

*Escuela Técnica Superior de
Ingeniería de Caminos, Canales y Puertos*
UNIVERSIDAD DE CANTABRIA

Desarrollo de un Metamodelo de viento de alta resolución: Aplicaciones de predicción de corto plazo y de estimación de recurso eólico

Trabajo realizado por:
Álvaro Valcarce Ballester

Dirigido:
Paula Camus Braña
Fernando Javier Méndez Incera

Titulación:
**Máster Universitario en
Ingeniería de Caminos, Canales y
Puertos**

Santander, julio de 2025

Resumen

Título: Desarrollo de un Metamodelo de viento de alta resolución: Aplicaciones de predicción de corto plazo y de estimación de recurso eólico.
Autor: Álvaro Valcarce Ballester
Directores: Paula Camus Braña, Fernando Méndez Incera
Titulación: Máster en Ingeniería de Caminos, Canales y Puertos
Palabras clave: Energía, Aerogeneradores, Metamodelos, Predicción, Castríos, Viento.

Introducción

La transición energética global hacia fuentes renovables ha situado a la energía eólica como una de las alternativas más prometedoras en el proceso de descarbonización. En este contexto, la capacidad para predecir con precisión la disponibilidad del recurso eólico se vuelve esencial tanto para la planificación de la producción como para la operación eficiente de los parques eólicos. La variabilidad natural del viento y la complejidad topográfica de muchas localizaciones hacen que los modelos convencionales de predicción presenten limitaciones significativas, especialmente cuando se requiere una resolución temporal y espacial elevada.

Este trabajo se enmarca dentro del esfuerzo por superar esas limitaciones mediante el desarrollo de un metamodelo de viento de alta resolución. El enfoque propuesto combina simulaciones numéricas, reducción de dimensionalidad y técnicas de interpolación multivariable con el objetivo de generar predicciones horarias de viento y potencia con una antelación operativa útil. Particularmente, se centra en el parque eólico de Castríos, ubicado en una zona de alta complejidad orográfica, lo cual supone un desafío ideal para validar la robustez y aplicabilidad del modelo.

A lo largo del documento se detallan los objetivos perseguidos, la caracterización de la zona de estudio, la metodología empleada, los principales resultados obtenidos y las reflexiones derivadas del desarrollo e implementación del modelo. Finalmente, se proponen líneas futuras de investigación que permitirán seguir mejorando la precisión y eficiencia de la herramienta desarrollada.

Objetivos

El principal objetivo de este trabajo es el desarrollo e implementación de un metamodelo de viento de alta resolución que permita estimar, con antelación de entre 24 y 72 horas, la velocidad del viento en un parque eólico de localización compleja como es Castríos. Este modelo se plantea como una herramienta que, mediante técnicas de reducción de dimensionalidad y algoritmos de interpolación

multivariable, logre reproducir campos de viento horarios a partir de predictores de baja resolución como los ofrecidos por los reanálisis ERA5 o los modelos de predicción meteorológica del ECMWF. El carácter híbrido del modelo lo convierte en una opción ágil y eficiente para aplicaciones operacionales, permitiendo reconstrucciones rápidas con una carga computacional contenida.

De forma complementaria, se busca aplicar este metamodelo a la estimación de potencia eólica en cada aerogenerador y en el conjunto del parque, contribuyendo a una mejor planificación de la oferta energética y a la reducción de penalizaciones por desajustes entre la producción esperada y la real.

El objetivo final es avanzar hacia un sistema predictivo operativo capaz de generar series horarias de viento y potencia estimada a escala de parque eólico.

Zona de estudio

La zona de estudio seleccionada para el desarrollo de este trabajo es el parque eólico de Castríos, situado en la comarca de Espinosa de los Monteros, en la provincia de Burgos. Este emplazamiento, de alta complejidad topográfica, representa un entorno ideal para poner a prueba la robustez del metamodelo. Con una altitud media superior a los 1.300 metros y una configuración montañosa que condiciona fuertemente el comportamiento del viento, Castríos presenta un reto adicional a la modelización atmosférica habitual.

El parque cuenta con 24 aerogeneradores de potencia nominal de 1100 kW, cuya distribución a lo largo del relieve genera notables diferencias en la exposición al viento. Esto ha motivado un análisis detallado de la variabilidad espacial y temporal del recurso eólico, así como la necesidad de disponer de predicciones a escala de aerogenerador. Asimismo, se han considerado datos de observaciones locales, simulaciones numéricas y productos de reanálisis para caracterizar de forma integral la zona de estudio.

Metodología

El enfoque metodológico del trabajo se basa en la construcción de un metamodelo híbrido de viento utilizando tres componentes principales: reducción de dimensionalidad por PCA (Principal Component Analysis), selección de casos mediante MDA (Maximum Dissimilarity Algorithm) y reconstrucción de los campos de viento mediante funciones de base radial (RBF).

Primero, se seleccionan días representativos a partir del conjunto de vientos de baja resolución (ERA5), aplicando MDA para maximizar la diversidad del subconjunto. Luego, a través del modelo atmosférico WRF en resoluciones de 1 km (Cantabria) y 100 m (Santander) se obtienen campos horarios de viento que servirán como predictando.

Posteriormente, se aplica PCA sobre los datos de viento de alta resolución para reducir su dimensionalidad y poder reconstruir las series completas usando RBF. Este proceso se entrena con el conjunto representativo y se aplica luego para estimar viento en todo el periodo de estudio. Finalmente, se valida el modelo mediante comparación con datos medidos históricos y se transforma la velocidad del viento en potencia generada usando las curvas de potencia de los aerogeneradores.

Resultados

Los resultados obtenidos permiten distinguir dos aplicaciones clave del metamodelo desarrollado. En primer lugar, se ha logrado generar una reconstrucción del viento a alta resolución espacial en toda la región de Cantabria con una malla de 1 km y, en el caso de Santander, con una resolución aún más detallada de 100 m. Esta capacidad de reconstrucción precisa convierte al metamodelo en una herramienta valiosa para la evaluación del recurso eólico en fases preliminares de diseño y planificación de nuevos parques eólicos en la región.

En segundo lugar, se ha construido un metamodelo adaptado al parque eólico de Castrís que permite realizar predicciones horarias de viento y, en consecuencia, de potencia generada para cada aerogenerador. Las predicciones obtenidas presentan una reconstrucción más acertada del viento respecto a los valores de entrada de los modelos de reanálisis, lo cual se traduce en una estimación más realista de la producción energética del parque. Esta mejora en la precisión supone un avance significativo en el camino hacia una predicción operativa fiable del recurso eólico a escala de parque.

Reflexión

El desarrollo de un metamodelo de viento a alta resolución como el presentado en este trabajo supone una alternativa viable y eficiente frente a simulaciones numéricas completas, cuyo coste computacional resulta inasumible para aplicaciones operacionales diarias. La combinación de técnicas de selección, reducción e interpolación ha demostrado ser adecuada para generar reconstrucciones fidedignas del viento en zonas complejas.

Asimismo, la adaptación del modelo al caso de Castrís ha puesto de manifiesto la importancia de disponer de buenos datos de entrenamiento, así como la necesidad de seguir mejorando en la representación de eventos extremos.

En este sentido, el modelo constituye una herramienta valiosa no solo para predicción, sino también para la reconstrucción de series incompletas o la evaluación del recurso eólico en el pasado.

Líneas futuras de investigación

Entre las posibles extensiones del trabajo destacan las siguientes:

- Incorporar nuevas variables predictoras, como dirección del viento, temperatura o humedad, que podrían mejorar la calidad de las predicciones
- Aplicar el modelo a otras zonas de interés con diferentes características topográficas
- Realizar el proceso con un modelo de alta resolución (100 m) en la zona del parque eólico.
- Automatizar el proceso de predicción y generación de curvas de potencia horarias para su uso operativo por parte de las empresas explotadoras.

En particular, se plantea como siguiente paso inmediato la generación de predicciones completas para una semana entera, incluyendo tanto las series de viento como de potencia horaria por aerogenerador, con el objetivo de ofrecer una herramienta predictiva integral que permita mejorar la gestión diaria del recurso eólico en parques de ubicación compleja.

Abstract

Title: Development of a High-Resolution Wind Metamodel: Applications for Short-Term Forecasting and Wind Resource Assessment

Author: Álvaro Valcarce Ballester

Directors: Paula Camus Braña, Fernando Méndez Incera

Degree: Master's degree in civil engineering

Key words: Energy, Wind Turbines, Metamodels, Forecasting, Castríos, Wind.

Introduction

The global energy transition toward renewable sources has positioned wind energy as one of the most promising alternatives in the decarbonization process. In this context, the ability to accurately predict the availability of the wind resource becomes essential for both production planning and the efficient operation of wind farms. The natural variability of wind and the topographical complexity of many locations result in significant limitations for conventional prediction models, especially when high temporal and spatial resolution is required.

This work is part of an effort to overcome these limitations through the development of a high-resolution wind metamodel. The proposed hybrid approach combines numerical simulations, dimensionality reduction, and multivariable interpolation techniques with the goal of generating hourly wind and power forecasts with operationally useful lead times. The focus is placed particularly on the Castríos wind farm, located in an area of high orographic complexity, making it an ideal setting to validate the robustness and applicability of the model.

Throughout the document, the objectives, the characterization of the study area, the methodology used, the main results obtained, and the reflections derived from the development and implementation of the model are detailed. Finally, future lines of research are proposed that aim to continue improving the accuracy and efficiency of the developed tool.

Objectives

The main objective of this work is the development and implementation of a high-resolution wind metamodel that enables the estimation of wind speed 24 to 72 hours in advance at a wind farm located in a complex terrain such as Castríos. This model is conceived as a tool that, through dimensionality reduction techniques and multivariable interpolation algorithms, can reconstruct hourly wind fields from low-resolution predictors such as those provided by ERA5 reanalysis or ECMWF weather forecasting models. The hybrid nature of the model makes it a flexible and

efficient option for operational applications, allowing for fast reconstructions with a contained computational load.

In addition, the aim is to apply this metamodel to the estimation of wind power at each wind turbine and for the entire wind farm, contributing to improved energy supply planning and reducing penalties due to discrepancies between expected and actual production. The ultimate goal is to move toward an operational predictive system capable of generating hourly series of wind and estimated power at wind farm scale.

Study Area

The selected study area for the development of this work is the Castríos wind farm, located in the region of Espinosa de los Monteros, in the province of Burgos. This site, with high topographical complexity, represents an ideal environment to test the robustness of the metamodel. With an average elevation exceeding 1,300 meters and a mountainous configuration that strongly influences wind behavior, Castríos presents an additional challenge for traditional atmospheric modeling.

The wind farm consists of 24 wind turbines with a nominal power of 1100 kW each, distributed along the terrain in such a way that significant differences in wind exposure arise. This has prompted a detailed analysis of the spatial and temporal variability of the wind resource, as well as the need for turbine-scale forecasts. Local observations, numerical simulations, and reanalysis products have been considered to comprehensively characterize the study area.

Methodology

The methodological approach of this work is based on the construction of a hybrid wind metamodel using three main components: dimensionality reduction through PCA (Principal Component Analysis), case selection using MDA (Maximum Dissimilarity Algorithm), and reconstruction of wind fields using Radial Basis Functions (RBF).

First, representative days are selected from the set of low-resolution wind data (ERA5), applying MDA to maximize the diversity of the subset. Then, using the WRF atmospheric model at resolutions of 1 km (Cantabria) and 100 m (Santander), hourly wind fields are obtained that will serve as the target variables.

Next, PCA is applied to the high-resolution wind data to reduce its dimensionality and reconstruct the complete series using RBF. This process is trained with the representative subset and then applied to estimate wind conditions across the entire study period. Finally, the model is validated by comparing its output with historical measurement data, and wind speed is transformed into generated power using the wind turbines' power curves.

Results

The results allow us to distinguish two key applications of the developed metamodel. First, a high-resolution spatial reconstruction of wind has been successfully generated across the entire Cantabria region with a 1 km mesh, and for Santander, with an even finer resolution of 100 m. This reconstruction capability makes the metamodel a valuable tool for assessing the wind resource during early stages of design and planning for new wind farms in the region.

Second, a metamodel tailored to the Castríos wind farm has been built, allowing for hourly forecasts of wind and, consequently, power generation for each wind turbine. The obtained forecasts show a more accurate reconstruction of the wind compared to the input values from reanalysis models, resulting in a more realistic estimate of the park's energy production. This improvement in precision represents a significant step toward a reliable operational forecast of wind resources at the wind farm scale.

Reflection

The development of a high-resolution wind metamodel such as the one presented in this work represents a viable and efficient alternative to full numerical simulations, whose computational cost is prohibitive for daily operational applications. The combination of selection, reduction, and interpolation techniques has proven suitable for generating reliable wind reconstructions in complex areas.

Moreover, adapting the model to the case of Castríos has highlighted the importance of having high-quality training data, as well as the need to continue improving the representation of extreme events. In this sense, the model serves as a valuable tool not only for forecasting but also for reconstructing incomplete series or assessing historical wind resource availability.

Future Lines of Research

Possible extensions of this work include:

- Incorporate new predictor variables, such as wind direction, temperature, or humidity, which could improve the quality of the forecasts.
- Applying the model to other areas of interest with different topographical characteristics.
- Implementing the process using a high-resolution (100 m) model in the wind farm zone.
- Automating the forecasting and hourly power curve generation process for operational use by energy companies.

As an immediate next step, the goal is to generate complete forecasts for an entire week, including both wind and hourly power series per turbine, to offer an

integrated predictive tool for improving the daily management of wind resources in complex terrain wind farms.

Índice

Resumen	2
Abstract	6
1. Introducción	16
1.1 Relevancia	16
1.2 Objetivos	18
2. Sitio de estudio	19
2.1 Parque eólico Castríos	20
2.1.1. Características geográficas	20
2.1.2. Características mecánicas	21
2.1.3. Datos históricos	23
3. Métodos y Herramientas	25
3.1 Bluemath-tk	25
3.2 Técnicas de Data Science: MDA, PCA, RBF	25
3.2.1. MDA (Maximum Dissimilarity Algorithm)	25
3.2.2. PCA (Principal Component Analysis)	28
3.2.3. RBF (Radial Basis Functions)	29
3.3 Cluster Computacional	29
4. Datos históricos	31
4.1 ECMWF Reanalysis v5: ERA5	32
4.2 Boya Augusto G. Linares	32
4.3 Red Vigía Virgen del mar	33
4.4 CMT	34
4.5 Parayas	35
4.6 Castríos	37
4.7 IFS - ECMWF's	37
5. Metamodelos	39
5.1 Calibración del metamodelo	39
5.1.1. Selección de casos representativos de vientos diarios LR	41
5.1.2. Modelización numérica de los vientos HR a partir de los casos de viento LR	43
5.1.3. Construcción y ajuste del metamodelo	44
5.1.4. Conclusiones	45
5.2 Evaluación de la precisión del metamodelo	45
5.2.1. Evaluación en la estación Augusto G. Linares	45
5.2.2. Evaluación de la estación de Virgen del Mar	47

5.2.3.	Evaluación de la estación de CMT.....	50
5.2.4.	Evaluación de la estación de Parayas	53
5.3	Estudio de la Rugosidad del terreno.....	56
5.4	Parametrización del emplazamiento del parque eólico	59
5.4.1.	Rugosidad del parque eólico	59
5.4.2.	Comparativa entre el metamodelo y los datos históricos.....	62
6.	Predicción	70
6.1	Predicción de viento	71
6.2	Predicción de potencia	74
6.2.1.	Desarrollo de red neuronal	79
6.2.2.	Resultados	81
7.	Conclusiones y futuras líneas de investigación	86
7.1	Conclusiones.....	86
7.2	Futuras líneas de investigación	87
8.	Bibliografía.....	89

Índice de figuras

Ilustración 1. Modelos climáticos globales (ERA5), que cubren un periodo de más de 80 años (1940-2024). Observaciones en el parque eólico de Castríos limitadas a 10 años de datos (2014-2024).	19
Ilustración 2. Parque Eólico Castríos. Fuente: Ayuntamiento Espinosa de los monteros	20
Ilustración 3. Ubicación de los aerogeneradores. Fuente: CASTRIOS S.L.	21
Ilustración 4. Especificaciones Técnicas de los aerogeneradores. Fuente: Wind Power.....	22
Ilustración 5. Curva de generación de energía. Fuente: Wind Power	22
Ilustración 6. Variabilidad espacial de las velocidades medias (panel izquierdo) y de las velocidades máximas (panel derecho) registradas durante enero de 2022 en los aerogeneradores del parque eólico de Castríos.....	23
Ilustración 7. Velocidad media percibida por cada aerogenerador.	24
Ilustración 8. Maximum dissimilarity selection. Fuente: Camus et al. (2011).....	27
Ilustración 9. Ubicación de las estaciones con registros históricos de viento.	31
Ilustración 10. Datos históricos AGL.	33
Ilustración 11. Datos históricos VM.....	34
Ilustración 12. Datos históricos CMT.	34
Ilustración 13. Datos históricos Parayas.	35
Ilustración 14. Serie histórica de datos en Santander.....	36
Ilustración 15. Datos históricos Aerogenerador 11.	37
Ilustración 16. Descripción del modelo y su entrenamiento. Fuente: ECMWF (European Centre for Medium-Range Weather Forecasts)	38
Ilustración 17. Puntos de interés.....	40
Ilustración 18. Procedimiento aplicado.	41
Ilustración 19. Varianza acumulada.	41
Ilustración 20. Representación del viento de baja resolución real (izquierda), asumido (central) y su diferencia (derecha).....	42
Ilustración 21. Casos MDA considerados (izquierda) y sus correspondientes vientos espaciales para los 16 primeros casos (derecha).....	43
Ilustración 22. Campos horarios sobre el dominio de Cantabria (resolución 1 km) y Santander (resolución 100 m).....	44
Ilustración 23. Comparativa entre un día de fuertes vientos reales(izquierda), simulados(centro) y error(derecha).....	44
Ilustración 24. Correlación entre ERA5-AGL de manera horaria (izquierda) y diaria (derecha).....	46
Ilustración 25. Comparativa ERA5-AGL diaria para el primer trimestre de 2014..	46
Ilustración 26. Dirección del viento medida y la estimada por ERA5.	47
Ilustración 27. Correlación entre ERA5-VM de manera horaria (izquierda) y diaria (derecha).....	47
Ilustración 28. Datos históricos VM - ERA5.....	48
Ilustración 29. Comparativa Metamodelo - Datos Históricos VM.....	48
Ilustración 30. Metamodelo - Datos Históricos VM. Primer trimestre de 2013	49
Ilustración 31. Dirección del viento medida en campo (VM) y la predicha por el metamodelo	49

Ilustración 32. Correlación entre ERA5-CMT de manera horaria (izquierda) y diaria (derecha).....	50
Ilustración 33. Datos Históricos CMT - ERA5.....	51
Ilustración 34. Comparativa Metamodelo - Datos Históricos CMT.	51
Ilustración 35. Metamodelo - Datos Históricos CMT. Primer trimestre 2020.	52
Ilustración 36. Dirección del viento medida en campo (CMT) (izquierda) y la predicha por el metamodelo (derecha).	52
Ilustración 37. Correlación horaria ERA5 – Parayas.....	53
Ilustración 38. Datos Históricos Parayas - ERA5.....	54
Ilustración 39. Comparativa Metamodelo - Datos Históricos Parayas.....	54
Ilustración 40. Metamodelo - Comparación Histórica Parayas. Primer trimestre 2000	55
Ilustración 41. Dirección del viento medida en campo (Parayas) (izquierda) y la predicha por el metamodelo (derecha).	55
Ilustración 42. Comparación de los campos de viento generados por distintos modelos en un instante representativo. Arriba a la izquierda: viento interpolado de ERA5. Arriba a la derecha: reconstrucción con el metamodelo de 1 km en la región de Cantabria. Abajo a la izquierda: zoom de esa misma reconstrucción en la zona de Santander. Abajo a la derecha: reconstrucción con el metamodelo de 100 m. ..	57
Ilustración 43. Rugosidad modelo a 100 m resolución (izquierda) y de 1 km de resolución (derecha).	58
Ilustración 44. Topografía a la izquierda y Rugosidad a la derecha del parque eólico "Castríos".....	59
Ilustración 45. Malla de 1 km de resolución con los elementos del parque eólico "Castríos".....	59
Ilustración 46. Gráfica del viento según elevación. Fuente: Suisse Éole	60
Ilustración 47. Perfiles logarítmicos para el coeficiente de rugosidad obtenido.....	62
Ilustración 48. Aerogenerador con punto del modelo de referencia.....	63
Ilustración 49. Comparativa diaria HR 1km-Aerogeneradores Grupo 1.....	64
Ilustración 50. Comparativa horaria en las dos primeras semanas de 2014 en el Grupo 1.	65
Ilustración 51. Comparativa diaria HR 1km-Aerogeneradores Grupo 2.....	66
Ilustración 52. Comparativa horaria en las dos primeras semanas de 2014 en el Grupo 2.	67
Ilustración 53. Comparativa diaria HR 1km-Aerogenerador Grupo 3.....	67
Ilustración 54. Comparativa horaria en las dos primeras semanas de 2014 en el Grupo 3.	68
Ilustración 55. Comparativa diaria HR 1km-Aerogenerador Grupo 4.....	69
Ilustración 56. Comparativa horaria en las dos primeras semanas de 2014 en el Grupo 4.	69
Ilustración 57. Comparativa entre el metamodelo original de 1km de resolución y el independiente para cada aerogenerador.....	71
Ilustración 58. Comparación de medidas reales y predichas del 19 al 22 de junio.	73
Ilustración 59. Velocidad del viento real vs predicha por los diferentes predictores del 23 al 29 de junio.....	74

Ilustración 60. Potencia real vs predicha por los diferentes modelos en la semana del 23 al 29 de junio.....	78
Ilustración 61. Potencia generada vs predicha por los modelos de manera diaria.....	79
Ilustración 62. Curva de potencia elaborada en base a las técnicas de Machine Learning.....	81
Ilustración 63. Potencia real vs predicha por los diferentes modelos en la semana del 23 al 29 de junio con modelo de red neuronal.....	84
Ilustración 64. Potencia generada vs la predicha de manera diaria con modelo de red neuronal.	85
Ilustración 65. Producción del 25 de junio a las 12h.....	85

Índice de tablas

Tabla 1. Coeficiente de rugosidad.....	61
--	----

1. Introducción

1.1 Relevancia

En la actualidad, una parte muy significativa de la energía producida en España proviene de fuentes renovables; una tendencia en constante crecimiento impulsada por las políticas de emisión cero predominantes. Como consecuencia, los parques eólicos, tanto terrestres como marinos, están adquiriendo una relevancia cada vez mayor al generar una energía limpia y económica[1].

Los parques eólicos, en particular aquellos situados en zonas montañosas, son altamente sensibles a las variaciones del viento, lo que puede ocasionar, en plazos de tiempo reducidos, pasar de operar a plena capacidad a detenerse por razones de seguridad. Esta situación conlleva un impacto económico significativo para las empresas explotadoras, ya que estos parques deben pronosticar con 24 horas de antelación la energía que generarán.

El mercado diario de electricidad, conocido como Acoplamiento Único Diario (SDAC, por sus siglas en inglés), permite la realización de transacciones eléctricas para las 24 horas del día siguiente mediante ofertas de compra y venta. Este mercado, acoplado con Europa desde 2014, determina diariamente los precios y volúmenes de energía a las 12:00 CET, en función de la oferta y la demanda. Los precios varían según la capacidad de interconexión entre zonas, lo que puede ocasionar diferencias tarifarias si la capacidad resulta insuficiente. En este contexto, la predicción de la energía desempeña un papel crucial, ya que permite anticipar las fluctuaciones en la oferta y la demanda, optimizando así la gestión y planificación del sistema eléctrico. En particular, en el sector eólico, la estimación precisa de estas variaciones del viento representa un desafío significativo para los operadores de parques eólicos, dado que cualquier desviación respecto a la producción comprometida puede dar lugar a penalizaciones económicas [2].

En los últimos años, el avance en la comprensión de la variabilidad del viento ha impulsado el desarrollo de diversas soluciones destinadas a optimizar la gestión de la energía eólica. Estas incluyen no solo nuevos protocolos y sistemas innovadores de recopilación de datos, sino también modelos numéricos avanzados que permiten una mejor predicción de la producción energética. Además, la mejora en los sistemas de predicción de la energía eólica es una consecuencia directa de los avances en las capacidades de modelado atmosférico. En conjunto, estos desarrollos conforman un enfoque integral para la gestión de los parques eólicos, combinando métodos científicos de vanguardia con aplicaciones prácticas que permiten optimizar la producción y minimizar los riesgos económicos asociados a las fluctuaciones del viento [3].

Los modelos numéricos de predicción, como los desarrollados por ECMWF o AEMET [4] constituyen herramientas fundamentales para la estimación del viento y, en consecuencia, de la producción energética. Sin embargo, estos modelos suelen ofrecer resoluciones espaciales que no son suficientes para aplicaciones a escala de parque eólico.

Para solventar esta limitación, existen distintas técnicas de downscaling que permiten aumentar la resolución espacial de las predicciones de viento. El downscaling dinámico, basado en modelos numéricos anidados, proporciona resultados precisos, pero requiere un elevado coste computacional. Como alternativa, el downscaling estadístico se basa en establecer relaciones estadísticas entre vientos de baja resolución y vientos locales de alta resolución a partir de datos históricos. No obstante, esta técnica depende en gran medida de la disponibilidad de dichos datos históricos. Una tercera aproximación es el downscaling híbrido, que combina técnicas estadísticas o de aprendizaje automático con simulaciones numéricas. Este enfoque, basado en metamodelos, ha sido tradicionalmente utilizado en la predicción de fenómenos como mareas y corrientes y representa una opción prometedora para mejorar las predicciones de viento a escala local.

Por tanto, el objetivo de este Trabajo de Fin de Máster (TFM) es desarrollar un metamodelo para la predicción del viento en un parque eólico, con especial énfasis en lograr una alta resolución espacial. Este aspecto es especialmente relevante en regiones montañosas, donde la orografía compleja y la rugosidad del terreno suponen un reto adicional para la modelización atmosférica.

Las metodologías híbridas, ampliamente empleadas en el ámbito de la oceanografía operacional, han demostrado ser eficaces en la mejora de predicciones en entornos costeros complejos. En particular, en el grupo GeoOcean se han desarrollado metamodelos que combinan simulaciones numéricas con técnicas estadísticas y de aprendizaje automático, como es el caso de HySwash, un modelo híbrido para la predicción de procesos de oleaje en zonas someras[5].

Estas metodologías suelen seguir una secuencia de pasos bien definida:

- 1- Selección de datos históricos de vientos a baja resolución
- 2- Identificación de casos representativos, que incluyan eventos extremos poco frecuentes
- 3- Simulación numérica de dichos casos mediante modelos avanzados de alta resolución
- 4- Reconstrucción de la variable de respuesta en series completas utilizando técnicas de reducción de dimensionalidad e interpolación, como Análisis de Componentes Multivariantes (MDA) o funciones de base radial (RBF).

El objetivo de este Trabajo de Fin de Máster (TFM) es adaptar y aplicar este enfoque híbrido al ámbito de la predicción eólica, con el desarrollo de un metamodelo específico para la estimación del viento a escala de parque eólico. Dado que se trata

de un modelo novedoso, no se dispone de antecedentes directos que empleen esta metodología. No obstante, el uso conjunto de técnicas como el Algoritmo de selección basado en la Máxima Disimilitud (MDA) y las Funciones de Base Radial (RBF), inspiradas en los desarrollos previos del grupo GeoOcean, constituye una aportación original y relevante dentro de este campo de estudio.

La combinación de estas metodologías permitirá el desarrollo de metamodelos capaces de aproximar los resultados de simulaciones dinámicas con un esfuerzo computacional significativamente menor. Esto facilita la predicción eficiente de procesos atmosféricos, lo que contribuirá a mejorar la estimación del viento en una determinada región.

