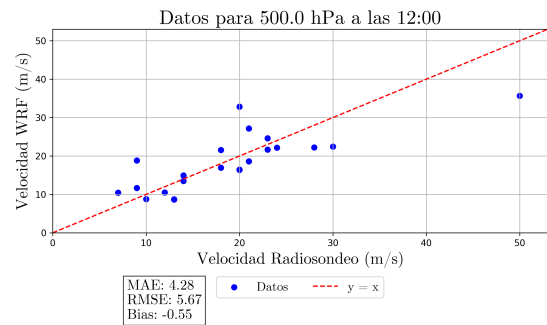


Error a las 12:00 UTC

Figura 4.18: Comparación de los valores obtenidos por el modelo y los del radiosondeo para velocidad a 400 hPa.

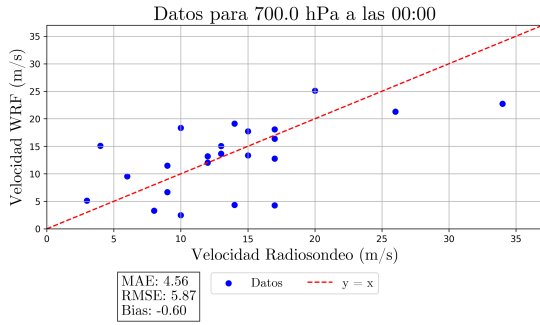


Error a las 00:00 UTC

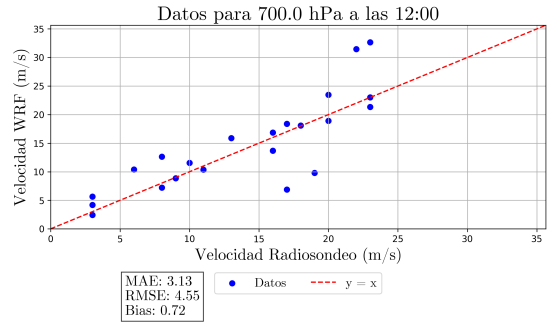


Error a las 12:00 UTC

Figura 4.19: Comparación de los valores obtenidos por el modelo y los del radiosondeo para velocidad a 500 hPa.

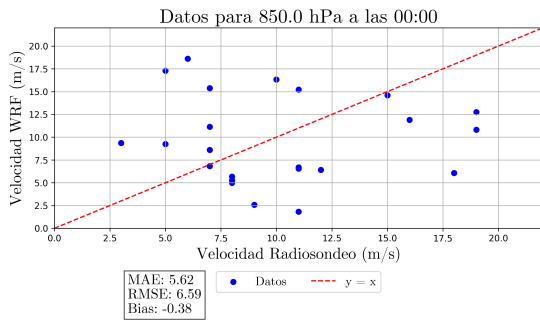


Error a las 00:00 UTC

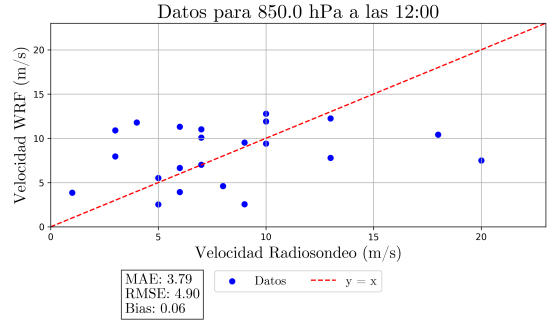


Error a las 12:00 UTC

Figura 4.20: Comparación de los valores obtenidos por el modelo y los del radiosondeo para velocidad a 700 hPa.

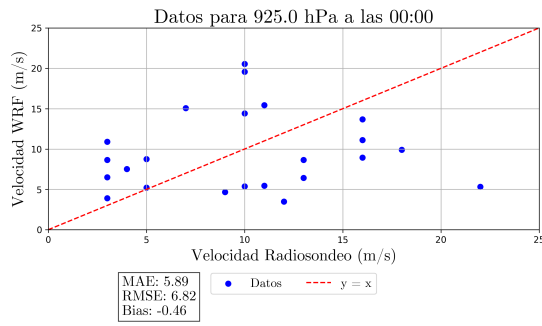


Error a las 00:00 UTC

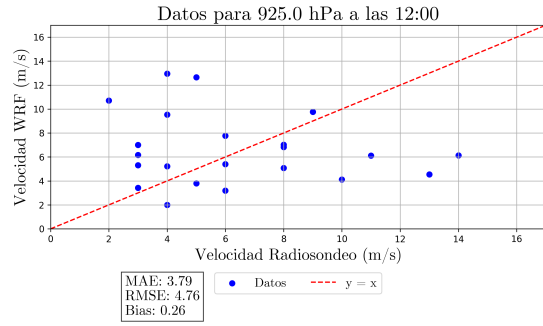


Error a las 12:00 UTC

Figura 4.21: Comparación de los valores obtenidos por el modelo y los del radiosondeo para velocidad a 850 hPa.



Error a las 00:00 UTC



Error a las 12:00 UTC

Figura 4.22: Comparación de los valores obtenidos por el modelo y los del radiosondeo para velocidad a 925 hPa.

Niveles de presión (hPa)	00:00 horas			12:00 horas		
	MAE (hPa)	RMSE (hPa)	Bias (hPa)	MAE (hPa)	RMSE (hPa)	Bias (hPa)
300	4.27	5.51	-0.29	2.8	3.64	1.04
400	4.05	5.76	0.06	3.72	4.37	-0.01
500	4.96	7.05	-1.42	4.2	5.67	-0.55
700	4.56	5.87	-0.60	3.13	4.55	0.72
850	5.62	6.59	-0.38	3.79	4.90	00.06
925	5.89	6.82	-0.46	3.79	4.76	0.26

Tabla 4.9: Valores de MAE, RMSE y Bias para las velocidades en cada nivel de presión.

En lo que respecta a la velocidad del viento, el comportamiento de los errores sigue una tendencia similar a la descrita previamente para las direcciones. Se aprecia, nuevamente, una menor dispersión en los niveles altos de la troposfera. Esto se traduce en errores notablemente inferiores en los niveles de presión en torno a los 500 hPa y superiores, donde el flujo atmosférico tiende a ser más uniforme y mejor representado por los modelos.

Por el contrario, en los niveles de presión atmosférica inferiores, particularmente en 850 y 925 hPa, se vuelven a observar un mayor grado de dispersión en la estimación de la velocidad del viento. Esta diferencia, coherente con los resultados obtenidos en el estudio de la dirección del viento, responde a la mayor complejidad en la modelización de los procesos cercanos a la superficie: la fricción, la rugosidad del terreno, y los efectos térmicos locales afectan directamente a la magnitud del viento, introduciendo una mayor variabilidad que es compleja de simular con exactitud.

Capítulo 5

Metodología híbrida para la predicción de generación de energía eólica

En este capítulo se presentan los resultados del trabajo realizado para desarrollar una metodología híbrida para la predicción horaria de energía eólica basada en la integración de las salidas del modelos WRF junto con datos del reanálisis ERA5 y algoritmos de aprendizaje automático.

De manera específica, las tareas que se discuten en los siguientes apartados son las siguientes:

- Analizar comparativamente la capacidad predictiva de los modelos WRF y ERA5, evaluando su rendimiento individual en la predicción de la potencia generada por un parque eólico real situado en Valladolid, España.
- Aplicar y comparar diferentes algoritmos de regresión de Machine Learning (árboles de decisión, Random Forest, SVR, redes neuronales, entre otros) para estimar la producción horaria de energía eólica a partir de las variables meteorológicas proporcionadas por WRF y ERA5 por separado, así como su combinación.
- Explorar la complementariedad entre modelos WRF y ERA5, comprobando si el uso conjunto de ambos conjuntos de datos mejora significativamente el rendimiento del sistema de predicción frente al uso de un único modelo.
- Evaluar cuantitativamente la precisión de las predicciones mediante métricas estándar (MAE, RMSE, R^2), con el fin de identificar la configuración óptima del sistema híbrido.

- Ofrecer una base metodológica robusta que sirva de referencia para futuras aplicaciones de sistemas híbridos de predicción en el ámbito de las energías renovables.

El trabajo realizado y los resultados obtenidos han sido publicados en el artículo Hybridizing Machine Learning algorithms with numerical models for accurate wind power forecasting (Abad-Santiago et al., 2025) [103].

5.1. Metodología

En esta sección se detalla la metodología empleada en el enfoque híbrido de predicción que se propone en este trabajo. Primero, se presentan los dos conjuntos de datos meteorológicos obtenidos de modelos numéricos —el modelo mesoescalar WRF y el reanálisis ERA5—, cuyas características se describen en las secciones ?? y 5.1.2, respectivamente. A continuación, en la sección 5.1.3, se explican los métodos específicos de regresión basados en Machine Learning que se han utilizado para integrar dicha información meteorológica y estimar la producción de energía eólica.

5.1.1. Modelo mesoescalar WRF

La ejecución del modelo WRF se ha realizado sobre la base de la configuración que se ha expuesto en los apartados anteriores de esta memoria de tesis doctoral, esto es, las simulaciones se realizaron mediante una configuración de dominios anidados, con resolución creciente hacia la zona de estudio. Se siguió el procedimiento estándar de preprocesado con WPS y la posterior ejecución del modelo dinámico (real.exe y wrf.exe). Es importante destacar que las simulaciones se llevaron a cabo sin asimilación de datos. Solo se utilizaron las condiciones atmosféricas disponibles a las 00:00 UTC del día inicial como punto de partida, a partir del cual el modelo evoluciona libremente para predecir las siguientes horas.

En lo que respecta a las fuentes externas de datos, se emplearon las siguientes:

- Datos globales NCEP GDAS/FNL (Global Data Assimilation System / Final Analyses) [86]:
 - Resolución espacial: 0,25°.
 - Intervalo temporal de 6 horas.
 - Este conjunto de datos proporciona análisis y pronósticos de la troposfera global, incluyendo variables como la temperatura, la presión atmosférica,

la humedad y el viento. Son fundamentales para poder inicializar el modelo así como para actualizar de las condiciones de contorno durante el transcurso de las simulaciones.

- Campos estáticos proporcionados por NCAR [104]:
 - Resolución: 30 segundos de arco ($\approx 1 \text{ km}$).
 - Incluyen información geográfica como topografía, uso del suelo, clasificación de coberturas, máscaras tierra-agua y albedo. Estos datos son esenciales para definir correctamente las características físicas del terreno dentro del dominio de simulación.

Para el desarrollo de la metodología híbrida se consideró como periodo de estudio todo el año 2022 por ser éste el periodo de tiempo de información sobre generación de energía eólica de que se disponía.

En lo que respecta a la selección de variables meteorológicas con capacidad de aportar información relevante para obtener predicciones adecuadas, se han tenido en cuenta aquellas que tienen influencia sobre un aerogenerador para ser capaz de generar energía eléctrica. Así, las tres componentes del viento en las direcciones u , v y w (este-oeste, norte-sur y vertical) resultan fundamentales para caracterizar el comportamiento del flujo de aire en torno a las turbinas. Además de la velocidad, la densidad del aire influye de forma decisiva en la energía cinética disponible, y esta densidad depende, a su vez, de la temperatura del aire, del contenido de vapor de agua en la atmósfera, caracterizable mediante el punto de rocío y la presión atmosférica a nivel de superficie.

Por tanto, para capturar adecuadamente las condiciones que determinan la producción de energía en el parque eólico analizado, se han considerado las siguientes variables generadas por el modelo numérico WRF:

Además de lo indicado anteriormente, la elección de estas variables, expresadas en su conjunto en la Tabla 5.1, es coherente con la aproximación seguida en estudios similares [105]. En consecuencia, representan un conjunto robusto de predictores que recogen los principales factores que afectan a la conversión de la energía del viento en energía eléctrica.

5.1.2. Modelo de reanálisis ERA5

El uso de modelos de reanálisis se ha consolidado como una herramienta fundamental en estudios atmosféricos y de energías renovables, gracias a su capacidad para proporcionar reconstrucciones consistentes y detalladas del estado pasado de la

Tabla 5.1: Variables consideradas del modelo WRF

Variable	Descripción	Unidad
T2	Temperatura del aire a 2 metros sobre la superficie	K
P2	Presión a 2 metros sobre la superficie	Pa
DP2	Punto de rocío a 2 metros sobre la superficie	K
P	Presión en diferentes niveles	Pa
U	Componente zonal (este-oeste) del viento por nivel	m/s
V	Componente meridional (norte-sur) del viento por nivel	m/s
W	Componente vertical del viento por nivel	m/s

atmósfera. Entre los distintos productos existentes, ERA5 (ECMWF Reanalysis v5), desarrollado por el Centro Europeo de Predicción a Medio Plazo (ECMWF) [106], representa la quinta generación de reanálisis y es uno de los conjuntos de datos más avanzados y utilizados en la actualidad.

ERA5 combina observaciones meteorológicas (in situ, satelitales, radiosondas, etc.) con un modelo numérico de predicción mediante sofisticadas técnicas de asimilación de datos (4D-Var). Como resultado, ofrece una estimación coherente del estado de la atmósfera a nivel global con:

- Resolución espacial de $0,25^\circ \times 0,25^\circ$ ($\approx 31 \text{ km}$)
- Resolución temporal horaria desde 1940 hasta la actualidad
- 137 niveles verticales, desde la superficie terrestre hasta los $0,01 \text{ hPa}$

Estas características convierten a ERA5 en una fuente ideal para estudios climáticos, validaciones de modelos numéricos de predicción y uso en aplicaciones diversas. Para el caso de su uso en problemas de análisis de generación de energía eólica, ERA5 resulta interesante toda vez que proporciona variables clave como la velocidad y dirección del viento a 10 y 100 m del nivel de la superficie [107, 108], además de parámetros atmosféricos relevantes en estudios de generación de energía eléctrica mediante aerogeneradores, como la temperatura del aire, la presión atmosférica o punto de rocío.

Diversos estudios han validado su precisión frente a medidas observadas, mostrando que ERA5 mejora significativamente respecto a productos anteriores similares, como ERA-Interim, especialmente en regiones con buena cobertura de observaciones. Sin embargo, su principal limitación es la latencia temporal, ya que los datos

suelen estar disponibles con un retraso de varios días, lo que lo convierte en una herramienta retrospectiva y no predictiva.

Por ello, combinar ERA5 con modelos numéricos como WRF, que permiten proyectar condiciones futuras, puede presentar una oportunidad excelente para aumentar la fiabilidad de los sistemas de predicción de energía eólica, maximizando la utilidad de ambos tipos de modelos: la robustez histórica de ERA5 y la capacidad de pronóstico de WRF.

Las variables que genera el reanálisis ERA5 y que se han considerado como relevantes para la generación de energía eólica, son: $T2$, P , $DP2$, $U10$, $V10$, $U100$ y $V100$ y se detallan en la Tabla 5.2:

Tabla 5.2: Variables consideradas del modelo ERA5

Variable	Descripción	Unidad
T2	Temperatura del aire a 2 metros sobre la superficie	K
P	Presión en superficie	Pa
DP2	Punto de rocío a 2 metros sobre la superficie	K
U10	Componente zonal (este-oeste) del viento a 10 metros de altura	m/s
V10	Componente meridional (norte-sur) del viento a 10 metros de altura	m/s
U100	Componente zonal del viento a 100 metros de altura	m/s
V100	Componente meridional del viento a 100 metros de altura	m/s

Estas variables, al igual que las seleccionadas para el modelo WRF, son clave para evaluar el estado de la atmósfera y su impacto en la producción eólica, ya que, como se ha indicado anteriormente, la temperatura y la presión influyen directamente en la densidad del aire, mientras que el punto de rocío proporciona información sobre el contenido de vapor de agua en la atmósfera. Por su parte, los componentes del viento a distintas alturas permiten analizar tanto la intensidad como la dirección del viento, factores esenciales en el rendimiento de los aerogeneradores y en la estimación precisa de la producción energética.

5.1.3. Técnicas de Machine Learning aplicadas

En el campo del Machine Learning (ML), o aprendizaje automático, existen múltiples algoritmos que pueden aplicarse a tareas de regresión, cada uno con carac-

terísticas particulares. En este trabajo se han utilizado los siguientes modelos para abordar el problema de predicción de la generación de energía eléctrica mediante tecnología eólica:

Árboles de regresión

Un árbol de regresión (o regression tree, en inglés) [109] es un modelo que predice una variable continua dividiendo progresivamente el conjunto de datos en subconjuntos más pequeños, cada uno con características similares. Estas divisiones se basan en condiciones sobre las variables de entrada, del tipo "¿la velocidad del viento es mayor que 5 m/s?", y se organizan en una estructura con forma de árbol: en cada bifurcación (o nodo), el modelo decide cómo separar los datos, y al final de cada rama (o hoja) se asigna una predicción. Esta predicción suele ser el valor medio de la variable de salida para los datos que han llegado hasta esa hoja.

Su principal ventaja radica en la interpretabilidad: un árbol de regresión permite seguir fácilmente la lógica de decisión desde la raíz hasta una predicción final. Además, pueden capturar relaciones no lineales y manejar tanto variables numéricas como categóricas sin necesidad de transformación previa. Sin embargo, son modelos altamente sensibles a pequeñas variaciones en los datos, lo que puede derivar en estructuras muy distintas con cambios mínimos en el conjunto de entrenamiento. Esta alta varianza los hace proclives al sobreajuste, especialmente cuando el árbol no se regula adecuadamente mediante técnicas como la poda o el control de la profundidad máxima.

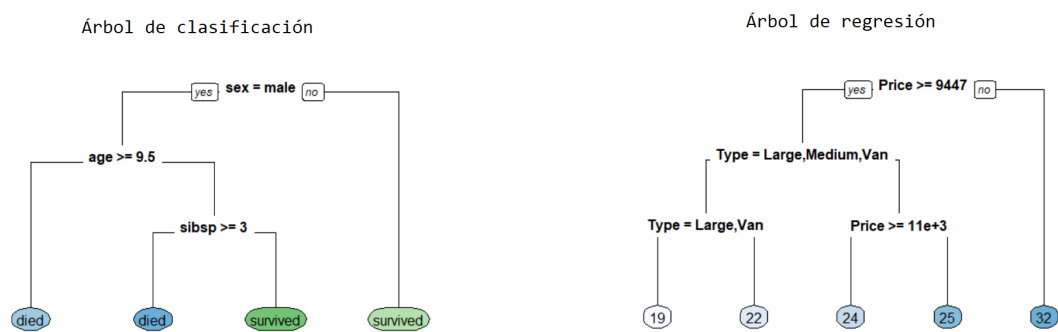


Figura 5.1: Esquema de ejemplo del funcionamiento de los Árboles de Regresión

Random Forests

Random Forest [110] es un algoritmo de aprendizaje automático basado en la combinación de múltiples árboles de decisión, cada uno entrenado sobre diferen-

tes subconjuntos aleatorios del conjunto de datos original. A diferencia de un solo árbol (los árboles de regresión expuestos en el apartado anterior), que puede ser inestable y sensible a pequeñas variaciones en los datos, un bosque aleatorio agrega las predicciones de muchos árboles independientes para obtener un resultado final más robusto. En problemas de regresión, esta agregación se realiza promediando las salidas de todos los árboles.

El proceso se basa en dos ideas clave: por un lado, el muestreo aleatorio de los datos (técnica conocida como "bagging"), y por otro, la selección aleatoria de variables en cada división del árbol, lo que introduce diversidad entre los modelos individuales. Esta diversidad es precisamente lo que permite reducir la varianza general del sistema, haciendo que el modelo general sea menos propenso al sobreajuste que un solo árbol.

Random Forest es especialmente útil en problemas con muchas variables de entrada y relaciones no lineales complejas. Funciona bien sin necesidad de escalar los datos y es bastante resistente al ruido y a la presencia de valores atípicos. Su principal desventaja es que, al tratarse de un modelo compuesto por muchos árboles, pierde la facilidad de interpretación que tienen los árboles individuales, y puede requerir mayor capacidad de cálculo si se trabaja con grandes volúmenes de datos.

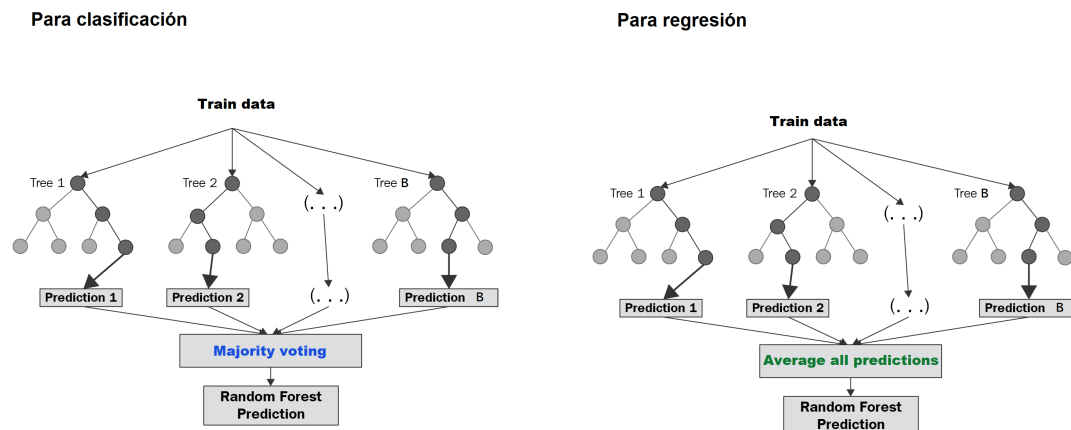


Figura 5.2: Esquema de ejemplo del funcionamiento de los Random Forests

Support Vector Regression (SVR)

Support Vector Regression [111] es una técnica derivada de las máquinas de vectores soporte (SVM), orientada a resolver problemas de regresión. A diferencia de otros modelos que intentan minimizar directamente el error entre las predicciones y los valores reales, SVR introduce un margen de tolerancia, llamado epsilon, dentro

del cual los errores no se penalizan. Es decir, el modelo busca una función que pronostique los datos de forma suficientemente precisa, pero permitiendo pequeñas desviaciones si estas se encuentran dentro de ese margen epsilon aceptable.

La idea fundamental es encontrar una función (normalmente una recta o una curva en un espacio transformado) que se aleje lo menos posible de todos los puntos, pero sin necesidad de pasar exactamente por ellos. Para conseguirlo, SVR transforma el espacio de entrada a uno de mayor dimensión mediante funciones núcleo (o "kernels"), lo que le permite capturar relaciones no lineales entre las variables. A partir de ahí, se optimiza un problema que equilibra la complejidad del modelo (que se quiere mantener baja) y la precisión (que se quiere alta).

SVR es una técnica recomendable cuando se trabaja con conjuntos de datos de tamaño moderado y relaciones complejas entre variables. No obstante, su rendimiento depende en gran medida de la elección de los hiperparámetros, como el ancho del margen ϵ , el parámetro de regularización y el tipo de kernel. Además, su coste computacional crece rápidamente con el tamaño del conjunto de datos, lo que puede hacer que no sea la opción más eficiente para problemas muy grandes o con muchas variables.

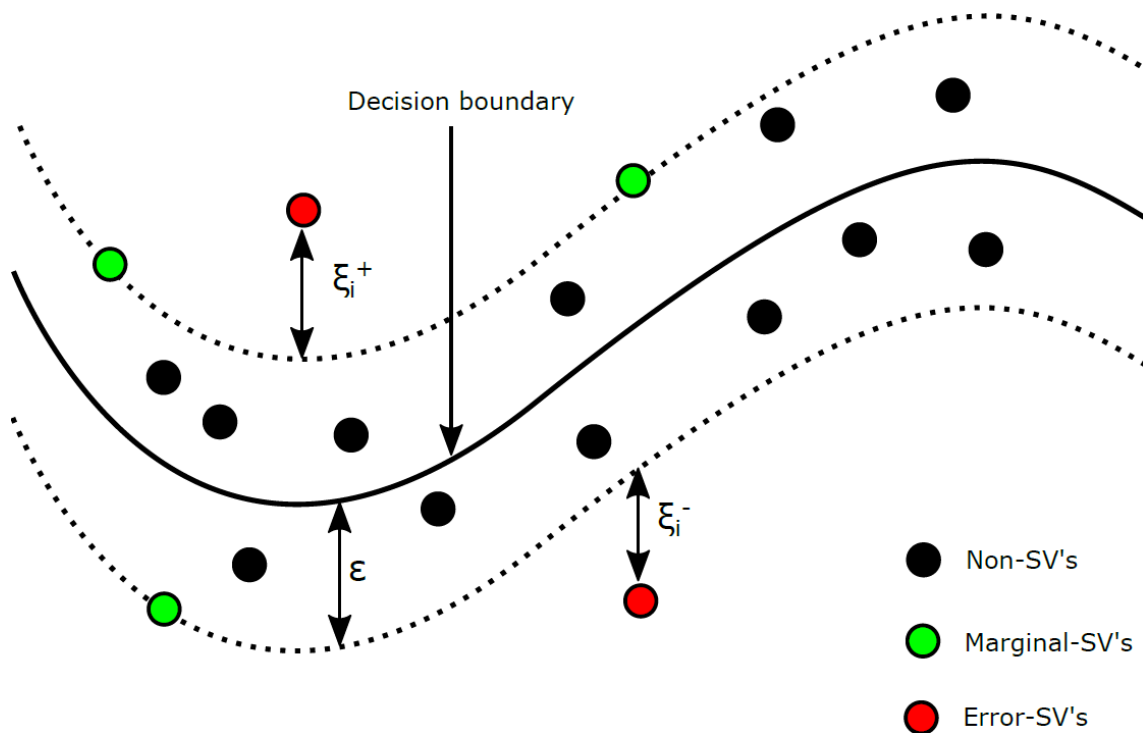


Figura 5.3: Esquema de ejemplo de funcionamiento del SVR

Multi-Layer Perceptron (MLP)

El perceptrón multicapa, o MLP por sus siglas en inglés, [112] es una de las arquitecturas más representativas de las redes neuronales artificiales. Se compone de varias capas de nodos (también llamados neuronas), organizadas en una estructura secuencial: una capa de entrada, una o varias capas ocultas, y una capa de salida. Cada neurona recibe valores de las anteriores, aplica una combinación lineal seguida de una función de activación no lineal (como ReLU o sigmoide), y transmite el resultado a la siguiente capa. Esta estructura le permite modelar relaciones complejas y no lineales entre las variables de entrada y la salida deseada.

Durante el entrenamiento, el MLP ajusta los pesos de las conexiones internas mediante un proceso iterativo llamado retropropagación del error (backpropagation), que calcula cómo influye cada peso en el error global y lo corrige utilizando técnicas de optimización como el descenso de gradiente. Esta capacidad de aprendizaje automático lo convierte en un modelo muy versátil, capaz de aproximar funciones arbitrariamente complejas si se le da la arquitectura adecuada y suficiente información.

Su principal fortaleza es precisamente esa flexibilidad, lo que lo hace especialmente adecuado para problemas donde las relaciones entre variables son altamente no lineales o difíciles de expresar mediante modelos más tradicionales. Sin embargo, los MLP también presentan algunas limitaciones importantes: requieren una cantidad considerable de datos para generalizar correctamente, son sensibles a la elección de la arquitectura (número de capas, neuronas por capa, función de activación, etc.), y su entrenamiento puede ser costoso desde el punto de vista computacional. Además, al tratarse de un modelo de tipo "caja negra", su interpretación no es sencilla, lo que puede dificultar la trazabilidad de las decisiones en contextos donde la transparencia es importante.

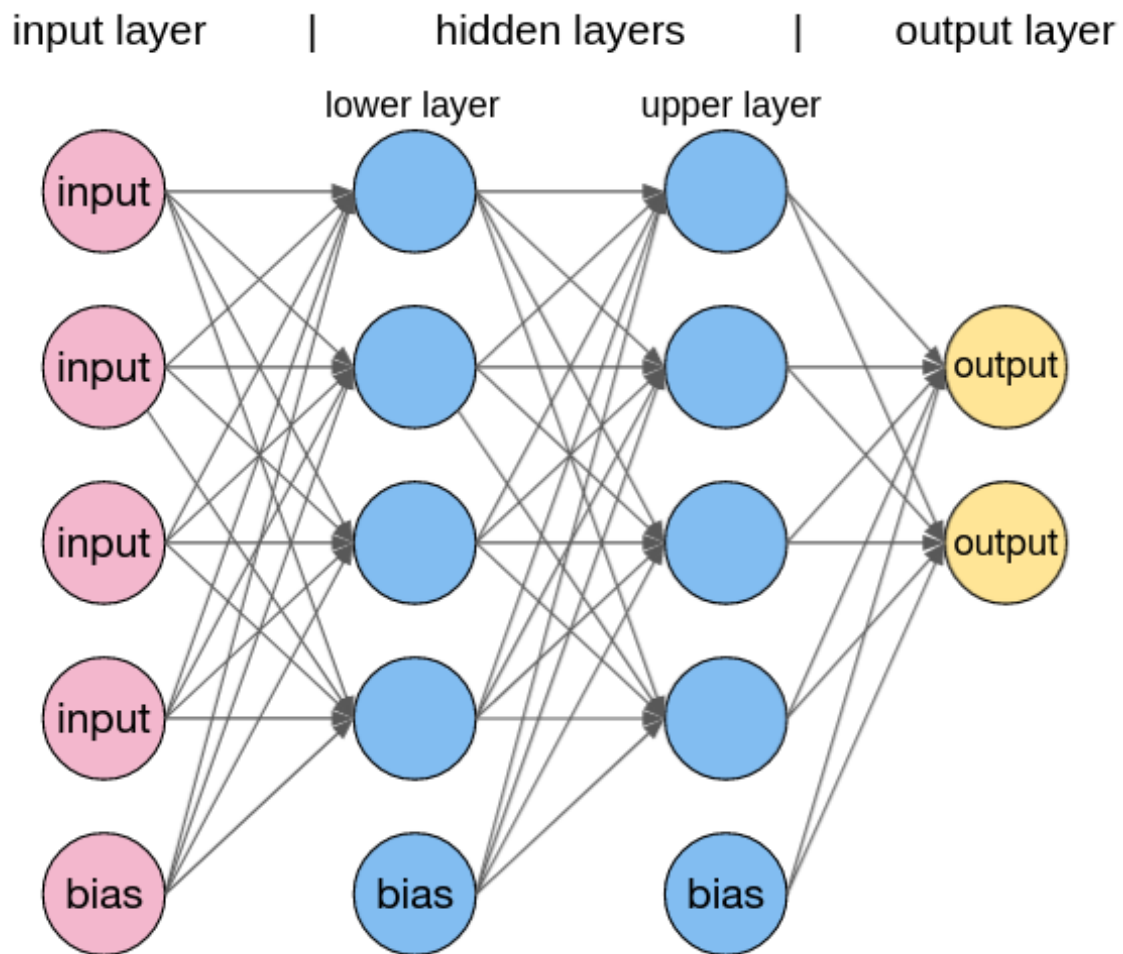


Figura 5.4: Esquema de ejemplo de funcionamiento de MLP

Extreme Learning Machines (ELM)

Extreme Learning Machines [113] son una variante simplificada de las redes neuronales de una sola capa oculta, diseñadas con el objetivo de reducir drásticamente el tiempo de entrenamiento. La principal característica que las distingue de otros modelos neuronales es que los pesos entre la capa de entrada y la capa oculta se asignan de forma aleatoria y no se actualizan durante el proceso de aprendizaje. Únicamente se ajustan los pesos que conectan la capa oculta con la capa de salida, lo cual se hace resolviendo directamente un sistema de ecuaciones lineales.

Esta estrategia elimina la necesidad de iteraciones complejas como las que se utilizan en el descenso de gradiente, lo que convierte al entrenamiento de ELM en un proceso extremadamente rápido y eficiente desde el punto de vista computacional. A pesar de su simplicidad, las ELM han demostrado ser capaces de lograr buenos resultados en tareas de regresión y clasificación, especialmente en problemas de alta dimensionalidad o donde el tiempo de entrenamiento es un factor limitante.

Sin embargo, el hecho de que parte de sus parámetros se fijen al azar introduce cierta inestabilidad en el modelo. El rendimiento final puede variar dependiendo de la inicialización aleatoria, y no siempre se garantiza que se alcance una representación óptima de los datos. Además, al igual que ocurre con otros modelos basados en redes neuronales, su interpretabilidad es limitada, lo que dificulta la comprensión del proceso de toma de decisiones subyacente. En resumen, ELM ofrece una alternativa rápida y efectiva en muchos casos, aunque con menos control y menos capacidad de explicación que otros modelos más complejos.

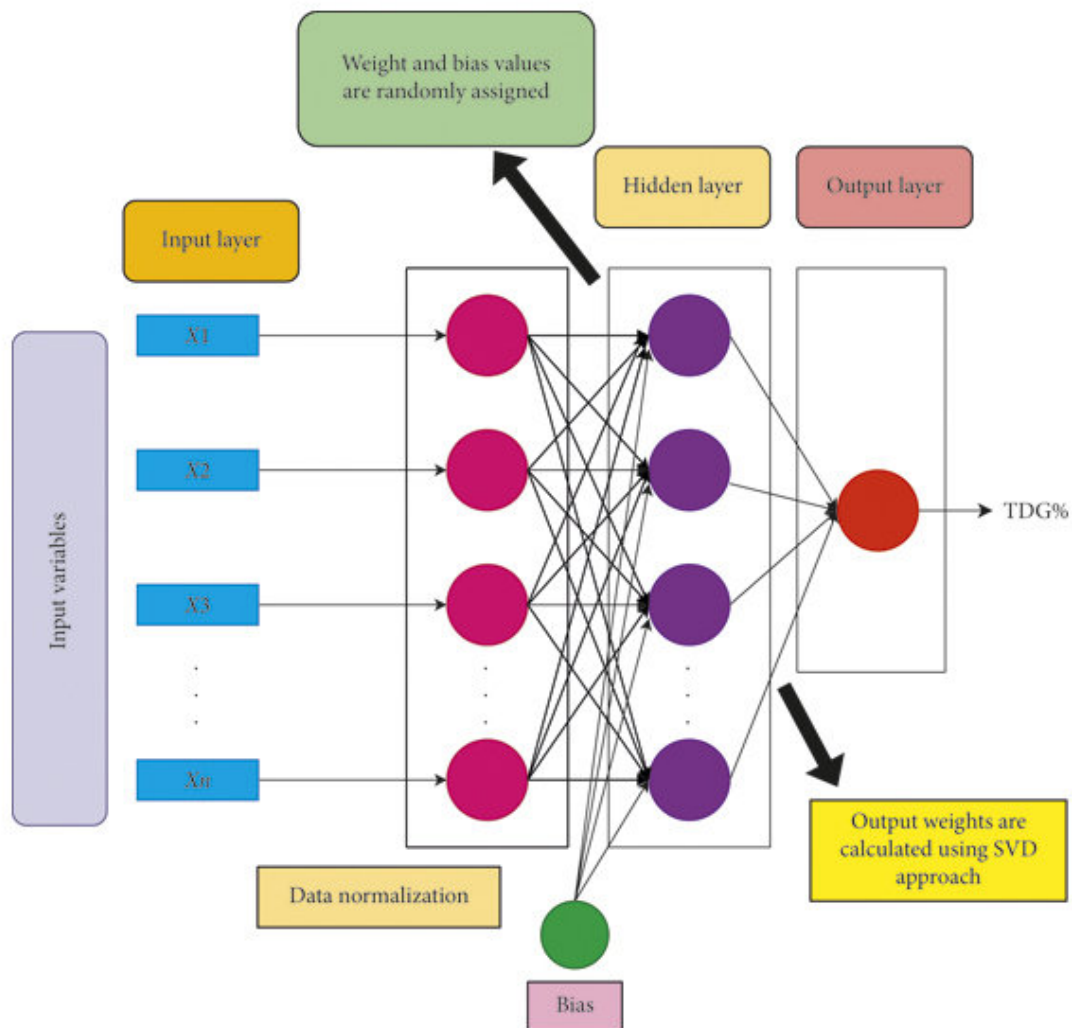


Figura 5.5: Esquema de ejemplo de funcionamiento de ELM

Least Squares Support Vector Regression (LSSVR)

Least Squares Support Vector Regression [114] es una variante de los modelos SVR que modifica la formulación clásica del problema de optimización para simpli-

ficar su resolución. Mientras que el SVR tradicional utiliza una función de pérdida con margen y optimiza un problema con restricciones que requieren métodos de programación cuadrática, LSSVR reemplaza estas restricciones por una penalización basada en mínimos cuadrados. Esto convierte el problema original en un sistema de ecuaciones lineales, mucho más rápido de resolver.

El objetivo del modelo sigue siendo el mismo: encontrar una función que se ajuste a los datos con la mayor precisión posible sin sobreajustar. Al igual que en SVR, se utiliza una función núcleo (kernel) para proyectar los datos a un espacio de mayor dimensión, donde la relación entre las variables de entrada y la salida puede ser lineal, aunque no lo sea en el espacio original. Esto le permite capturar relaciones no lineales de forma muy eficaz.

La principal ventaja de LSSVR frente a su versión estándar es la ganancia en eficiencia computacional, que resulta especialmente útil cuando se trabaja con conjuntos de datos grandes. Sin embargo, esta simplificación también tiene un coste: al no mantener un margen explícito, el modelo puede ser más sensible al ruido en los datos. Además, requiere una cuidadosa selección de hiperparámetros, como el coeficiente de regularización y el parámetro del kernel, para evitar problemas de sobreajuste o falta de generalización. En definitiva, LSSVR es un modelo potente y eficiente para tareas de regresión, especialmente útil cuando se busca un compromiso entre precisión y coste computacional.

Light Gradient Boosting Machine (LGBM)

Light Gradient Boosting Machine [115] es un algoritmo de aprendizaje automático basado en la técnica de boosting por gradiente, específicamente diseñado para ser altamente eficiente tanto en tiempo de entrenamiento como en consumo de memoria. Su funcionamiento se basa en construir un conjunto de árboles de decisión de forma secuencial, donde cada nuevo árbol intenta corregir los errores cometidos por la suma de los árboles anteriores. A diferencia de los métodos de bagging, como Random Forest, que buscan reducir la varianza combinando modelos independientes, el boosting se enfoca en reducir el sesgo mediante una optimización iterativa.

LGBM incorpora varias mejoras respecto a implementaciones anteriores de boosting, como XGBoost. Una de las más destacadas es el uso de histogramas discretizados, que agrupan los valores de las variables para reducir el número de comparaciones, acelerando así el proceso de entrenamiento. Además, utiliza una técnica llamada hoja a hoja ("leaf-wise") en la construcción de los árboles, lo que permite un ajuste más fino a los datos en comparación con el crecimiento nivel a nivel ("level-wise") que utilizan otros algoritmos.

Esta combinación de velocidad, eficiencia y capacidad de modelado hace que LGBM sea especialmente adecuado para conjuntos de datos grandes, con muchas variables o con estructuras complejas no lineales. Sin embargo, su rendimiento depende en gran medida de la correcta selección de los hiperparámetros, y como ocurre con otros métodos de boosting, existe el riesgo de sobreajuste si se construyen árboles demasiado profundos o se entrena durante demasiadas iteraciones. Además, su interpretación es menos intuitiva que la de modelos más simples, lo que puede suponer una limitación en contextos donde la transparencia del modelo es relevante.

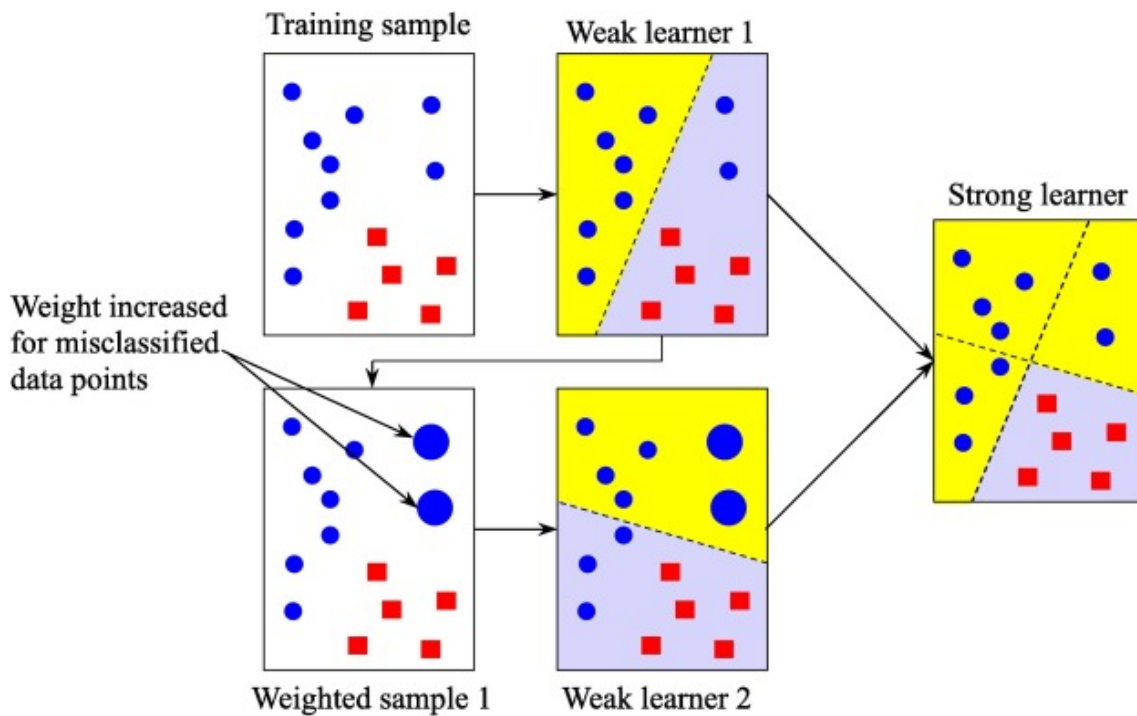


Figura 5.6: Esquema de ejemplo de funcionamiento de LGBM

5.1.4. Preprocesado y ajuste de hiperparámetros

Antes del entrenamiento de los modelos, se realizó un proceso de estandarización de las variables predictoras. Esta transformación consiste en restar la media y dividir entre la desviación estándar de cada variable, de modo que todas las características tengan media cero y varianza unitaria. Esta etapa resulta especialmente importante en modelos sensibles a la escala de los datos, como los basados en kernels (SVR, LSSVR) o redes neuronales (MLP), ya que evita que variables con diferentes órdenes de magnitud dominen el proceso de aprendizaje. La fórmula que se ha utilizado es la siguiente:

$$x' = \frac{x - \bar{x}}{\sigma} \quad (5.1)$$

donde x es el vector de características original, $\bar{x}$ representa la media de las características y σ su desviación estándar.