1.2 Objetivos

El objetivo principal de este proyecto es desarrollar un metamodelo de alta resolución que permita predecir el viento en el emplazamiento de un parque eólico a lo largo de un periodo de tiempo con una antelación de 24, 36, 48 y 72 horas. El parque eólico de Castrís, situado en el norte de la provincia de Burgos ha sido seleccionado como el parque de estudio. La razón para elegir este parque radica, por un lado, por el reto que supone la predicción meteorológica en un emplazamiento complejo de alta montaña y, por otro, por la posibilidad de acceder a sus registros históricos, lo cual facilitará la validación de la metodología en el corto plazo.

Los objetivos específicos son:

- Desarrollo de un metamodelo para la reconstrucción de vientos a alta resolución. Este desarrollo se estructura en dos fases diferenciadas:
 - i) Calibración del metamodelo a partir de datos históricos representativos.
 - ii) Validación mediante la comparación entre los resultados obtenidos y observaciones independientes, con el objetivo de evaluar su capacidad predictiva.
- Aplicación del metamodelo para la predicción del viento a alta resolución en la ubicación del parque eólico. Una vez calibrado y validado, el metamodelo se emplea para generar series temporales de viento con alta resolución espacial en el emplazamiento específico del parque.
- Aplicación de las predicciones de viento a la estimación de la potencia eólica generada por el parque. A partir de las series de viento obtenidas, se estiman los valores de producción energética, lo que permite evaluar el recurso eólico disponible y su aprovechamiento en términos de generación eléctrica.

Estos objetivos buscan mejorar la comprensión de la variabilidad del viento en entornos complejos y proporcionar una herramienta práctica para la predicción de la producción eólica, optimizando así la gestión de los parques eólicos y minimizando los riesgos económicos asociados a las fluctuaciones del viento.

2. Sitio de estudio

Para el desarrollo de este TFM, se plantea la construcción de un metamodelo que relacione datos de viento a alta resolución (predictando) con datos a baja resolución espacial (predictor), provenientes de reanálisis atmosféricos globales. Este metamodelo tiene como objetivo mejorar la representación local del recurso eólico en zonas donde los datos observacionales son escasos o discontinuos.

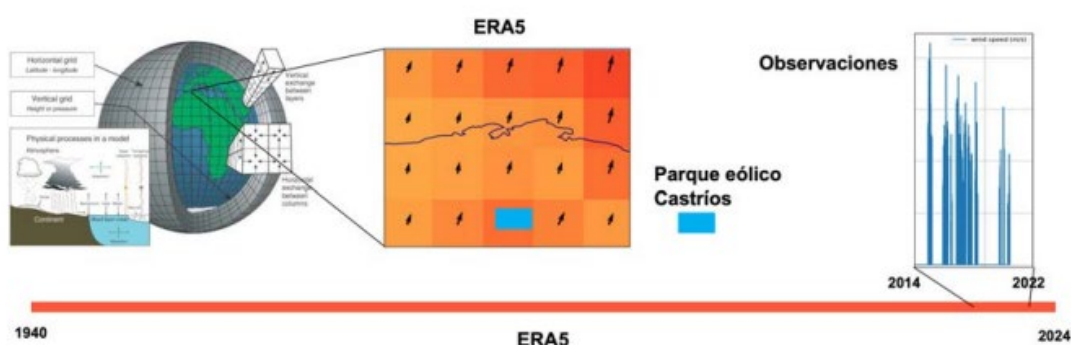


Ilustración 1. Modelos climáticos globales (ERA5), que cubren un periodo de más de 80 años (1940-2024). Observaciones en el parque eólico de Castríos limitadas a 10 años de datos (2014-2024).

La ilustración anterior representa el planteamiento metodológico aplicado. Los datos de baja resolución proceden del reanálisis ERA5, el cual proporciona series temporales horarias continuas desde 1940 a escala global con una resolución espacial de 0.25° , como se explica más en detalle más adelante. Sin embargo, esta resolución de 0.25° o 30 km no es adecuada para la caracterización precisa del recurso eólico en zonas donde el terreno juega un papel importante. Por este motivo, es necesario el uso de simulaciones numéricas de alta resolución, como es el modelo WRF que permite trabajar con simulaciones del orden de 1 km o inferiores.

En el contexto de este TFM, el parque eólico de Castríos es el área principal de aplicación del metamodelo. Para evaluar el valor añadido del metamodelo, se realiza una etapa de validación previa en la zona de Santander, donde si se disponen de datos históricos más longevos y una malla de simulación de 100 m de resolución. Además, como se aprecia en la ilustración anterior, los datos de viento de Castríos presentan ciertas discontinuidades temporales por lo que las validaciones permitirán justificar la capacidad del metamodelo para reproducir el recurso eólico cuando se dispone de datos de entrenamiento de mayor resolución espacial. En Castríos se cuenta con simulaciones de viento a una resolución espacial de 1 km. Si bien hubiera sido óptimo disponer de una malla de 100 m, como en el caso de Santander, la ausencia de esta resolución obliga a trabajar con la malla de 1 km. No obstante, se plantea como una posible mejora futura la aplicación de una malla de 100 m, siempre que los datos disponibles a 1 km demuestren ser de calidad suficiente para la reproducción del recurso eólico en el parque.

A continuación, se realiza una descripción en detalle del parque eólico de Castríos, objeto de este TFM, en el cual se aplicará el metamodelo desarrollado.

2.1 Parque eólico Castríos

2.1.1. Características geográficas

El Parque Eólico “Castríos” se encuentra en el término municipal de Espinosa de los Monteros, en la provincia de Burgos, dentro de la comunidad autónoma de Castilla y León. Este enclave, situado a una altitud superior a los 1.350 metros sobre el nivel del mar, se caracteriza por su relieve montañoso y la presencia de vientos favorables para la generación de energía renovable.



Ilustración 2. Parque Eólico Castríos. Fuente: Ayuntamiento Espinosa de los monteros

La construcción del parque eólico comenzó en el año 2004 y contó con una inversión cercana a los 25 millones de euros. Su desarrollo respondía a la necesidad de incrementar la producción de energía limpia y renovable, contribuyendo a la reducción de emisiones de gases de efecto invernadero y al cumplimiento de los objetivos nacionales e internacionales en materia de transición energética. Su puesta en marcha no solo fortaleció el sistema eléctrico de la región, sino que también supuso un impulso económico para el entorno local, generando empleo y mejorando infraestructuras de acceso a la zona.

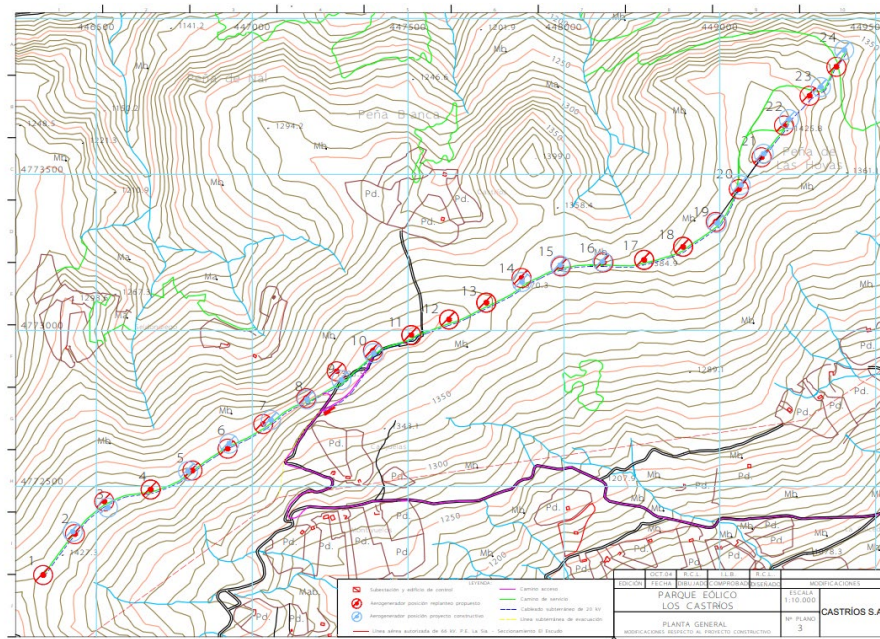


Ilustración 3. Ubicación de los aerogeneradores. Fuente: CASTRIOS S.L.

Desde su entrada en funcionamiento hasta el año 2025, la explotación del parque ha transcurrido dentro de lo esperado, aunque su ubicación en una zona de alta montaña ha supuesto un desafío constante debido a los fenómenos climatológicos extremos, que complican considerablemente las labores de mantenimiento. No obstante, gracias a una planificación eficiente y a un equipo de profesionales altamente capacitados, el parque ha logrado mantener su operatividad dentro de los parámetros establecidos inicialmente [6].

2.1.2. Características mecánicas

El parque cuenta con un total de 24 aerogeneradores, modelo MADE AE-61, con una potencia nominal de 1.100 kW cada uno, lo que suma una capacidad total instalada de 26,4 MW. Estos aerogeneradores están diseñados con una torre de 55 metros de altura y un rotor de 61 metros de diámetro, optimizando la captación del viento en una zona con condiciones climáticas propicias para la producción de energía eólica [6].

- Número de palas: 3
- Diámetro del rotor: 61 m
- Altura del buje: 55 m
- Potencia nominal: 1.100 kW
- Orientación: Barlovento
- Tª ambiente de diseño: -20°/50°
- Tª ambiente de operación: -10°/40°
- Control de potencia: Por entrada en pérdida
- Velocidad de arranque: 3,5 m/s
- Velocidad de parada: 25 m/s

Ilustración 4. Especificaciones Técnicas de los aerogeneradores. Fuente: Wind Power

En términos de producción energética, el parque ha mostrado una variabilidad acorde con las fluctuaciones meteorológicas y operativas. Desde 2007 hasta 2023, su generación anual ha oscilado entre los 47,2 mil y los 63,9 mil MWh, alcanzándose un máximo en el año 2022, con una producción de 63.955 MWh generados.

La producción de los aerogeneradores depende directamente de la cantidad de viento que reciben. Por esta razón, resulta fundamental predecir con precisión la velocidad del viento en cada ubicación donde se encuentran instalados, para minimizar los desvíos. En caso de discrepancias entre la producción real y la estimada, es necesario asumir costos asociados (penalizaciones) tanto por infraproducción como por sobreproducción.

A continuación, se presenta la gráfica que muestra la potencia generada por los aerogeneradores en función de la velocidad del viento.

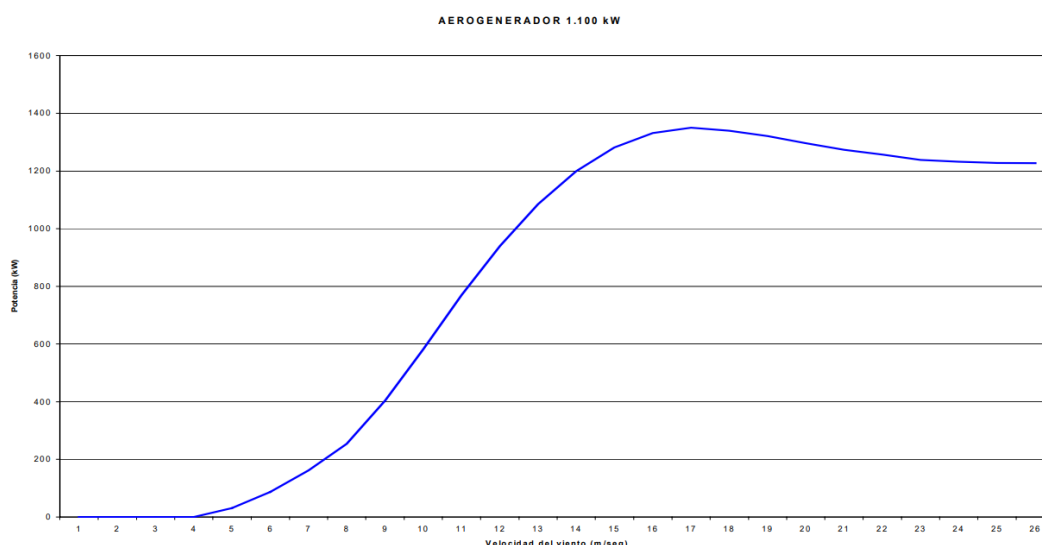


Ilustración 5. Curva de generación de energía. Fuente: Wind Power

Como se puede observar, la diferencia entre prever una velocidad de 14 m/s o 17 m/s no es significativa, ya que en ambos casos el aerogenerador operaría a su máxima potencia. Sin embargo, por razones de seguridad, es obligatorio detener el aerogenerador cuando la velocidad del viento supera los 25 m/s. En este sentido, la diferencia entre predecir 23 m/s y 26 m/s es crucial, ya que implica estar operando a máxima potencia o, por el contrario, no generar energía en absoluto [7]. Igualmente, en situaciones de velocidades más bajas, la diferencia puede estar entre que la máquina se acople a la red o que no, si bien en estos casos la producción es también pequeña. Con velocidades intermedias, en las que puede haber variaciones notables de producción, es cuando puede ser más interesante dotarse de una herramienta de predicción lo más precisa posible.

La predicción de estos picos de viento representa un gran desafío para los operadores de los distintos parques eólicos, no solo en este caso particular, sino en toda España y Europa.

2.1.3. Datos históricos

Con el objetivo de predecir con mayor precisión la velocidad del viento en los emplazamientos deseados, es fundamental comprender su comportamiento en la zona de estudio. Para ello, se utilizarán los registros históricos de cada aerogenerador, lo que permitirá analizar cuáles de los 24 dispositivos generan mayor o menor cantidad de energía.

Tal como se muestra a continuación, las diferencias en la producción entre los distintos aerogeneradores pueden llegar a ser significativas, lo que resalta la importancia de un análisis detallado para optimizar la generación de energía en cada ubicación.

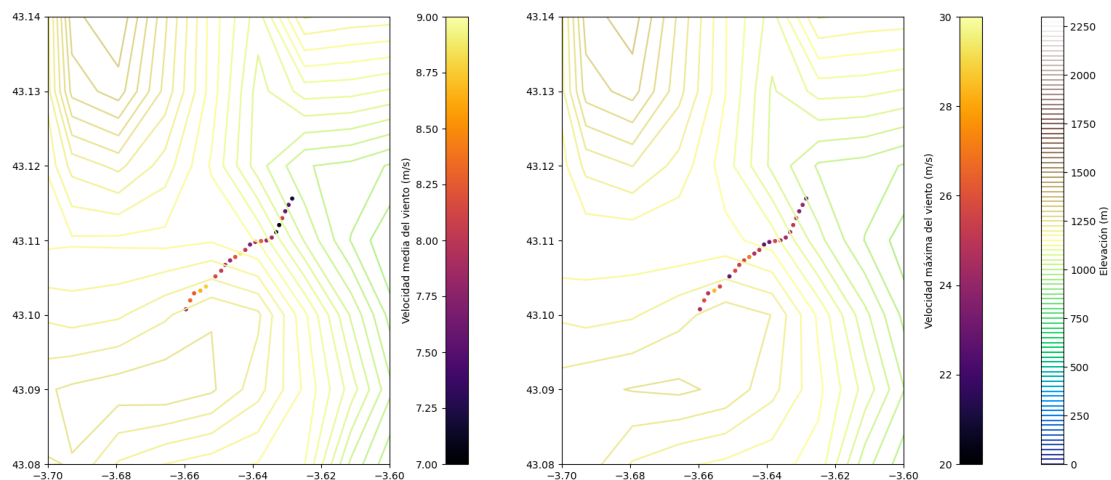


Ilustración 6. Variabilidad espacial de las velocidades medias (panel izquierdo) y de las velocidades máximas (panel derecho) registradas durante enero de 2022 en los aerogeneradores del parque eólico de Castrios.

Respecto a la ubicación física de los aerogeneradores en el emplazamiento, se puede observar que aquellos situados en los picos de las montañas tienen un rendimiento superior en términos de generación de energía en comparación con los que se encuentran algo más resguardados, es decir, parcialmente cubiertos o protegidos por las formaciones montañosas circundantes. Esta diferencia en el rendimiento se debe principalmente a los efectos del viento sobre la topografía local.

En las cumbres y picos de las montañas, los aerogeneradores están expuestos a vientos más constantes y de mayor velocidad, ya que las corrientes de aire no se ven afectadas por barreras naturales como las que se encuentran en las laderas o valles. En cambio, los aerogeneradores situados en zonas más bajas o protegidas por la montaña experimentan vientos más turbulentos y menos consistentes, lo que puede disminuir la eficiencia en la conversión de energía eólica.

Adicionalmente, la presencia de elementos como las colinas, rocas o vegetación en la zona intermedia de la montaña puede generar un efecto de sombra de viento que altera la uniformidad y la velocidad del mismo, lo que ocasiona una menor producción de energía en estos emplazamientos.

Por lo tanto, los aerogeneradores ubicados en las cumbres, con acceso directo a vientos más fuertes y constantes, son los que logran mayores tasas de producción de energía en comparación con aquellos en zonas más protegidas o bajas dentro de la misma área montañosa.

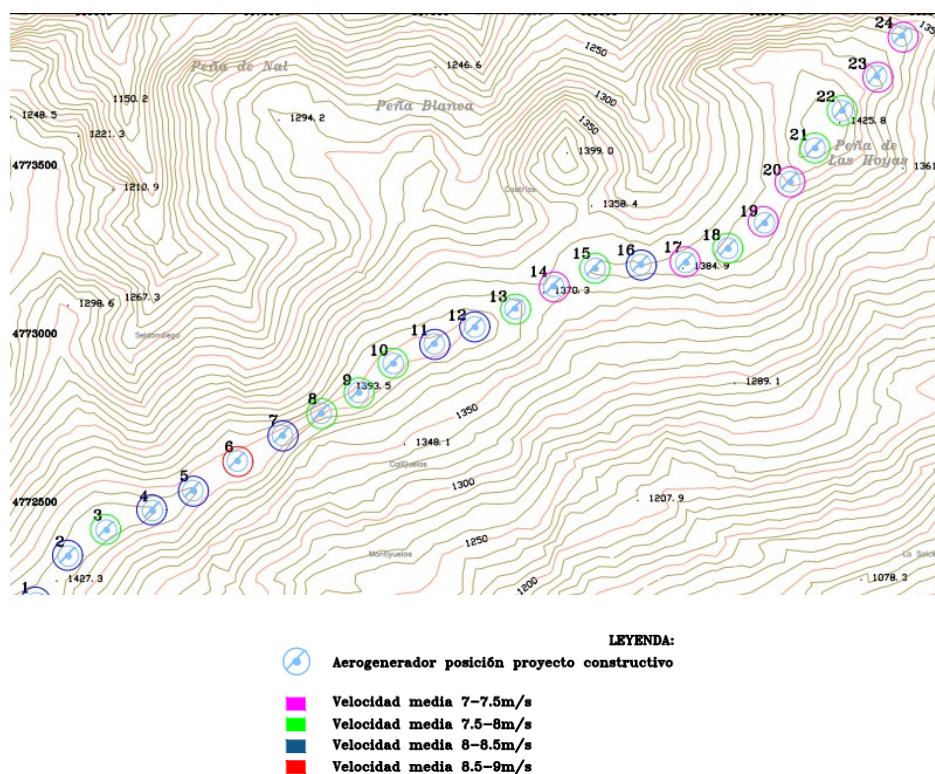


Ilustración 7. Velocidad media percibida por cada aerogenerador.

3. Métodos y Herramientas

En este apartado se hará una breve descripción a los diferentes métodos utilizados en este estudio. Cada herramienta se describirá en términos de sus características generales y capacidades.

3.1 Bluemath-tk

En los últimos años el grupo de investigación GeoOcean de la UC ha promovido la creación de un Consorcio de Universidades y Centros de Investigación para el desarrollo de la biblioteca BlueMath, que ofrece múltiples herramientas en el lenguaje de programación Python y es accesible para toda la comunidad científica. Dentro de Bluemath se engloban diferentes técnicas matemáticas para el análisis y la caracterización de las dinámicas costeras.

Estas herramientas se enmarcan dentro del campo de la Ciencia de Datos e incluyen desde técnicas de clasificación, métodos de reducción de dimensionalidad (como el Análisis de Componentes Principales), hasta diseños de experimentos [8]. También se considerarán métodos de interpolación multidimensional (por ejemplo, Radial Basis Functions) o redes neuronales para el desarrollo de metamodelos, en caso de emplearse estrategias híbridas para el modelado de inundación costera.

BlueMath tiene como objetivo proporcionar herramientas (toolkit), métodos (downscaling híbrido, downscaling estadístico y downscaling dinámico) y servicios climáticos (pronóstico a corto plazo, pronóstico estacional, hindcast, evaluación probabilística, variabilidad climática, cambio climático) para abordar problemas a diferentes escalas espaciales (global, regional, local y alta resolución) que involucran diferentes procesos físicos (viento, olas, marea meteorológica y astronómica, ciclones tropicales, nivel del mar, morfodinámica, etc.)[9].

Para la realización de este trabajo, se ha participado en la elaboración de código que ahora está accesible para el público en general, contribuyendo al progreso de la tecnología de BlueMath. Esta colaboración, junto con el desarrollo de herramientas y métodos dentro del marco de BlueMath, ha sido esencial para la realización de este Trabajo Fin de Máster (TFM).

3.2 Técnicas de Data Science: MDA, PCA, RBF

El objetivo de esta sección es describir las diferentes técnicas de ciencia de datos (para selección, clasificación, interpolación, etc.) utilizadas a lo largo del proyecto.

3.2.1. MDA (Maximum Dissimilarity Algorithm)

El Algoritmo de Máxima Disimilitud (MDA, por sus siglas en inglés) es una técnica utilizada para seleccionar un subconjunto representativo de tamaño M a partir de una base de datos más grande que contiene N vectores o muestras. El propósito

principal del algoritmo es garantizar que los elementos seleccionados en el subconjunto representen de forma adecuada la diversidad presente en el conjunto total de datos.

El proceso comienza con una muestra de datos $X = \{x_1, x_2, \dots, x_N\}$, donde cada x_i es un vector n -dimensional. Inicialmente, el subconjunto está vacío. Para comenzar la selección, se elige el primer vector que se incluirá en el subconjunto. Este vector inicial es aquel que presenta la mayor suma de disimilitudes respecto a todos los demás vectores de la base de datos, lo que significa que es el vector más diferente del resto.

A partir de ahí, el algoritmo añade elementos de forma iterativa. En cada paso, se calcula la disimilitud entre los vectores que aún no han sido seleccionados y los que ya forman parte del subconjunto. Se elige aquel vector que tenga la mayor disimilitud mínima con respecto al subconjunto existente, siguiendo el criterio conocido como MaxMin. Este proceso garantiza que cada nuevo vector añadido aporte la máxima diversidad posible al subconjunto.

El algoritmo se detiene una vez que se han seleccionado M vectores. De esta forma, se obtiene un subconjunto de datos que es diverso y representativo del conjunto original. Esta técnica fue introducida por Kennard y Stone en 1969, y desde entonces han surgido múltiples variantes que modifican aspectos como la forma de inicialización o la medida exacta de disimilitud utilizada. En este caso específico, se aplica la versión MaxMin, que es una de las más comunes y efectivas.

Este tipo de selección es particularmente útil en áreas como el aprendizaje automático, la validación de modelos y la reducción de datos, donde trabajar con conjuntos representativos más pequeños puede mejorar la eficiencia sin sacrificar calidad en los resultados.

Por ejemplo, si el subconjunto está compuesto por R vectores ($R \leq M$), la disimilitud entre el vector x_i de la muestra de datos (es decir, uno de los $N - R$ vectores aún no seleccionados) y el vector v_j perteneciente al subconjunto de tamaño R se calcula de la siguiente manera:

$$d_{ij} = ||x_i - v_j||, \quad \text{donde } i = 1, \dots, N - R \text{ y } j = 1, \dots, R$$

La disimilitud $d_{i,\text{subconjunto}}$ entre el vector x_i y el subconjunto se define como:

$$d_{i,\text{subconjunto}} = \min ||x_i - v_j||, \quad \text{donde } i = 1, \dots, N - R \text{ y } j = 1, \dots, R$$

Es decir, se toma la mínima disimilitud entre el vector x_i y todos los vectores del subconjunto. El siguiente vector seleccionado para incluir en el subconjunto es aquel que tenga el mayor valor de $d_{i,\text{subconjunto}}$, es decir, el que sea más distinto (en el peor de los casos) respecto a los elementos ya seleccionados[10].

MDA tiene una complejidad temporal esperada de $O(M^2N)$ para seleccionar un subconjunto de M -miembros de una base de datos de N -miembros. Sin embargo, en este proyecto se ha considerado un algoritmo más eficiente con una complejidad de $O(MN)$ desarrollado por Polinsky et al. (1996). En esta versión más eficiente, la distancia $d_{i,subset}$ no requiere calcular la distancia entre todos los vectores, d_{ij} . En su lugar, para la selección del vector $(R+1)$ -ésimo, $d_{i,subset}$ se define como la distancia mínima entre el vector x_i y el último vector añadido al subconjunto, y la distancia mínima entre x_i y los $R-1$ vectores anteriores del subconjunto [10]:

$$d_{i,subset} = \min(d_{i,R}, d_{i,subset}(R-1))$$

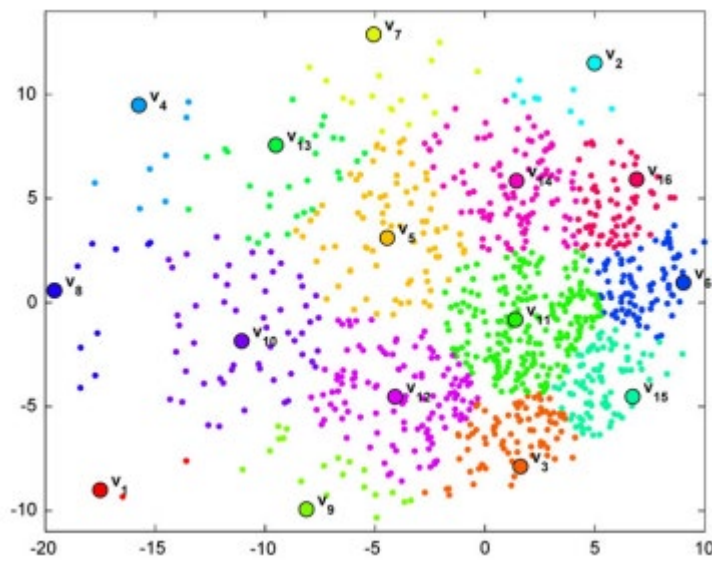


Ilustración 8. Maximum dissimilarity selection. Fuente: Camus et al. (2011)

La ilustración anterior muestra un ejemplo donde se ha utilizado MDA para obtener un subconjunto representativo dentro de una base de datos. Los vectores en el subconjunto están representados por puntos más grandes y están numerados en el orden de selección. El primer vector seleccionado es el más disímil al resto de los datos, típicamente ubicado en el borde del espacio de datos. Las selecciones posteriores incluyen puntos que son progresivamente menos similares, resultando en un subconjunto que está uniformemente distribuido a lo largo del dominio de datos. Aunque MDA no es una técnica de agrupamiento, se considera que cada punto de datos está representado por el vector más cercano en el subconjunto, por lo que se muestran en el mismo color.

En resumen, el Algoritmo de Máxima Disimilitud (MDA) selecciona un subconjunto representativo de un conjunto de datos más grande eligiendo iterativamente los elementos más disímiles. Comenzando con el vector que tiene la mayor suma de disimilitudes en relación con los demás, MDA añade elementos que maximizan la diversidad del subconjunto. Este proceso asegura que el subconjunto seleccionado

represente efectivamente la variabilidad de los datos originales, facilitando el manejo y análisis eficiente de los datos en procesos posteriores.