En cuanto al ajuste de los hiperparámetros, se empleó una estrategia basada en búsqueda aleatoria (randomized search) combinada con validación cruzada de cinco pliegues (5-fold cross-validation). En esta estrategia, el conjunto de entrenamiento se divide en cinco partes iguales, entrenando el modelo en cuatro de ellas y validando en la quinta. Este procedimiento se repite cinco veces, utilizando cada subconjunto como validación en una ocasión distinta, y se calcula el error medio sobre todas las particiones. Este enfoque permite obtener estimaciones robustas del rendimiento esperado en datos no vistos, al tiempo que reduce el riesgo de sobreajuste asociado a una única partición.

Para cada modelo se evaluaron diez combinaciones aleatorias de hiperparámetros, extraídas de un rango razonable para cada uno de ellos. La combinación con menor error medio de validación fue seleccionada para el entrenamiento final. Una vez optimizados, los modelos se entrenaron de nuevo sobre la totalidad del conjunto de entrenamiento y posteriormente se evaluaron sobre el conjunto de prueba reservado.

La aproximación seguida busca garantizar un equilibrio adecuado entre la exploración del espacio de hiperparámetros y la eficiencia computacional. Aunque existen técnicas más exhaustivas como la búsqueda en malla (grid search), su coste crece rápidamente con el número de parámetros y valores posibles, en este estudio, la búsqueda aleatoria ha demostrado ser suficientemente eficaz para encontrar configuraciones competitivas en todos los modelos considerados. En la Tabla 5.3 se muestran los valores seleccionados en este proceso.

5.1.5. Caso de estudio y variables predictoras consideradas

Este trabajo se ha centrado en un parque eólico situado en La Mudarra, municipio de la provincia de Valladolid (Castilla y León, España) (coordenadas: 41° 49' 57,6" N 4° 53' 39,2" O). Esta instalación, ubicada en la comarca de Montes Torozos, cuenta con 12 aerogeneradores Siemens-Gamesa SG 3.4-132, cada uno con una potencia nominal de 3.465 kW y un diámetro de rotor de 132 metros, lo que suma una capacidad total instalada de 41.580 kW.

El entorno geográfico de la zona se caracteriza por un relieve predominantemente llano, con ligeras variaciones asociadas a la acción erosiva de pequeños cursos fluviales. A una altitud media de 842 metros sobre el nivel del mar, la ubicación

Tabla 5.3: Rangos de búsqueda definidos para la optimización de hiperparámetros en cada modelo de Machine Learning.

Modelo	Hiperparámetros considerados
RT (Árboles de Regresión)	Profundidad máxima (<code>max_depth</code>): 1–50 Muestras mínimas por hoja (<code>min_samples_leaf</code>): 1–50
RF (Random Forest)	Número de árboles (<code>n_estimators</code>): 100–600 Bootstrap (<code>bootstrap</code>): True / False
SVR (Support Vector Regression)	Regularización (<code>C</code>): 10–500 Margen (<code>epsilon</code>): 0–1 Núcleo (<code>kernel</code>): rbf
LSSVR (Least Squares SVR)	Regularización (<code>C</code>): 1–10 Parámetro del núcleo (<code>gamma</code>): 10^{-4} –1
MLP (Perceptrón Multicapa)	Número de capas ocultas: 3 Neuronas por capa: 20–600 Activación: relu, sigmoid Optimizador: adam Tamaño de lote (<code>batch_size</code>): 16, 32, 64 Épocas: 100–500
ELM (Extreme Learning Machines)	Tamaño de la capa oculta (<code>hidden_size</code>): 100–1000
LGBM (LightGBM)	Número de hojas (<code>num_leaves</code>): 20–200 Iteraciones (<code>n_estimators</code>): 50–500

combina condiciones topográficas favorables para la instalación de parques eólicos con un clima típicamente continental. Esta región, al norte del valle del río Duero, se encuentra en una zona de transición entre la meseta y los valles agrícolas de Tierra de Campos, lo que la convierte en un enclave representativo de entornos de interior peninsular.

Para este estudio se ha considerado una resolución temporal de una hora, tanto en las variables predictoras como en los datos de producción de energía eléctrica del parque eólico. La serie temporal completa se corresponde con el año 2022 para el cual se dispone de registros continuos de potencia generada. Con el fin de entrenar y validar la metodología de predicción, los datos se han dividido de forma secuencial: el 70 % inicial se ha destinado al entrenamiento, y el 30 % restante se ha reservado para la evaluación de cada modelo. El objetivo predictivo ha sido la estimación de la potencia generada a una hora vista ($h + 1$), a partir de las condiciones atmosféricas observadas en la hora actual (h).

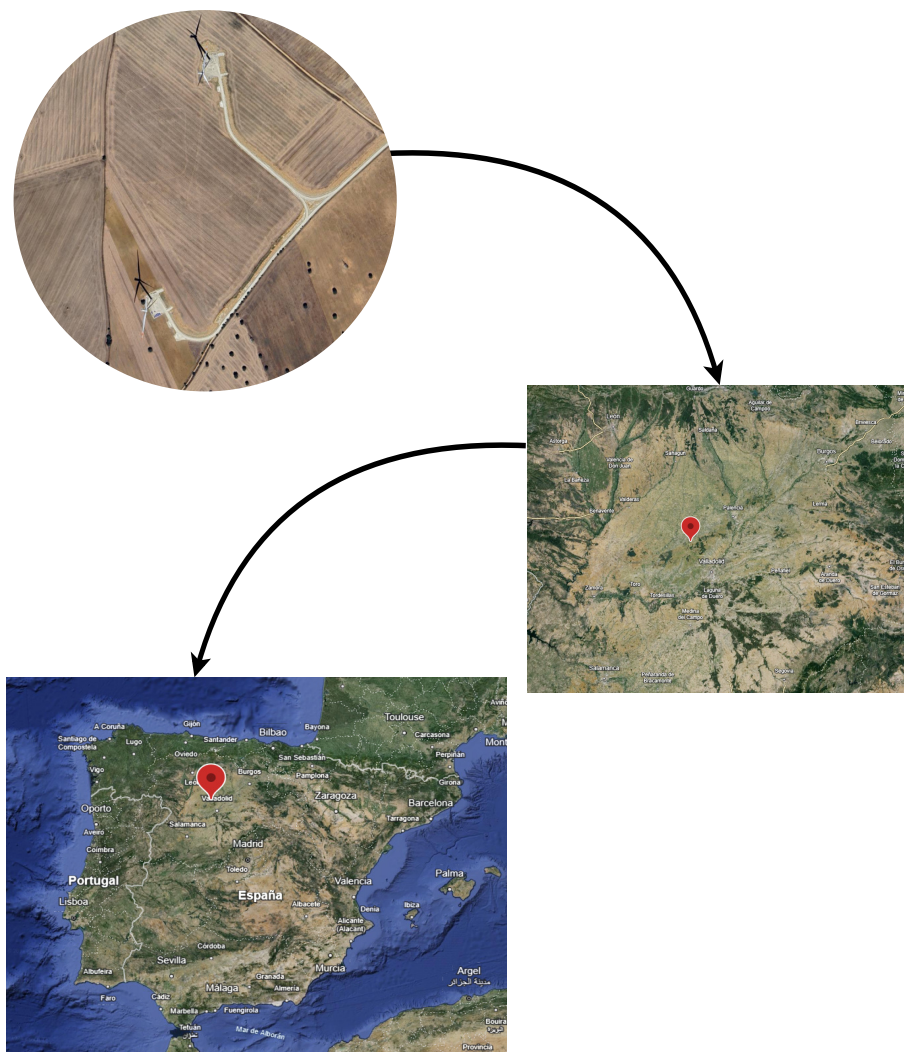


Figura 5.7: Ubicación del parque eólico considerado para la predicción de la energía eléctrica.

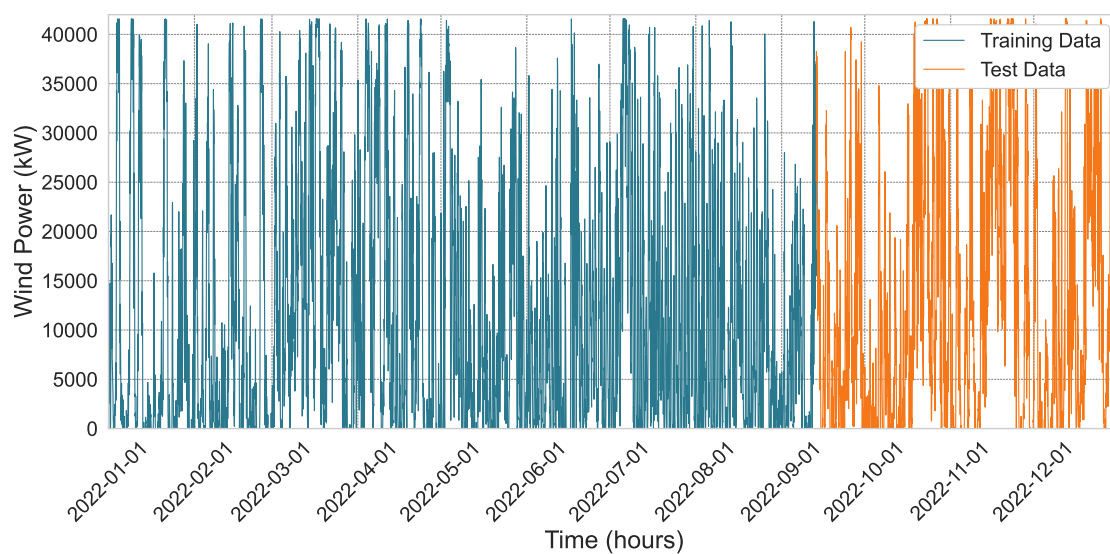


Figura 5.8: Producción eólica media horaria en 2022 y su distribución en entrenamiento y test para los modelos ML

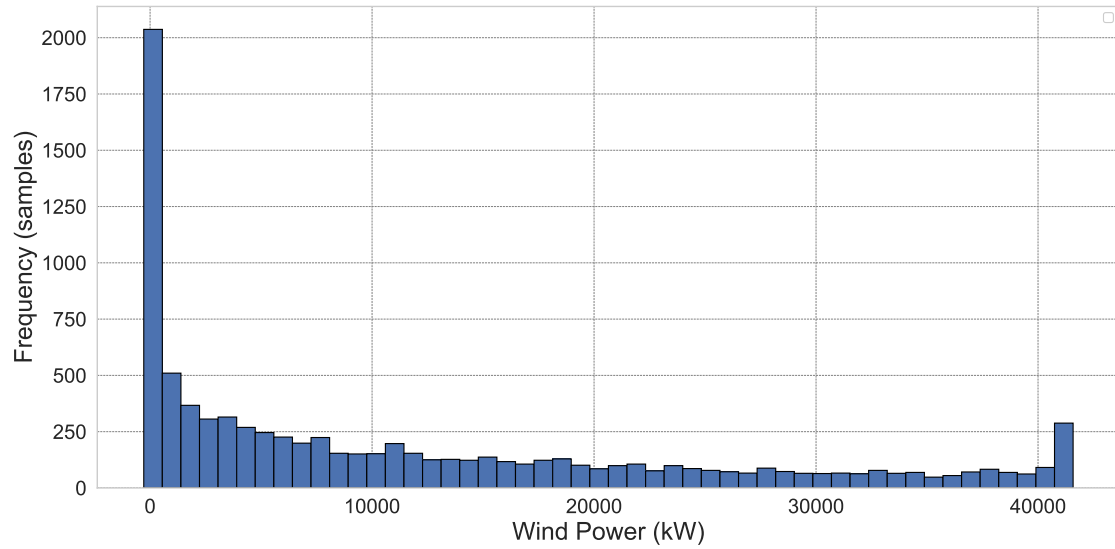


Figura 5.9: Distribución de la producción eólica media horaria en 2022.

Las variables predictoras empleadas en este estudio son las correspondientes al reanálisis del ERA5 y las simulaciones numéricas generadas por el modelo WRF expuestas anteriormente y que tienen influencia directa en la generación de energía eléctrica mediante la tecnología eólica por tener efecto sobre el rendimiento de los aerogeneradores.

En el caso de ERA5, la resolución espacial se limita a una malla de seis nodos que cubren la zona de estudio. Esta malla resulta de combinar dos latitudes ($42,0^\circ$ y $41,75^\circ$) con tres longitudes ($\sim 5,25^\circ$, $\sim 5,00^\circ$ y $\sim 4,75^\circ$). Cada nodo proporciona una única lectura por variable y hora, lo que implica un total de 6 puntos geográficos por cada instante temporal.

Por su parte, el modelo WRF ofrece una malla más densa, con un total de 20 nodos distribuidos en cinco longitudes ($\sim 5,333^\circ$, $\sim 5,156^\circ$, $\sim 4,978^\circ$, $\sim 4,800^\circ$ y $\sim 4,622^\circ$) y cuatro latitudes ($41,552^\circ$, $41,731^\circ$, $41,910^\circ$ y $42,090^\circ$). A esto se suma una mayor riqueza vertical, ya que el modelo proporciona valores en 29 niveles de presión, lo que permite representar con mayor detalle el perfil atmosférico, especialmente en lo que respecta al comportamiento del viento en altura.

Esta diferencia se traduce en una notable disparidad en el volumen de datos disponibles por hora: mientras que ERA5 proporciona 42 valores por hora (una lectura por variable y nodo), WRF genera 119 valores por hora y nodo, resultando en más de 2.300 variables por hora en la configuración empleada. Además, es importante considerar la latencia temporal de cada fuente: ERA5 presenta un retraso aproximado de una semana en la disponibilidad de datos, mientras que WRF permite obtener simulaciones con una antelación de 12 a 24 horas, lo que refuerza su aplicabilidad

en escenarios operativos.

En resumen, la fusión de ambas fuentes no solo aporta una mayor riqueza informativa y resolución espacial, sino que también permite combinar la robustez retrospectiva de ERA5 con la capacidad de predicción a corto plazo de WRF, ofreciendo una base sólida para el entrenamiento de modelos híbridos de aprendizaje automático aplicados a la predicción horaria de energía eólica.

Tabla 5.4: Coordenadas espaciales de los nodos considerados en cada modelo.

Modelo	Latitudes	Longitudes
ERA5	41.75, 42.00	-5.25, -5.00, -4.75
WRF	41.552, 41.731, 41.910, 42.090	-5.333, -5.156, -4.978, -4.800, -4.622

5.1.6. Métricas de evaluación del rendimiento

Para evaluar el rendimiento de las diferentes aproximaciones para estimar la generación horaria de potencia eléctrica, se han empleado tres métricas ampliamente utilizadas en problemas de regresión: el error absoluto medio (MAE), la raíz del error cuadrático medio (RMSE) y el coeficiente de determinación (R^2). Las dos primeras métricas ya han sido descritas con detalle en secciones anteriores de esta memoria de tesis doctoral (ver ecuaciones 4.1, 4.2 y 4.3).

El coeficiente de determinación (R^2) se introduce aquí como una métrica complementaria que permite valorar hasta qué punto el modelo ha sido capaz de capturar la estructura subyacente de los datos. En términos generales, (R^2) indica qué proporción de la variabilidad de la variable que se desea predecir puede ser explicada por el modelo. Si el modelo logra predecir perfectamente todos los valores observados, (R^2) tomará el valor 1. Un valor de 0 implica que el modelo no mejora respecto a una predicción constante basada en la media de los datos. Si el valor es negativo, el modelo rinde incluso peor que esa predicción trivial.

$$R^2 = \frac{\sum_{i=1}^n (y_i - E[y])(\hat{y}_i - E[\hat{y}])}{\sqrt{\sum_{i=1}^n (y_i - E[y])^2} \sqrt{\sum_{i=1}^n (\hat{y}_i - E[\hat{y}])^2}} \quad (5.2)$$

Donde $\hat{y}_i$ representa los valores predichos (proporcionados por el modelo), y_i los valores observados (reales), $E[y]$ la media de los valores observados, $E[\hat{y}]$ la media de los valores predichos y n es el número total de muestras.

A diferencia del MAE o el RMSE, que cuantifican directamente el error en las mismas unidades que la variable objetivo (en este caso, kilovatios), el (R^2) proporciona una medida relativa, que ayuda a entender no solo si el modelo acierta, sino

también si comprende la lógica y los patrones que rigen el comportamiento de la variable de interés.

Esta combinación de métricas permite una evaluación integral del sistema de predicción: se cuantifica tanto el error absoluto cometido como la capacidad del modelo para explicar la variabilidad de los datos. Este enfoque es especialmente útil en estudios como el presente, donde se comparan aproximaciones con diferentes arquitecturas y fuentes de información, y se busca no solo minimizar el error, sino también identificar configuraciones capaces de generalizar lo mejor posible ante distintas situaciones atmosféricas.

5.2. Resultados obtenidos

En esta sección se presentan los resultados de predicción obtenidos por los distintos algoritmos de aprendizaje automático. Como ya se ha indicado, los modelos se entrenaron utilizando el 70 % de los datos, correspondientes a la parte inicial de la serie temporal, mientras que el 30 % restante se utilizó como conjunto de prueba para la validación. La variable objetivo fue la estimación de la producción horaria de potencia eléctrica generada por el parque eólico, obtenida a partir de los datos meteorológicos procedentes de cada uno de los modelos numéricos por separado y mediante su uso combinado.

En la Tabla [5.5](#) se muestran las métricas de error obtenidas por cada modelo de Machine Learning utilizando tres configuraciones distintas de entrada: únicamente datos de WRF, únicamente datos de ERA5 y la combinación de ambos conjuntos. Las métricas presentadas son el MAE, el RMSE y el coeficiente de determinación (R^2), promediadas en cinco ejecuciones independientes para cada modelo, con el fin de mitigar los efectos de aleatoriedad que introducen algunos algoritmos.

Se observa que los errores obtenidos son del mismo orden de magnitud para todos los modelos, independientemente del tipo de entrada. Esta consistencia general se mantiene entre las distintas técnicas, aunque con algunas diferencias significativas en cuanto a rendimiento. En particular, el modelo Random Forest (RF) destaca por su rendimiento consistente, alcanzando sistemáticamente los valores más bajos de RMSE y los valores más altos de (R^2) en las tres configuraciones. Asimismo, el modelo LGBM también muestra un comportamiento notable, especialmente en términos de MAE cuando se emplean variables del modelo WRF.

Por otro lado, aunque los valores obtenidos con ERA5 y WRF por separado son similares, se observa una mejora clara al combinar ambos conjuntos de datos. En todos los casos, los errores se reducen y el valor de (R^2) alcanza su máximo, lo que

confirma que la combinación de entradas permite capturar mejor la variabilidad de la producción de potencia eléctrica mediante tecnología eólica.

Esta mejora puede visualizarse de forma más clara en las figuras correspondientes: 5.10, 5.11 y 5.12, donde se representan mediante gráficos de barras los errores medios (MAE y RMSE) y el valor de (R^2) para cada configuración de entrada. Al considerar el MAE medio, se observa que el uso exclusivo de WRF introduce un aumento del error de aproximadamente un 2 % con respecto a ERA5. Sin embargo, al combinar ambas fuentes de datos, el error disminuye en torno a un 4 %, lo cual representa una mejora significativa. Esta diferencia se acentúa aún más si se analizan los mejores casos individuales: el error de WRF supera en un 4 % al de ERA5, pero la combinación reduce el error un 6 % adicional respecto a este último.

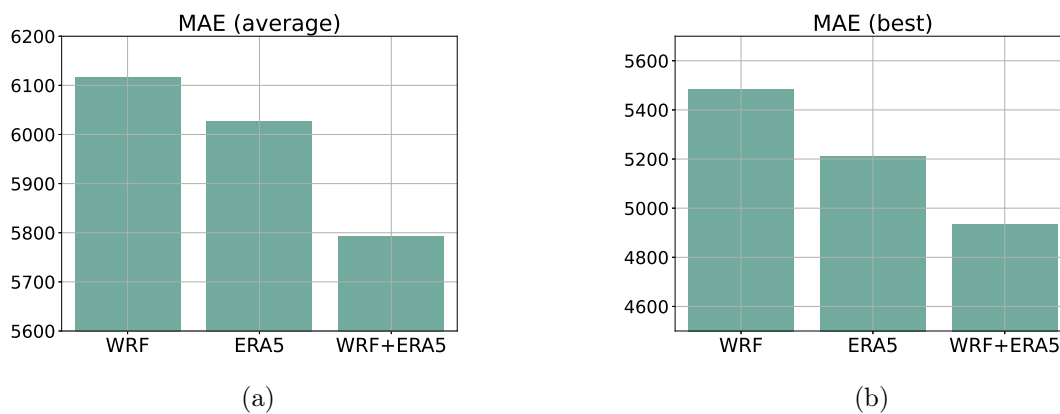


Figura 5.10: Representación gráfica del *MAE* (promedio (a) y el mejor (b)) para cada caso de estudio.

Una tendencia similar se observa con el RMSE: los errores medios individuales de WRF y ERA5 son muy próximos, mientras que su combinación los reduce aproximadamente un 5 %. En el mejor caso, la diferencia se amplía aún más: WRF muestra un error un 3 % mayor que ERA5, y la combinación consigue disminuir el RMSE en un 9 %.

En cuanto al coeficiente de determinación (R^2), se aprecia que la media de los modelos con entradas combinadas mejora en 4 % respecto a las configuraciones individuales. En el mejor de los casos, ERA5 obtiene un valor de (R^2) un 3 % superior al de WRF, pero al fusionar ambos conjuntos, se obtiene una mejora adicional de otro 4 %.

Finalmente, las Figuras 5.13, 5.14 y 5.15 muestran la comparación entre los valores reales de producción de potencia eléctrica y las predicciones generadas por los modelos que obtuvieron los mejores resultados en cada configuración: LGBM para WRF, RF para ERA5, y nuevamente LGBM para la combinación. Cada punto

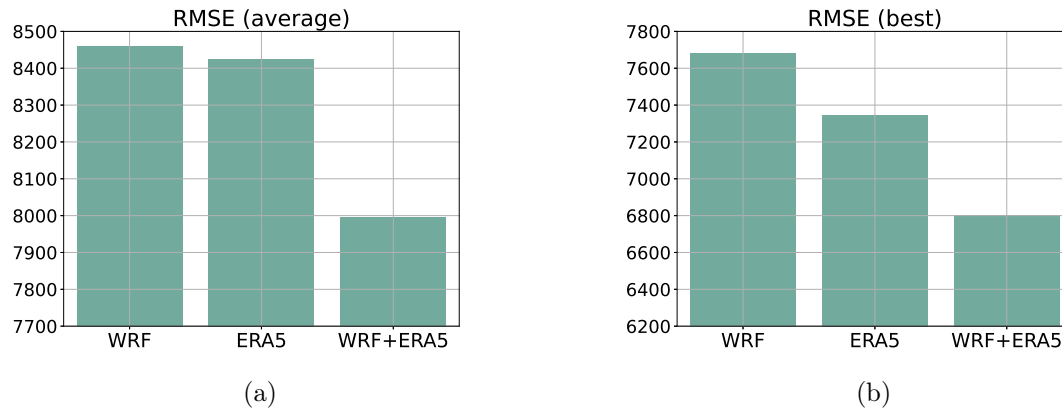


Figura 5.11: Representación gráfica del $RMSE$ (promedio (a) y el mejor (b)) para cada caso de estudio.

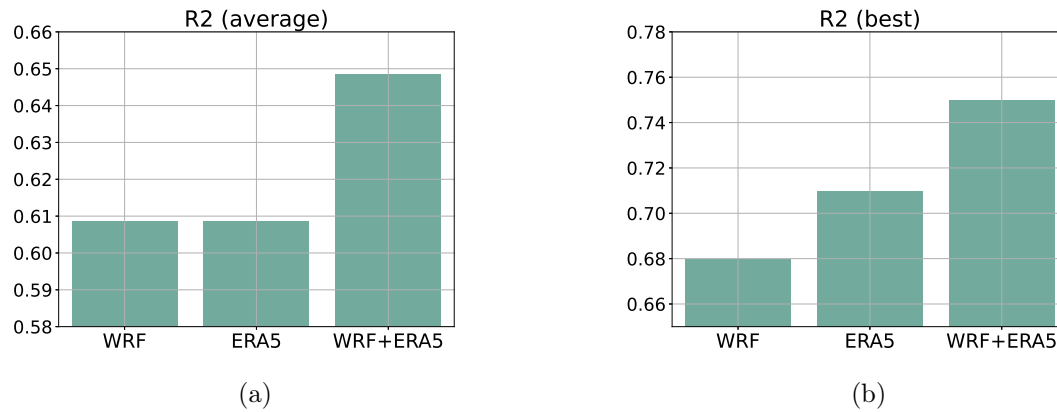


Figura 5.12: Representación gráfica del R^2 (promedio (a) y el mejor (b)) para cada caso de estudio.

en los gráficos representa la producción horaria del 30 % de los datos reservado para validación.

En general, los modelos son capaces de seguir las tendencias de la producción del parque, aunque presentan cierta dificultad para capturar con precisión eventos intensos y rápidos o situaciones en las que se alcanza la capacidad máxima de generación.

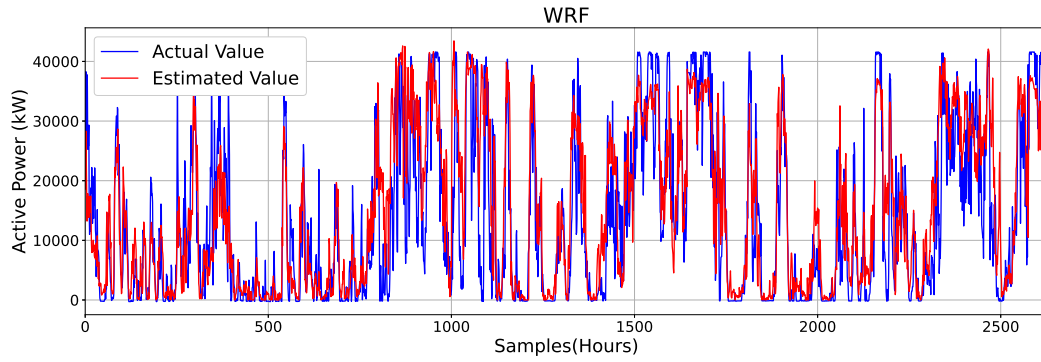


Figura 5.13: Comparación entre los valores reales de potencia del parque eólico y las predicciones generadas por los mejores escenarios para el modelo *LGBM* en cada hora del conjunto de datos de prueba con *WRF*.

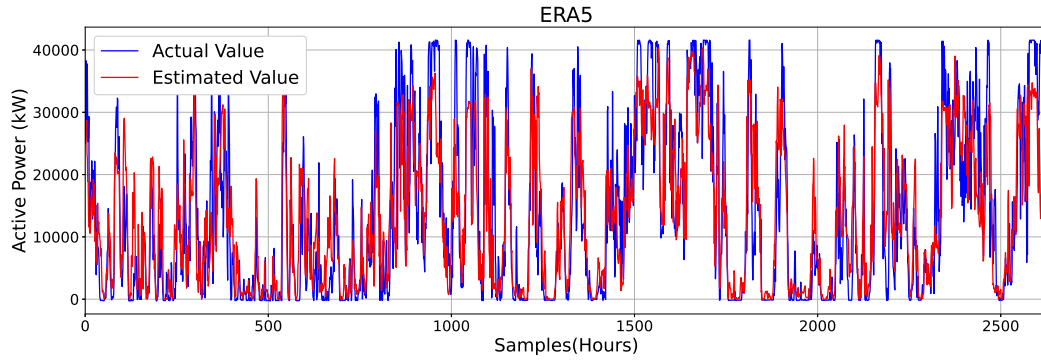


Figura 5.14: Comparación entre los valores reales de potencia del parque eólico y las predicciones generadas por los mejores escenarios para el modelo *RF* en cada hora del conjunto de datos de prueba con *ERA5*.

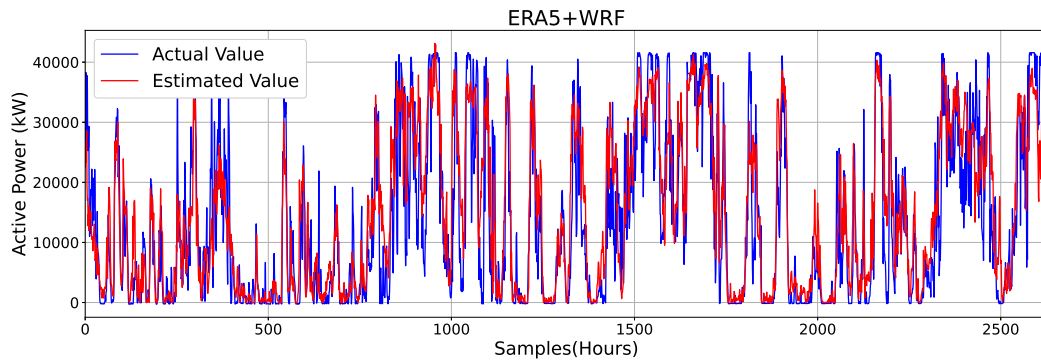


Figura 5.15: Comparación entre los valores reales de potencia del parque eólico y las predicciones generadas por los mejores escenarios para el modelo *LGBM* en cada hora del conjunto de datos de prueba del modelo de mezcla.

Tabla 5.5: Comparisons between the different errors obtained.

WRF				ERA5				ERA5+WRF			
	MAE	RMSE	R2		MAE	RMSE	R2		MAE	RMSE	R2
RT	6178.75	8792.31	0.58	RT	5702.61	8219.74	0.63	RT	6040.88	8686.44	0.59
RF	5603.47	7683.12	0.68	RF	5210.78	7346.91	0.71	RF	4988.88	6803.96	0.75
SVR	6930.67	9227.61	0.54	SVR	6509.47	9230.23	0.54	SVR	6746.17	9023.44	0.56
MLP	6692.16	9422.98	0.52	MLP	7621.40	10793.26	0.37	MLP	6312.39	8758.30	0.58
ELM	6171.63	8210.96	0.63	ELM	6062.45	7999.92	0.65	ELM	6134.34	8332.69	0.62
LSSVR	5757.95	8009.44	0.65	LSSVR	5540.15	7678.94	0.68	LSSVR	5393.73	7448.39	0.70
LGBM	5485.33	7876.25	0.66	LGBM	5535.64	7712.33	0.68	LGBM	4934.27	6929.28	0.74
Average	6117.14	8460.38	0.61	Average	6026.07	8425.90	0.61	Average	5792.95	7997.50	0.65
Best	5485.33	7683.12	0.68	Best	5210.78	7346.91	0.71	Best	4934.27	6803.96	0.75

5.3. Discusión de resultados

La predicción precisa de la generación de energía eólica es fundamental para optimizar el rendimiento y la fiabilidad de los parques eólicos. Una buena estimación permite una mejor integración de la energía renovable en la red eléctrica, reduce los costes operativos y mejora la planificación y gestión de los recursos energéticos. Además, una predicción fiable contribuye a la estabilidad del sistema y, por tanto, favorece la transición hacia un modelo energético más sostenible. Dada la naturaleza variable e intermitente del recurso eólico, el uso de técnicas avanzadas como el aprendizaje automático, combinadas con distintas fuentes meteorológicas, se presenta como una vía prometedora para mejorar la calidad de las predicciones.

En este trabajo se ha propuesto un enfoque de predicción híbrido que combina modelos meteorológicos (mesoescalares y de reanálisis) con técnicas de regresión basadas en aprendizaje automático. Los resultados obtenidos muestran una mejora sustancial en la precisión cuando se utilizan conjuntamente las salidas de WRF y ERA5 como entradas del modelo, en comparación con el uso de cualquiera de ellas por separado. Este hecho refuerza la idea de que ambos conjuntos de datos aportan información complementaria: mientras que ERA5 ofrece una base sólida y coherente derivada de observaciones históricas, WRF añade capacidad predictiva a corto plazo y una mayor resolución espacial y vertical.

El uso de modelos de Machine Learning ha demostrado ser una herramienta eficaz para gestionar y extraer valor de conjuntos de datos meteorológicos complejos y de gran volumen. Estos modelos no solo permiten automatizar el proceso de aprendizaje y ajuste, sino que también son capaces de filtrar y priorizar las variables más relevantes en función de su impacto sobre la producción eólica. Además, una vez entrenados, estos sistemas pueden integrarse fácilmente en flujos operativos, mejorando la reactividad y adaptabilidad de los sistemas de gestión energética.

Otro aspecto importante es la disponibilidad creciente de datos abiertos y gratuitos como ERA5, que facilita el entrenamiento de modelos sin necesidad de infraestructuras propias de observación. Además, el progreso constante en el ámbito del hardware y la computación paralela está reduciendo los tiempos de entrenamiento y permitiendo ejecutar modelos más complejos en menos tiempo, lo que abre nuevas posibilidades para aplicaciones operativas.

La comparación entre los modelos construidos a partir de ERA5 y WRF de forma independiente también ha sido reveladora. Aunque ambos obtienen resultados similares en términos generales, se aprecia una ligera ventaja del modelo basado en ERA5, probablemente debido a la robustez estadística de los datos históricos de reanálisis. No obstante, debe tenerse en cuenta que WRF permite simular condicio-

nes futuras con una latencia mucho menor, lo que lo hace idóneo para aplicaciones en tiempo real. Así, la utilización conjunta de ambos modelos no solo mejora la precisión de las predicciones, sino que permite obtener una solución más equilibrada entre disponibilidad temporal, resolución espacial y capacidad de proyección.

En definitiva, la estrategia propuesta de combinar datos meteorológicos de distinta naturaleza con algoritmos de aprendizaje automático no solo ha demostrado mejorar la precisión de las predicciones horarias de energía eólica en un caso de estudio concreto, sino que también representa un enfoque versátil y escalable aplicable a otros contextos, regiones y horizontes temporales.

Conclusiones

El trabajo presentado en esta memoria de tesis doctoral evidencia la eficacia de los modelos híbridos que combinan simulaciones numéricas de la atmósfera con técnicas de Machine Learning para la predicción precisa de la generación de energía eléctrica mediante tecnología eólica. Esta aproximación permite aprovechar las fortalezas de los modelos físicos, como el WRF y el ERA5, junto con la capacidad de adaptación entornos predictivos de los algoritmos de aprendizaje automático, logrando así una mejora sustancial en la estimación de la producción energética a corto plazo en parques eólicos.

Se ha validado el sistema desarrollado utilizando datos reales de un parque eólico ubicado en La Mudarra, municipio de la provincia de Valladolid (Castilla y León, España), mostrando que los algoritmos de Machine Learning entrenados conjuntamente con datos del WRF y ERA5 superan significativamente en precisión a los entrenados con un único modelo numérico. En particular, los algoritmos Random Forest y LGBM han ofrecido los mejores resultados, alcanzando valores óptimos en métricas como MAE , $RMSE$ y R^2 . Este hecho confirma que el enfoque híbrido no solo mejora la capacidad predictiva, sino que también permite capturar mejor las variaciones locales debidas a la topografía o condiciones atmosféricas particulares.

El modelo ha sido sometido a una validación rigurosa, tanto en el artículo base como en la parte aplicada en el radiosondeo en LEMD, con una doble evaluación diaria durante el mes de julio, considerando diferentes niveles de presión atmosférica. Todo ello ha evidenciado que el modelo WRF es capaz de modelar correctamente el comportamiento del viento en la Península Ibérica, al menos en el interior, y que el comportamiento general del sistema híbrido es robusto y fiable.

Más allá de los buenos resultados obtenidos, una de las principales aportaciones de esta tesis radica en el valor añadido del enfoque híbrido frente a otras estrategias comúnmente empleadas. A diferencia de los modelos puramente físicos, que dependen de una representación exhaustiva de los procesos atmosféricos, o de los modelos puramente estadísticos, que requieren grandes volúmenes de datos históricos homogéneos, la metodología propuesta permite combinar la capacidad explicativa de los

modelos numéricos con la capacidad adaptativa y de corrección de errores sistemáticos propia del Machine Learning. Esto da lugar a un sistema predictivo más equilibrado, capaz de generalizar mejor ante condiciones meteorológicas variables y con un menor sesgo estructural.

Asimismo, este trabajo se enmarca plenamente en el contexto de la transición energética hacia fuentes renovables, donde la variabilidad e incertidumbre inherente a recursos como el viento representa uno de los principales retos para la planificación y operación del sistema eléctrico. En este sentido, las metodologías avanzadas de predicción, como la aquí desarrollada, son fundamentales para facilitar una mayor penetración de la energía eólica en la red, mejorar la fiabilidad del suministro eléctrico y reducir la dependencia de fuentes fósiles. La capacidad de anticiparse con precisión a la producción eólica horaria permite optimizar el despacho energético, disminuir los costes por desvíos y aumentar la eficiencia global del sistema.

Entre las contribuciones principales de esta tesis se encuentran:

- La integración eficaz de modelos numéricos con algoritmos de ensamble como Random Forest o LGBM, lo cual permite una predicción más precisa y menos sensible a la variabilidad del viento.
- La validación detallada de dicha integración en un entorno real, con una resolución temporal horaria y un enfoque específico sobre un parque eólico en operación.
- La comparación directa de los modelos individuales frente al modelo combinado, demostrando cuantitativamente las ventajas del enfoque híbrido.

Asimismo, se han planteado diferentes líneas futuras de trabajo. En primer lugar, el estudio de la reducción de dimensionalidad mediante técnicas de selección de variables podría permitir optimizar aún más los tiempos de entrenamiento y predicción, sin comprometer la calidad de las estimaciones. En segundo lugar, sería interesante aplicar esta metodología a regiones con topografías más complejas —como entornos montañosos, costeros o marinos— tanto para validar la robustez del sistema como para estudiar el efecto de las condiciones locales en el comportamiento del modelo híbrido.

Además, se destaca el potencial del modelo WRF no solo como herramienta predictiva, sino también por su capacidad para abordar una amplia gama de aplicaciones complementarias. Entre ellas, su uso en simulaciones retrospectivas mediante técnicas de asimilación inversa de datos, resulta particularmente prometedor.

Por otro lado, el modelo WRF cuenta con la ventaja de estar diseñado para ejecutarse en arquitecturas de computación en paralelo, utilizando tecnologías como

MPI (Message Passing Interface), lo que abre la puerta a simulaciones más detalladas y eficientes gracias al aprovechamiento de hardware moderno y de alto rendimiento. La mejora en las capacidades computacionales puede traducirse directamente en una mayor resolución espacial y temporal, lo que a su vez refuerza la precisión del sistema híbrido propuesto en esta tesis.

Finalmente, otra línea de trabajo interesante sería la incorporación de técnicas más avanzadas de redes neuronales profundas, como pueden ser las LSTM o modelos híbridos basados en aprendizaje secuencial, lo cual permitiría extender la capacidad de predicción a horizontes temporales más amplios, teniendo en cuenta la evolución temporal del viento y las dependencias dinámicas a lo largo del tiempo.

En conclusión, esta tesis aporta una base sólida para el desarrollo de herramientas de predicción eólica más precisas, robustas y adaptables, que pueden ser de gran utilidad en el contexto de la transición energética hacia fuentes renovables. El enfoque híbrido aquí propuesto representa un paso significativo hacia la mejora operativa de los parques eólicos y la planificación eficiente del sistema eléctrico en escenarios dominados por la variabilidad atmosférica. En este proceso, el modelo WRF no solo destaca por su capacidad para generar predicciones a corto plazo con alta resolución, sino también por su versatilidad como herramienta científica. Su arquitectura modular, escalable y optimizada para la computación en paralelo permite afrontar simulaciones complejas en entornos reales, y abre la puerta a aplicaciones complementarias como el análisis retrospectivo de eventos atmosféricos mediante técnicas de asimilación de datos. Esta flexibilidad convierte al WRF en un aliado estratégico no solo para la predicción, sino también para la investigación de fenómenos meteorológicos pasados y la mejora continua de los modelos híbridos basados en aprendizaje automático.