3.2.2. PCA (Principal Component Analysis)

El Análisis de Componentes Principales (PCA) se emplea para reducir la dimensionalidad de los datos, lo que ayuda a eliminar la redundancia y facilita los procesos de selección e interpolación. Aunque las técnicas MDA y RBF son capaces de manejar datos multivariados, PCA simplifica aún más los datos al reducir sus dimensiones mientras preserva sus características esenciales. Al proyectar los datos en un espacio de menor dimensión, PCA captura la varianza más significativa dentro de los datos de muestra. Esta técnica extrae las máximas correlaciones posibles de los campos espaciales mientras mantiene la diversidad de las situaciones representadas [11].

Dada una variable espaciotemporal $X_i(x_i, t_i)$ donde x representa las variables espaciales estandarizadas y t_i representa el tiempo, se aplica PCA para obtener un nuevo espacio de d dimensiones. Los vectores propios, también conocidos como funciones ortogonales empíricas (EOFs), de la matriz de covarianza de los datos definen los ejes de este nuevo espacio. El objetivo de PCA es encontrar el número mínimo de EOFs que puedan transformar los datos originales en componentes principales (PCs) mientras explican la máxima varianza necesaria para el problema en cuestión.

Los datos originales pueden expresarse como una combinación lineal de EOFs y PCs:

$$X_i(x_i, t_i) = (EOF_1)x * PC_1(t_i) + (EOF_2)x * PC_2(t_i) + \dots + (EOF_d)x * PC_d(t_i)$$

Después de aplicar PCA, los datos se definen por los componentes principales:

$$X_{EOFi} = \{PC_1, PC_2, \dots, PC_d\}; \quad i = 1, \dots, N$$

Cada PC representa una fracción de la variabilidad en los datos, listada en orden decreciente de varianza explicada. Los datos originales pueden reconstruirse utilizando un cierto número de PCs. Cuantos más PCs se consideren, mayor será la varianza explicada y menor el error de reconstrucción.

El criterio para seleccionar el número adecuado de PCs se basa en el error cuadrático medio de reconstrucción (RMSE). Al aumentar el número de PCs y evaluar el RMSE, podemos determinar el número óptimo de componentes que equilibre la varianza explicada y la precisión de la reconstrucción.

En resumen, PCA reduce la dimensionalidad de los datos proyectándolos en un espacio de menor dimensión que captura la varianza más significativa, definida por los vectores propios de la matriz de covarianza de los datos. Esta reducción facilita

el manejo de los datos en procesos posteriores mientras mantiene las características críticas de los datos originales [11].

3.2.3. RBF (Radial Basis Functions)

Las Funciones de Base Radial (RBF) son una técnica de interpolación efectiva utilizada para reconstruir series temporales, particularmente adecuada para datos dispersos y multivariados. La interpolación RBF se ha aplicado con éxito en muchos campos, a menudo produciendo mejores resultados que otros métodos de interpolación. Al comparar métodos para interpolar datos bidimensionales dispersos, RBF ha producido consistentemente los resultados más precisos.

Suponiendo que tenemos una función de valores reales $f = f(x)$ que se desea aproximar. Se nos dan M puntos de datos dispersos $\{x_1, \dots, x_M\}$ de dimensión n y los valores reales asociados a la función $\{f_1, \dots, f_M\}$, donde $f_i = f(x_i)$ para $i = 1, \dots, M$. El método de interpolación RBF aproxima la función como una suma ponderada de funciones base radialmente simétricas centradas en los puntos de datos. La función de aproximación se da por:

$$RBF(x) = p(x) + \sum_{j=1}^M a_j \phi(\|x - x_j\|)$$

Aquí, ϕ es la función de base radial y $\| * \|$ denota la norma euclidiana. El término $p(x)$ es una base polinómica formada por un conjunto de monomios, con coeficientes $\{b_0, b_1, \dots, b_n\}$. Los coeficientes a_j de la función de base radial y los coeficientes del polinomio b se determinan aplicando las restricciones de interpolación $RBF(x_i) = f(x_i)$ para todos los puntos de datos dados [11].

3.3 Cluster Computacional

El Servicio de Supercomputación de Santander (SSC) en la Universidad de Cantabria, a través del IFCA (Instituto de Física de Cantabria), proporciona servicios de computación de alto rendimiento (HPC) a investigadores de la Universidad de Cantabria y otros centros de investigación nacionales e internacionales. El superordenador del SSC, llamado Altamira, es uno de los 14 nodos de la Red Española de Supercomputación (RES), coordinada por el Centro de Supercomputación de Barcelona (BSC), y está integrado como una Infraestructura Científica y Técnica Singular (ICTS) (Servicios Científicos Técnicos de Investigación de la Universidad de Cantabria, s.f.).

Altamira I

La primera versión de Altamira (Altamira I) se instaló en 2007 como un nodo de la Universidad de Cantabria dentro de la Red Española de Supercomputación. Este

clúster consistía en 256 servidores IBM PowerPC JS20 con un total de 512 núcleos, utilizando parte de Marenostrium I previamente instalado en el BSC.

Altamira II

En 2012, Altamira se actualizó con la instalación de un nuevo sistema compuesto por 240 servidores IBM-Idataplex dx360m4. Este sistema se actualizó aún más en 2017 con la integración de nuevos nodos de Marenostrium 3. Altamira II alcanzó la posición 358 en la lista TOP 500 de los superordenadores más potentes del mundo.

Altamira II+

Actualmente, Altamira cuenta con un clúster de 400 servidores, cada uno equipado con 2 procesadores Intel Sandybridge Xeon E5-2670 y 64 GB de RAM. El sistema incluye un sistema de almacenamiento compartido GPFS con más de 2 PB de capacidad de datos y un sistema de respaldo en cinta que puede almacenar datos durante hasta 20 años. Todos los nodos están interconectados a través de una red Mellanox Infiniband FDR de baja latencia a 40 Gbps, formando una topología FAT Tree. En total, la infraestructura consta de más de 2500 núcleos y 10 TB de RAM, alcanzando una capacidad de 50 TFlops. Los usuarios acceden a los servidores a través de un sistema de colas llamado SLURM para gestionar el acceso compartido a los nodos, o a través de la plataforma de computación en la nube OpenStack.

Altamira proporciona aproximadamente 20 millones de horas de computación al año a más de 140 usuarios de la Universidad de Cantabria y otras instituciones nacionales, así como a investigadores de la Red Nacional de Supercomputación. Altamira contribuye con alrededor de 4.5 millones de horas en cada período de 4 meses a la RES, proporcionando el 25% de los recursos de computación totales del SSC. Desde 2021, el Servicio de Supercomputación de Santander ha mantenido un sistema de gestión de calidad certificado bajo la norma ISO 9001:2015.

Partición GeoOcean

En cuanto a equipos e infraestructura, el grupo GeoOcean gestiona una sección dedicada de este clúster. Esta partición incluye 4 nodos, cada uno equipado con 96 núcleos y 512 GB de RAM. Además, este clúster ofrece almacenamiento distribuido con una capacidad de 58 TB. Este recurso dedicado permite al grupo GeoOcean realizar simulaciones avanzadas y análisis pertinentes a su investigación en oceanografía y geociencias.

4. Datos históricos

Para el desarrollo del metamodelo es fundamental disponer de datos históricos que permitan calibrar y validar los resultados obtenidos una vez implementado dicho modelo. En este contexto, se opta por trabajar con diferentes puntos de referencia disponibles en la ciudad de Santander, ya que en esta zona se cuenta tanto con la malla de resolución espacial de 1 km como con la de 100 m.

Los datos históricos proceden de cuatro estaciones representativas: la boya Augusto G. Linares (punto rojo), en una zona litoral; una boya ubicada en la Virgen del Mar (punto negro), en una zona costera; la estación meteorológica de Cueto (punto morado), situada en una zona terrestre del municipio de Santander; y la estación de Parayas (punto rosa), emplazada en una zona más interior en el municipio de Camargo. Cabe destacar que el parque eólico de Castríos no presenta similitud directa con ninguna de estas estaciones, dado su emplazamiento en una zona montañosa, alejada de las tres estaciones referidas. No obstante, su inclusión permitirá realizar una primera aproximación respecto a la precisión de la predicción en condiciones topográficas distintas.

Ubicación de las estaciones con registros históricos de viento

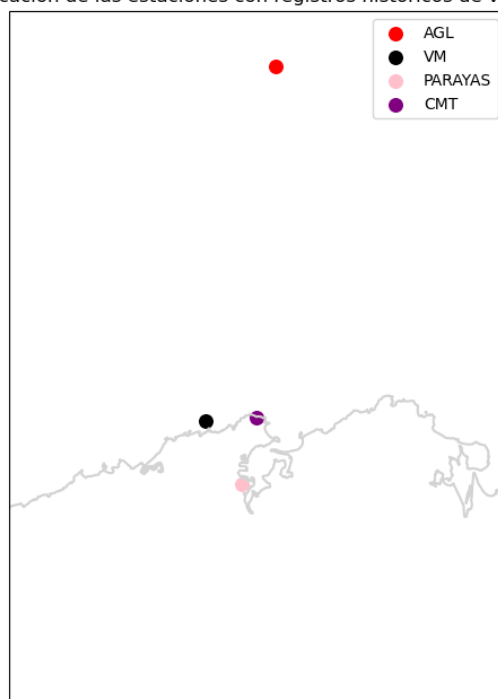


Ilustración 9. Ubicación de las estaciones con registros históricos de viento.

Como se puede apreciar se tienen dos en mar y dos en tierra para poder observar cómo afecta la topografía al modelo, ya que en el emplazamiento montañoso será muy significativo.

Es importante señalar que muchos de los dispositivos empleados para el registro del viento han presentado periodos sin datos, ya sea debido a fallos técnicos o a labores de mantenimiento. En este sentido, el desarrollo del metamodelo no solo tiene como objetivo la predicción del recurso eólico, sino que también se plantea como una herramienta útil para la reconstrucción de series históricas en aquellos intervalos de tiempo en los que no se dispone de registros.

Asimismo, se han incluido los datos históricos del parque eólico objeto de estudio, los cuales también presentan numerosos vacíos temporales. Estos vacíos pueden deberse tanto a condiciones meteorológicas extremas como a tareas de mantenimiento, que impidieron la recogida continua de información. La aplicación del metamodelo en este contexto permitirá evaluar su capacidad para completar y mejorar la calidad de las series temporales incompletas.

4.1 ECMWF Reanalysis v5: ERA5

ERA5 es un reanálisis climático de alta calidad producido por el Copernicus Climate Change Service (C3S), que ofrece datos climáticos globales desde 1950 hasta la fecha, con actualizaciones mensuales. Este reanálisis pertenece a la quinta versión de la serie "ERA" y es una herramienta ampliamente utilizada en estudios climáticos, predicción meteorológica y modelado atmosférico debido a su alta resolución tanto espacial como temporal. Los datos proporcionados por ERA5 son especialmente valiosos para el análisis de la variabilidad climática y la predicción del clima en escalas tanto regionales como globales [12].

Una de las principales características de ERA5 es su capacidad para ofrecer datos de reanálisis, lo que significa que la información se genera a partir de modelos numéricos que integran observaciones reales obtenidas de estaciones meteorológicas, satélites y boyas, entre otros. Este enfoque permite reconstruir de manera coherente las condiciones atmosféricas pasadas, lo que resulta crucial para estudios sobre el cambio climático y para la mejora de modelos predictivos [13].

Los datos de ERA5 cubren una amplia gama de variables atmosféricas, como temperatura, velocidad del viento, humedad, presión atmosférica, radiación solar y otros parámetros clave [14].

Tal como se detallará en el apartado correspondiente al metamodelo, estos datos constituirán el conjunto de baja resolución que actuará como predictor en el proceso de modelado. Posteriormente, serán ajustados mediante el uso de datos de alta resolución.

4.2 Boya Augusto G. Linares

La boya oceanográfica Augusto González de Linares forma parte del sistema de observación marina gestionado por el Instituto de Español de Oceanografía, orientado al monitoreo en tiempo real de las condiciones meteo-oceanográficas en

el litoral cantábrico. Esta plataforma está estratégicamente ubicada en una zona litoral próxima a Santander y está equipada con tecnología avanzada que permite el registro continuo de variables clave para el estudio del medio marino. Entre los parámetros registrados se encuentran la velocidad y dirección del viento, así como características del oleaje, como la altura significativa, el período y la dirección. Los datos obtenidos constituyen una fuente de información fundamental para la caracterización del entorno marítimo, siendo de especial relevancia en estudios de modelización, predicción y validación en el ámbito de la energía eólica marina.

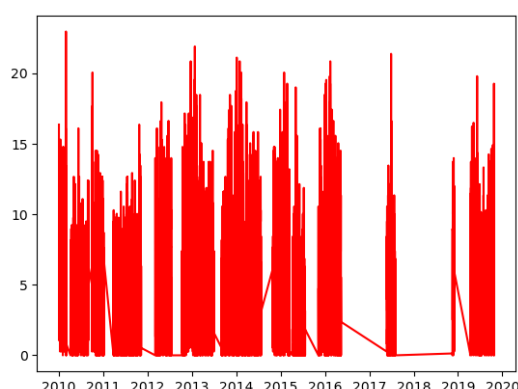


Ilustración 10. Datos históricos AGL.

La boyeta Augusto G. Linares presenta datos desde el 2010 hasta el 2020 pero como se aprecia, suele presentar un mes a lo largo del año completamente parado e incluso durante la etapa del 2016 al 2019 apenas estuvo operativo. Sin embargo, los datos de los que se disponen servirán para calibrar el modelo sin ningún tipo de influencia por el medio.

4.3 Red Vigía Virgen del mar

La Red de Vigilancia Integral de las Aguas (VIGÍA) constituye una plataforma de monitoreo en tiempo real orientada a la observación sistemática de la calidad ambiental y las condiciones físico-oceanográficas de las aguas costeras de Cantabria. Este sistema se sustenta, entre otros componentes, por una boyeta oceanográfica instalada en el litoral próximo a la Virgen del Mar (Santander), la cual está equipada con sensores de alta precisión. Dichos sensores permiten la medición continua de parámetros como el oleaje (altura significativa, dirección y período) y el viento (velocidad y dirección)[15].

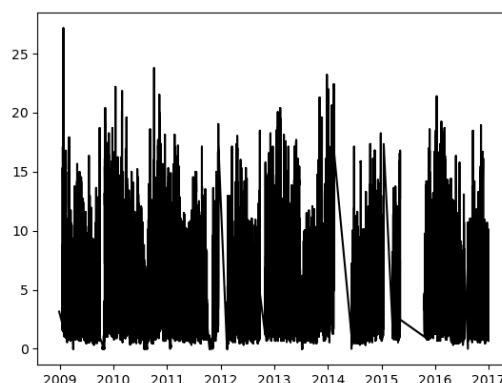


Ilustración 11. Datos históricos VM.

La red VIGÍA ha proporcionado datos de manera relativamente constante durante el periodo comprendido entre los años 2009 y 2017, con la excepción de un breve intervalo entre 2014 y 2015 en el que no se llevó a cabo recolección de datos. En este contexto, la boya utilizada en el presente estudio permitirá validar el metamodelo en comparación con la boya AGL, lo que posibilitará analizar la influencia de la topografía y otros factores en los resultados obtenidos.

4.4 CMT

La Estación Meteorológica de Santander-CMT (Centro Meteorológico Territorial) constituye una de las principales fuentes de datos meteorológicos en la región de Cantabria. Gestionada por la Agencia Estatal de Meteorología (AEMET), esta estación forma parte de la red nacional de observación y se encuentra ubicada en la ciudad de Santander, en una zona costera que la convierte en un punto estratégico para el monitoreo atmosférico. Su instrumental permite la medición continua de variables como la temperatura, la humedad relativa, la velocidad y dirección del viento, la presión atmosférica y la precipitación.

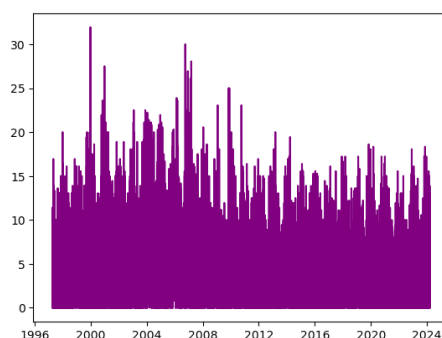


Ilustración 12. Datos históricos CMT.

Los datos proporcionados por la estación CMT son los más consistentes entre los cuatro puntos de referencia considerados en el presente estudio. Esta estación ofrece una serie temporal continua desde 1996 hasta 2024, con registros

prácticamente sin interrupciones. Dado su emplazamiento en tierra firme, los datos de esta estación se emplearán para la validación del metamodelo en comparación con los registros de la estación de Parayas.

4.5 Parayas

La estación meteorológica ubicada en el Aeropuerto Seve Ballesteros-Santander, conocido comúnmente como Aeropuerto de Parayas, constituye un punto de referencia clave para la observación meteorológica en la región norte de España. Gestionada por la Agencia Estatal de Meteorología (AEMET), esta estación forma parte de la red nacional de observación y proporciona datos continuos y de alta calidad sobre variables atmosféricas como temperatura, humedad relativa, velocidad y dirección del viento, presión atmosférica y precipitaciones. Su localización, en una zona abierta próxima a la costa, le permite registrar con precisión las condiciones meteorológicas características del entorno cantábrico, siendo una fuente fundamental tanto para la elaboración de pronósticos como para el desarrollo de estudios climáticos y ambientales en Cantabria.



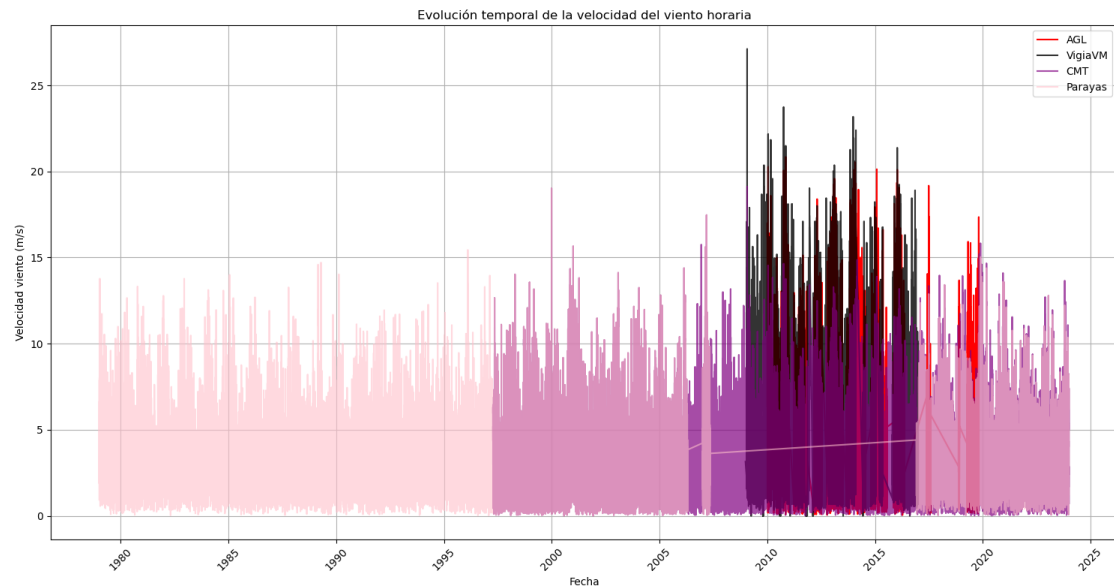
Ilustración 13. Datos históricos Parayas.

La estación meteorológica de Parayas constituye el punto de observación más antiguo del que se dispone en la zona de Santander, con registros que abarcan desde 1980 hasta 2023. No obstante, presenta una interrupción en la serie temporal entre el año 2008 y 2016. En este contexto, el metamodelo desarrollado en el presente estudio se utilizará para completar los vacíos de datos correspondientes a ambos periodos de inactividad.

En conclusión, en la zona de Santander se dispone de datos desde 1980 hasta 2025, que se complementan entre sí para conformar una serie temporal amplia y consolidada. Esto resulta especialmente útil para los procesos de calibración y simulación planteados en este proyecto.

La combinación de las diferentes fuentes permite analizar con detalle la disponibilidad y cobertura temporal de los datos, así como identificar las discontinuidades presentes en cada serie, como se muestra en la siguiente ilustración. Este análisis es clave para asegurar la calidad del estudio, ya que facilita

la detección de periodos con datos faltantes o inconsistentes y permite gestionar adecuadamente estas lagunas, garantizando así una mayor fiabilidad en los resultados obtenidos.



Evolución temporal de la velocidad del viento horaria (m/s) de ERA5 vs Observaciones

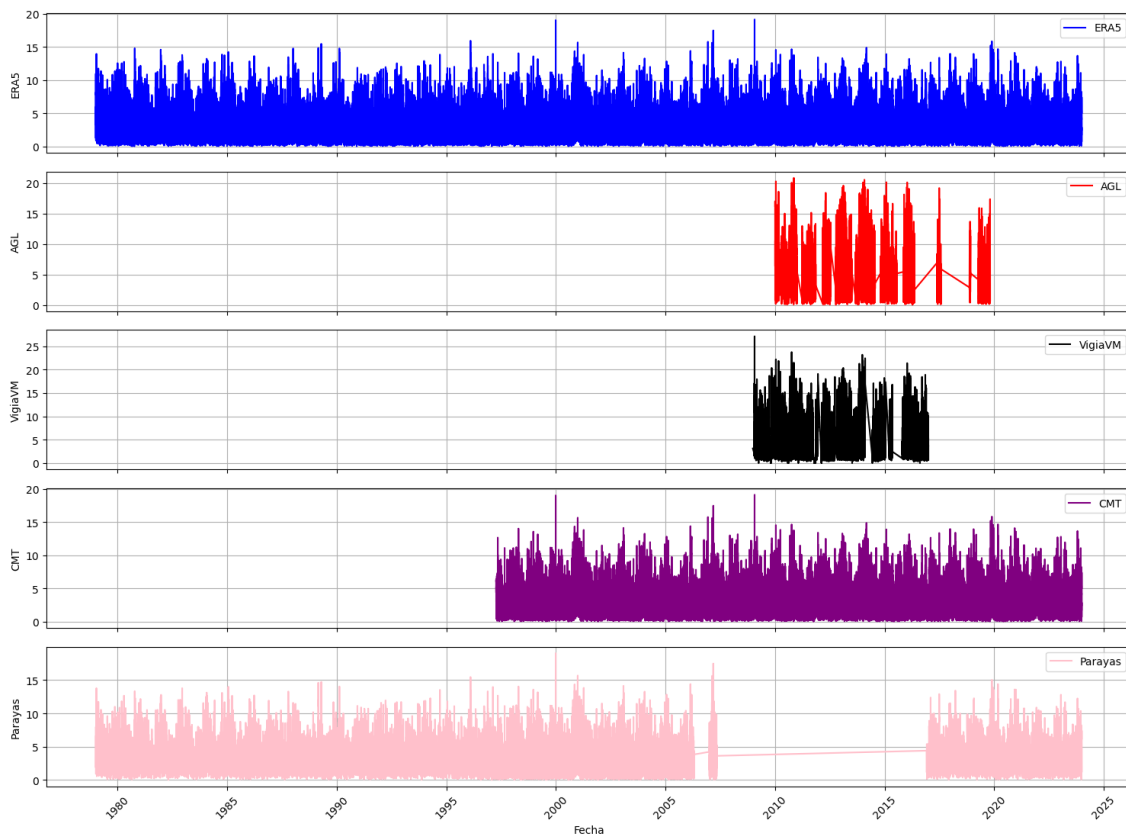


Ilustración 14. Serie histórica de datos en Santander.

4.6 Castríos

El parque eólico de Castríos está constituido por un total de 24 aerogeneradores, cada uno de los cuales dispone de un sensor propio para la medición de la velocidad, ubicado en el rotor. Al igual que en los casos anteriores, las mediciones obtenidas serán utilizadas tanto para la calibración del metamodelo como para la reconstrucción de datos en aquellos intervalos temporales en los que no se disponga de registros, ya sea por fallos técnicos, mantenimiento u otras causas a través del metamodelo.

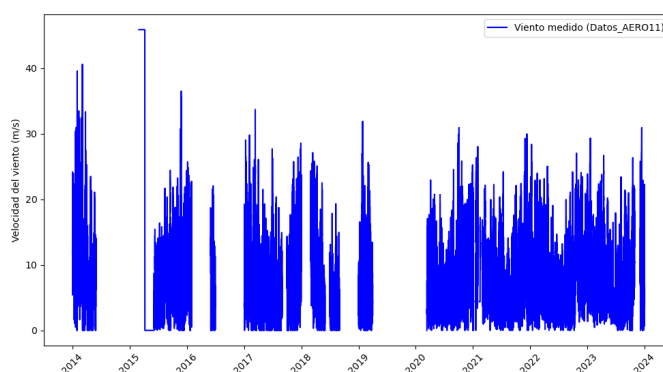


Ilustración 15. Datos históricos Aerogenerador 11.

Cabe destacar que, en las primeras etapas de operación del parque, los promotores no otorgaban prioridad a la conservación y gestión de los datos de viento, lo que dio lugar a prolongados periodos sin información debido a la falta de reparación de los anemómetros tras su avería. Como se puede observar, los registros históricos del parque presentan largos periodos de tiempo en los que no se registró la velocidad del viento, siendo ya en los últimos años cuando se tiene un registro más constante. Esta circunstancia refuerza la necesidad de disponer de un modelo robusto que reconstruya las series temporales ausentes con la mayor precisión posible de cara a futuras predicciones. En este sentido, la reconstrucción histórica facilitará una mejor evaluación del recurso eólico, ya que contar con un mayor volumen de datos permitirá un ajuste más preciso del metamodelo y, por tanto, una mejora en la calidad de las predicciones.

4.7 IFS - ECMWF's

En las últimas décadas, el Integrated Forecasting System (IFS), desarrollado por el European Centre for Medium-Range Weather Forecasts (ECMWF), se ha consolidado como uno de los sistemas de predicción meteorológica numérica más avanzados y precisos a escala global. Este sistema forma la base operativa de las predicciones del ECMWF y se emplea también en múltiples servicios meteorológicos nacionales e internacionales, incluidos los componentes del programa Copernicus.

El IFS combina un modelo atmosférico global espectral, un sistema de asimilación de datos de última generación y una serie de parametrizaciones físicas que

representan procesos subescalares. La atmósfera está acoplada con modelos de superficie y océano, permitiendo simulaciones realistas de las interacciones entre diferentes componentes del sistema climático. Actualmente, el modelo opera con una resolución horizontal de hasta 9 km y 137 niveles verticales, extendiéndose desde la superficie hasta la estratopausa [16].

El sistema se basa en un enfoque de asimilación 4D-Var (cuatro dimensiones variacionales), lo que le permite integrar datos de observación de múltiples fuentes incluyendo satélites, radares, radiosondas y estaciones terrestres de manera coherente y continua en el tiempo. Esta integración mejora sustancialmente el estado inicial del modelo, clave para lograr predicciones fiables. Además, el IFS admite configuraciones deterministas y probabilísticas, lo que permite cuantificar la incertidumbre asociada a los pronósticos[17].

Las salidas del IFS se generan dos veces al día (00 y 12 UTC), aunque las versiones de ensemble (ENS) se ejecutan también a las 06 y 18 UTC con menor resolución. Gracias a su política de datos abiertos desde 2020, muchas de estas predicciones están disponibles públicamente a través del portal del ECMWF y los servicios Copernicus.