Bibliografía

- [1] Peter Bauer, Alan Thorpe y Gilbert Brunet. “The quiet revolution of numerical weather prediction”. En: *Nature* 525.7567 (2015), págs. 47-55.
- [2] IREA Irena. “Future of wind: Deployment, investment, technology, grid integration and socio-economic aspects”. En: *Abu Dhabi* (2019).
- [3] Richard S. J. Tol. *The economic impact of weather and climate*. 2021. arXiv: [2102.13110 \[econ.GN\]](https://arxiv.org/abs/2102.13110). URL: <https://arxiv.org/abs/2102.13110>.
- [4] Monica Billio et al. *The impact of Climate on Economic and Financial Cycles: A Markov-switching Panel Approach*. 2020. arXiv: [2012.14693 \[econ.EM\]](https://arxiv.org/abs/2012.14693). URL: <https://arxiv.org/abs/2012.14693>.
- [5] IPCC Climate Change. “The physical science basis”. En: *(No Title)* (2013).
- [6] Gregor Giebel et al. “The state-of-the-art in short-term prediction of wind power: A literature overview”. En: (2011).
- [7] Sergio Martin-Martinez et al. “Wind Power Variability and Singular Events”. En: *Advances in Wind Power*. InTech, nov. de 2012. ISBN: 9789535108634. DOI: [10.5772/52654](https://doi.org/10.5772/52654). URL: <http://dx.doi.org/10.5772/52654>.
- [8] Aoife M Foley et al. “Current methods and advances in forecasting of wind power generation”. En: *Renewable energy* 37.1 (2012), págs. 1-8.
- [9] Organización Meteorológica Mundial. *Clima*. 2025. URL: <https://wmo.int/topics/climate> (visitado 17-04-2025).
- [10] National Centers for Environmental Information (NCEI). *Global Forecast System (GFS)*. Accessed: 2024-11-26. 2024. URL: <https://www.ncei.noaa.gov/products/weather-climate-models/global-forecast>.
- [11] Franco Molteni et al. “The new ECMWF seasonal forecast system (System 4)”. En: *European Centre for Medium-Range Weather Forecasts* 656 (2011), págs. 1-49. URL: <https://www.ecmwf.int/node/9213>.

- [12] Jennifer E Kay et al. “The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability”. En: *Bulletin of the American Meteorological Society* 96.8 (2015), págs. 1333-1349.
- [13] Andreas Klamt. “The COSMO and COSMO-RS solvation models”. En: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 1.5 (2011), págs. 699-709.
- [14] Aleš Farda et al. “Model ALADIN as regional climate model for Central and Eastern Europe”. En: *Studia Geophysica et geodaetica* 54 (2010), págs. 313-332.
- [15] Faith A. Morrison. *An Introduction to Fluid Mechanics*. New York, NY, USA: Cambridge University Press, 2013. ISBN: 978-1-107-00353-8. URL: <https://www.cambridge.org/9781107003538>.
- [16] Paul A. Tipler. “Capítulo 15-4: Temperatura”. En: *Física para la Ciencia y la Tecnología*. 3^a edición. Vol. Tomo 1. Traducción de la obra original: Physics for Scientists and Engineers, Third Edition, publicada por Worth Publishers, Inc., New York. Barcelona, España: Editorial Reverté, S.A., 1992. Cap. 15. ISBN: 84-291-4366-1.
- [17] Markku Rummukainen. “State-of-the-art with regional climate models”. En: *WIREs Climate Change* 1.1 (2010), págs. 82-96. DOI: <https://doi.org/10.1002/wcc.8> eprint: <https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcc.8> URL: <https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcc.8>.
- [18] John L. McGregor. “Recent developments in variable-resolution global climate modelling”. En: *Climatic Change* 129.3 (2015), págs. 369-380. ISSN: 1573-1480. DOI: [10.1007/s10584-013-0866-5](https://doi.org/10.1007/s10584-013-0866-5) URL: <https://doi.org/10.1007/s10584-013-0866-5>.
- [19] PJ Minnett et al. “Half a century of satellite remote sensing of sea-surface temperature”. En: *Remote Sensing of Environment* 233 (2019), pág. 111366.
- [20] Francois Louchet. *How far can we trust climate change predictions?* 2022. arXiv: [2204.11619 \[physics.ao-ph\]](https://arxiv.org/abs/2204.11619) URL: <https://arxiv.org/abs/2204.11619>.
- [21] Hugh Morrison et al. “Confronting the challenge of modeling cloud and precipitation microphysics”. En: *Journal of advances in modeling earth systems* 12.8 (2020), e2019MS001689.

- [22] S George H Philander. “El Nino southern oscillation phenomena”. En: *Nature* 302.5906 (1983), págs. 295-301.
- [23] Mark Z Jacobson. *Fundamentals of atmospheric modeling*. Cambridge university press, 1999.
- [24] AP Van Ulden y GJ Van Oldenborgh. “Large-scale atmospheric circulation biases and changes in global climate model simulations and their importance for climate change in Central Europe”. En: *Atmospheric Chemistry and Physics* 6.4 (2006), págs. 863-881.
- [25] Jun Hu, Julien Emile-Geay y Judson Partin. “Correlation-based interpretations of paleoclimate data – where statistics meet past climates”. En: *Earth and Planetary Science Letters* 459 (2017), págs. 362-371. ISSN: 0012-821X. DOI: <https://doi.org/10.1016/j.epsl.2016.11.048>. URL: <https://www.sciencedirect.com/science/article/pii/S0012821X16306823>.
- [26] Coupled Model Intercomparison Project (CMIP). *Coupled Model Intercomparison Project (CMIP): Advancing climate science through coordinated experiments and model intercomparisons*. <https://www.wcrp-climate.org/wgcm-cmip>. Coordinated by the World Climate Research Programme (WCRP). 2023. URL: <https://www.wcrp-climate.org/wgcm-cmip>.
- [27] Miguel B Araujo et al. “Validation of species–climate impact models under climate change”. En: *Global change biology* 11.9 (2005), págs. 1504-1513.
- [28] Ricardo Vinuesa et al. *Opportunities for machine learning in scientific discovery*. 2024. arXiv: [2405.04161 \[cs.LG\]](https://arxiv.org/abs/2405.04161). URL: <https://arxiv.org/abs/2405.04161>.
- [29] César França. *AI empowering research: 10 ways how science can benefit from AI*. 2023. arXiv: [2307.10265 \[cs.GL\]](https://arxiv.org/abs/2307.10265). URL: <https://arxiv.org/abs/2307.10265>.
- [30] Vladimir M Krasnopolsky, Michael S Fox-Rabinovitz y Dmitry V Chalikov. “New approach to calculation of atmospheric model physics: Accurate and fast neural network emulation of longwave radiation in a climate model”. En: *Monthly Weather Review* 133.5 (2005), págs. 1370-1383.
- [31] Martin G Schultz et al. “Can deep learning beat numerical weather prediction?” En: *Philosophical Transactions of the Royal Society A* 379.2194 (2021), pág. 20200097.
- [32] Sue Ellen Haupt, William P Mahoney y Keith Parks. “Wind power forecasting”. En: *Weather matters for energy*. Springer, 2014, págs. 295-318.

- [33] Tom Beucler et al. *Next-Generation Earth System Models: Towards Reliable Hybrid Models for Weather and Climate Applications*. 2024. arXiv: [2311.13691](https://arxiv.org/abs/2311.13691) [physics.aos-ph]. URL: <https://arxiv.org/abs/2311.13691>.
- [34] Eloy Peña-Asensio et al. “Deep machine learning for meteor monitoring: Advances with transfer learning and gradient-weighted class activation mapping”. En: *Planetary and Space Science* 238 (nov. de 2023), pág. 105802. ISSN: 0032-0633. DOI: [10.1016/j.pss.2023.105802](https://doi.org/10.1016/j.pss.2023.105802). URL: <http://dx.doi.org/10.1016/j.pss.2023.105802>.
- [35] Hans Hersbach et al. “The ERA5 global reanalysis”. En: *Quarterly Journal of the Royal Meteorological Society* 146.730 (2020), págs. 1999-2049.
- [36] Md Saydur Rahman et al. *Comparative Evaluation of Weather Forecasting using Machine Learning Models*. 2024. arXiv: [2402.01206](https://arxiv.org/abs/2402.01206) [cs.LG]. URL: <https://arxiv.org/abs/2402.01206>.
- [37] Zied Ben-Bouallegue et al. *The rise of data-driven weather forecasting*. 2023. arXiv: [2307.10128](https://arxiv.org/abs/2307.10128) [physics.aos-ph]. URL: <https://arxiv.org/abs/2307.10128>.
- [38] Weather Research and Forecasting Model. *Weather Research and Forecasting Model (WRF)*. Accessed: 2024-11-28. 2024. URL: <https://www.mmm.ucar.edu/weather-research-and-forecasting-model>.
- [39] National Center for Atmospheric Research. *National Center for Atmospheric Research*. Accessed: 2024-11-28. 2024. URL: <https://ncar.ucar.edu/>.
- [40] National Oceanic and Atmospheric Administration. *National Oceanic and Atmospheric Administration*. Accessed: 2024-11-28. 2024. URL: <https://www.noaa.gov/>.
- [41] National Centers for Environmental Prediction. *National Centers for Environmental Prediction*. Accessed: 2024-11-28. 2024. URL: <https://www.ncep.noaa.gov/>.
- [42] University Corporation for Atmospheric Research. *University Corporation for Atmospheric Research*. Accessed: 2024-11-28. 2024. URL: <https://www.ucar.edu/>.
- [43] National Weather Service. *National Weather Service*. Accessed: 2024-11-28. 2024. URL: <https://www.weather.gov/>.
- [44] National Hurricane Center. *National Hurricane Center*. Accessed: 2024-11-28. 2024. URL: <https://www.nhc.noaa.gov/>.

- [45] Climate Prediction Center. *Climate Prediction Center*. Accessed: 2024-11-28. 2024. URL: <https://www.cpc.ncep.noaa.gov/>.
- [46] William C Skamarock et al. “A description of the advanced research WRF version 4”. En: *NCAR tech. note ncar/tn-556+ str* 145 (2019).
- [47] Rapid Refresh Forecast System. *Rapid Refresh Forecast System (RAP)*. Accessed: 2024-11-28. 2024. URL: <https://rapidrefresh.noaa.gov/>.
- [48] North American Mesoscale Forecast System. *North American Mesoscale Forecast System (NAM)*. Accessed: 2024-11-28. 2024. URL: <https://www.nco.ncep.noaa.gov/pmb/products/nam/>.
- [49] Jason Otkin, Hung-Lung Huang y Axel Seifert. “A comparison of microphysical schemes in the WRF model during a severe weather event”. En: *Papers delivered at 7th WRF users’ workshop, Boulder, CO, USA*. Citeseer. 2006, págs. 19-22.
- [50] Song-You Hong y Jeong-Ock Jade Lim. “The WRF single-moment 6-class microphysics scheme (WSM6)”. En: *Asia-Pacific Journal of Atmospheric Sciences* 42.2 (2006), págs. 129-151.
- [51] John S Kain y J Michael Fritsch. “Convective parameterization for mesoscale models: The Kain-Fritsch scheme”. En: *The representation of cumulus convection in numerical models*. Springer, 1993, págs. 165-170.
- [52] MICHAEL Tiedtke. “A comprehensive mass flux scheme for cumulus parameterization in large-scale models”. En: *Monthly weather review* 117.8 (1989), págs. 1779-1800.
- [53] Erik Price et al. “GPU-accelerated longwave radiation scheme of the Rapid Radiative Transfer Model for General Circulation Models (RRTMG)”. En: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7.8 (2014), págs. 3660-3667.
- [54] Pedro A Jiménez et al. “A revised scheme for the WRF surface layer formulation”. En: *Monthly weather review* 140.3 (2012), págs. 898-918.
- [55] KBRR Hariprasad et al. “Numerical simulation and intercomparison of boundary layer structure with different PBL schemes in WRF using experimental observations at a tropical site”. En: *Atmospheric Research* 145 (2014), págs. 27-44.
- [56] NOAH Land Surface Model. *NOAH Land Surface Model*. Accessed: 2024-11-28. 2024. URL: <https://ral.ucar.edu/model/noah-multiparameterization-land-surface-model-noah-mp-lsm>.

- [57] Geir Evensen. “The ensemble Kalman filter: Theoretical formulation and practical implementation”. En: *Ocean dynamics* 53 (2003), págs. 343-367.
- [58] NCAR Command Language (NCL). *The NCAR Command Language (NCL)*. Accessed: 2024-11-28. 2024. URL: <https://www.ncl.ucar.edu/>.
- [59] Grid Analysis and Display System (GrADS). *The Grid Analysis and Display System (GrADS)*. Accessed: 2024-11-28. 2024. URL: <http://cola.gmu.edu/grads/>.
- [60] Hunter, John D. *Matplotlib: A 2D Graphics Environment*. Available: <https://matplotlib.org/>, Accessed: 2024-11-28. 2007. DOI: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55).
- [61] PyNGL Development Team. *PyNGL: A Python Language for Scientific Visualization*. Accessed: 2024-11-28. 2024. URL: <https://www.pyngl.ucar.edu/>.
- [62] Message Passing Interface Forum. *Message Passing Interface (MPI)*. Accessed: 2024-11-28. 2024. URL: <https://www.mpi-forum.org/>.
- [63] Holger Blume et al. “OpenMP-based parallelization on an MPCore multiprocessor platform—A performance and power analysis”. En: *Journal of Systems Architecture* 54.11 (2008), págs. 1019-1029.
- [64] Xin Zhang et al. “Development of an efficient regional four-dimensional variational data assimilation system for WRF”. En: *Journal of Atmospheric and Oceanic Technology* 31.12 (2014), págs. 2777-2794.
- [65] Lisa Bengtsson et al. “The HARMONIE–AROME model configuration in the ALADIN–HIRLAM NWP system”. En: *Monthly Weather Review* 145.5 (2017), págs. 1919-1935.
- [66] Krishna K Osuri et al. “Customization of WRF-ARW model with physical parameterization schemes for the simulation of tropical cyclones over North Indian Ocean”. En: *Natural Hazards* 63 (2012), págs. 1337-1359.
- [67] Isidora Jankov et al. “Influence of initial conditions on the WRF–ARW model QPF response to physical parameterization changes”. En: *Weather and Forecasting* 22.3 (2007), págs. 501-519.
- [68] Julien Crétat et al. “Uncertainties in simulating regional climate of Southern Africa: sensitivity to physical parameterizations using WRF”. En: *Climate dynamics* 38 (2012), págs. 613-634.
- [69] Ali Abdolali et al. “Evaluating the accuracy and uncertainty of atmospheric and wave model hindcasts during severe events using model ensembles”. En: *Ocean Dynamics* 71 (2021), págs. 217-235.

- [70] Ghassan J Alaka Jr et al. “Lifetime Performance of the Operational Hurricane Weather Research and Forecasting Model (HWRF) for North Atlantic Tropical Cyclones”. En: *Bulletin of the American Meteorological Society* 105.6 (2024), E932-E961.
- [71] Youngjun Son, Emanuele Di Lorenzo y Jian Luo. “WRF-Hydro-CUFA: A scalable and adaptable coastal-urban flood model based on the WRF-Hydro and SWMM models”. En: *Environmental Modelling & Software* 167 (2023), pág. 105770.
- [72] Andrés Merino et al. “WRF hourly evaluation for extreme precipitation events”. En: *Atmospheric Research* 274 (2022), pág. 106215.
- [73] Dhana Lakshmi Douluri y Arun Chakraborty. “Assessment of WRF-ARW model parameterization schemes for extreme heavy precipitation events associated with atmospheric rivers over West Coast of India”. En: *Atmospheric Research* 249 (2021), pág. 105330.
- [74] Mukesh Kumar et al. “Evaluating the performance of WRF in simulating winds and surface meteorology during a Southern California wildfire event”. En: *Frontiers in Earth Science* 11 (2024), pág. 1305124.
- [75] B. S. Giese et al. “An ensemble of ocean reanalyses for 1815-2013 with sparse observational input”. En: *Journal of Geophysical Research: Oceans* 121 (2016), págs. 6891-6910. DOI: [10.1002/qj.776](https://doi.org/10.1002/qj.776).
- [76] S. Hirahara, M. Ishii e Y. Fukuda. “Centennial-scale sea surface temperature analysis and its uncertainty”. En: *Journal of Climate* 27 (2014), págs. 57-75. DOI: [10.1002/qj.776](https://doi.org/10.1002/qj.776).
- [77] G. P. Compo et al. “The Twentieth Century Reanalysis Project”. En: *Quarterly Journal of the Royal Meteorological Society* 137 (2011), págs. 1-28. DOI: [10.1002/qj.776](https://doi.org/10.1002/qj.776).
- [78] G. P. Compo, J. S. Whitaker y P. D. Sardeshmukh. “Feasibility of a 100 year reanalysis using only surface pressure data”. En: *Bulletin of the American Meteorological Society* 87 (2006), págs. 175-190. DOI: [10.1002/qj.776](https://doi.org/10.1002/qj.776).
- [79] J. S. Whitaker et al. “Reanalysis without radiosondes using ensemble data assimilation”. En: *Monthly Weather Review* 132 (2004), págs. 1190-1200. DOI: [10.1002/qj.776](https://doi.org/10.1002/qj.776).
- [80] Steven Elbert Peckham. “WRF/Chem version 3.3 user’s guide”. En: (2012).

- [81] WR Sessions et al. "An investigation of methods for injecting emissions from boreal wildfires using WRF-Chem during ARCTAS". En: *Atmospheric Chemistry and Physics* 11.12 (2011), págs. 5719-5744.
- [82] Wei Wei et al. "A WRF-Chem model study of the impact of VOCs emission of a huge petro-chemical industrial zone on the summertime ozone in Beijing, China". En: *Atmospheric Environment* 175 (2018), págs. 44-53.
- [83] J Li, Alex Mahalov y P Hyde. "Impacts of agricultural irrigation on ozone concentrations in the Central Valley of California and in the contiguous United States based on WRF-Chem simulations". En: *Agricultural and Forest Meteorology* 221 (2016), págs. 34-49.
- [84] José J Camata et al. "Applying future Exascale HPC methodologies in the energy sector". En: (2016).
- [85] Frederik Kurzrock et al. "A review of the use of geostationary satellite observations in regional-scale models for short-term cloud forecasting". En: *Meteorologische Zeitschrift* 27.4 (2018), págs. 277-298.
- [86] National Centers for Environmental Prediction, National Weather Service, NOAA, U.S. Department of Commerce. *NCEP GFS 0.25 Degree Global Forecast Grids Historical Archive*. Boulder CO, 2015. URL: <https://doi.org/10.5065/D65D8PWK>.
- [87] C. Otte y D. Vidal-Madjar. "Estimation of land surface temperature with NOAA 9 data". En: *Remote Sensing of Environment* 40 (1992), págs. 27-41.
- [88] A. M. Pérez Burgos. "Estimación de la Temperatura Superficial Mediante Imágenes de Satélite Utilizando un Modelo Monocanal". 150 pp. Tesis Doctoral. Universidad de Valladolid, 1994.
- [89] A. Romo. "Localización de bancos de atún rojo en el Mediterráneo mediante Teledetección". 272 pp. Tesis Doctoral. Universidad de Valladolid, 2004.
- [90] Gabriel Cuevas-Figueroa, Peter K. Stansby y Timothy Stallard. "Accuracy of WRF for prediction of operational wind farm data and assessment of influence of upwind farms on power production". En: *Energy* 254 (2022), pág. 124362. ISSN: 0360-5442. DOI: <https://doi.org/10.1016/j.energy.2022.124362>. URL: <https://www.sciencedirect.com/science/article/pii/S0360544222012658>.

- [91] Pedro A. Jiménez y Jimmy Dudhia. “On the Ability of the WRF Model to Reproduce the Surface Wind Direction over Complex Terrain”. En: *Journal of Applied Meteorology and Climatology* 52.7 (2013), págs. 1610-1617. DOI: [10.1175/JAMC-D-12-0266.1](https://doi.org/10.1175/JAMC-D-12-0266.1). URL: <https://journals.ametsoc.org/view/journals/apme/52/7/jamc-d-12-0266.1.xml>.
- [92] Mozhgan Ghazi Mirsaeed, Mojtaba Zoljoodi y Zahra and Seifari. “Verification of WRF model On Accuracy and Precision of Various Schemes and Evaluation of Precipitation Forecast in Iran”. En: *Geographical Researches* 28.2 (2013). eprint: <http://georesearch.ir/article-1-484-en.pdf>. URL: <http://georesearch.ir/article-1-484-en.html>.
- [93] Manju Mohan y Ankur Prabhat Sati. “WRF model performance analysis for a suite of simulation design”. En: *Atmospheric Research* 169 (2016), págs. 280-291. ISSN: 0169-8095. DOI: <https://doi.org/10.1016/j.atmosres.2015.10.013>. URL: <https://www.sciencedirect.com/science/article/pii/S0169809515003348>.
- [94] William Y. Y. Cheng y W. James Steenburgh. “Evaluation of Surface Sensible Weather Forecasts by the WRF and the Eta Models over the Western United States”. En: *Weather and Forecasting* 20.5 (2005), págs. 812-821. DOI: [10.1175/WAF885.1](https://doi.org/10.1175/WAF885.1). URL: https://journals.ametsoc.org/view/journals/wefo/20/5/waf885_1.xml.
- [95] Juan J. Ruiz, Celeste Saulo y Julia Nogués-Paegle. “WRF Model Sensitivity to Choice of Parameterization over South America: Validation against Surface Variables”. En: *Monthly Weather Review* 138.8 (2010), págs. 3342-3355. DOI: [10.1175/2010MWR3358.1](https://doi.org/10.1175/2010MWR3358.1). URL: <https://journals.ametsoc.org/view/journals/mwre/138/8/2010mwr3358.1.xml>.
- [96] Maria Grazia De Giorgi, Antonio Ficarella y Marco Tarantino. “Error analysis of short term wind power prediction models”. En: *Applied Energy* 88.4 (2011), págs. 1298-1311. ISSN: 0306-2619. DOI: <https://doi.org/10.1016/j.apenergy.2010.10.035>. URL: <https://www.sciencedirect.com/science/article/pii/S030626191000437X>.
- [97] Patrícia Tuchtenhagen et al. “WRF model assessment for wind intensity and power density simulation in the southern coast of Brazil”. En: *Energy* 190 (2020), pág. 116341. ISSN: 0360-5442. DOI: <https://doi.org/10.1016/j.energy.2019.116341>. URL: <https://www.sciencedirect.com/science/article/pii/S0360544219320365>.
- [98] D. S. Wilks. *Statistical Methods in Atmospheric Sciences*. 2006.

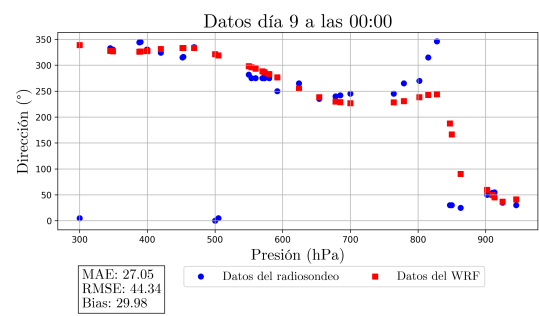
- [99] Alfredo Rodríguez et al. “Artículo Original Pronóstico numérico a corto plazo de la rapidez del viento para los parques eólicos de Gibara I y II Short-term numerical forecast of wind speed for the Gibara I and II wind farms”. En: 28 (jul. de 2022).
- [100] World Meteorological Organization. *Guide to Instruments and Methods of Observation. Volume III – Observing Systems*. WMO-No. 8. Available online. Accessed 2025-04-17. Geneva, Switzerland: WMO, 2023. URL: <https://library.wmo.int/viewer/68661/?offset=>.
- [101] Ian T. Jolliffe y David B. Stephenson, eds. *Forecast Verification: A Practitioner’s Guide in Atmospheric Science*. 2.^a ed. Chichester, UK: Wiley-Blackwell, 2012. ISBN: 9781119960003. DOI: [10.1002/9781119960003](https://doi.org/10.1002/9781119960003). URL: <https://onlinelibrary.wiley.com/doi/book/10.1002/9781119960003>.
- [102] Yudong Tian et al. “Performance Metrics, Error Modeling, and Uncertainty Quantification”. En: *Monthly Weather Review* 144.2 (2016), págs. 607-613. DOI: [10.1175/MWR-D-15-0087.1](https://doi.org/10.1175/MWR-D-15-0087.1). URL: <https://journals.ametsoc.org/view/journals/mwre/144/2/mwr-d-15-0087.1.xml>.
- [103] Álvaro Abad-Santiago et al. “Hybridizing Machine Learning Algorithms With Numerical Models for Accurate Wind Power Forecasting”. En: *Expert Systems* 42.2 (2025). e13830 EXSY-Sep-24-4142.R1, e13830. DOI: <https://doi.org/10.1111/exsy.13830>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/exsy.13830>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/exsy.13830>.
- [104] University Corporation for Atmospheric Research (UCAR). *WRF Preprocessing System (WPS) Geographical Data*. https://www2.mmm.ucar.edu/wrf/users/download/get_sources_wps_geog.html. Accessed: April 28, 2024. 2024.
- [105] Axel Kleidon. “Physical limits of wind energy within the atmosphere and its use as renewable energy: From the theoretical basis to practical implications”. En: *Meteorologische Zeitschrift* 30.3 (jul. de 2021), págs. 203-225. ISSN: 0941-2948. DOI: [10.1127/metz/2021/1062](https://doi.org/10.1127/metz/2021/1062). URL: <http://dx.doi.org/10.1127/metz/2021/1062>.
- [106] Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J-N. “ERA5 hourly data on single levels from 1940 to present.” En: *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)* (2023). URL: <https://doi.org/10.24381/cds.adbb2d47>.

- [107] Jaume Ramon et al. “What global reanalysis best represents near-surface winds?” En: *Quarterly Journal of the Royal Meteorological Society* 145.724 (2019), págs. 3236-3251.
- [108] Rajat Kanti Samal. “Assessment of wind energy potential using reanalysis data: A comparison with mast measurements”. En: *Journal of Cleaner Production* 313 (2021), pág. 127933.
- [109] Wei-Yin Loh. “Classification and regression trees”. En: *Wiley interdisciplinary reviews: data mining and knowledge discovery* 1.1 (2011), págs. 14-23.
- [110] Leo Breiman. “Random forests”. En: *Machine learning* 45.1 (2001), págs. 5-32.
- [111] Mariette Awad et al. “Support vector regression”. En: *Efficient learning machines: Theories, concepts, and applications for engineers and system designers* (2015), págs. 67-80.
- [112] Marius-Constantin Popescu et al. “Multilayer perceptron and neural networks”. En: *WSEAS Transactions on Circuits and Systems* 8.7 (2009), págs. 579-588.
- [113] Jian Wang et al. “A review on extreme learning machine”. En: *Multimedia Tools and Applications* 81.29 (2022), págs. 41611-41660.
- [114] Haifeng Wang y Dejin Hu. “Comparison of SVM and LS-SVM for regression”. En: *2005 International conference on neural networks and brain*. Vol. 1. IEEE. 2005, págs. 279-283.
- [115] Pei Wang, Kabir Nagrecha y Nuno Vasconcelos. “Gradient-based algorithms for machine teaching”. En: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, págs. 1387-1396.

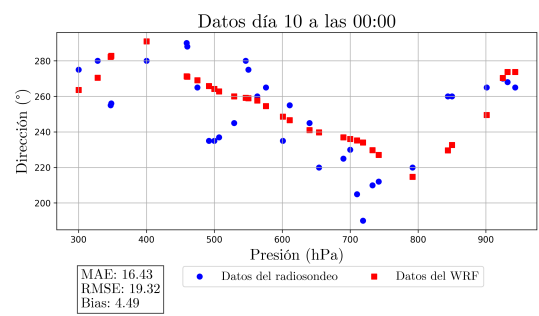
Anexo

Gráficas Validación

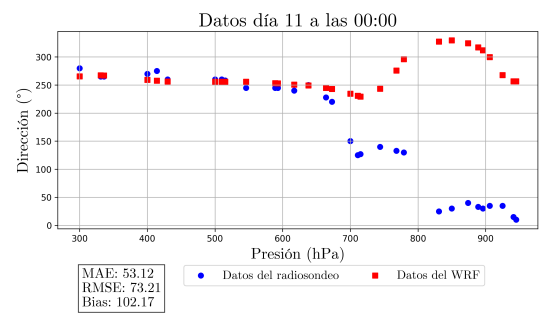
Gráficas por días a las 00:00 para dirección



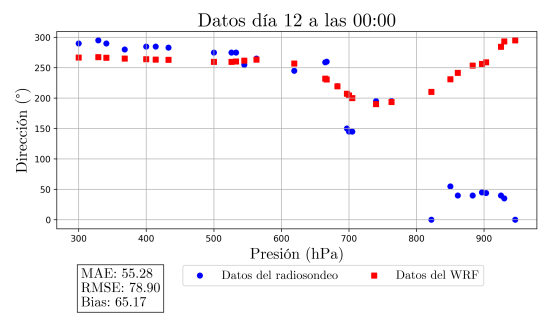
Gráfica para la dirección del día 9



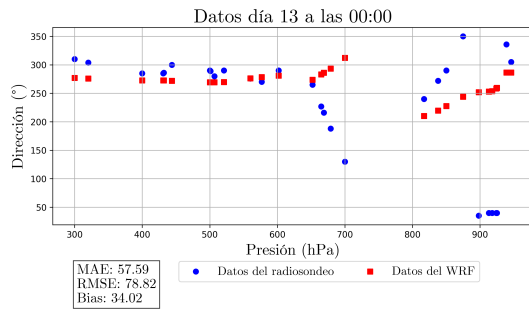
Gráfica para la dirección del día 10



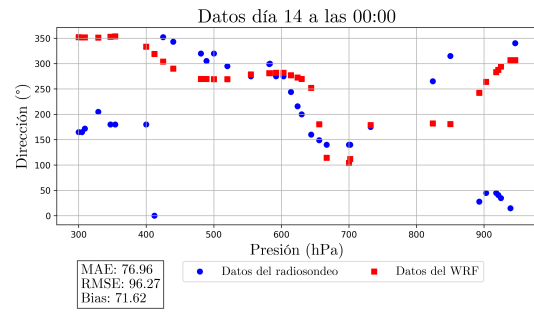
Gráfica para la dirección del día 11



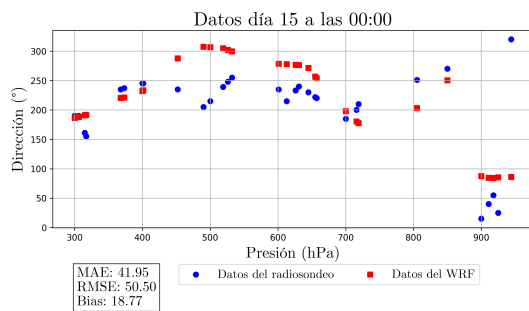
Gráfica para la dirección del día 12



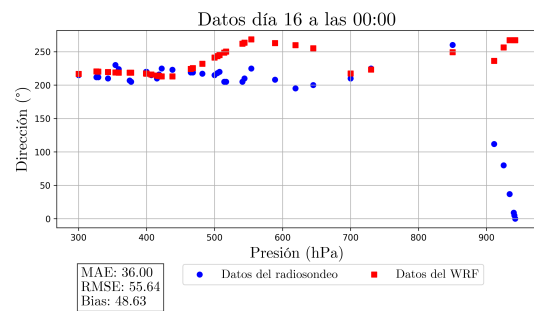
Gráfica para la dirección del día 13



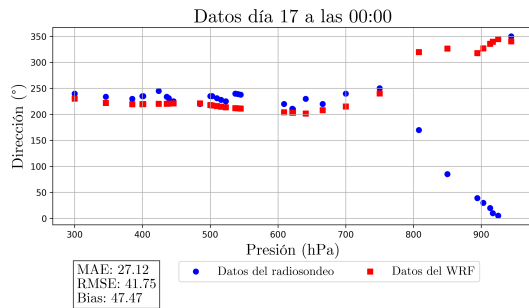
Gráfica para la dirección del día 14



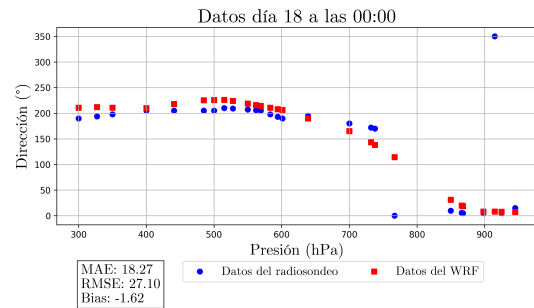
Gráfica para la dirección del día 15



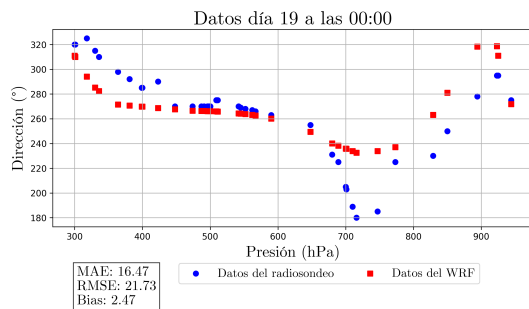
Gráfica para la dirección del día 16



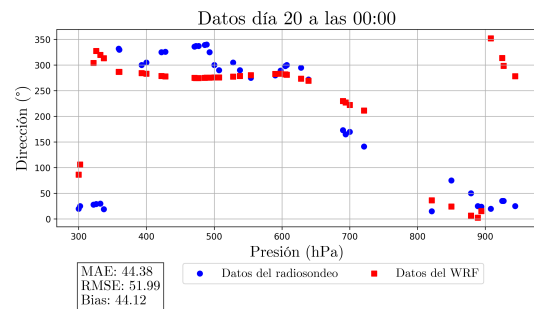
Gráfica para la dirección del día 17



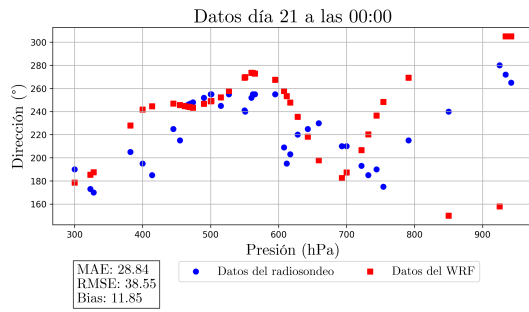
Gráfica para la dirección del día 18



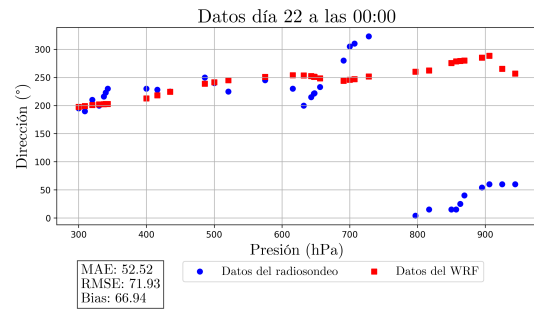
Gráfica para la dirección del día 19



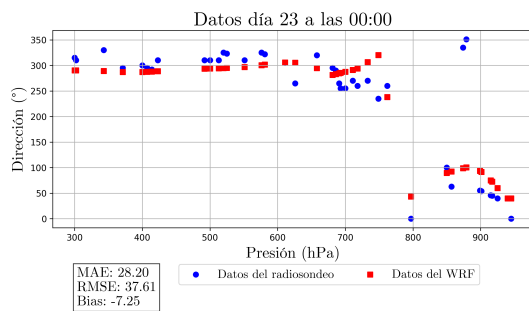
Gráfica para la dirección del día 20



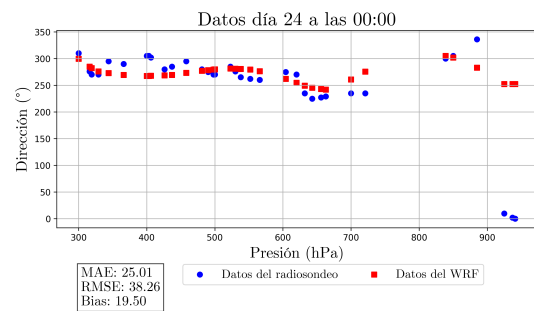
Gráfica para la dirección del día 21



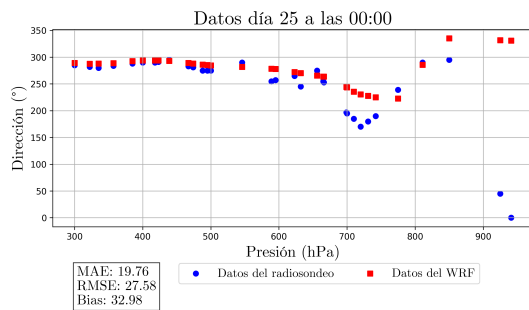
Gráfica para la dirección del día 22



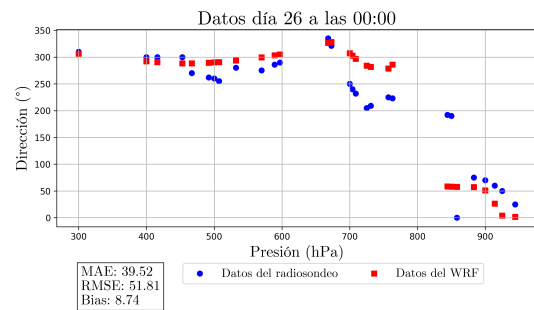
Gráfica para la dirección del día 23



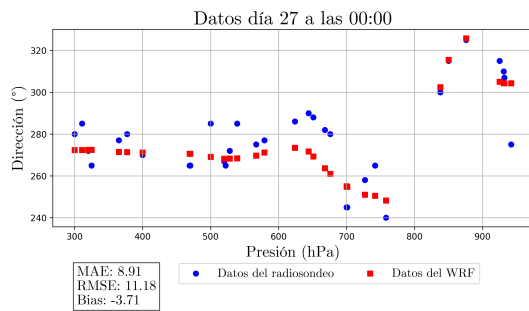
Gráfica para la dirección del día 24



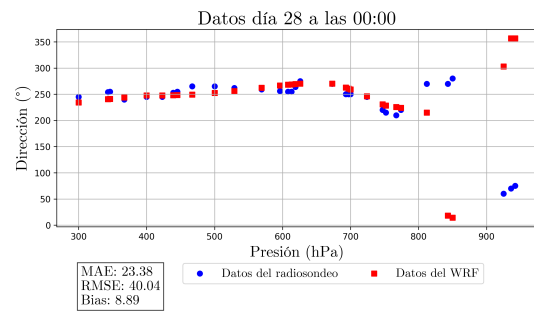
Gráfica para la dirección del día 25



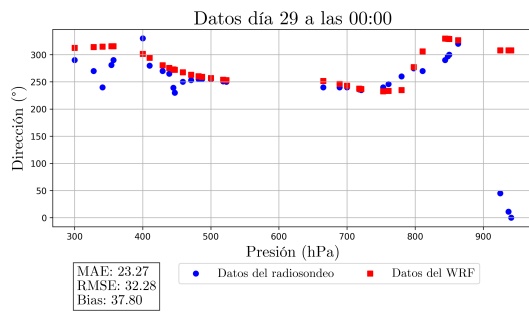
Gráfica para la dirección del día 26



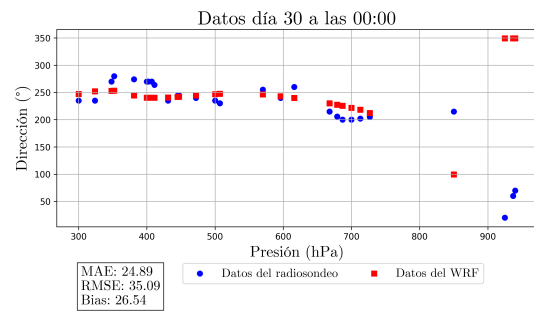
Gráfica para la dirección del día 27



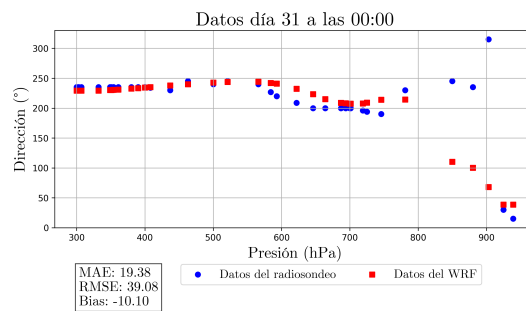
Gráfica para la dirección del día 28



Gráfica para la dirección del día 29

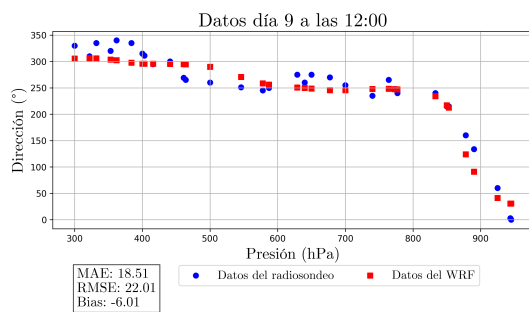


Gráfica para la dirección del día 30

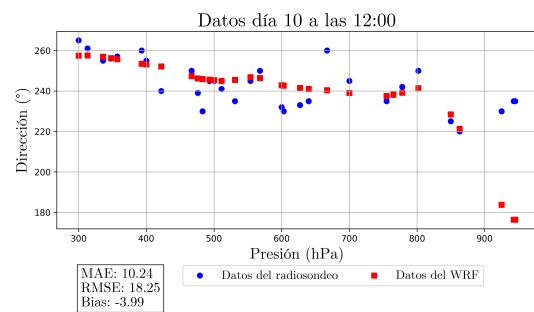


Gráfica para la dirección del día 31

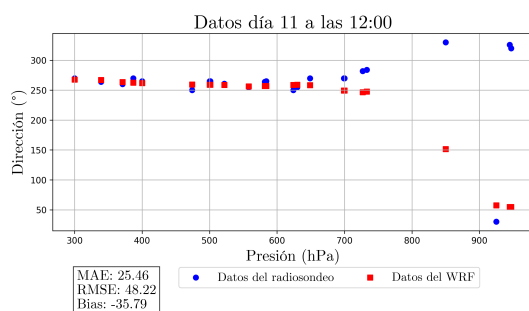
Gráficas por días a las 12:00 para dirección