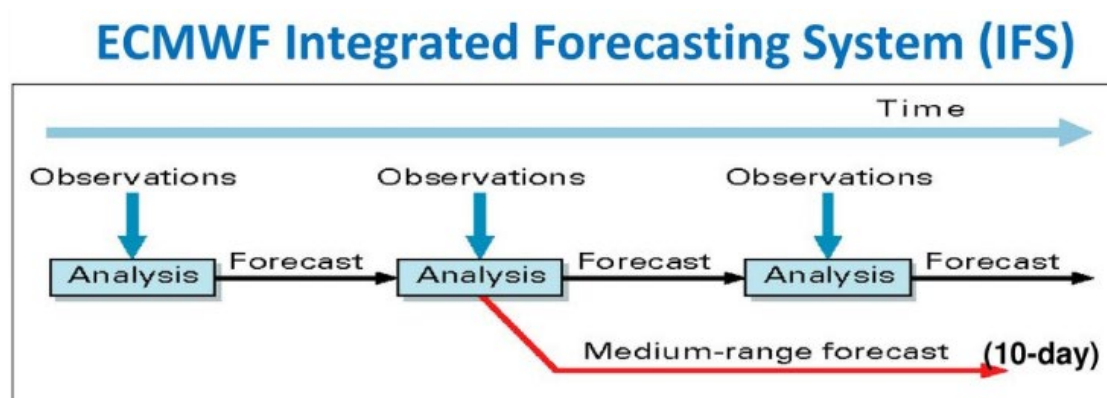


Ilustración 16. Descripción del modelo y su entrenamiento. Fuente: ECMWF (European Centre for Medium-Range Weather Forecasts)

En este trabajo se ha utilizado un conjunto de datos que contiene pronósticos a 15 días de variables del modelo atmosférico generado por el sistema de previsión integrado (IFS) del ECMWF, con una resolución espacial de 0.25° . A estos datos se les denomina casi en tiempo real (NRT), ya que se publican dos veces al día poco después de la emisión oficial de los datos de pronóstico en tiempo real del ECMWF, de los cuales constituyen un subconjunto. Con ellos se realiza la predicción a 24, 48 y 72 horas vista a través del metamodelo desarrollado.

5. Metamodelos

5.1 Calibración del metamodelo

Como ya se ha señalado, el objetivo del presente trabajo es predecir el viento en un emplazamiento, en este caso, en un parque eólico, utilizando la metodología previamente explicada. Para ello es necesario el desarrollo de un metamodelo que reconstruya el campo de vientos horarios a alta resolución (HR)(predictando) a partir de vientos de baja resolución (LR)(predictor), lo que permitirá reconstruir el registro histórico del recurso eólico. Esto proporcionará una base sólida para ajustar el metamodelo y, en consecuencia, mejorar la precisión de las predicciones futuras.

Es importante comprender el significado de los términos predictor y predictando en el contexto del presente estudio, ya que constituyen elementos fundamentales en el desarrollo del metamodelo propuesto. El predictor hace referencia a la variable o conjunto de variables independientes que se utilizan como entrada del modelo para realizar la predicción. En este caso específico, se trata de los datos de viento de baja resolución (en este caso los datos de ERA5), los cuales, aunque menos detallados espacial o temporalmente, están disponibles de forma más extensa y constituyen la base sobre la que se construye el metamodelo.

Por su parte, el predictando corresponde a la variable dependiente, es decir, aquella que se desea estimar a partir del predictor. En este trabajo, el predictando está representado por los campos de viento horarios a alta resolución (en este caso los vientos horarios sobre Cantabria a 1km y Santander a 100m).

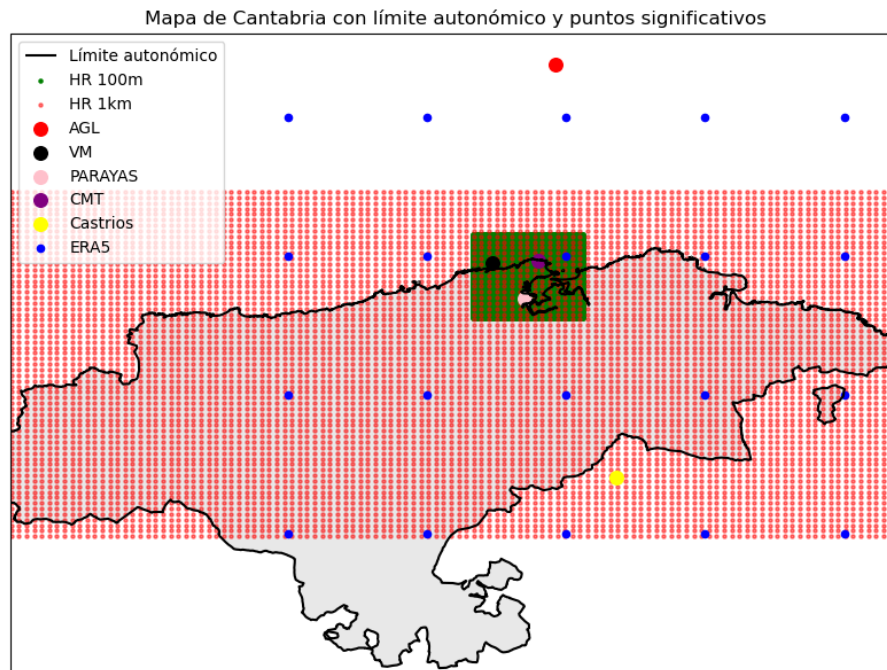


Ilustración 17. Puntos de interés.

La metodología para desarrollar el metamodelo consiste en los siguientes pasos:

- 1- Se aplica un Análisis de Componentes Principales (PCA) a los datos de LR y HR, con el objetivo de reducir la dimensionalidad y preservar las características más significativas de los datos, facilitando su manejo en etapas posteriores.
- 2- Selección de muestras representativas: A continuación, se emplea el Algoritmo de Máxima Disimilitud (MDA) sobre el conjunto de datos reducido por PCA, para seleccionar un subconjunto representativo. Este subconjunto de casos (los 6000 mencionados anteriormente) se elige de forma que maximice la diversidad y cubra de manera eficiente el dominio de los datos.
- 3- Reconstrucción utilizando datos de alta resolución: Los puntos seleccionados mediante MDA se comparan con los datos de alta resolución (ya sea a 1 km o 100 m) para realizar la reconstrucción sobre el resto de los puntos. Este paso permite hacer estimaciones más precisas basadas en un subconjunto representativo de los datos.

Una vez finalizada la etapa de calibración se pasa a la etapa de validación a través de los datos históricos mencionados en el apartado anterior. Esta etapa consistirá en:

- 4- Reconstrucción histórica y evaluación de la habilidad del metamodelo: Finalmente, se aplica el metamodelo para reconstruir el viento a alta resolución (Cantabria 1 km y Santander 100 m) durante todo el periodo

cubierto por ERA5 (1979-2025). La calidad y precisión del modelo se evalúan mediante la comparación de estas reconstrucciones con los datos históricos de las diferentes estaciones mencionadas anteriormente.

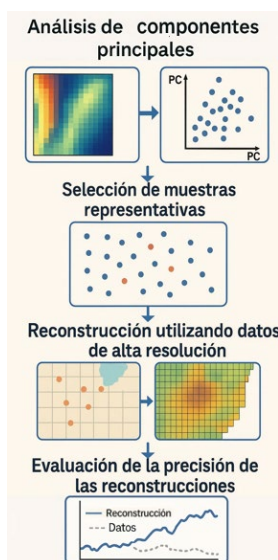


Ilustración 18. Procedimiento aplicado.

5.1.1. Selección de casos representativos de vientos diarios LR

Se utiliza la base de reanálisis histórico ERA5 (1979-2024) con vientos horarios a 0.25 grados de resolución para la definición de los vientos LR. Para ello se establece un dominio espacial que abarca gran parte de Cantabria, que comprende 20 nodos de ERA5 (puntos azules en la Ilustración 17. Puntos de interés.).

La selección de casos representativos se realiza sobre los vientos LR a escala diaria. Previamente, se aplica la técnica de Análisis de Componentes Principales (Principal Component Analysis, PCA) a los vientos diarios HR para reducir la dimensionalidad de los datos y simplificar la selección de casos a simular. En este caso, se han elegido las 37 primeras componentes principales (PCs), que explican el 99% de la varianza de los vientos LR.

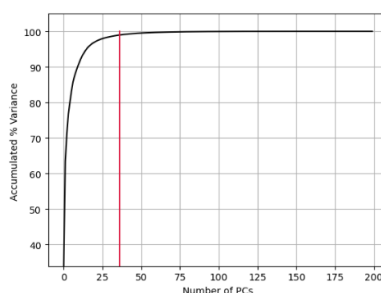


Ilustración 19. Varianza acumulada.

A continuación, se muestra la recreación del viento de baja resolución frente aquella con el 99% de la varianza en la zona de Santander.

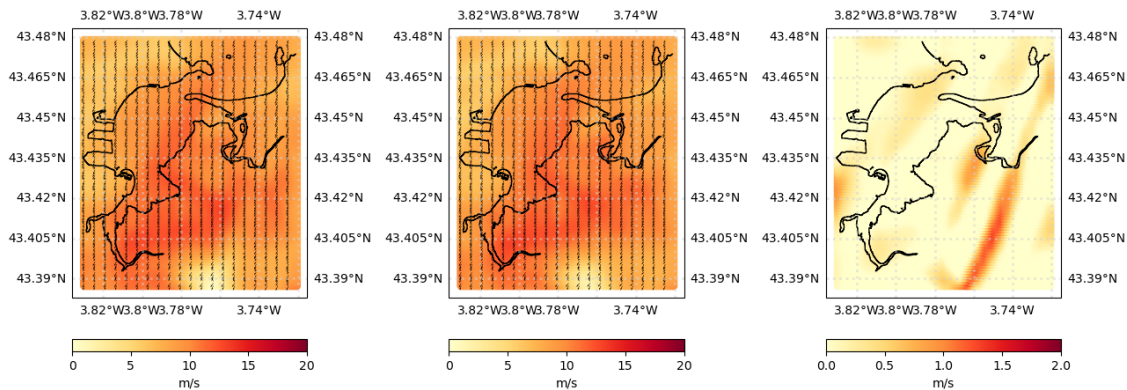


Ilustración 20. Representación del viento de baja resolución real (izquierda), asumido (central) y su diferencia (derecha).

Como se puede observar al reducir más del 80% de las dimensiones el error que se presenta es mínimo. Esto nos permite reducir la dimensionalidad de los datos y simplificar la selección de casos a simular como se expuso previamente.

En esta etapa, se seleccionaron 365 días representativos a partir de los datos diarios de ERA5 disponibles desde 1979 hasta 2024 (más de 16.000 días en total), con el objetivo de construir un conjunto diverso para la reconstrucción del campo de viento. Estos 365 días fueron posteriormente simulados con WRF en alta resolución, generando un total de 8760 campos horarios de viento.

Una vez reducidas de manera significativa las dimensiones de los datos de baja resolución, se procede a la aplicación del algoritmo de Máxima Disimilitud (Maximum Dissimilarity Algorithm, MDA). Con el objetivo de entrenar el modelo, se seleccionó un subconjunto representativo de 6000 casos a partir del total de 8760 disponibles procedentes de las simulaciones en alta resolución (HR). Si bien esta cifra puede considerarse elevada, especialmente teniendo en cuenta que quedarán menos de 3000 casos disponibles para la fase de evaluación interna del modelo, dicha elección se justifica por la necesidad de realizar una reconstrucción altamente precisa del campo de viento. Además, es importante señalar que la validación final se llevará a cabo utilizando datos de campo, lo que refuerza la decisión de priorizar una selección amplia y representativa en esta etapa del proceso.

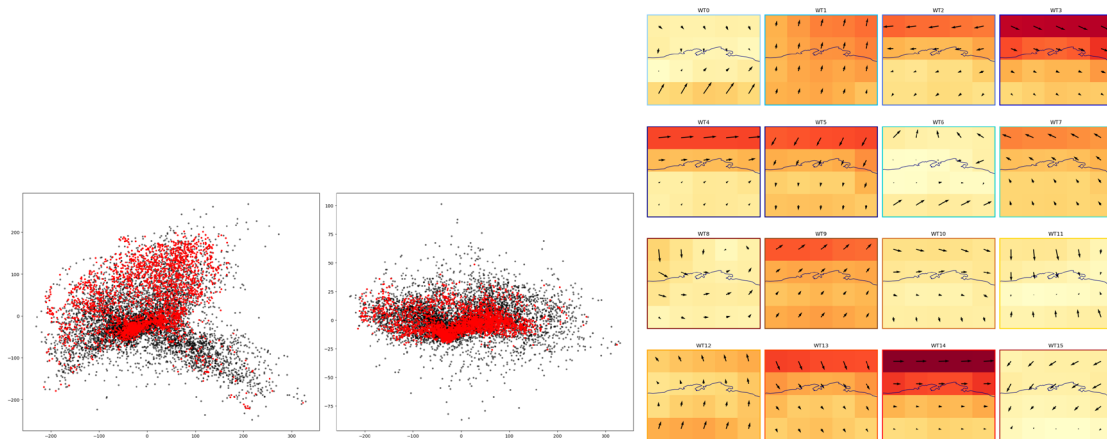


Ilustración 21. Casos MDA considerados (izquierda) y sus correspondientes vientos espaciales para los 16 primeros casos (derecha).

5.1.2. Modelización numérica de los vientos HR a partir de los casos de viento LR

Esta selección de casos de vientos diarios LR se simula con el modelo Weather Research and Forecasting (WRF) a escala diaria. El modelo WRF se fuerza con las condiciones horarias de diferentes variables atmosféricas (temperatura, geopotencial, ...), obtenidas a partir del reanálisis ERA5, a escala horaria. Este sistema de cálculo numérico para simulación atmosférica (NWP) requiere de datos de horas previas para poder representar de manera precisa las condiciones en el instante de interés en el metamodelo desarrollado se tiene en cuenta las 24 horas anteriores cada 6 horas. Con ello se obtienen campos de viento a alta resolución para el dominio de Cantabria a 1 km y de Santander a 100 m.

El grupo GeoOcean dispone de datos de vientos de alta resolución en la zona de Santander, con una resolución de 100 metros, y en la zona de Cantabria, con una resolución de 1 kilómetro. Además, cuentan con datos históricos de AEMET en Santander y con las observaciones de boyas mencionadas previamente.

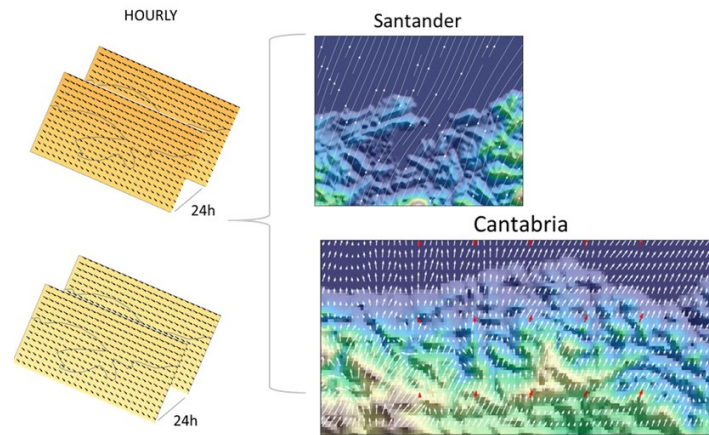


Ilustración 22. Campos horarios sobre el dominio de Cantabria (resolución 1 km) y Santander (resolución 100 m).

5.1.3. Construcción y ajuste del metamodelo

La reconstrucción histórica de los vientos HR sobre Santander y Cantabria (predictando) se realiza mediante la aplicación de la técnica de interpolación multidimensional basada en funciones de base radial (Radial Basic Function, RBF), en función de los vientos LR (predictor). Esta técnica requiere un procesado del predictando consistente en la reducción de la dimensional de los mismos mediante PCA, quedándonos con un número de PCs que explican el 99% de la varianza de los datos. En el caso del predictando, el número de PCs seleccionados ha sido 23. La aplicación de PCA se realiza para la descomposición del predictor multidimensional (campo de viento HR) en variables unidimensionales (PCs) para el ajuste de una función RBF de forma independiente. Una vez ajustada la función RBF para cada PC, se reconstruye el campo espacial teniendo en cuenta la correspondiente función ortogonal empírica (Empirical Orthogonal Functions, EOFs) del predictando.

Una vez terminada la reconstrucción se selecciona un día de fuertes vientos en la comunidad para observar la reconstrucción quedando de la siguiente manera.

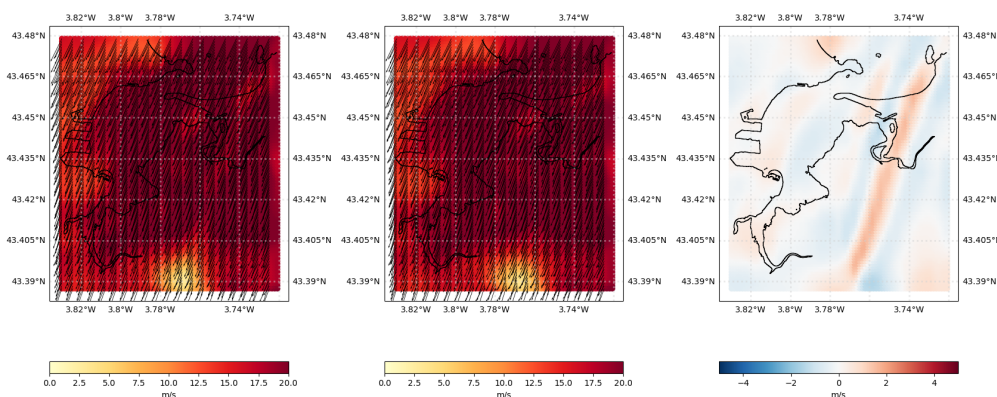


Ilustración 23. Comparativa entre un día de fuertes vientos reales(izquierda), simulados(centro) y error(derecha).

Como se aprecia el error es bastante pequeño teniendo una diferencia máxima de 3m/s entre el real y el reconstruido.

5.1.4. Conclusiones

En conclusión y como se puede observar en la Ilustración 18. Procedimiento aplicado, se implementó un proceso eficiente para reconstruir series de viento a alta resolución combinando PCA para reducir la dimensionalidad y MDA para seleccionar muestras representativas. Esto permitió realizar reconstrucciones precisas con un menor coste computacional. A continuación, se realiza la comparación con datos históricos que validó la capacidad del modelo para reproducir el viento en Cantabria (1 km) y Santander (100 m) durante el periodo 1979-2025.

5.2 Evaluación de la precisión del metamodelo

Como se muestra en la Ilustración 17. Puntos de interés y se expuso previamente, se cuenta con cuatro localizaciones que disponen de datos históricos, así como con dos modelos de alta resolución que serán empleados como referencia para ajustar el modelo ERA5 en función de dichos datos. La principal diferencia entre ambos modelos de alta resolución radica en la escala espacial de los datos: en el modelo de resolución de 100 metros, cada punto representa información agregada correspondiente a un área de 100 metros a su alrededor, mientras que, en el modelo de 1 kilómetro, cada punto abarca una superficie de 1 km a su alrededor.

Como es esperable, el modelo de 100 metros proporciona una mayor precisión espacial que el de 1 km. Sin embargo, contar con este último resulta útil como paso intermedio entre los modelos de alta resolución y el modelo ERA5, cuya resolución es de aproximadamente 30 km.

A continuación, se muestran las validaciones efectuadas.

5.2.1. Evaluación en la estación Augusto G. Linares

Los datos provenientes de fuentes como ERA5 ofrecen una representación más precisa del viento en zonas marítimas, donde la influencia de factores externos es mínima. Por esta razón, el estudio se aborda partiendo del entorno marítimo hasta llegar, finalmente, al parque eólico ubicado en zona montañosa. De este modo, siguiendo el orden mencionado para empezar, se encuentra la boya AGL.

Es importante destacar que la boya AGL no se encuentra dentro del dominio de las simulaciones a alta resolución (ni 100 m ni 1 km), por lo que su comparación se realizará únicamente frente a los datos de ERA5, ya sea en formato horario o agregado a escala diaria. Dado que el viento es una variable especialmente compleja de reproducir con alta precisión a escala horaria, es habitual recurrir a resoluciones temporales mayores, como la diaria, para una evaluación más robusta.

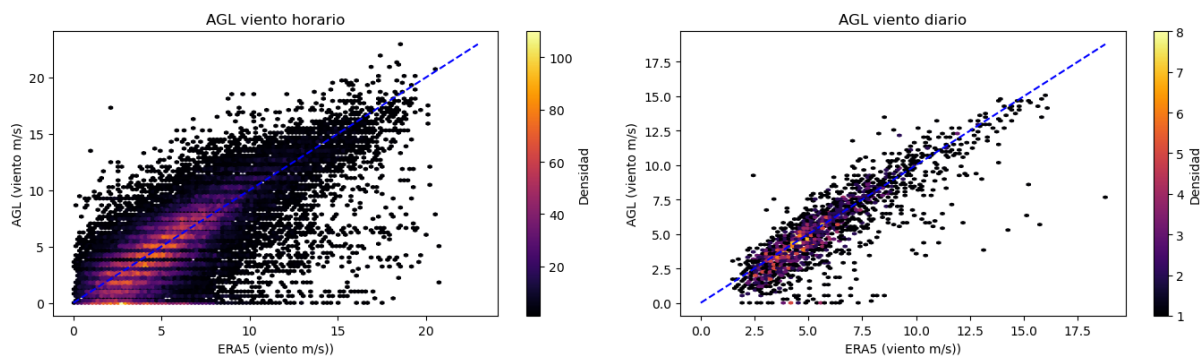


Ilustración 24. Correlación entre ERA5-AGL de manera horaria (izquierda) y diaria (derecha).

En esta primera estimación y como se observa en la ilustración anterior, el modelo ha reproducido los valores históricos de la boya con un alto grado de precisión, especialmente al analizar la media diaria, donde la mayoría de los datos se encuentran distribuidos a lo largo de la diagonal. En el caso de la comparación horaria, aunque se evidencia una mayor dispersión de los datos, se aprecia que la mayor densidad de los mismos también se sitúa en torno a la diagonal.

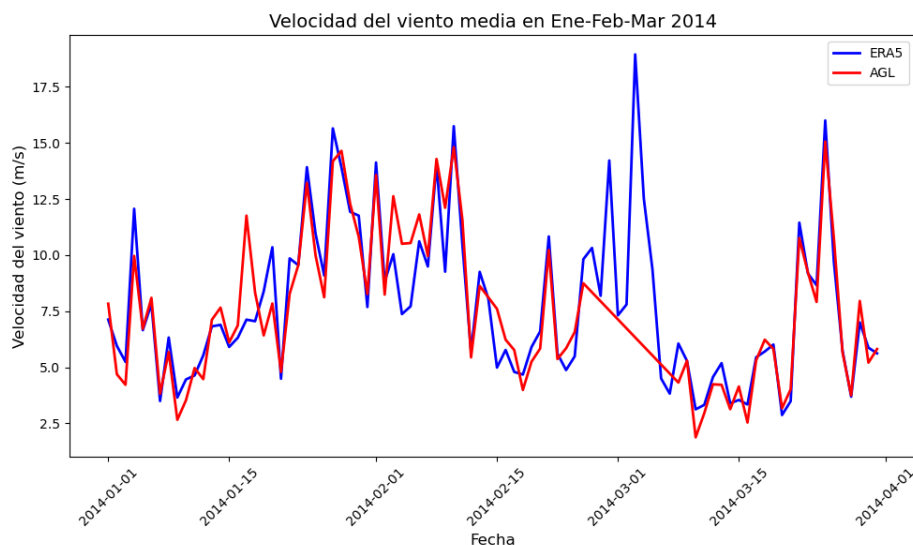


Ilustración 25. Comparativa ERA5-AGL diaria para el primer trimestre de 2014.

Se puede apreciar como los datos del reanálisis ERA5 son muy precisos en intervalos de tiempo donde hubo vientos de gran magnitud. Asimismo, como se mencionó anteriormente, es posible completar los períodos en los que la boya probablemente dejó de registrar datos debido a las condiciones adversas del temporal. Esto se evidencia claramente en la reconstrucción, donde se obtiene el pico de viento más alto del trimestre.

La Ilustración 26. Dirección del viento medida y la estimada por ERA5 muestra las rosas de viento obtenidas a partir de los registros de la boya AGL y de los datos del reanálisis ERA5, considerando únicamente los instantes coincidentes. Se observa

que los vientos predominantes provienen del noroeste y del este, direcciones que el reanálisis es capaz de reproducir adecuadamente.

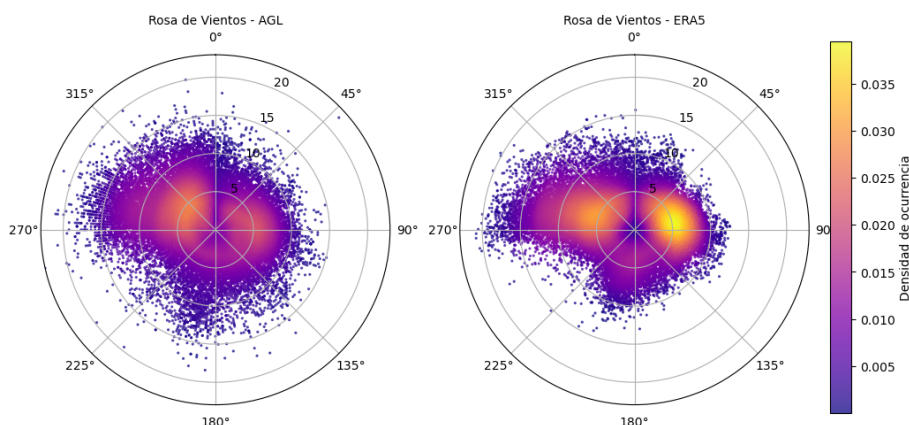


Ilustración 26. Dirección del viento medida y la estimada por ERA5.

5.2.2. Evaluación de la estación de Virgen del Mar

El siguiente punto para validar será el localizado en la Virgen del Mar, este servirá para compararlo con la boya AGL donde se observa la mejora por el aumento de resolución de las simulaciones numéricas.

Como se muestra a continuación, ERA5 infraestima el viento y esto se debe principalmente a que la boya se encuentra en una zona relativamente abrigada y el viento ya se ve afectado por factores costeros los cuales ERA5 no termina de recoger.

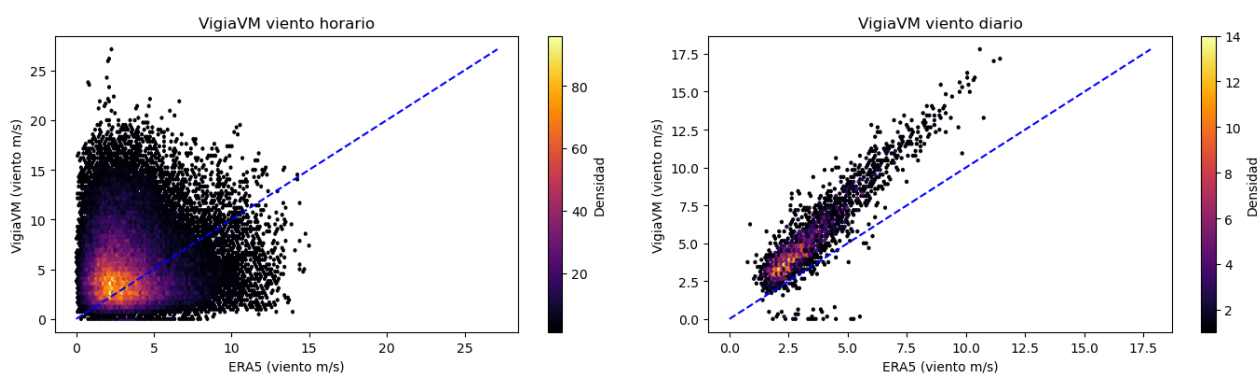


Ilustración 27. Correlación entre ERA5-VM de manera horaria (izquierda) y diaria (derecha).

Dado que la vigía de Virgen del Mar se sitúa aproximadamente a mitad de camino entre los dos puntos de muestreo del modelo, se podía intuir que estos datos serían una estimación, la cual será validada a continuación mediante modelos de alta resolución.

A continuación, se muestran los datos históricos medidos en la red Vigía respecto a los datos del punto ERA5 más cercano. En el cual se observa una clara infraestimación de los valores del viento.

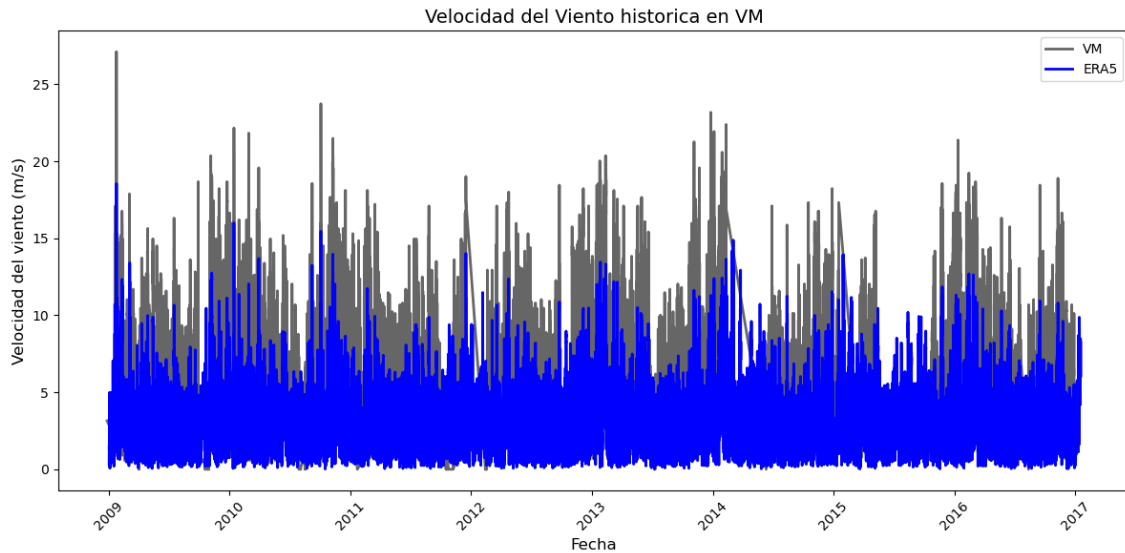


Ilustración 28. Datos históricos VM - ERA5.

En cuanto a los datos de alta resolución se observa una clara mejora en general sobre la infraestimación previa, aunque no se aprecian diferencias apenas entre el modelo de alta resolución de 100 m y 1 km.

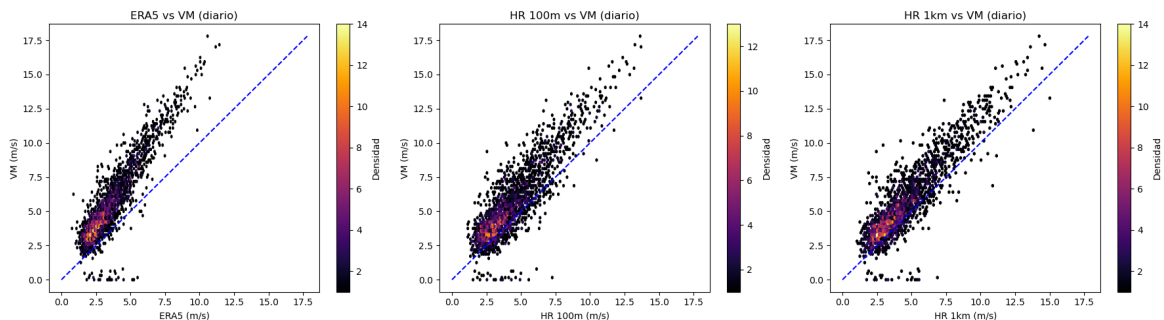


Ilustración 29. Comparativa Metamodelo - Datos Históricos VM

Con el objetivo de analizar la respuesta temporal de los diferentes metamodelos, se presenta una gráfica correspondiente al primer trimestre de 2013. En dicho periodo, se observa cómo los modelos de alta resolución son capaces de reproducir con mayor precisión los máximos de viento en comparación con las estimaciones proporcionadas por ERA5. Esta mejora en la representación de los picos es especialmente relevante para aplicaciones donde la caracterización de eventos extremos resulta crítica.

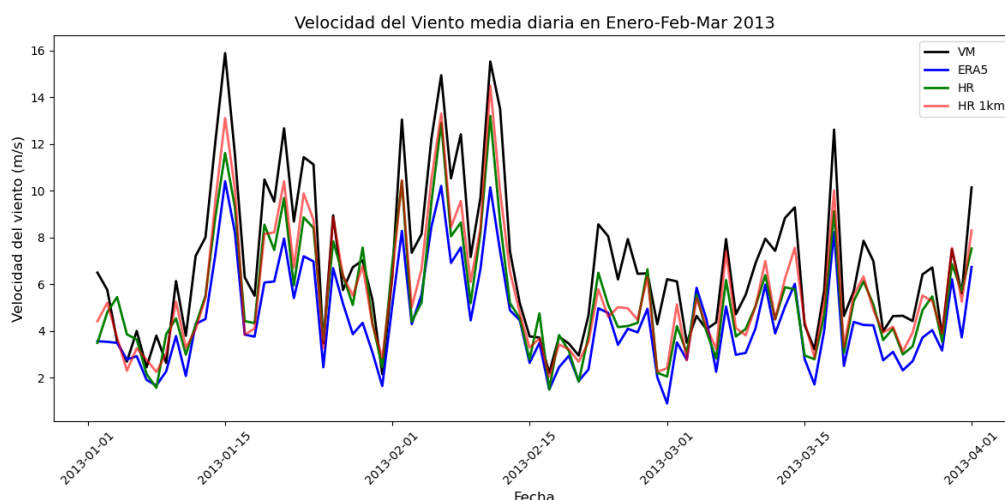


Ilustración 30. Metamodelo - Datos Históricos VM. Primer trimestre de 2013

Con el fin de validar la capacidad del metamodelo para reproducir no solo la magnitud del viento, sino también su componente direccional, se lleva a cabo un análisis comparativo entre las direcciones del viento registradas por la boya oceanográfica del sistema VIGÍA y las estimadas por el metamodelo en el mismo intervalo temporal. Este análisis permite identificar posibles sesgos angulares o desviaciones sistemáticas en la salida del metamodelo, lo cual resulta fundamental para determinar su aplicabilidad en escenarios de predicción y modelado de viento en alta resolución.

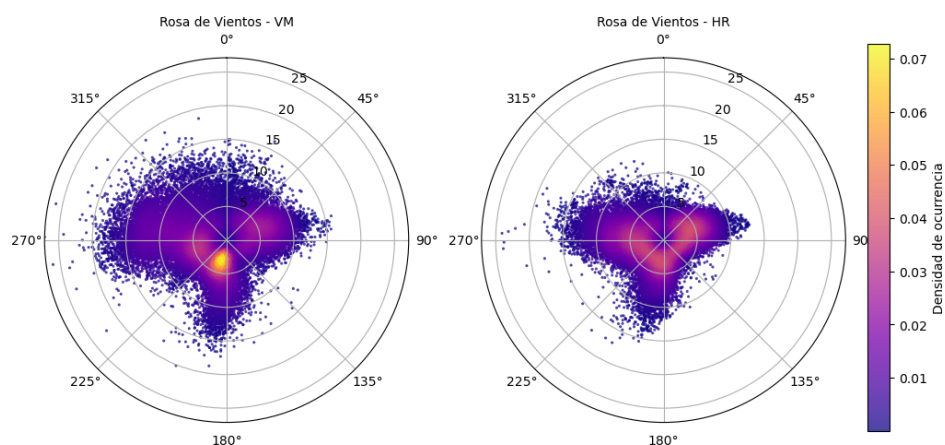


Ilustración 31. Dirección del viento medida en campo (VM) y la predicha por el metamodelo

Se observa un buen grado de concordancia general, aunque con ciertas discrepancias principalmente asociadas a la magnitud del viento como se pudo apreciar previamente.

En comparación con los datos medidos en la boya AGL, se observa que en este caso el modelo ERA5 no resulta suficiente para la reconstrucción de los datos, debido a una mayor influencia de factores costeros. Mientras que en un emplazamiento con

profundidad indefinida y alejado de la costa el modelo ERA5 era adecuado para la caracterización del recurso eólico, en el presente caso se evidencia la necesidad de utilizar metamodelos de alta resolución que permitan una caracterización más precisa del recurso.

5.2.3. Evaluación de la estación de CMT

Una vez finalizadas las validaciones del metamodelo en una zona marina, se procede a realizar las validaciones en zonas terrestres, donde el viento presenta una mayor heterogeneidad debido a la influencia de diversos procesos locales. En primer lugar, se llevará a cabo el estudio en la estación CMT, con el objetivo de compararlo posteriormente con los resultados obtenidos en Parayas.

Antes de mostrar los diagramas de dispersión, conviene destacar que los registros históricos de viento pueden presentar cierta inhomogeneidad, algo habitual en series largas como la del CMT, que cuenta con más de 20 años de datos. Aun así, esta posible inhomogeneidad no explica por completo las diferencias observadas. En el caso del CMT, se observa que ERA5 no representa bien los valores registrados, ya que la mayor parte de los datos se alejan de la diagonal. Además, la mayoría se sitúan por encima de esta línea, lo que indica que ERA5 tiende a subestimar las velocidades de viento medidas en esta estación.

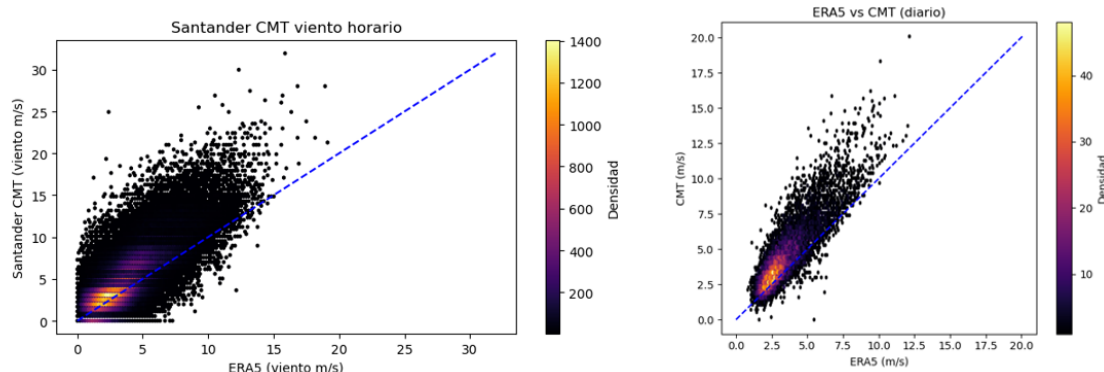


Ilustración 32. Correlación entre ERA5-CMT de manera horaria (izquierda) y diaria (derecha).

Para ilustrar estas diferencias a lo largo del tiempo, en la siguiente figura se representa la serie temporal completa de velocidad del viento registrada en el CMT comparada con los valores correspondientes del modelo ERA5. Esta visualización permite apreciar de forma clara la tendencia sistemática de ERA5 a subestimar los valores máximos de viento, especialmente en los eventos más intensos. La diferencia es especialmente notable en los primeros años del registro, donde los picos registrados por la estación superan con creces a los del modelo.

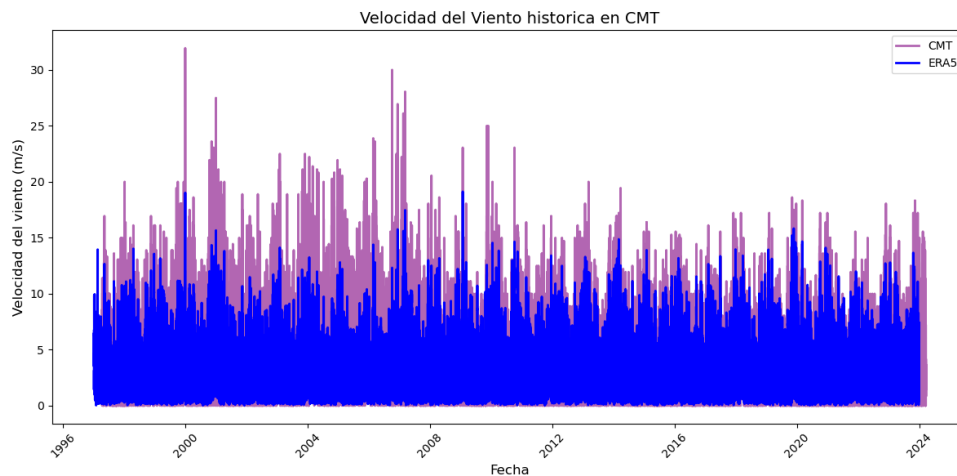


Ilustración 33. Datos Históricos CMT - ERA5.

Para evaluar el comportamiento de los modelos de alta resolución frente a las observaciones, se presentan a continuación las comparaciones entre los datos medidos en el CMT y las estimaciones obtenidas con ERA5, HR (resolución de 100 m) y HR 1 km (resolución de 1 km). Estas representaciones permiten visualizar claramente cómo los modelos de alta resolución ajustan mejor la distribución de los datos a lo largo de la diagonal, lo que indica una mayor capacidad para reproducir la variabilidad y magnitud de las velocidades del viento observadas. En particular, se observa una reducción en la dispersión y una mejor captura de los valores más altos respecto a ERA5.

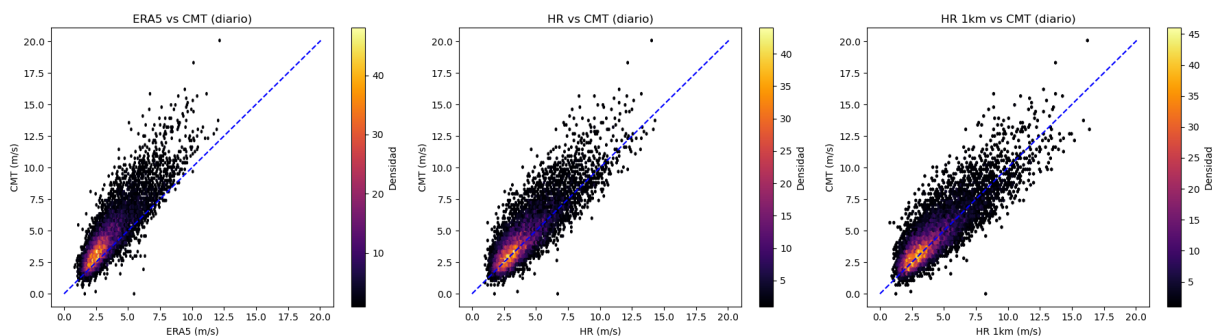


Ilustración 34. Comparativa Metamodelo - Datos Históricos CMT.

Al igual que en el caso anterior, se presenta un zoom al primer trimestre del año 200, donde se registraron vientos de alta magnitud. En el caso de los modelos de alta resolución, se aprecia una mejora significativa en la representación de los picos de viento, los cuales son capturados con mayor precisión. La diferencia respecto a los datos observados es considerablemente menor en comparación con las validaciones utilizadas previamente.

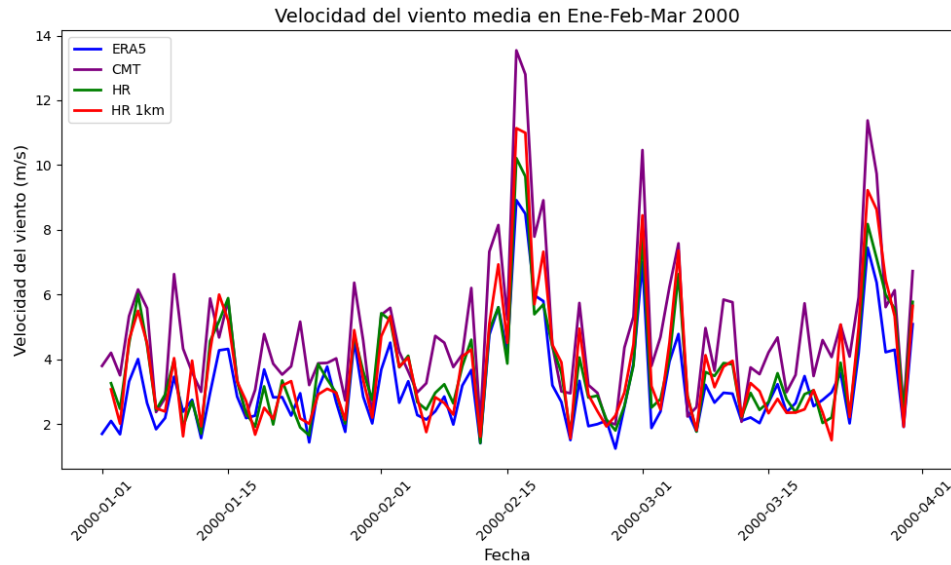


Ilustración 35. Metamodelo - Datos Históricos CMT. Primer trimestre 2020.

En el caso de las estaciones de CMT y Parayas, la dirección del viento está registrada con una resolución angular de 10° , lo que ha requerido una adaptación en el metamodelo para asegurar una comparación adecuada. Como se ha observado previamente, las magnitudes del viento registradas tienden a ser superiores a las simuladas por el modelo, lo que se refleja en una mayor densidad y longitud de los vectores en la rosa correspondiente a los datos observados (izquierda). Sin embargo, en términos de dirección, ambos conjuntos muestran una concordancia notable, con un claro predominio de flujos del oeste (270°). Esto indica que, a pesar de la subestimación en la magnitud, el metamodelo reproduce correctamente el patrón direccional del viento en la zona.

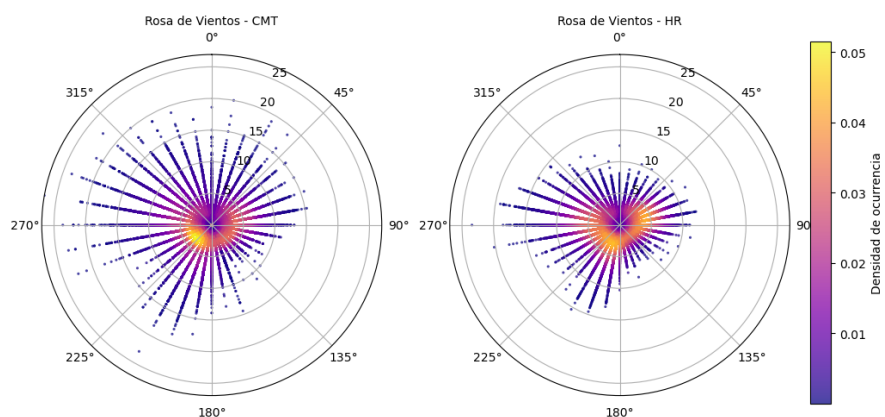


Ilustración 36. Dirección del viento medida en campo (CMT) (izquierda) y la predicha por el metamodelo (derecha).

5.2.4. Evaluación de la estación de Parayas

Por último, se analiza la estación de Parayas, la cual servirá para comparar sus resultados con los obtenidos en la estación CMT. Esta comparación permitirá validar tanto la correcta lectura de los datos en ambas estaciones como el nivel de precisión del metamodelo.

Como en los anteriores casos, el modelo de ERA5 subestima el valor del viento en el punto de estudio, aunque en este los valores medios sí son recogidos y son principalmente los picos los valores que no se terminan de asemejar de manera precisa.

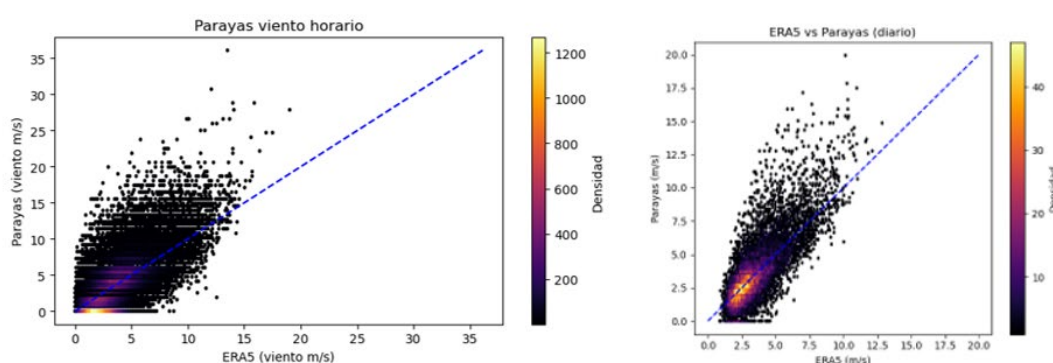


Ilustración 37. Correlación horaria ERA5 – Parayas.

La figura muestra la evolución histórica de la velocidad del viento en Parayas, comparando los datos registrados por la estación local con las estimaciones del modelo ERA5. A lo largo del tiempo, se puede apreciar que ERA5 tiende a representar velocidades máximas más bajas que las observadas, especialmente durante los eventos de viento más extremos. Esta diferencia es más pronunciada en los primeros años del registro, donde los picos medidos por la estación superan con claridad a los del modelo.

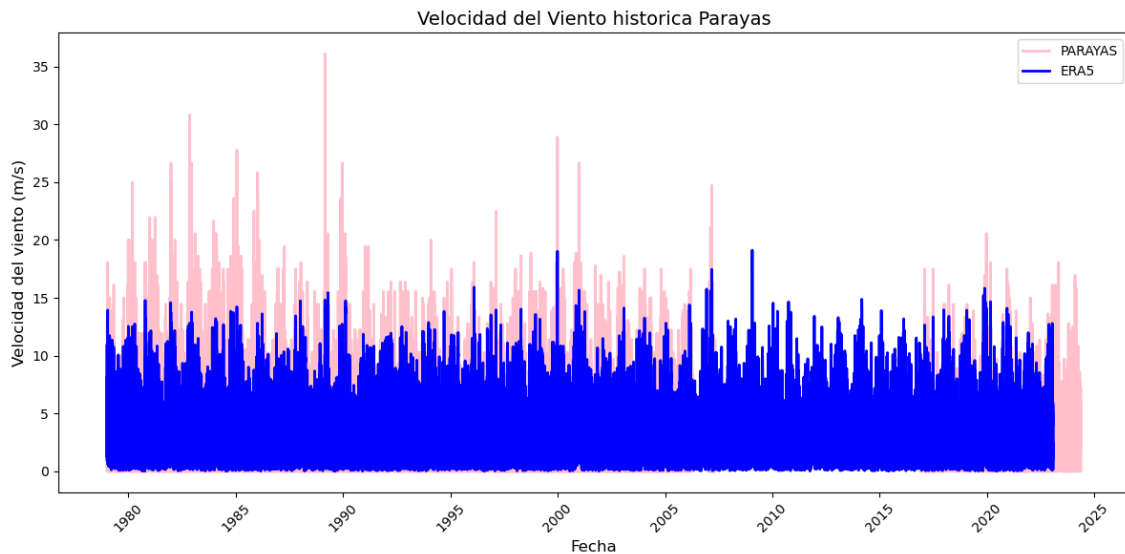


Ilustración 38. Datos Históricos Parayas - ERA5.

Para comparar cómo se ajustan los datos estimados por los distintos modelos con las mediciones reales de viento en Parayas, a continuación, se muestran tres gráficos. Cada uno de ellos enfrenta los valores registrados por la estación con los obtenidos por ERA5, el modelo de alta resolución a 100 metros y el modelo a 1 kilómetro.

En principio, se esperaría que el modelo de 100 metros fuera el más preciso, ya que trabaja con mayor nivel de detalle. Sin embargo, en los gráficos se ve que este modelo tiene más dispersión y errores que el de 1 kilómetro. Por su parte, el modelo de 1 kilómetro muestra una mejor relación con los datos reales. En cuanto a ERA5, vuelve a quedar claro que tiende a dar valores más bajos que los registrados por la estación, especialmente en los días con viento fuerte.

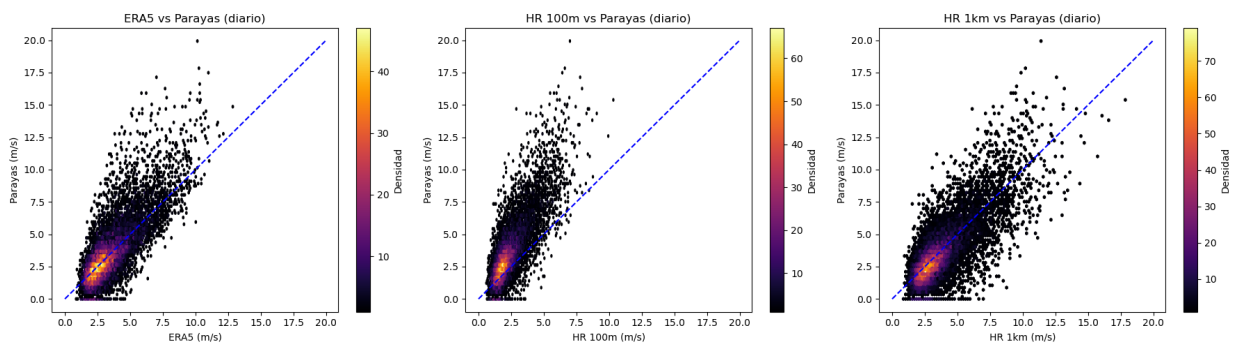


Ilustración 39. Comparativa Metamodelo - Datos Históricos Parayas.

Al igual que en el caso anterior, se muestra un detalle del primer trimestre del año 2000. En esta próxima ilustración se observa que tanto el modelo ERA5 como el modelo de alta resolución de 1 km reproducen de forma razonablemente adecuada la variabilidad y los picos máximos de velocidad del viento registrados, siguiendo en

términos generales la evolución temporal de los datos observados. En particular, ERA5 presenta una ligera subestimación, pero mantiene una buena correlación con las observaciones, mientras que el modelo HR 1 km muestra una mayor sensibilidad a los eventos de viento fuerte, reflejando con mayor precisión los máximos.

Contrariamente a lo esperado, el modelo de más alta resolución espacial (HR 100 m) no mejora la representación del viento; por el contrario, presenta una subestimación sistemática de la velocidad media y una mayor divergencia respecto a los datos observados, especialmente durante los eventos de viento intenso.

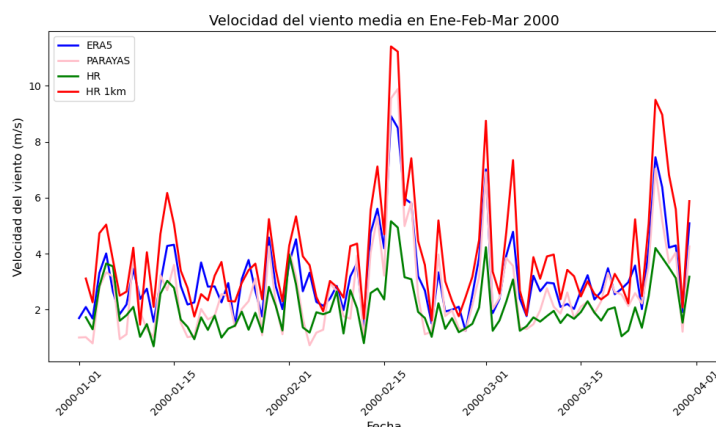


Ilustración 40. Metamodelo - Comparación Histórica Parayas. Primer trimestre 2000

Las direcciones del viento han sido elaboradas a partir de las salidas del modelo de alta resolución con malla de 100 metros, lo cual explica en parte la subestimación sistemática observada en la magnitud del viento en la siguiente ilustración.

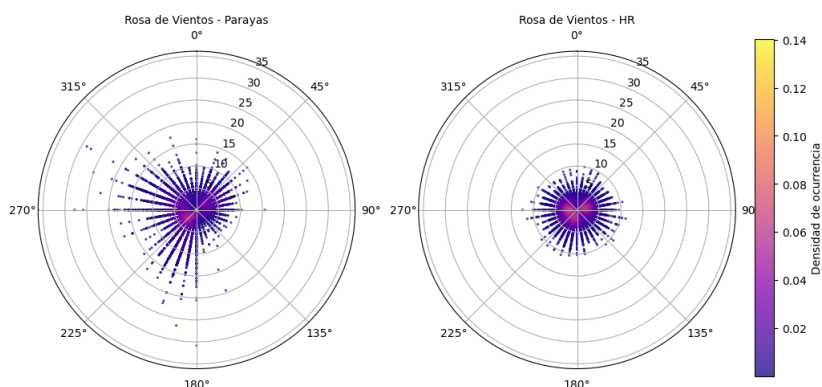


Ilustración 41. Dirección del viento medida en campo (Parayas) (izquierda) y la predicha por el metamodelo (derecha).