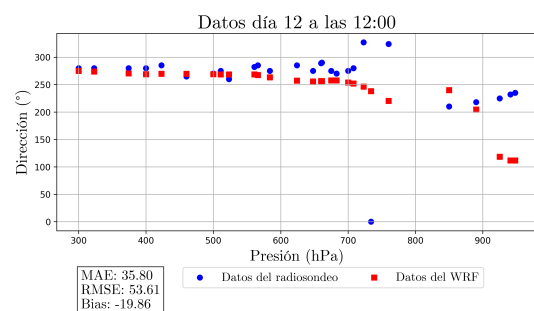
Gráfica para la dirección del día 9



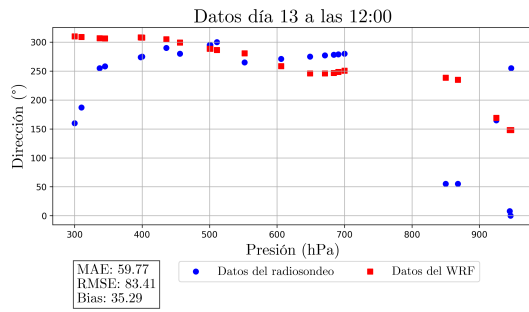
Gráfica para la dirección del día 10



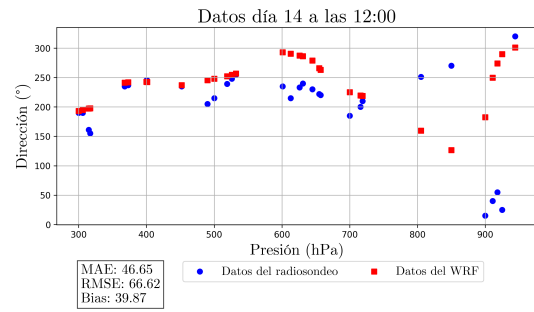
Gráfica para la dirección del día 11



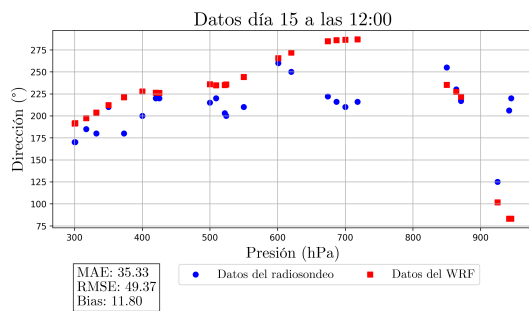
Gráfica para la dirección del día 12



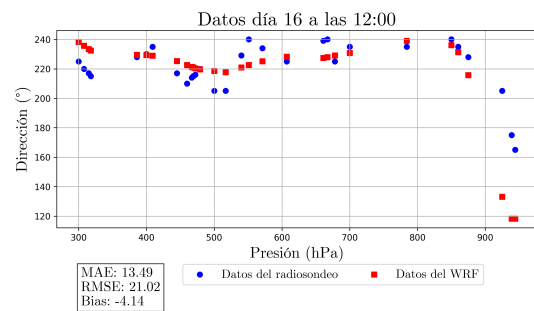
Gráfica para la dirección del día 13



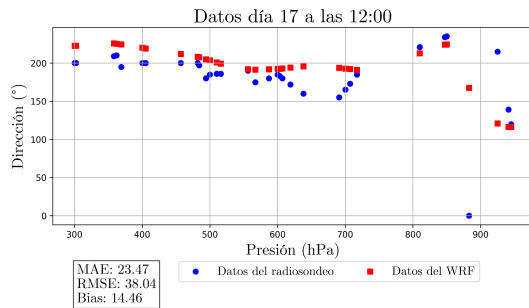
Gráfica para la dirección del día 14



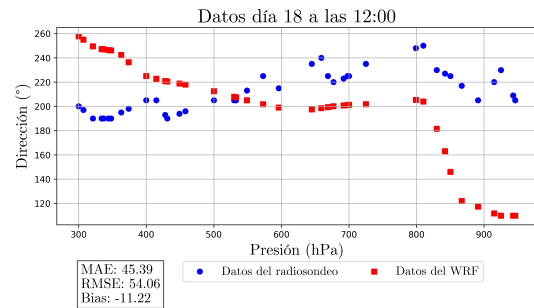
Gráfica para la dirección del día 15



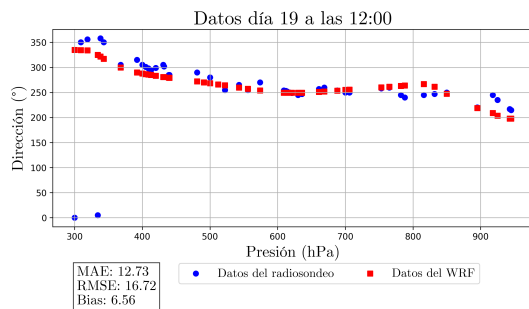
Gráfica para la dirección del día 16



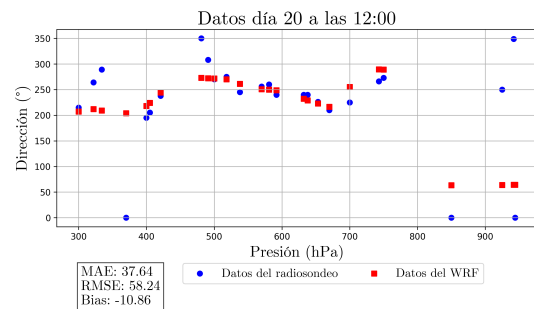
Gráfica para la dirección del día 17



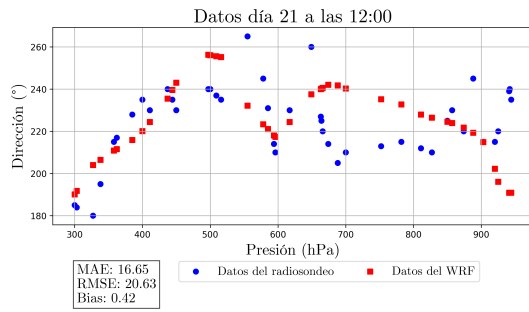
Gráfica para la dirección del día 18



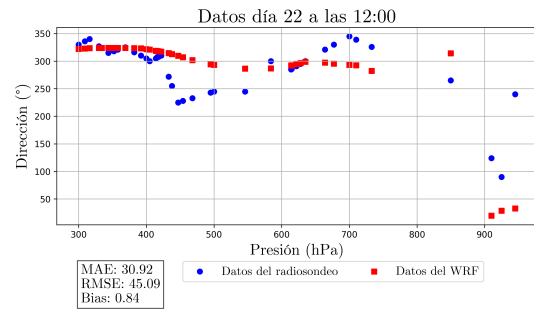
Gráfica para la dirección del día 19



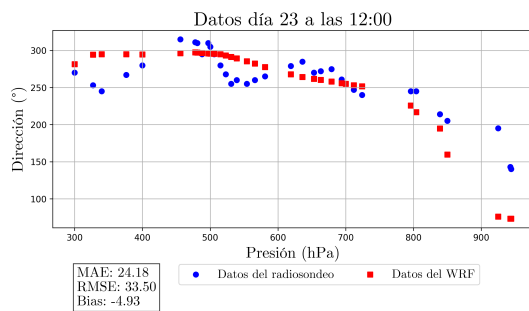
Gráfica para la dirección del día 20



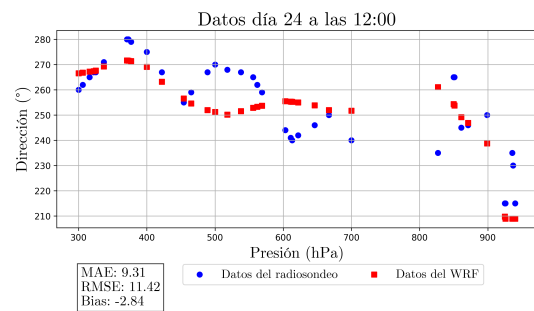
Gráfica para la dirección del día 21



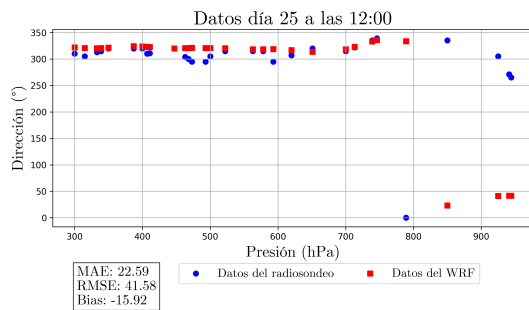
Gráfica para la dirección del día 22



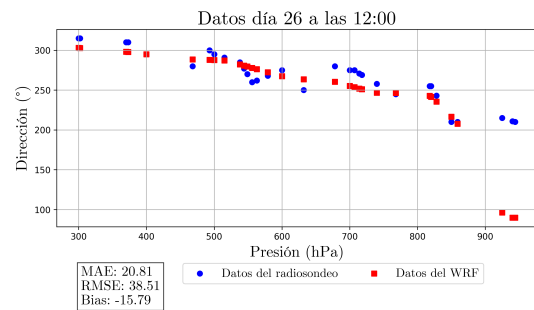
Gráfica para la dirección del día 23



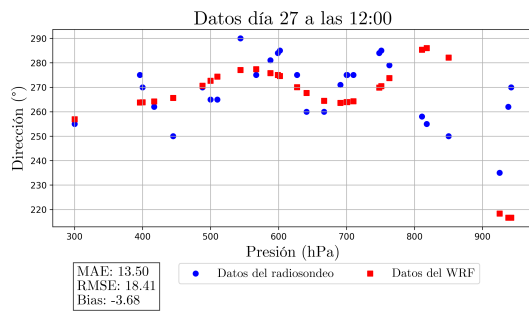
Gráfica para la dirección del día 24



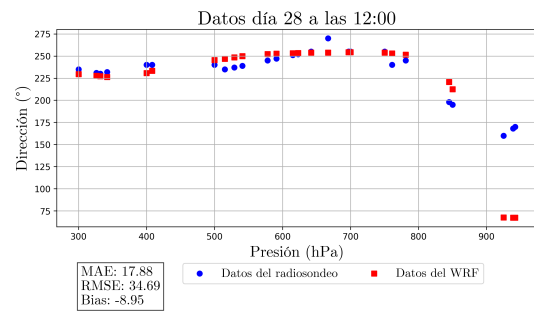
Gráfica para la dirección del día 25



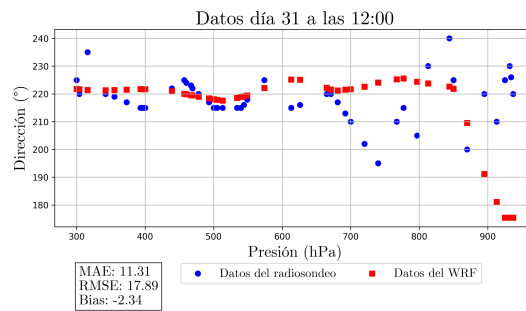
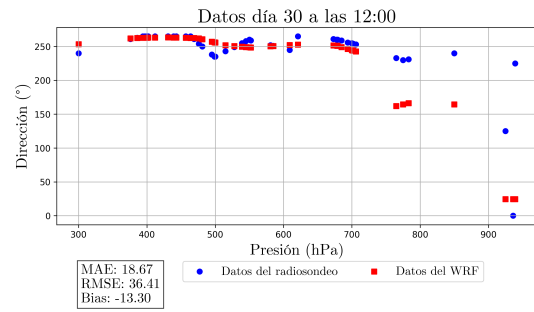
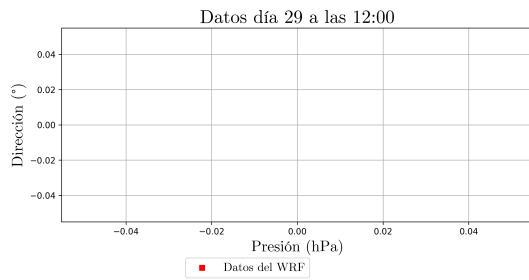
Gráfica para la dirección del día 26



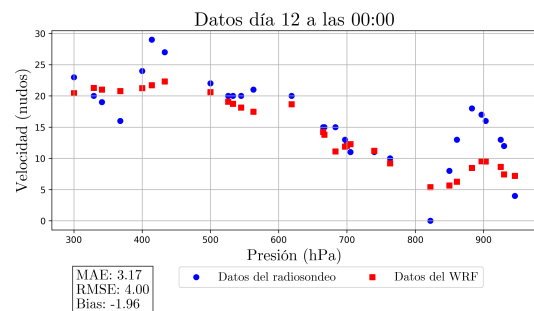
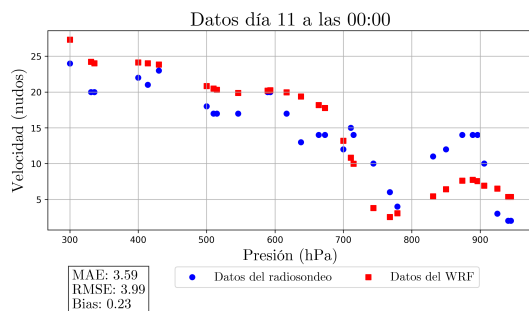
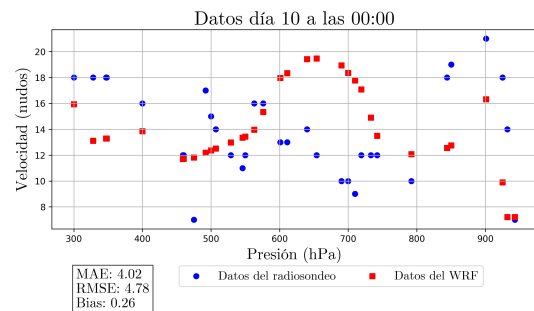
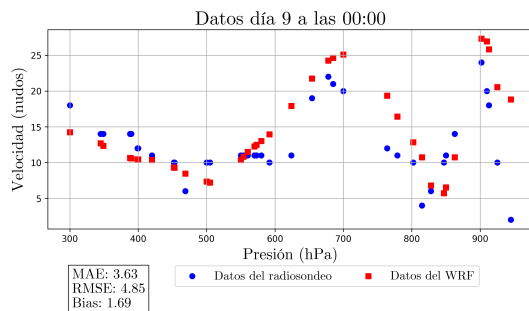
Gráfica para la dirección del día 27