Esta menor intensidad, ya detectada previamente en las comparaciones con datos observacionales, puede atribuirse a la mayor sensibilidad del modelo frente a efectos locales, como la fricción superficial, la rugosidad del terreno o el apantallamiento orográfico, todos ellos factores que se podrían explicar por la

cercana presencia del estuario en la zona donde se encuentra la estación meteorológica.

5.3 Estudio de la Rugosidad del terreno

Como se ha señalado anteriormente, el análisis de la rugosidad del terreno es un aspecto clave para la correcta estimación de la velocidad del viento, especialmente en aplicaciones relacionadas con la modelización y predicción eólica.

Es por ello que, a continuación, se mostrará un estudio sobre la rugosidad del terreno en los diferentes metamodelos, ya que comprender este parámetro y su influencia en la magnitud del viento es clave para una futura predicción precisa del recurso eólico en los aerogeneradores. Esto resulta especialmente relevante si se tiene en cuenta que los metamodelos utilizados proporcionan velocidades del viento a 10 metros sobre el nivel del suelo, mientras que los aerogeneradores se sitúan a una altura aproximada de 55 metros. Por tanto, el análisis de la rugosidad permitirá evaluar con mayor precisión cómo evoluciona la velocidad del viento en el perfil vertical y cómo puede corregirse la magnitud estimada en superficie para representar adecuadamente las condiciones reales en altura.

Con el objetivo de comprender por qué el metamodelo a 100 m tiende a subestimar los valores de viento, se ha analizado un instante representativo comparando los campos de viento generados por los distintos modelos. En la siguiente ilustración se muestran cuatro visualizaciones: en la parte superior izquierda, el viento interpolado de ERA5 (baja resolución); en la parte superior derecha, la reconstrucción con el metamodelo a 1 km para toda la región de Cantabria; en la parte inferior izquierda, un zoom local de esa misma reconstrucción centrado en la zona de Santander; y en la parte inferior derecha, el campo de viento generado por el metamodelo de 100 m.

Como puede observarse, la reconstrucción a 1 km ofrece una transición más suave y realista entre el mar y la tierra, capturando mejor los patrones locales que el modelo de ERA5, cuya resolución es más grosera. Esta mejora se traduce en una mejor representación del viento en ubicaciones como la boya VM.

Por el contrario, la reconstrucción a 100 m muestra velocidades de viento más bajas, sobre todo en la estación de Parayas. Aunque en VM y CMT las diferencias no son tan marcadas, en Parayas la subestimación es clara, lo que ayuda a explicar el comportamiento menos preciso de este modelo a mayor resolución.

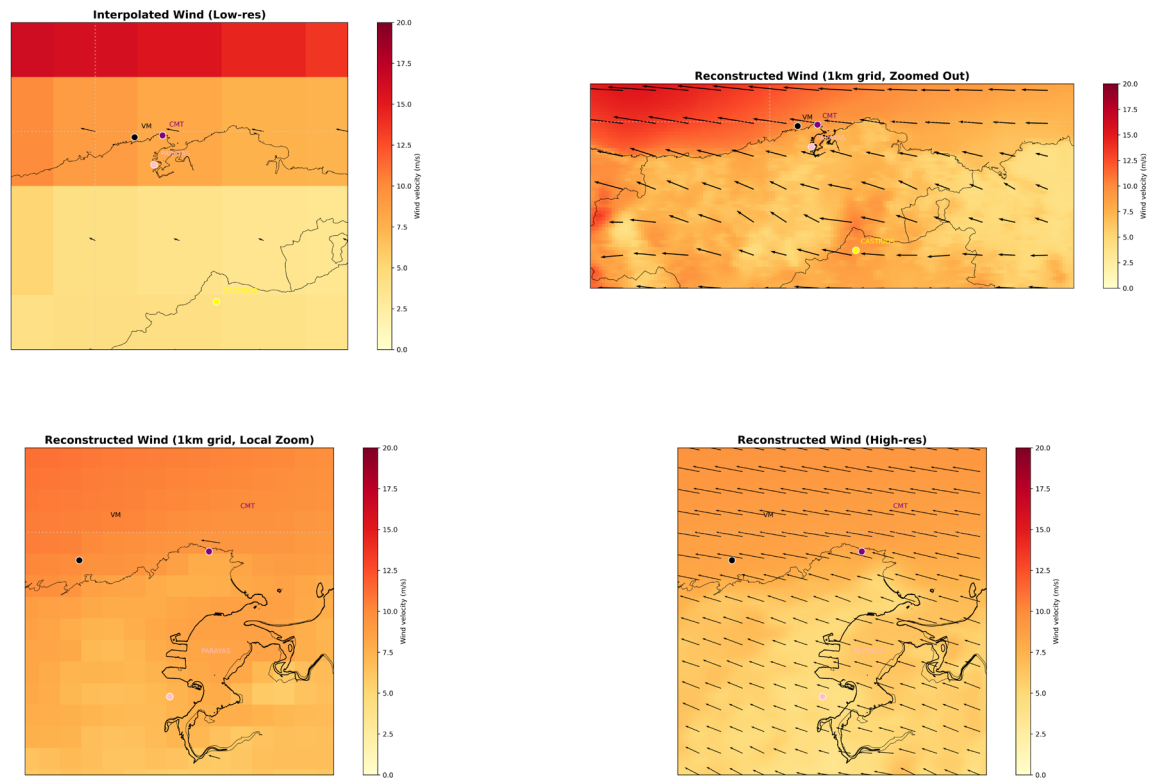


Ilustración 42. Comparación de los campos de viento generados por distintos modelos en un instante representativo. Arriba a la izquierda: viento interpolado de ERA5. Arriba a la derecha: reconstrucción con el metamodelo de 1 km en la región de Cantabria. Abajo a la izquierda: zoom de esa misma reconstrucción en la zona de Santander. Abajo a la derecha: reconstrucción con el metamodelo de 100 m.

En el presente estudio, se observa que la boya instalada en Virgen del Mar se encuentra a una altura de 3 metros, mientras que las estaciones meteorológicas de Parayas y el Centro Meteorológico Territorial (CMT) operan a 10 metros de altura. Esta discrepancia implica la necesidad de aplicar correcciones a los datos obtenidos por la boya, a fin de homogeneizarlos con los modelos numéricos, los cuales están calibrados para una altura de referencia estándar de 10 metros. Para ello, se utiliza una corrección logarítmica basada en la rugosidad del terreno, asumiendo un valor típico de 0.002 metros para superficies marítimas, este valor se justifica en Tabla 1. Coeficiente de rugosidad que se presenta más adelante.

Es importante destacar que este aspecto cobra aún mayor relevancia al simular el recurso eólico para el diseño y la evaluación del parque eólico proyectado, ya que las predicciones deben realizarse a una altura de 55 metros. Por ello, resulta fundamental contar con una buena caracterización del perfil vertical del viento, teniendo en cuenta tanto la rugosidad del terreno como las condiciones locales, para asegurar la precisión y fiabilidad de los resultados.

Es importante tener en cuenta que en el presente TFM, la predicción en la localización del parque eólico se hará con el modelo de 1 km de resolución. Este modelo no es el óptimo ya que para Santander se dispone de un modelo de alta

resolución de 100 m, como se enseñó previamente, pero servirá para una primera aproximación y la validación de si esta metodología puede ser aplicada en más profundidad en un futuro. Debido a esto, averiguar el porqué de las discrepancias entre el modelo de 100 m y 1 km es de vital importancia.

A continuación, se muestra la rugosidad que aplica cada modelo de alta resolución.

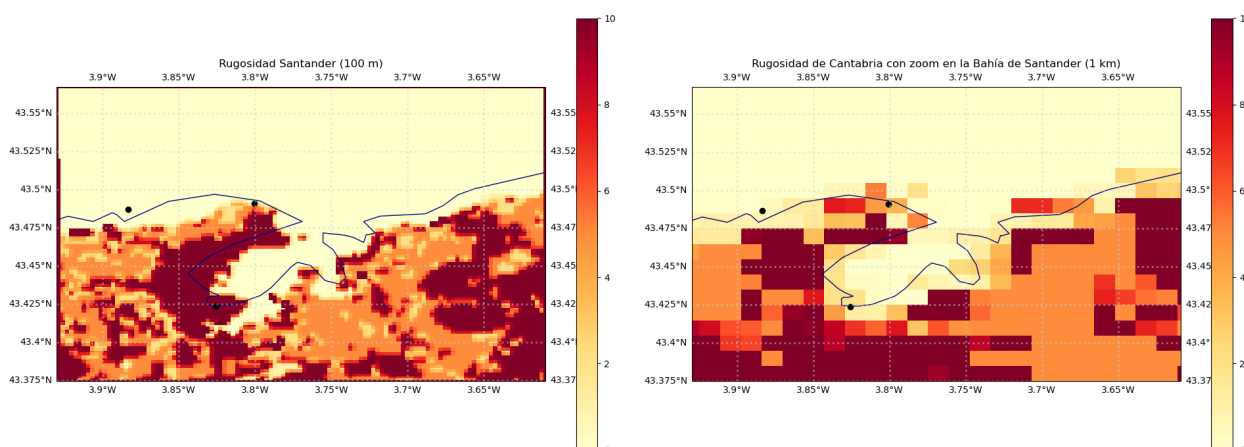


Ilustración 43. Rugosidad modelo a 100 m resolución (izquierda) y de 1 km de resolución (derecha).

Como puede observarse, el modelo con una resolución de 100 metros ofrece una representación mucho más detallada de la rugosidad del terreno, debido a su mayor resolución espacial. No obstante, se ha detectado un posible desplazamiento hacia la derecha en este modelo, ya que la zona del estuario de Santander aparece con una rugosidad anómalamente elevada. En contraste, el modelo con resolución de 1 kilómetro representa dicha área de manera más coherente con las características reales del terreno. Se considera que esta discrepancia podría ser la causa del error observado en los datos registrados por la estación meteorológica de Parayas que se encuentra justo en el límite de esa diferencia.

Al ampliar la visualización sobre la zona del parque eólico de Castríos, se observa que la superficie del parque queda contenida dentro de únicamente cuatro celdas de la malla de 1 km. Aunque esta resolución puede ser limitada para representar con precisión las diferencias locales entre aerogeneradores, se ha utilizado como punto de partida en ausencia de una malla más fina para esta región de Burgos.

El objetivo principal de este análisis no es solo identificar cuántas celdas cubren el parque, sino examinar si existen variaciones importantes de topografía y rugosidad dentro de esta zona. Estas características influyen directamente en la distribución espacial del viento y, por tanto, en el rendimiento del metamodelo. Además, resulta clave revisar los valores de rugosidad empleados en las simulaciones con WRF y considerar si estos pueden explicar las diferencias observadas con los datos reales. Finalmente, este análisis también permite estimar qué valores de rugosidad serían adecuados para extrapolar los resultados del metamodelo (referenciados a 10 metros) a la altura real de operación de los aerogeneradores, que en este caso es de 55 metros.

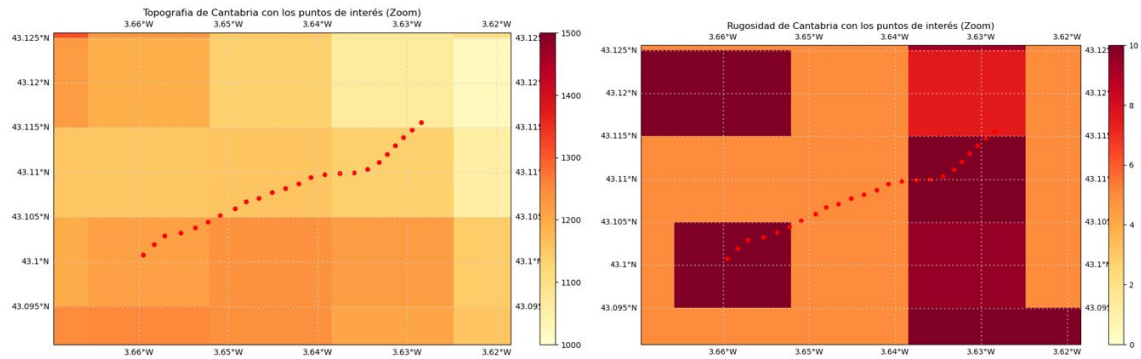


Ilustración 44. Topografía a la izquierda y Rugosidad a la derecha del parque eólico "Castríos".

5.4 Parametrización del emplazamiento del parque eólico

A continuación, se abordará el análisis del parque eólico objeto de este estudio. El parque se ha dividido en distintos grupos en función de la proximidad de cada aerogenerador a los nodos que conforman la malla de 1 km de resolución. Además de los 24 aerogeneradores, el parque cuenta también con un anemómetro independiente que servirá para la validación de los datos tanto observados como medidos en el propio aerogenerador.

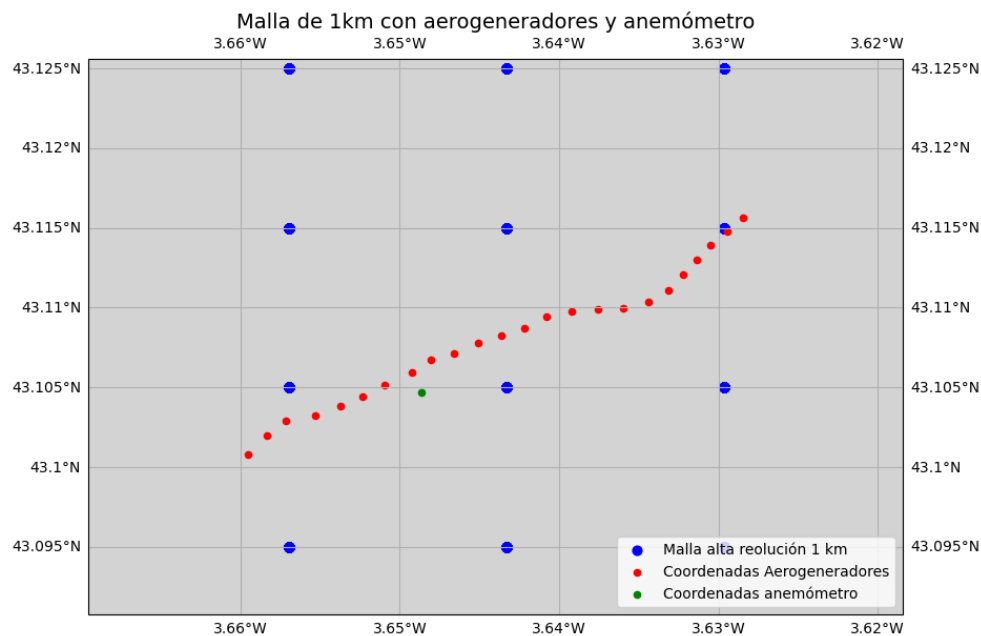


Ilustración 45. Malla de 1 km de resolución con los elementos del parque eólico "Castríos".

5.4.1. Rugosidad del parque eólico

Como se explicó anteriormente, conocer la rugosidad del terreno es clave para una correcta modelización del viento con el metamodelo. En la Ilustración 44.

Topografía a la izquierda y Rugosidad a la derecha del parque eólico "Castríos" se aprecia que el modelo de alta resolución de 1 km considera 3 rugosidades diferentes para todo el parque, y es que el valor de la rugosidad está directamente relacionado con el perfil de viento a distintas alturas.

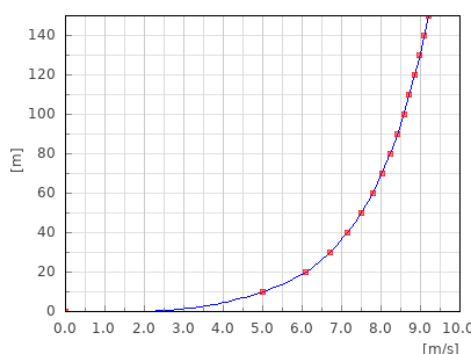


Ilustración 46. Gráfica del viento según elevación. Fuente: Suisse Éole

En este estudio, se adopta un perfil logarítmico del viento para caracterizar la variación de la velocidad del viento con la altura. A nivel del suelo, el viento es fuertemente afectado por la rugosidad del terreno y los obstáculos; sin embargo, a medida que se gana altitud, su velocidad aumenta hasta estabilizarse en niveles donde deja de estar influido por la superficie.

En ese tramo intermedio, el perfil logarítmico ofrece una estimación razonable de la velocidad del viento en función de la altura y de la rugosidad del terreno. Esta caracterización se expresa mediante la siguiente fórmula:

$$v_2 = v_1 \cdot \frac{\ln(h_2 / z_0)}{\ln(h_1 / z_0)}$$

Siendo:

- v_2 la velocidad en el punto 2
- v_1 la velocidad en el punto 1
- h_2 la altura en el punto 2
- h_1 la altura en el punto 1
- z_0 el coeficiente de rugosidad del terreno

El valor de z_0 puede ser estimado según la siguiente tabla:

Clase de rugosidad (z_0)	Longitud de rugosidad (m)	Tipos de superficies del terreno
0	0,0002	Superficies de agua: mar y lagos

Clase de rugosidad (z_0)	Longitud de rugosidad (m)	Tipos de superficies del terreno
0.5	0,0024	Terrenos abiertos con una superficie lisa, por ejemplo, hormigón, pistas de aeropuertos, césped cortado, etc.
1	0,03	Tierras agrícolas abiertas sin cercas ni setos, posiblemente con edificios muy dispersos y colinas muy suaves
1.5	0,055	Terreno agrícola con algunas casas y setos de 8 m de altura a una distancia de más de 1 km
2	0,1	Terreno agrícola con algunas casas y setos de 8 metros de altura a una distancia de unos 500 m
2.5	0,2	Terreno agrícola con muchas casas, arbustos y plantas, o setos de 8 m de altura a una distancia de unos 250 m
3	0,4	Pueblos, pequeñas ciudades, tierras agrícolas con muchos setos o setos altos, bosques y terrenos muy accidentados y accidentados
3.5	0,6	Ciudades más grandes con edificios altos
4	1,6	Grandes ciudades con edificios altos y rascacielos

Tabla 1. Coeficiente de rugosidad.

Dado que en el caso de un parque eólico en terreno montañoso no existe un valor específico que se ajuste directamente a las clases de rugosidad presentadas en la tabla anterior, se opta por ajustar un perfil logarítmico de vientos a los registros de vientos disponibles en el parque eólico a diferentes alturas. Esta aproximación se basa en la relación entre el anemómetro, ubicado a una altura de 10 metros, y el aerogenerador más cercano, situado a una distancia horizontal de 200 metros y una altura de 55 metros.

Para obtener el valor del coeficiente de rugosidad se considera que los datos más apropiados serán los datos provenientes del sureste ya que como se puede apreciar en la Ilustración 45. Malla de 1 km de resolución con los elementos del parque eólico de Castrís. es la dirección del viento en la que el anemómetro no tendría ningún tipo de apantallamiento por parte del aerogenerador. De esta manera se obtiene un coeficiente de rugosidad igual a 0,6702.

Es importante destacar que este valor final ha sido obtenido como un valor medio entre los diferentes instantes de tiempo ajustados.

A continuación, se muestra la relación entre el anemómetro y el aerogenerador con el coeficiente de rugosidad obtenido para unos datos aleatorios.

Perfiles logarítmicos vs. mediciones reales a 55 m (24 horas seleccionadas de forma aleatoria)

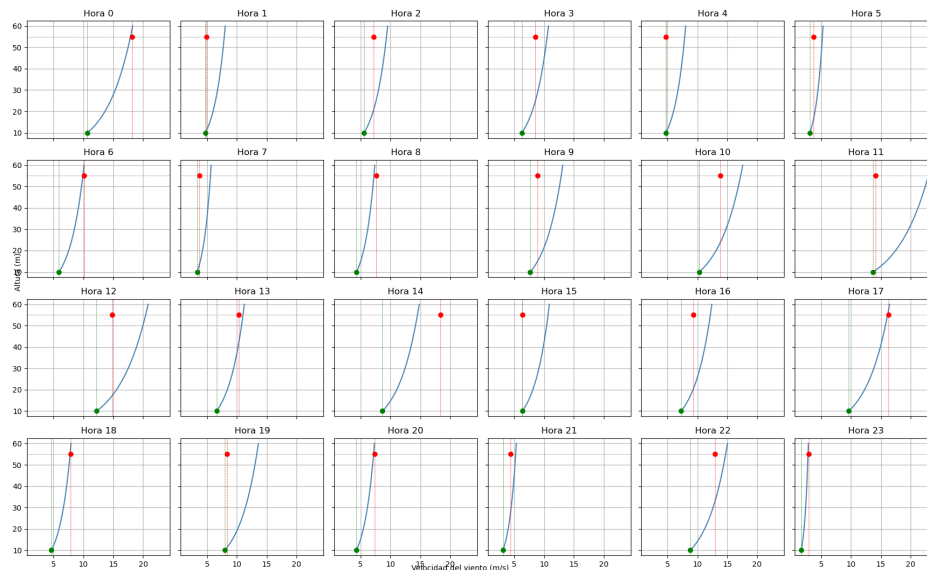


Ilustración 47. Perfiles logarítmicos para el coeficiente de rugosidad obtenido.

5.4.2. Comparativa entre el metamodelo y los datos históricos

En el presente apartado se realiza una comparativa entre los datos medidos por los aerogeneradores durante los últimos diez años y las estimaciones del viento proporcionadas por el metamodelo de alta resolución. Tal como se explicó anteriormente, este metamodelo se basa en una malla de resolución espacial de 1 km, lo que implica que múltiples aerogeneradores pueden estar asociados a un mismo punto de la malla, recibiendo por tanto una predicción idéntica. Esta limitación podrá ser abordada en el futuro mediante la implementación de un modelo con resolución de 100 metros.

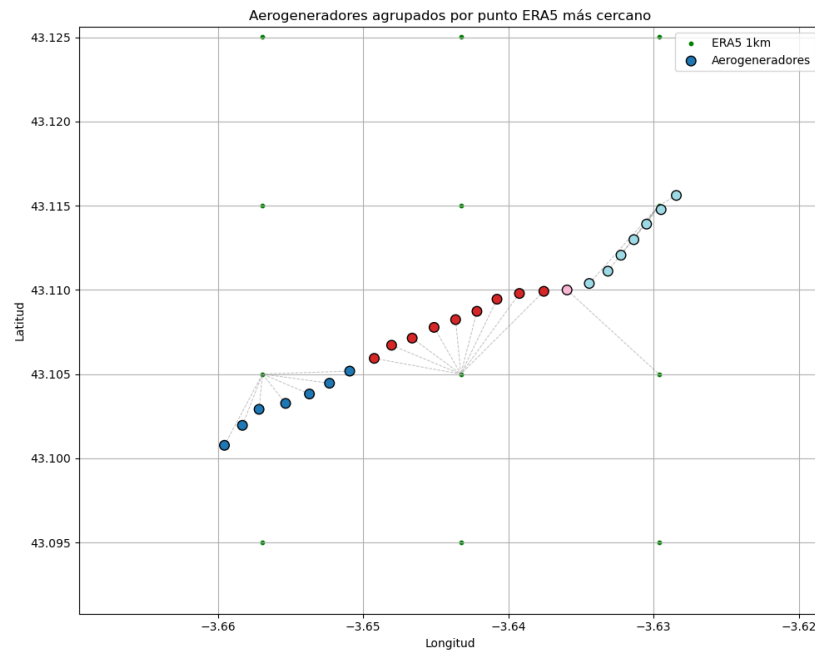


Ilustración 48. Aerogenerador con punto del modelo de referencia.

Se identifican un total de cuatro puntos de referencia del metamodelo para los 24 aerogeneradores que conforman el parque. El aerogenerador 23 se encuentra prácticamente coincidente con uno de los puntos del modelo, mientras que el aerogenerador 17 presenta una mayor distancia respecto al punto de referencia asignado. Esta diferencia espacial podría influir en la precisión de las futuras predicciones del recurso eólico. No obstante, para evaluar dicho impacto, resulta necesario analizar en primera instancia el metamodelo con todos los datos históricos disponibles.

5.4.2.1 Grupo 1 (Aerogeneradores 1-7)

A continuación, se muestra la comparación histórica tanto a escala horaria como diaria entre la reconstrucción del modelo de 1 km de resolución y los datos medidos por los aerogeneradores en los últimos 10 años.

En general, se observa una alta concentración de puntos alrededor de la diagonal para la mayoría de los aerogeneradores, especialmente en rangos de viento entre 5 y 15 m/s. Sin embargo, también pueden apreciarse ciertos patrones de dispersión, donde los valores reconstruidos tienden a sobreestimar o subestimar el viento medido en diferentes grados según el aerogenerador.

Densidad: Viento Medido vs Viento Reconstruido (Media Diaria Aerogeneradores 1 - 7)

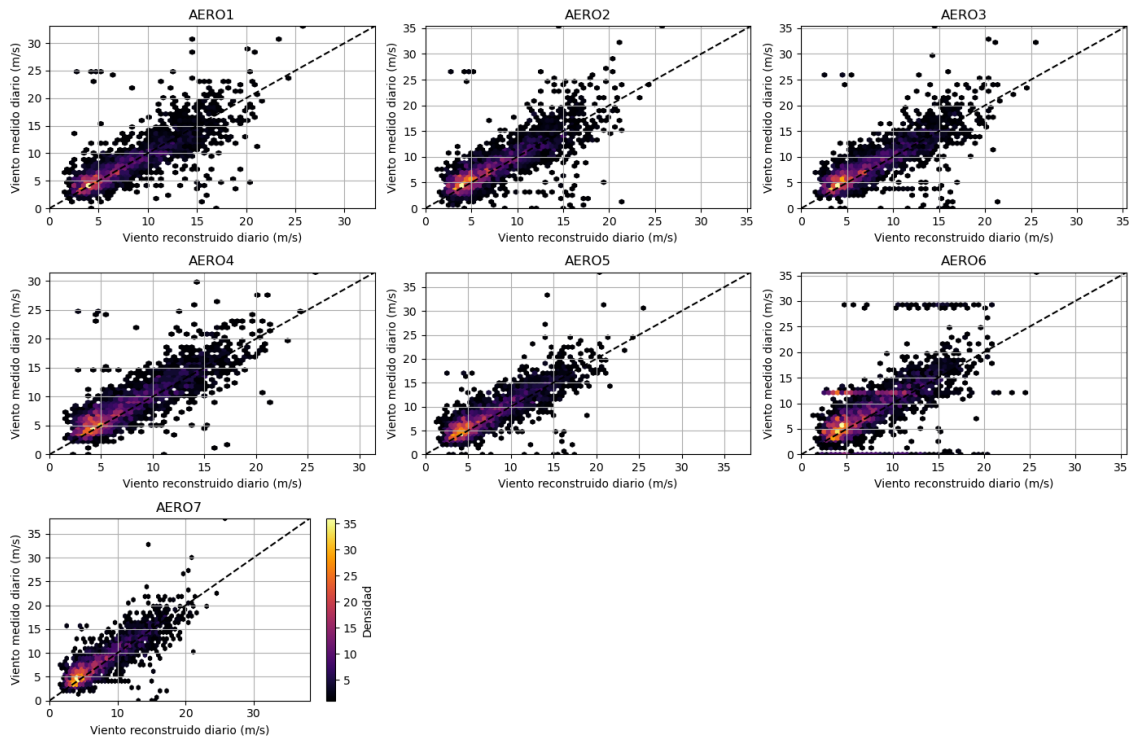


Ilustración 49. Comparativa diaria HR 1km-Aerogeneradores Grupo 1

En el siguiente caso se puede apreciar de manera más clara el comportamiento del metamodelo frente a los diferentes aerogeneradores y como cada aerogenerador se comporta de manera diferente debido a su variabilidad espacial por lo que será recomendable la aplicación de un modelo de 100 m de resolución en el futuro. Pero en general se ve un acertado del comportamiento del metamodelo a la hora de recrear situaciones pasadas en un entorno montañoso.

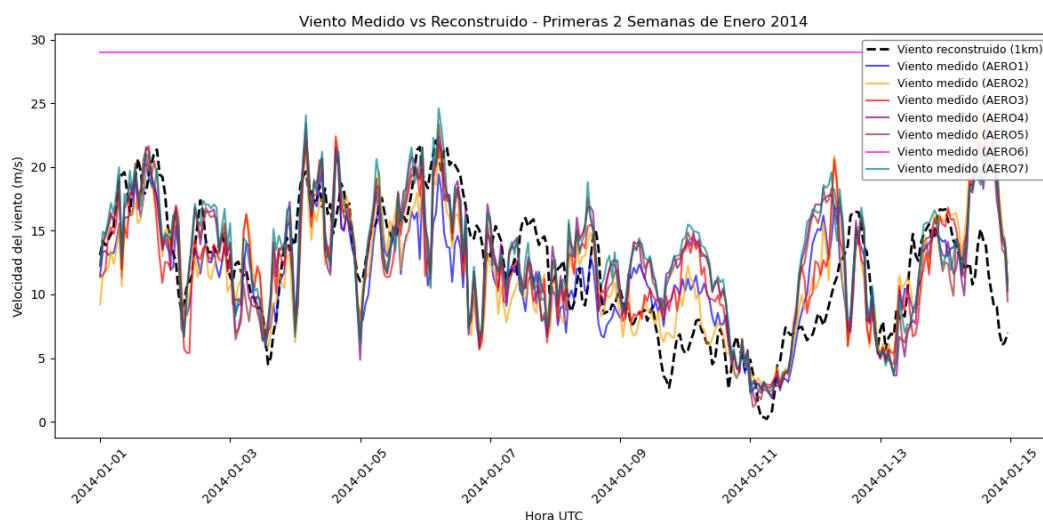


Ilustración 50. Comparativa horaria en las dos primeras semanas de 2014 en el Grupo 1.

5.4.2.2 Grupo 2 (Aerogeneradores 8-15)

El grupo 2 está formado desde el aerogenerador 8 hasta el 15, de nuevo se muestra la comparación histórica entre la reconstrucción del modelo de 1 km de resolución y los datos medidos por los aerogeneradores en los últimos 10 años.

En este caso, la reconstrucción también resulta bastante precisa en las zonas de mayor densidad. No obstante, se observan numerosos valores que se alejan de la diagonal, pero como puede apreciarse, estos suelen corresponder a datos aislados.

Densidad: Viento Medido vs Viento Reconstruido (Media Diaria Aerogeneradores 8 - 15)

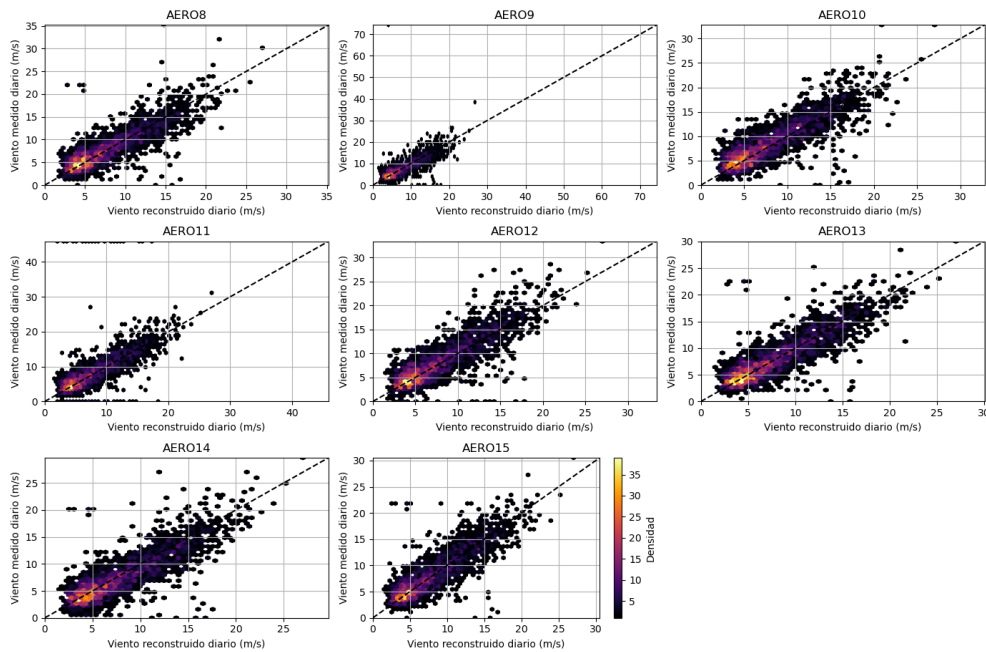


Ilustración 51. Comparativa diaria HR 1km-Aerogeneradores Grupo 2.

En la siguiente ilustración se observa que las series de viento medidas en estos aerogeneradores muestran una alta coherencia entre sí, lo que indica una menor variabilidad espacial en comparación con el grupo anterior (AERO1–AERO7). Esta homogeneidad sugiere que los aerogeneradores del grupo 2 están ubicados en una zona con condiciones de viento más uniformes lo cual favorece una mayor consistencia entre las mediciones.

A pesar de esta menor variabilidad espacial, se mantiene una discrepancia clara entre el viento reconstruido y el medido, especialmente en ciertos valles de la serie temporal.

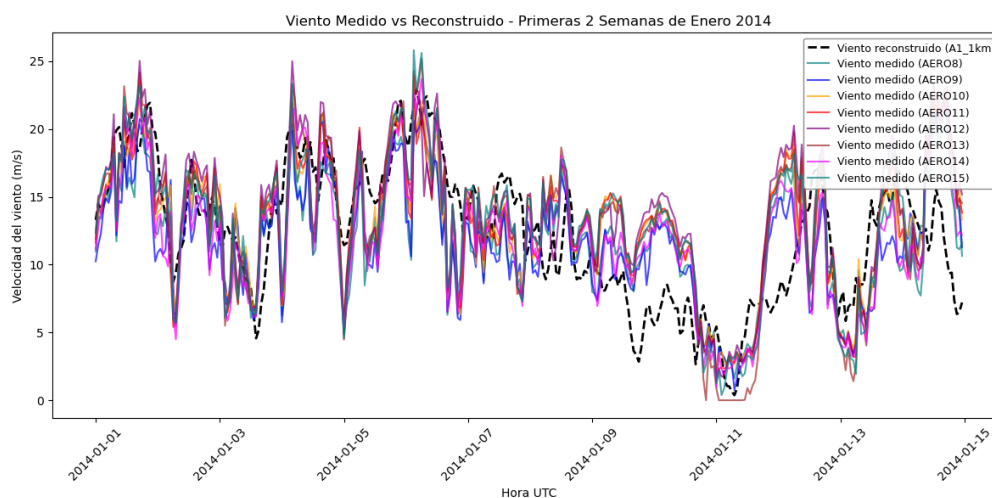


Ilustración 52. Comparativa horaria en las dos primeras semanas de 2014 en el Grupo 2.

5.4.2.3 Grupo 3 (Aerogenerador 16)

En el caso del grupo 3, esta únicamente compuesto por el aerogenerador 16.

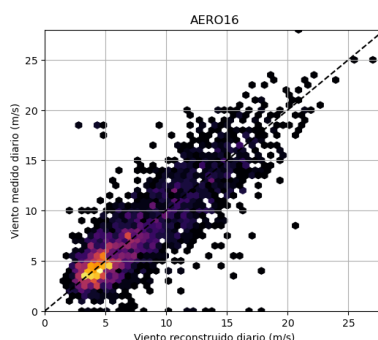


Ilustración 53. Comparativa diaria HR 1km-Aerogenerador Grupo 3.

De manera similar a los casos anteriores, la mayoría de los datos se concentran en la diagonal. No obstante, se observan algunos valores alejados de ella, pero siendo en su mayoría casos aislados.

En este caso, se observa de manera más evidente la discrepancia entre ambas series, especialmente en los valores máximos y mínimos, al no haber tantos aerogeneradores en el grupo. A pesar de que las tendencias generales siguen una evolución similar, las diferencias en la magnitud de los eventos son más marcadas que en los casos anteriores.

Este comportamiento puede atribuirse a la ubicación geográfica del aerogenerador 16, que corresponde al aerogenerador más alejado del dominio central del modelo de reconstrucción.

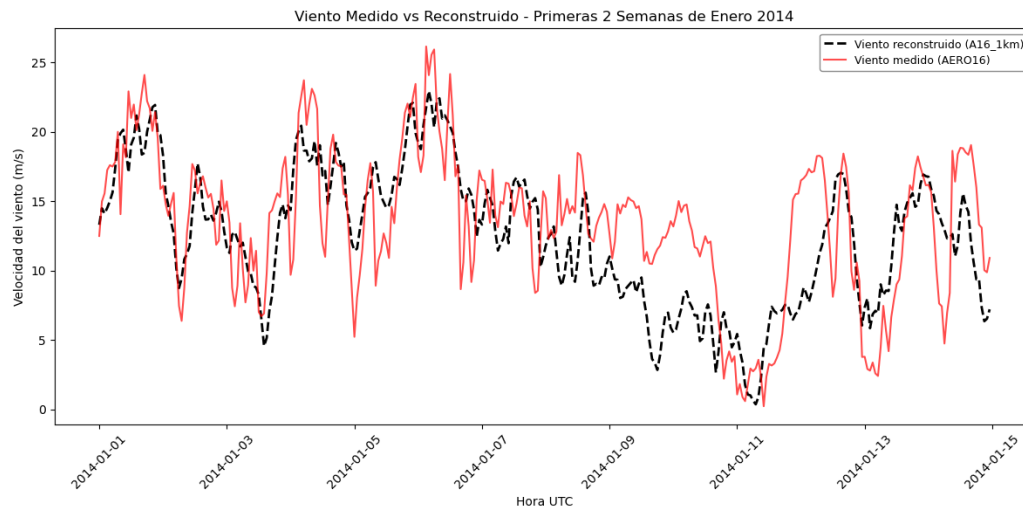


Ilustración 54. Comparativa horaria en las dos primeras semanas de 2014 en el Grupo 3.

5.4.2.4 Grupo 4 (Aerogeneradores 17-24)

Por último, se muestra la comparativa entre el viento reconstruido y el medido en el grupo 4 compuesto desde el aerogenerador 17 al 24.

En este caso, se puede observar como el aerogenerador 19 presenta una serie de anomalías. Es por ello que no se utiliza dicho aerogenerador para la validación de los datos. En los casos restantes, se observa que se sigue la tendencia anterior con la mayoría de datos sobre la diagonal y algún dato suelto disperso.

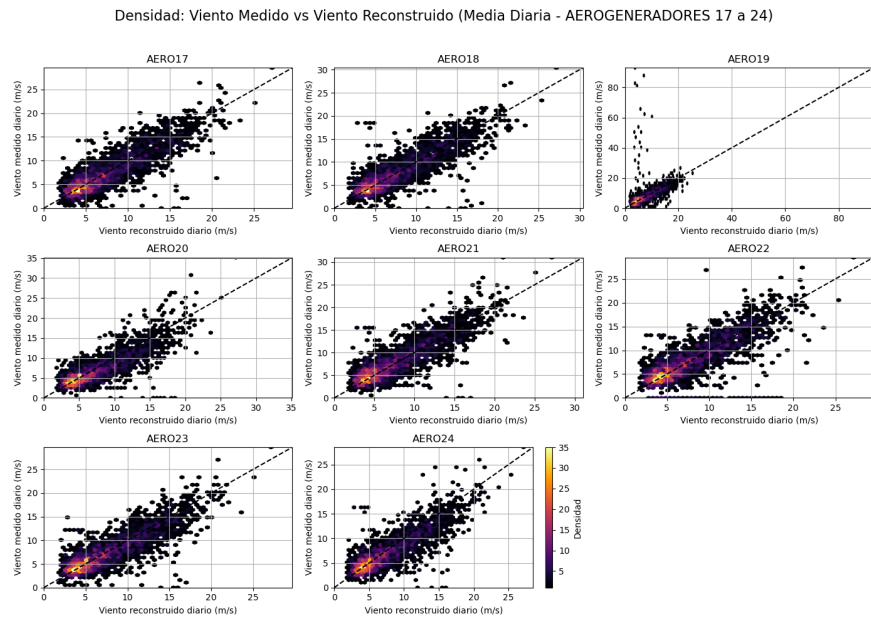


Ilustración 55. Comparativa diaria HR 1km-Aerogenerador Grupo 4.

En la siguiente ilustración, se observa una clara mejora con respecto a los anteriores. Esto se debe a que, como se explicó, el punto del metamodelo coincide con la ubicación de un aerogenerador, lo que proporciona una idea de cómo podría comportarse el metamodelo con una resolución de 100 metros.

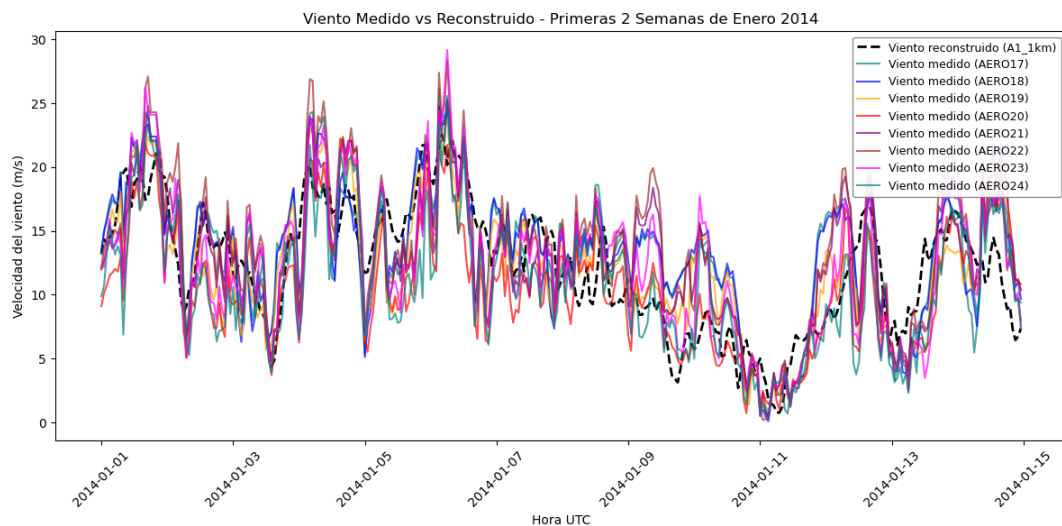


Ilustración 56. Comparativa horaria en las dos primeras semanas de 2014 en el Grupo 4.

6. Predicción

Una vez validado el metamodelo desarrollado, se da paso a la etapa de predicción del viento. Para ello, se emplean los datos provenientes del modelo AIFS previamente mencionado, a los cuales se les aplica el metamodelo elaborado.

Con el fin de mejorar la precisión de las predicciones, se propone la construcción de un segundo metamodelo. Este se basa en la estructura del primero, pero se ajusta utilizando los registros históricos específicos de cada aerogenerador, los mismos que fueron empleados en la validación del modelo original. Aunque esta estrategia no es habitual en la generación de metamodelos, ya que utilizar datos empleados en la validación puede derivar en una reconstrucción del fenómeno en lugar de una predicción genuina, se implementa en este caso con el objetivo de evaluar si dicha personalización permite obtener estimaciones más precisas. Asimismo, esta metodología posibilita la obtención de predicciones individualizadas para cada aerogenerador, en lugar de una estimación conjunta para grupos de aerogeneradores.

Como se muestra a continuación, este nuevo metamodelo es muy similar al original, observándose muy pocas diferencias. Entre estas, destaca una mejor apreciación de los picos y la eliminación de muchos valores residuales que superaron el primer filtro.

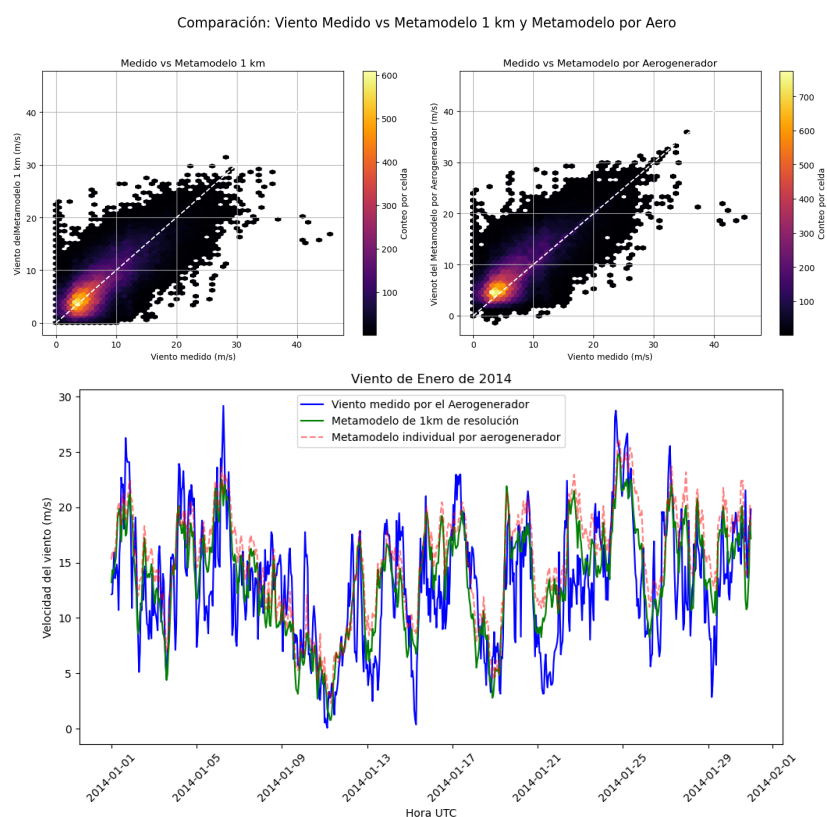


Ilustración 57. Comparativa entre el metamodelo original de 1km de resolución y el independiente para cada aerogenerador.

Ambos metamodelos se comparan, tanto entre sí como con el modelo de predicción actualmente utilizado en el parque eólico, y con los registros reales de viento correspondientes a los periodos de estudio.

Cabe señalar que una limitación relevante del presente trabajo radica en la resolución temporal de los datos disponibles, ya que, al utilizar un software de libre acceso, estos se presentan con una frecuencia de tres horas. En aplicaciones habituales de predicción eólica, se trabaja generalmente con datos horarios o incluso de mayor resolución. Por tanto, estos primeros metamodelos no solo permiten evaluar la viabilidad de la metodología propuesta, sino que también constituyen una aproximación inicial. En caso de que los resultados obtenidos sean satisfactorios, se valorará la incorporación de datos con mayor resolución temporal en etapas posteriores del estudio.

6.1 Predicción de viento

En esta sección se presentan los resultados obtenidos a partir de la aplicación de los distintos modelos de predicción del viento. Se analizan y comparan tres enfoques principales:

- Un metamodelo individualizado, que proporciona una predicción del viento específica para cada aerogenerador del parque.
- Un metamodelo agrupado dividido en 4 grupos, que ofrece una estimación general del viento para cada conjunto de aerogeneradores.
- El modelo operativo actual del parque, que entrega únicamente una media global del viento para todo el parque eólico, sin distinción entre aerogeneradores ni grupos.

El objetivo de esta comparación es evaluar el nivel de precisión y detalle que aporta cada enfoque de predicción, así como determinar si la incorporación de metamodelos, especialmente los individualizados por aerogenerador, permite mejorar la calidad de las predicciones respecto al modelo actualmente en uso.

A continuación, se exponen los resultados correspondientes a la primera semana en el que se pudo llevar a cabo la predicción, concretamente del 19 de junio al 22 de junio.

Dado que el metamodelo desarrollado genera predicciones individualizadas para cada uno de los 24 aerogeneradores, y con el fin de facilitar una interpretación clara y visualmente limpia, en la gráfica únicamente se representa la media de todos ellos. No obstante, también se dispone de las predicciones individuales para cada aerogenerador, que permiten un análisis más detallado en fases posteriores.

En primer lugar, se presenta la comparativa correspondiente a 24 horas vista, ya que es la más relevante para la operación del parque, siendo la que determina la estimación de producción eólica a corto plazo y la base para la planificación y venta de energía.

En esta ilustración se puede observar cómo, para los primeros cuatro días predichos, ambos metamodelos tienden a infraestimar la velocidad del viento en comparación con la media horaria real medida. Además, se incluye la serie del servicio contratado por Castríos, que también muestra valores generalmente más bajos que los datos reales.

Se evidencia que la variabilidad y los picos de viento capturados por la medición real no son del todo representados por los modelos, lo que sugiere que, aunque captan la tendencia general, no alcanzan a reflejar completamente la intensidad del viento registrado en esas fechas. Esto indica una necesidad de seguir ajustando los metamodelos para mejorar la precisión en escenarios de corto plazo.

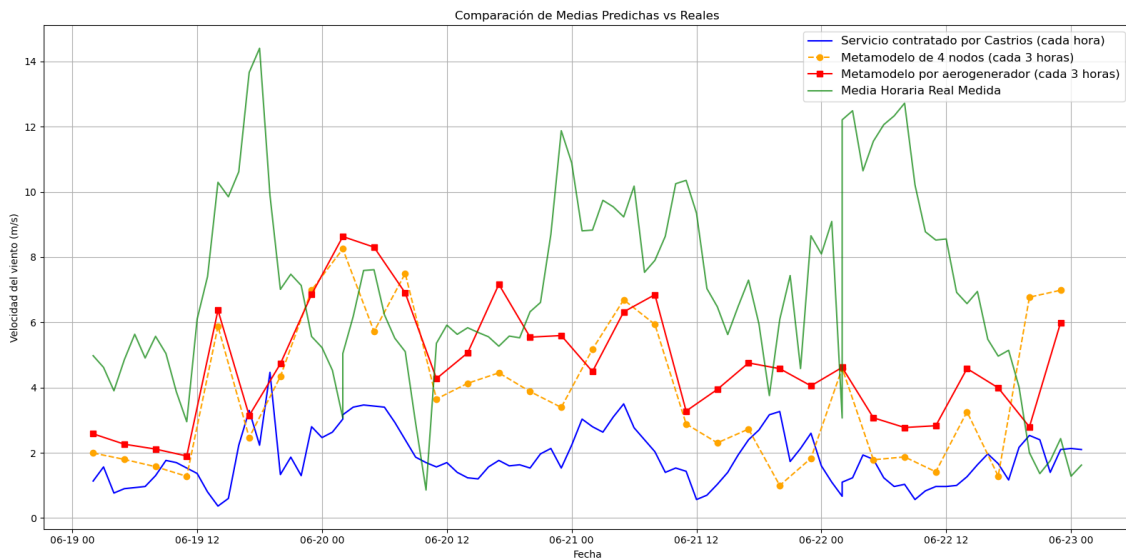


Ilustración 58. Comparación de medidas reales y predichas del 19 al 22 de junio.

Para la siguiente semana de predicciones, ya completamente procesada, se incluyen tanto las series horarias de viento como las estimaciones de potencia generada para cada hora del día. Este nivel de detalle horario es precisamente uno de los principales objetivos de este trabajo: ofrecer una herramienta que permita anticipar con mayor precisión la producción eólica en escalas temporales cortas, facilitando así una mejor planificación y gestión del recurso energético.

En relación con las predicciones de viento correspondientes a la segunda semana, se observa que ambos modelos tanto el desarrollado como el implementado actualmente en el parque presentan dificultades para capturar con precisión los picos de viento que se manifiestan únicamente durante una hora. No obstante, el metamodelo desarrollado muestra una mayor capacidad para reproducir dichos picos dentro de un intervalo razonable, siguiendo en gran medida la evolución temporal del viento observada. En particular, el modelo basado en los aerogeneradores logra ajustarse mejor a la dinámica del viento, mientras que el modelo basado en los cuatro nodos presenta un comportamiento más próximo al del sistema actualmente disponible en el parque.

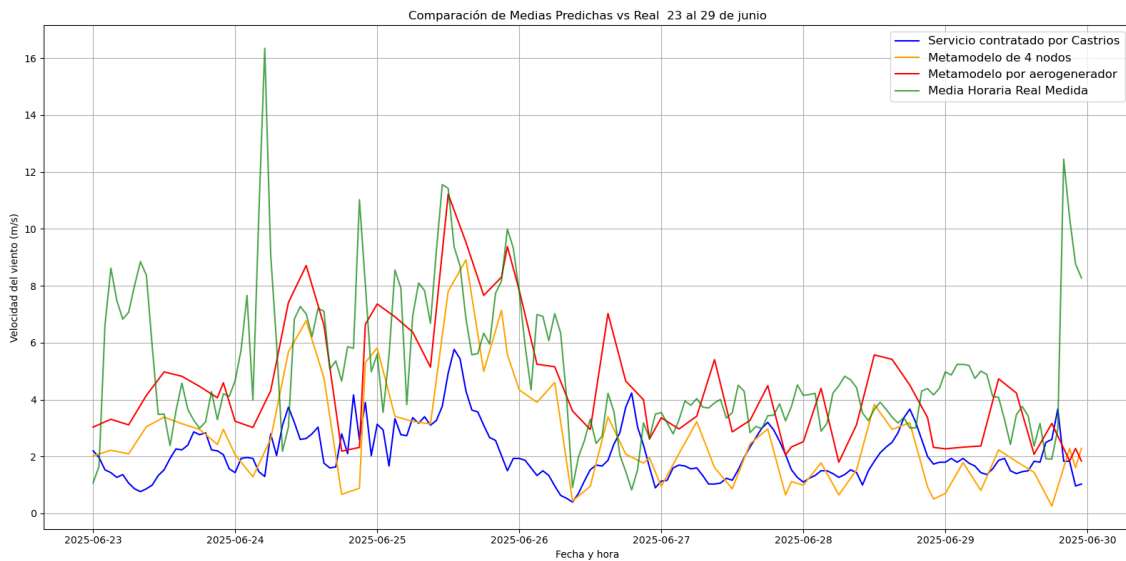


Ilustración 59. Velocidad del viento real vs predicha por los diferentes predictores del 23 al 29 de junio.

6.2 Predicción de potencia

Tal como se ha mencionado previamente, la estimación de la potencia generada por los aerogeneradores comienza a partir del 23 de junio. Dado que la operación de un parque eólico requiere conocer con precisión la cantidad de energía que se espera producir en cada hora, resulta indispensable contar con predicciones horarias de potencia. Sin embargo, como ya se indicó, los datos utilizados en este estudio provienen de modelos de libre acceso cuya resolución temporal es de tres horas. Para subsanar esta limitación y aproximar un comportamiento más realista del recurso, se ha aplicado una interpolación lineal entre los valores tres horarios, con el fin de obtener una serie continua con estimaciones horarias.

Este proceso de interpolación permite generar una base más adecuada sobre la cual aplicar los modelos de conversión de viento a potencia, que han sido desarrollados específicamente para cada aerogenerador. Así, se logra una predicción más detallada y operativamente útil, alineada con las necesidades reales de gestión y planificación energética del parque.