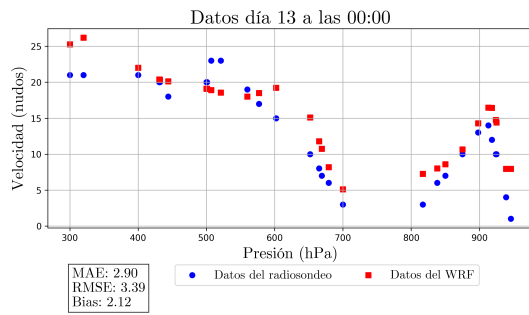


Gráfica para la dirección del día 28

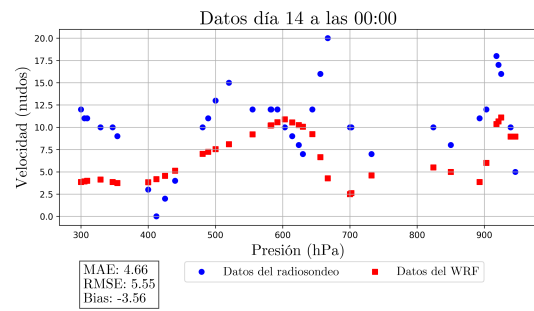


Gráficas por días a las 00:00 para velocidad

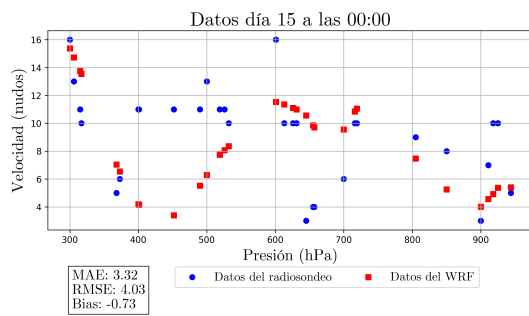




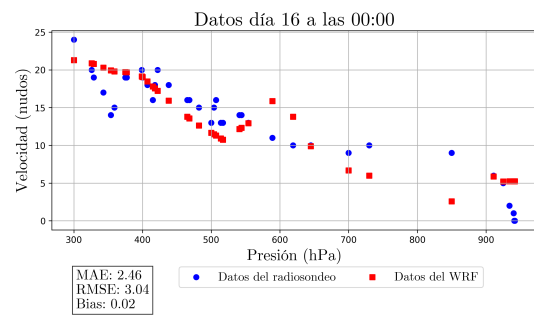
Gráfica para la velocidad del día 13



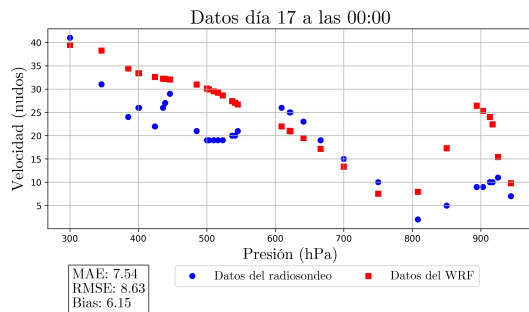
Gráfica para la velocidad del día 14



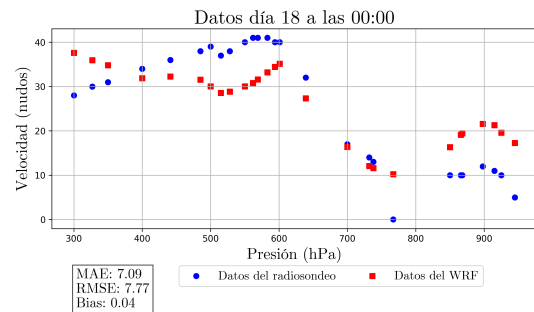
Gráfica para la velocidad del día 15



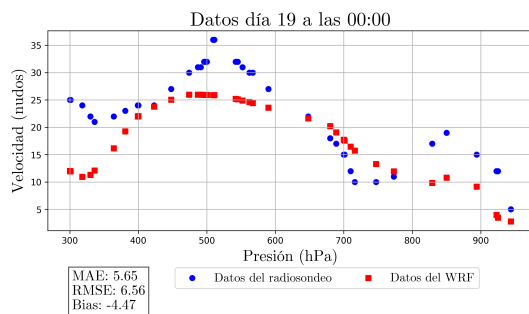
Gráfica para la velocidad del día 16



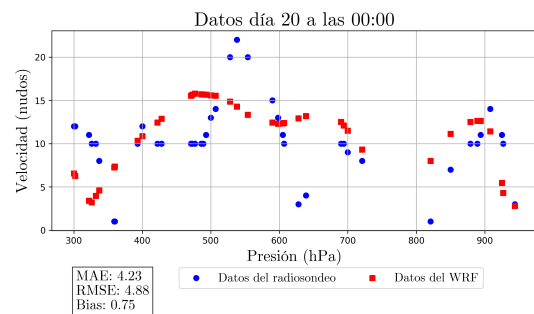
Gráfica para la velocidad del día 17



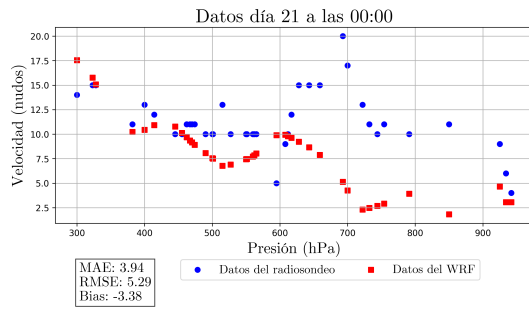
Gráfica para la velocidad del día 18



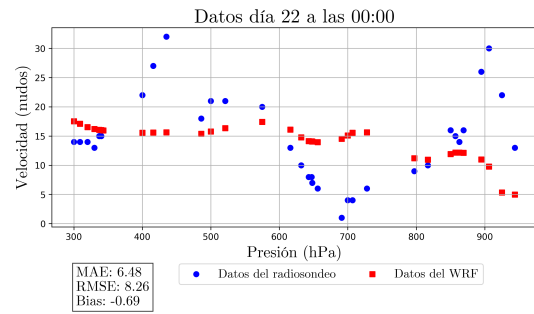
Gráfica para la velocidad del día 19



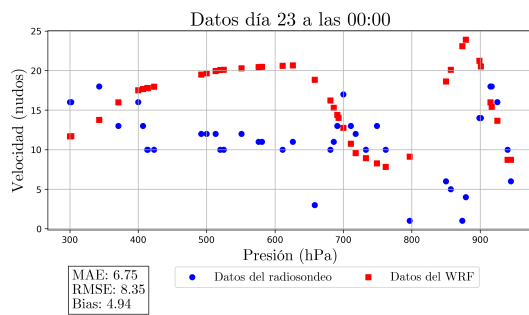
Gráfica para la velocidad del día 20



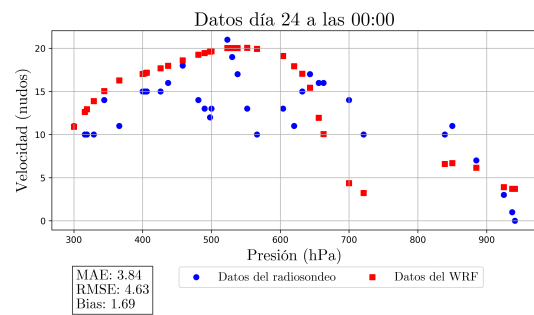
Gráfica para la velocidad del día 21



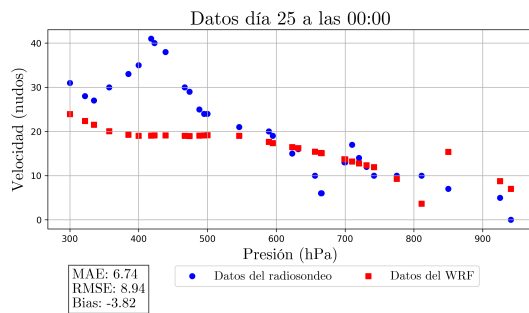
Gráfica para la velocidad del día 22



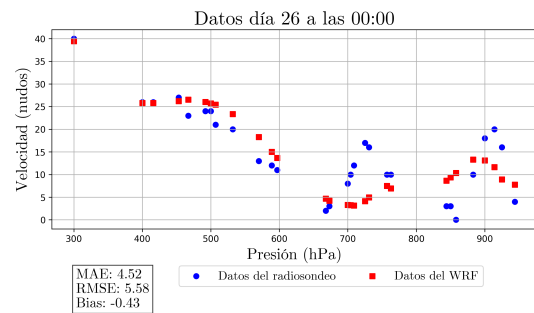
Gráfica para la velocidad del día 23



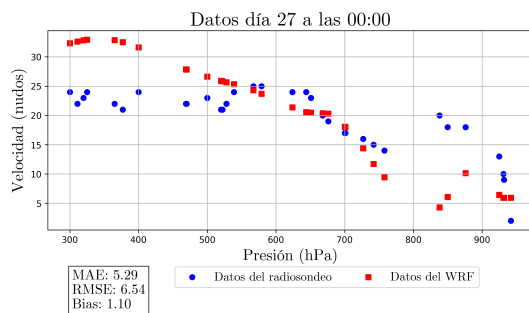
Gráfica para la velocidad del día 24



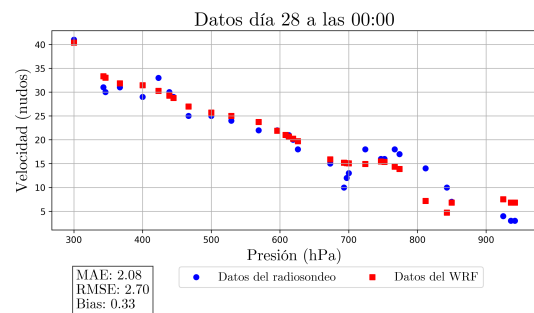
Gráfica para la velocidad del día 25



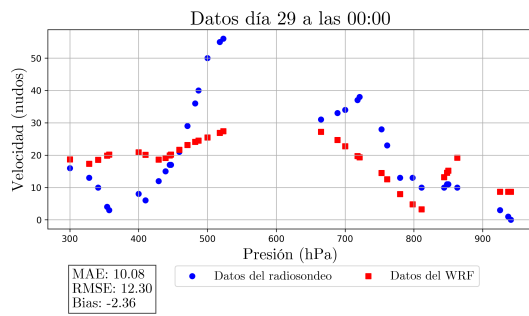
Gráfica para la velocidad del día 26



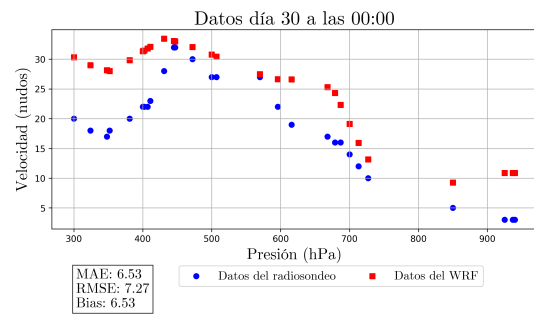
Gráfica para la velocidad del día 27



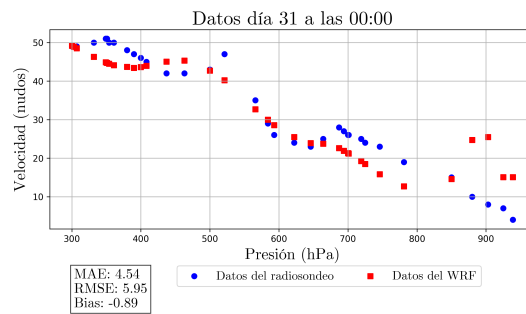
Gráfica para la velocidad del día 28



Gráfica para la velocidad del día 29

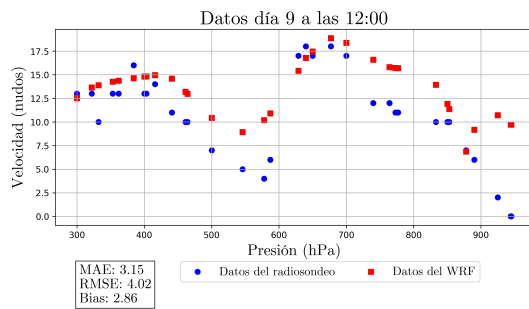


Gráfica para la velocidad del día 30

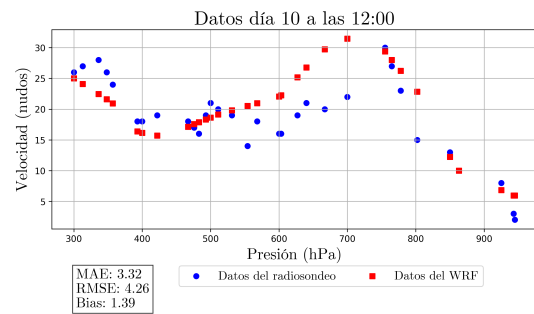


Gráfica para la velocidad del día 31

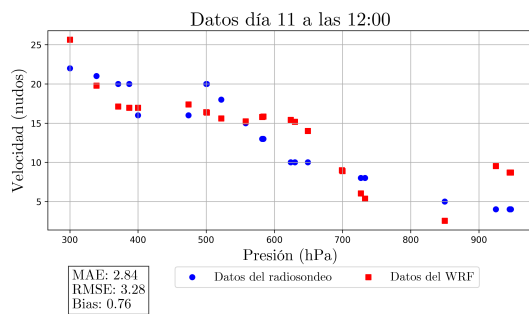
Gráficas por días a las 12:00 para la velocidad



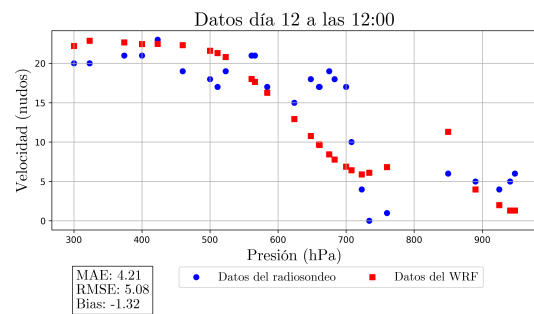
Gráfica para la velocidad del día 9



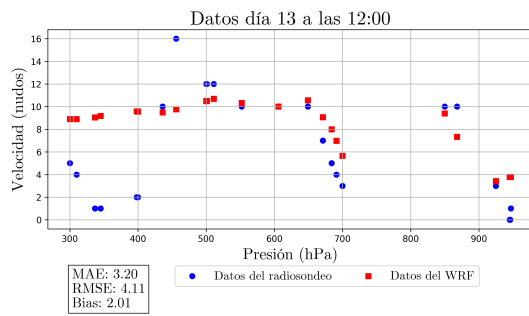
Gráfica para la velocidad del día 10



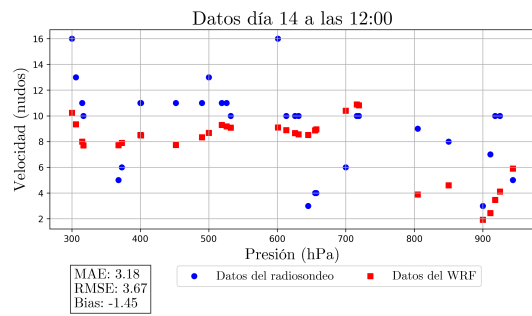
Gráfica para la velocidad del día 11



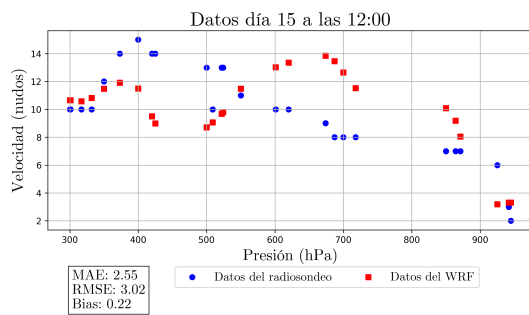
Gráfica para la velocidad del día 12



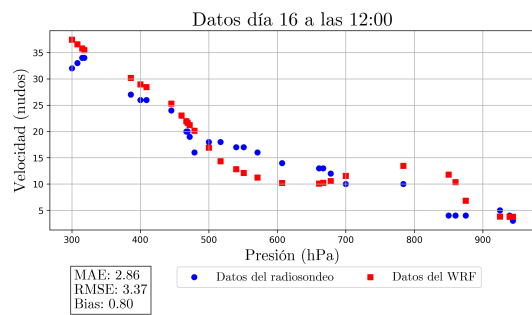
Gráfica para la velocidad del día 13



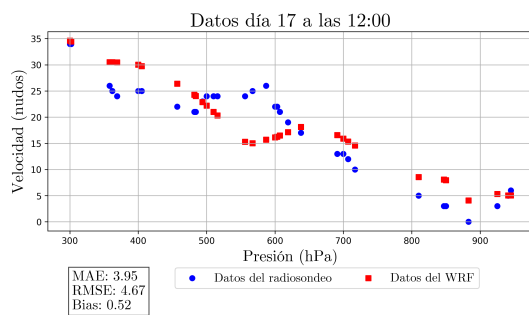
Gráfica para la velocidad del día 14



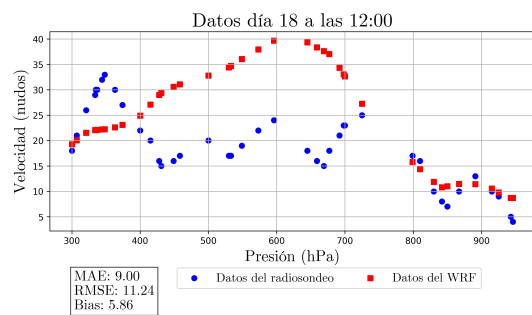
Gráfica para la velocidad del día 15



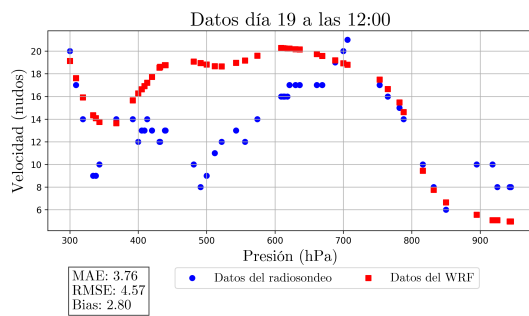
Gráfica para la velocidad del día 16



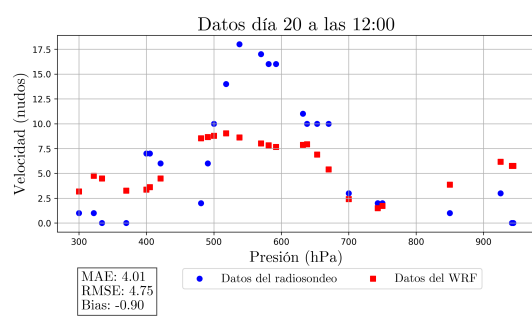
Gráfica para la velocidad del día 17



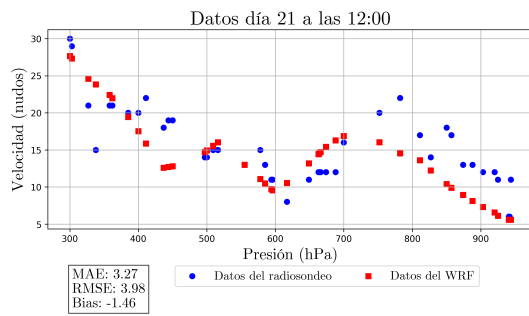
Gráfica para la velocidad del día 18



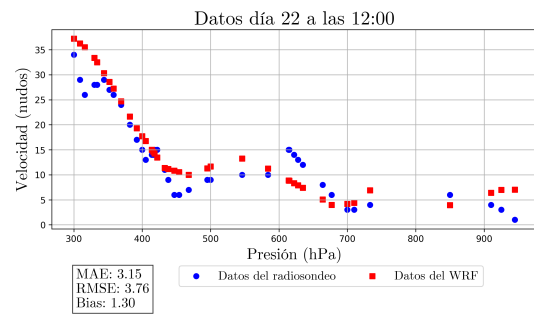
Gráfica para la velocidad del día 19



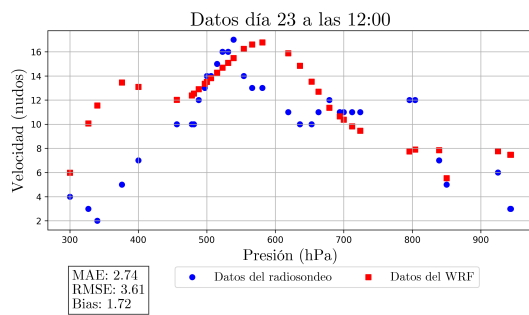
Gráfica para la velocidad del día 20



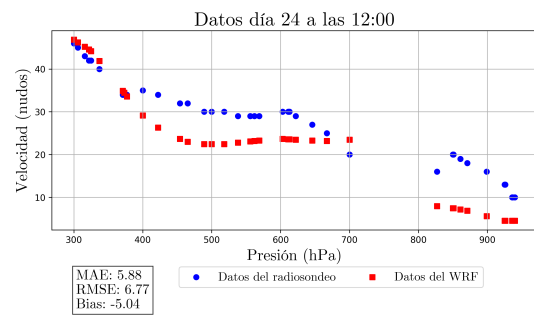
Gráfica para la velocidad del día 21



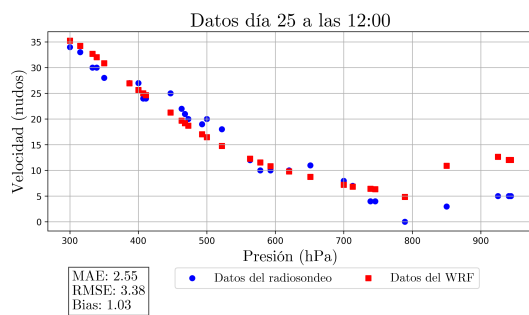
Gráfica para la velocidad del día 22



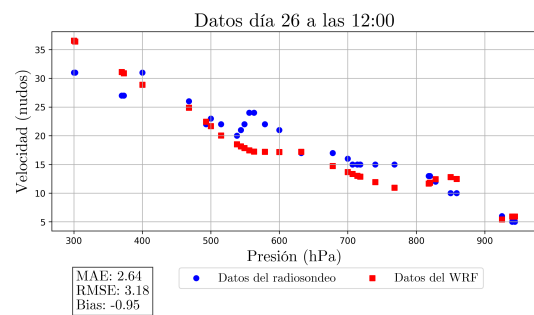
Gráfica para la velocidad del día 23



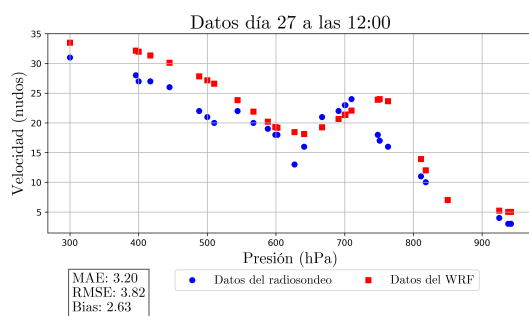
Gráfica para la velocidad del día 24



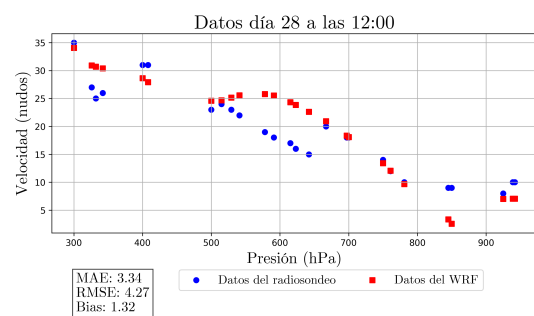
Gráfica para la velocidad del día 25



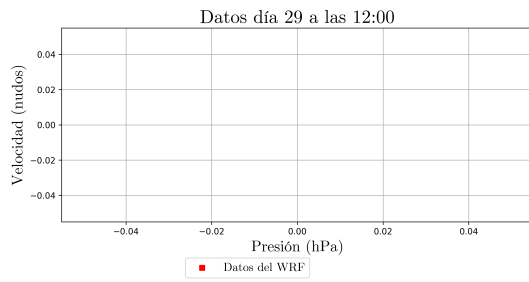
Gráfica para la velocidad del día 26



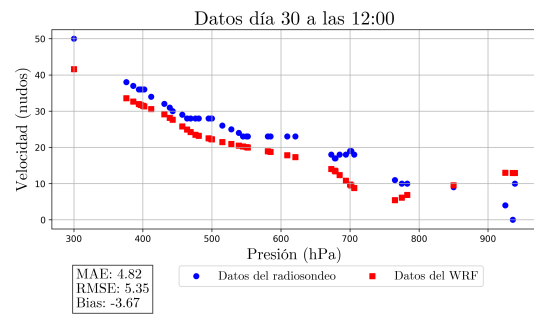
Gráfica para la velocidad del día 27



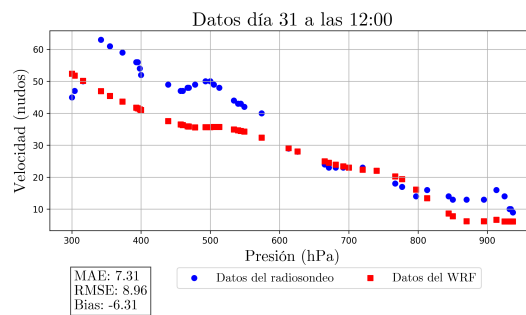
Gráfica para la velocidad del día 28



Gráfica para la velocidad del día 29



Gráfica para la velocidad del día 30



Gráfica para la velocidad del día 31