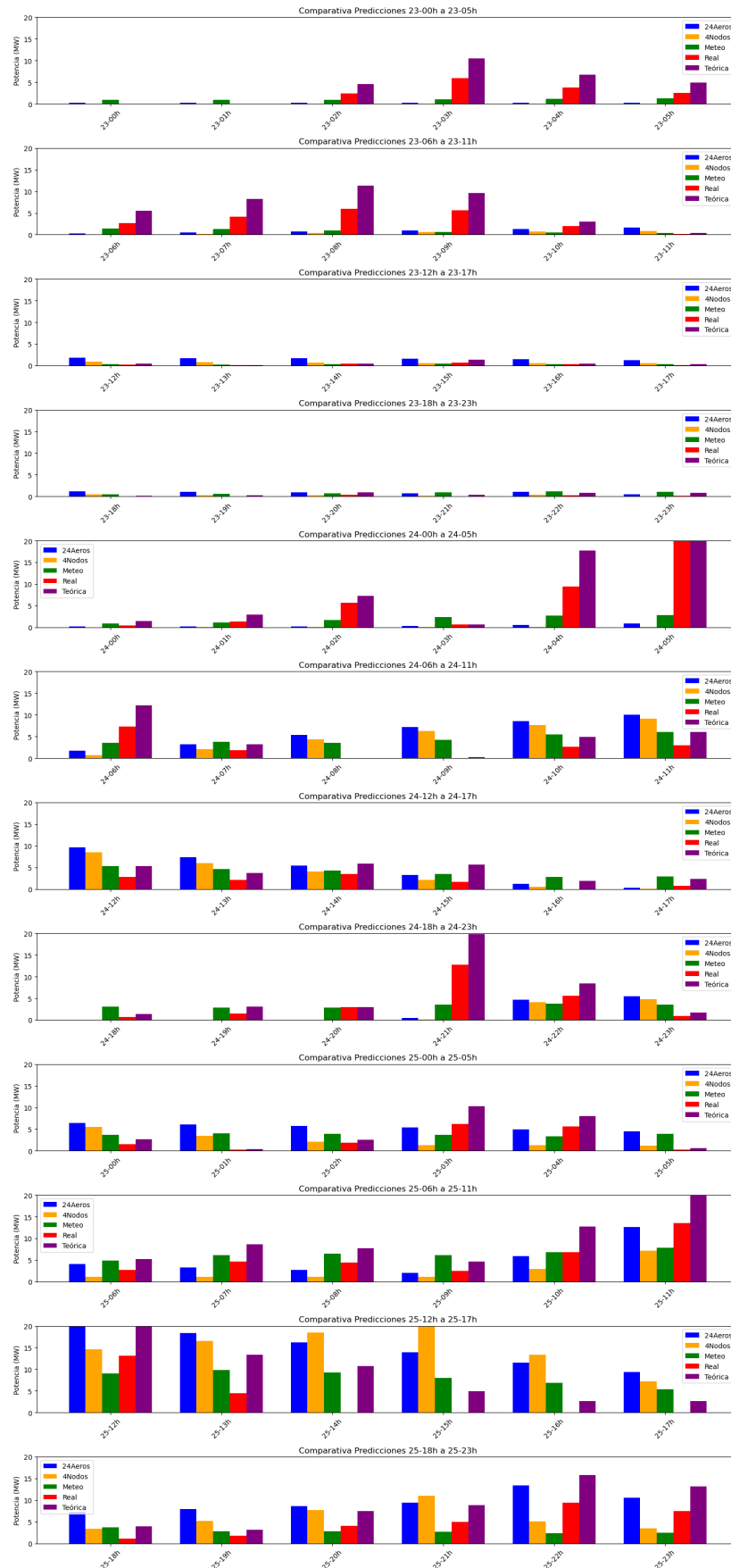
No obstante, es importante tener en cuenta ciertos factores que pueden afectar directamente la generación de potencia, más allá de la velocidad del viento. En particular, si en un determinado momento el precio de la electricidad cae por debajo de un umbral económico mínimo, es posible que el aerogenerador se detenga, aun cuando las condiciones de viento sean favorables. De igual forma, la realización de tareas de mantenimiento o paradas programadas puede implicar una ausencia total de generación, incluso en presencia de viento suficiente. Estas situaciones, que no dependen únicamente de las condiciones meteorológicas, deben considerarse en la interpretación de las predicciones de potencia, ya que representan limitaciones inherentes a la operación real del parque eólico.

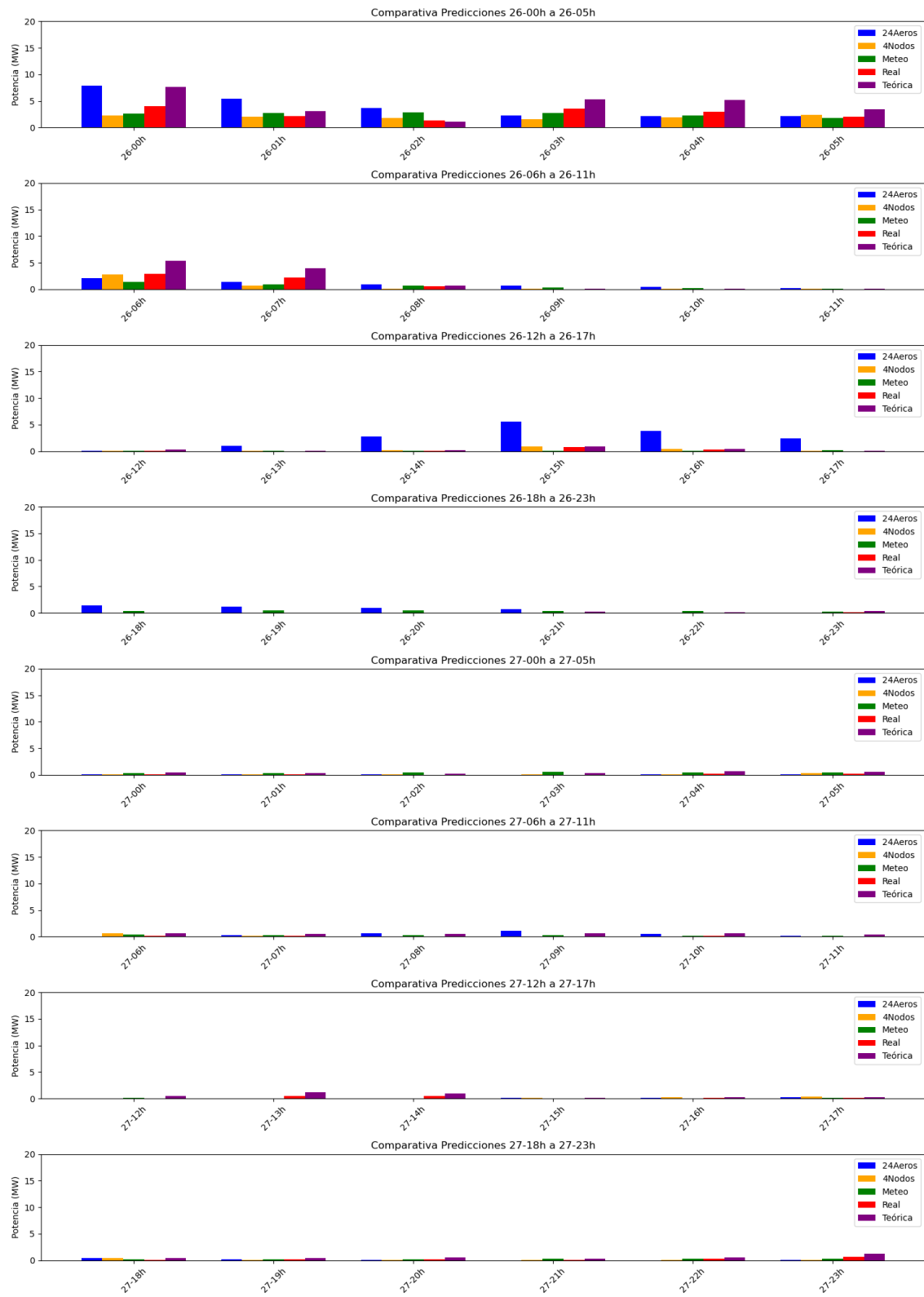
En este apartado, sin embargo, se presentan los resultados de producción eólica estimada sin aplicar aún dichas restricciones operativas. El objetivo es mostrar el potencial teórico de generación eléctrica en condiciones ideales, exclusivamente a partir de los datos de viento predicho y el comportamiento técnico de cada aerogenerador. Las consideraciones sobre interrupciones por mantenimiento o por razones económicas serán incorporadas en etapas posteriores del análisis, al realizar la comparativa con la energía realmente generada.

Al igual que en el caso anterior, se presentan los resultados obtenidos por los distintos enfoques de modelado, incluyendo el metamodelo individualizado, que estima la potencia generada para cada aerogenerador de forma específica, y el metamodelo general, que proporciona una predicción agrupada para el conjunto del parque. También se incorpora el modelo operativo actual del parque, basado en una media global de viento sin distinción entre aerogeneradores. Sin embargo, en esta sección dedicada a la predicción de potencia, se añade un cuarto enfoque: la estimación de la potencia teórica que debería haberse generado a partir de las velocidades de viento reales medidas, aplicando la correspondiente curva de conversión viento-potencia. Esta última referencia no se considera una predicción en sentido estricto, pero permite valorar el rendimiento potencial del parque bajo condiciones ideales, y sirve como punto de comparación clave para analizar el grado de ajuste de los modelos desarrollados.

Como puede observarse, la predicción horaria del viento presenta ciertas discrepancias respecto a los valores reales en algunas franjas temporales. Por otro lado, se podría intuir que, dado que el modelo contratado por el parque tiende a infraestimar la magnitud del viento, también subestimaría la producción de energía. Sin embargo, los datos muestran que el servicio contratado por el parque eólico de Castríos probablemente utiliza una curva de potencia propia, ajustada a partir de los numerosos años de operación, lo que permite compensar parcialmente dicha subestimación.

De forma general, y al finalizar el análisis diario, se aprecia que ambos metamodelos ofrecen estimaciones de producción similares. Tal como se mencionó anteriormente, la producción real siempre se mantiene por debajo de la teórica. Cabe destacar que, durante ciertos intervalos de tiempo, el parque ha permanecido parado debido a los bajos precios del mercado eléctrico, lo que explica la presencia de picos elevados en la producción teórica mientras que la producción real se mantiene en cero. Por este motivo, en el análisis comparativo se considerará tanto la serie de producción real como la teórica.





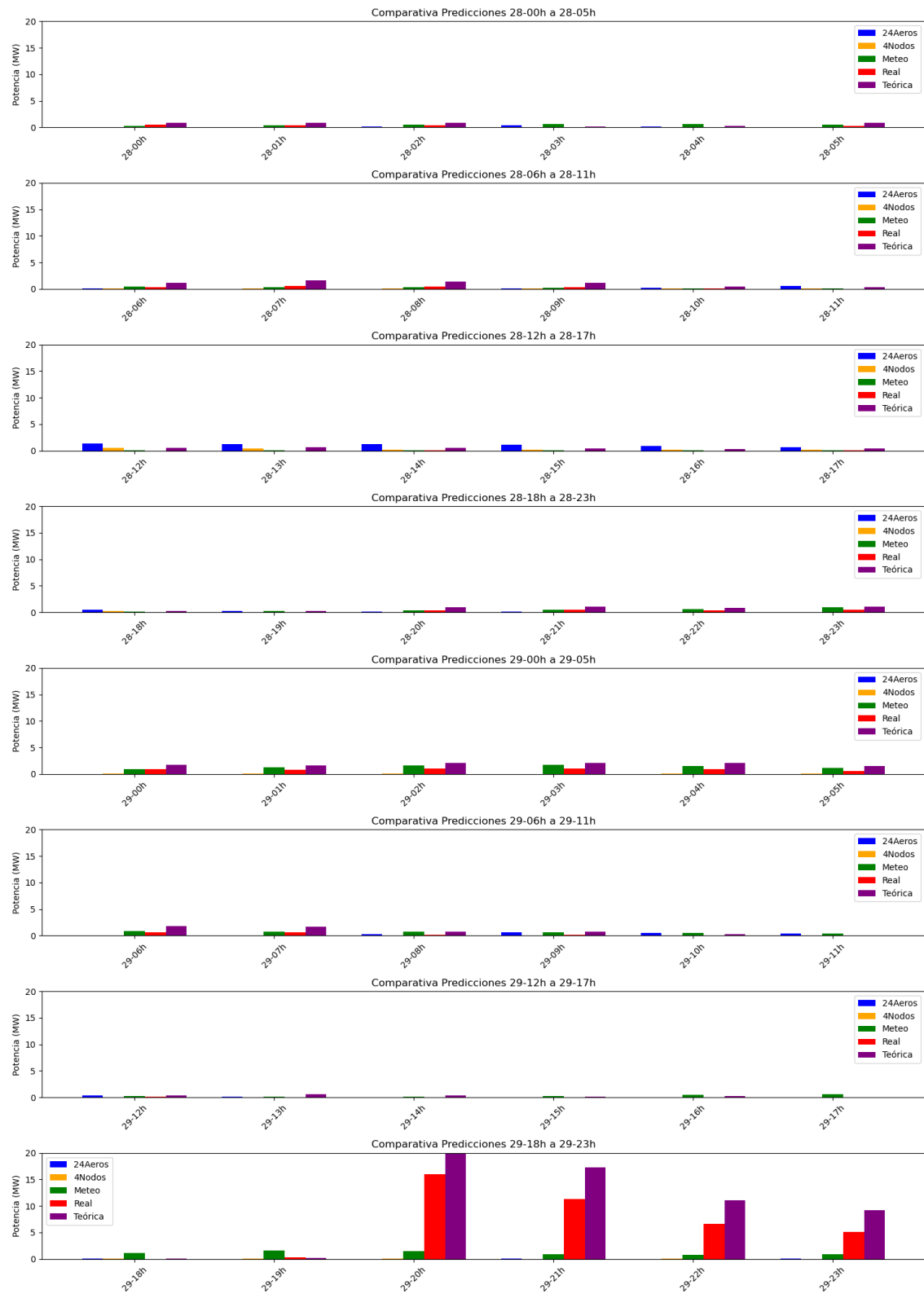


Ilustración 60. Potencia real vs predicha por los diferentes modelos en la semana del 23 al 29 de junio.

A continuación, se presenta la ilustración resumen correspondiente a la suma diaria de la potencia generada por cada metodología. En ella se aprecia que, a lo largo de los distintos días, los tres modelos de predicción ofrecen resultados consistentes entre sí, reflejando una evolución coherente de la producción estimada. No obstante, la comparación con la potencia real y teórica evidencia de forma más clara las discrepancias existentes entre ambas, lo que pone de manifiesto la necesidad de disponer de una curva de potencia específica que se adapte adecuadamente a los valores de viento generados por el metamodelo desarrollado. Esta adecuación permitiría mejorar la estimación de la producción energética en base a la simulación del viento, ajustándose con mayor precisión a las características reales del parque eólico.

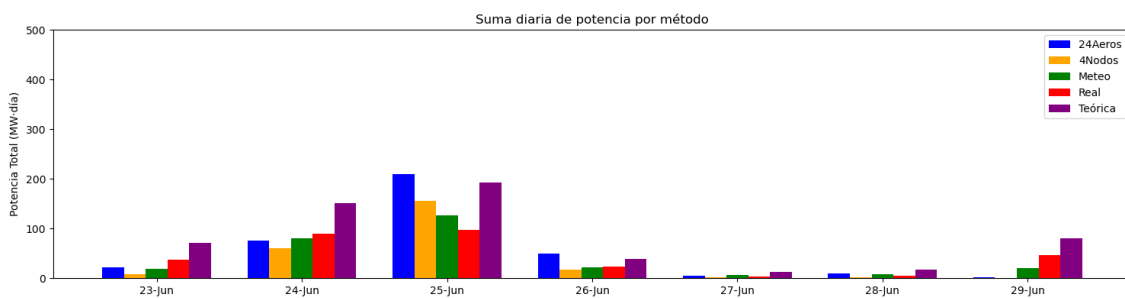


Ilustración 61. Potencia generada vs predicha por los modelos de manera diaria.

6.2.1. Desarrollo de red neuronal

Predecir la velocidad del viento es, de por sí, una tarea compleja debido a la alta variabilidad temporal y espacial del recurso eólico. Esta dificultad se ve aumentada cuando se observa que los aerogeneradores no siguen fielmente su curva de potencia teórica al trabajar con datos horarios de producción. Factores como el mantenimiento, las condiciones de carga, turbulencias locales o estrategias de control del parque pueden provocar desviaciones significativas respecto al comportamiento ideal.

Dado este escenario, estimar con precisión la potencia generada en un parque eólico a partir de series temporales de viento representa un reto adicional. Por ello, se propone el desarrollo de modelos de aprendizaje automático que permitan obtener una aproximación de la curva de potencia real de los aerogeneradores en función del viento predicho.

Para alcanzar este objetivo, se entrenan y comparan tres modelos supervisados: Árbol de Decisión (Decision Tree Regressor), Bosque Aleatorio (Random Forest Regressor) y una Red Neuronal Artificial (MLPRegressor). A continuación, se describe brevemente cada uno:

Los árboles de decisión son modelos predictivos que dividen recursivamente el espacio de características en regiones homogéneas respecto a la variable objetivo, mediante reglas de decisión basadas en los valores de las variables independientes. Existen dos tipos principales de árboles de decisión: los utilizados para clasificación,

que asignan una etiqueta de clase, y los de regresión, que predicen un valor numérico continuo. Los árboles de decisión son intuitivos y fáciles de interpretar, pero suelen ser propensos al sobreajuste, especialmente cuando se permiten profundidades elevadas.

Para mejorar la capacidad de generalización y reducir la varianza inherente a los árboles individuales, se desarrollaron los bosques aleatorios (random forests). Un bosque aleatorio es un conjunto de árboles de decisión, donde cada árbol es entrenado sobre una muestra aleatoria diferente de los datos y considera una selección aleatoria de las variables para dividir cada nodo. Esta aleatorización introduce diversidad entre los árboles, lo que permite que el conjunto generalice mejor que un único árbol [18].

Las Redes Neuronales Artificiales (ANN) son modelos computacionales utilizados en aprendizaje automático que están inspirados en el funcionamiento básico del cerebro, pero simplificados para procesar información. Están compuestas por capas de nodos o “neuronas” artificiales interconectadas, que reciben datos de entrada, los transforman mediante funciones matemáticas y producen una salida.

Las ANN aprenden ajustando los pesos de las conexiones entre nodos durante un proceso llamado entrenamiento, que consiste en minimizar el error entre las predicciones y los valores reales. Esto les permite modelar relaciones complejas y no lineales entre las variables de entrada y salida [19]. Son especialmente útiles para problemas donde las relaciones entre variables no son fáciles de capturar con modelos tradicionales, y por ello se utilizan para estimar la curva real de potencia de aerogeneradores en función del viento predicho.

Para el desarrollo de este proyecto, se opta por trabajar con redes neuronales. Aunque todas las redes empleadas presentan un desempeño similar en la fase de entrenamiento, con una precisión aproximada del 85%, se observa que la red neuronal ofrece una curva de potencia más depurada y suave, lo que facilita una mejor interpretación y estimación de la potencia real en función del viento predicho.

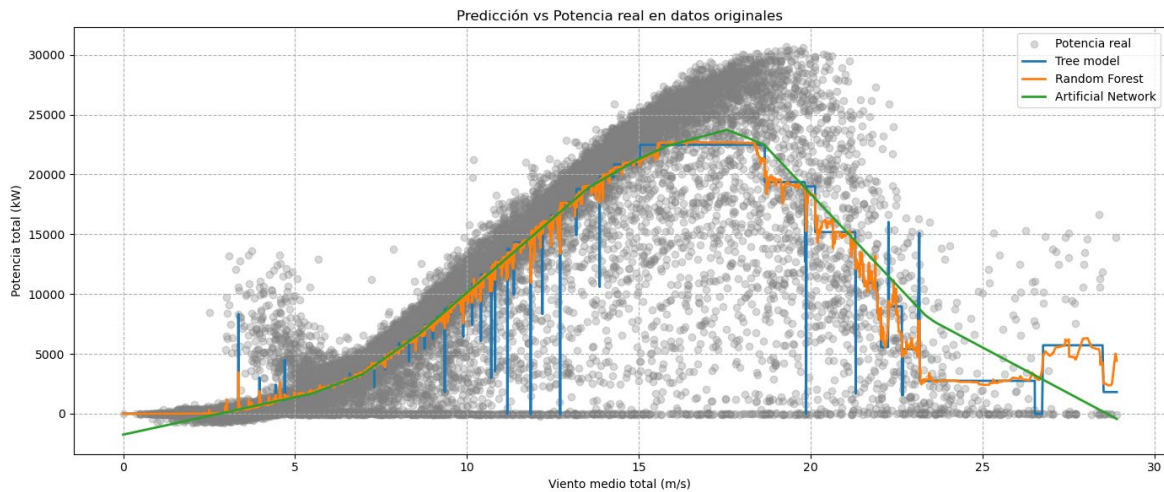


Ilustración 62. Curva de potencia elaborada en base a las técnicas de Machine Learning.

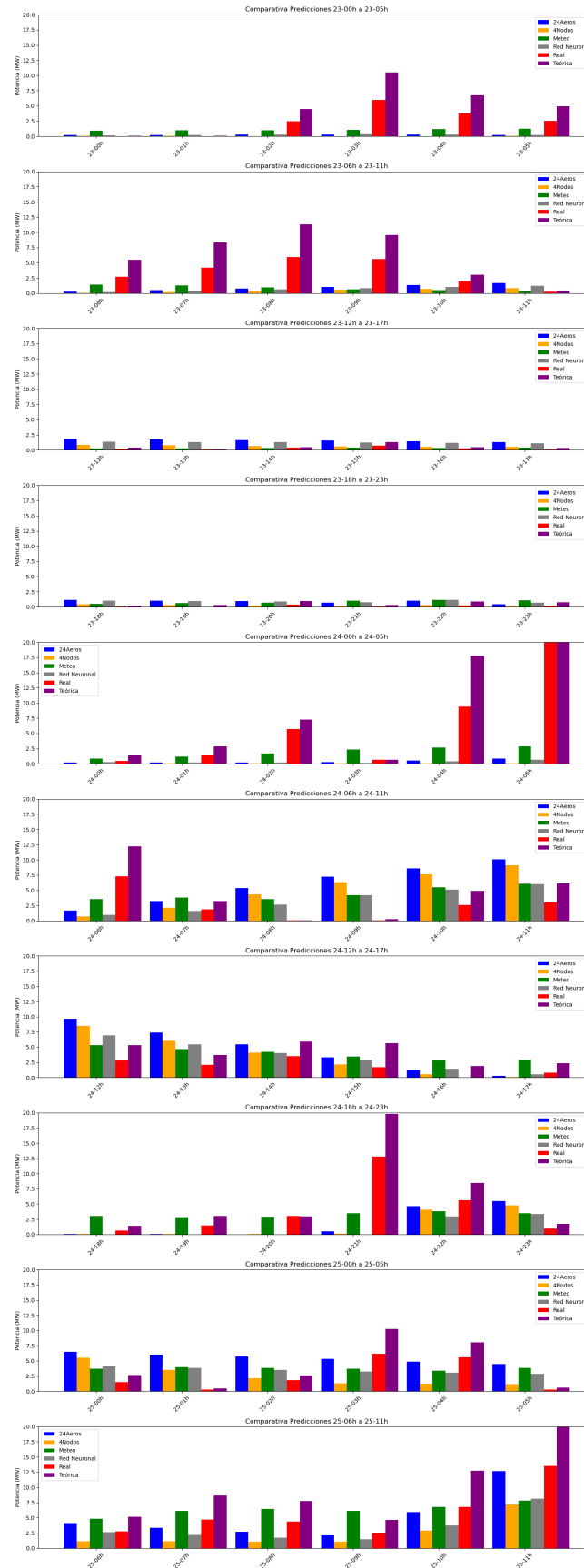
A pesar de trabajar con aerogeneradores cuyo viento de parada teórico es de 25 m/s, se observa que cuando se trabaja con datos horarios, la producción comienza a disminuir notablemente a partir de velocidades de viento cercanas a los 18 m/s. Esto indica que, en la práctica, el parque eólico deja de generar potencia antes de alcanzar el viento de parada nominal.

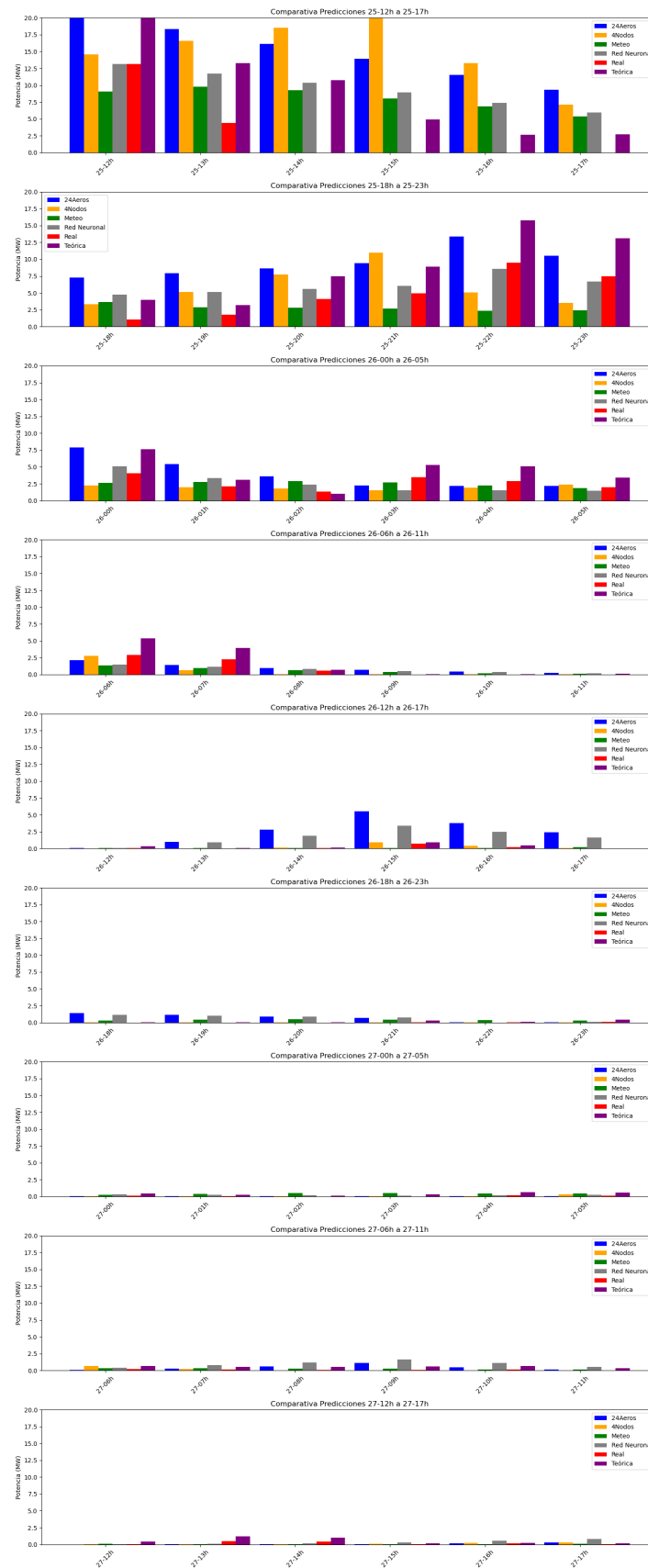
Asimismo, para velocidades de viento bajas, próximas a cero, la producción también es prácticamente nula, tal como se espera. La dispersión de los puntos en los datos reales evidencia la gran variabilidad en la potencia generada para un mismo valor de velocidad del viento, lo que refleja las condiciones reales y las múltiples influencias sobre la generación eléctrica en el parque.

Esta variabilidad y desviación respecto a la curva teórica hacen imprescindible el uso de modelos que puedan ajustarse a estas condiciones reales, como los presentados, para mejorar la estimación de la potencia producida en función del viento.

6.2.2. Resultados

A continuación, se presentan los resultados obtenidos al considerar la red neuronal desde el inicio. Se observa una clara mejora en los momentos del día en los que la predicción fue más acertada, especialmente en las predicciones correspondientes al 25 de junio.





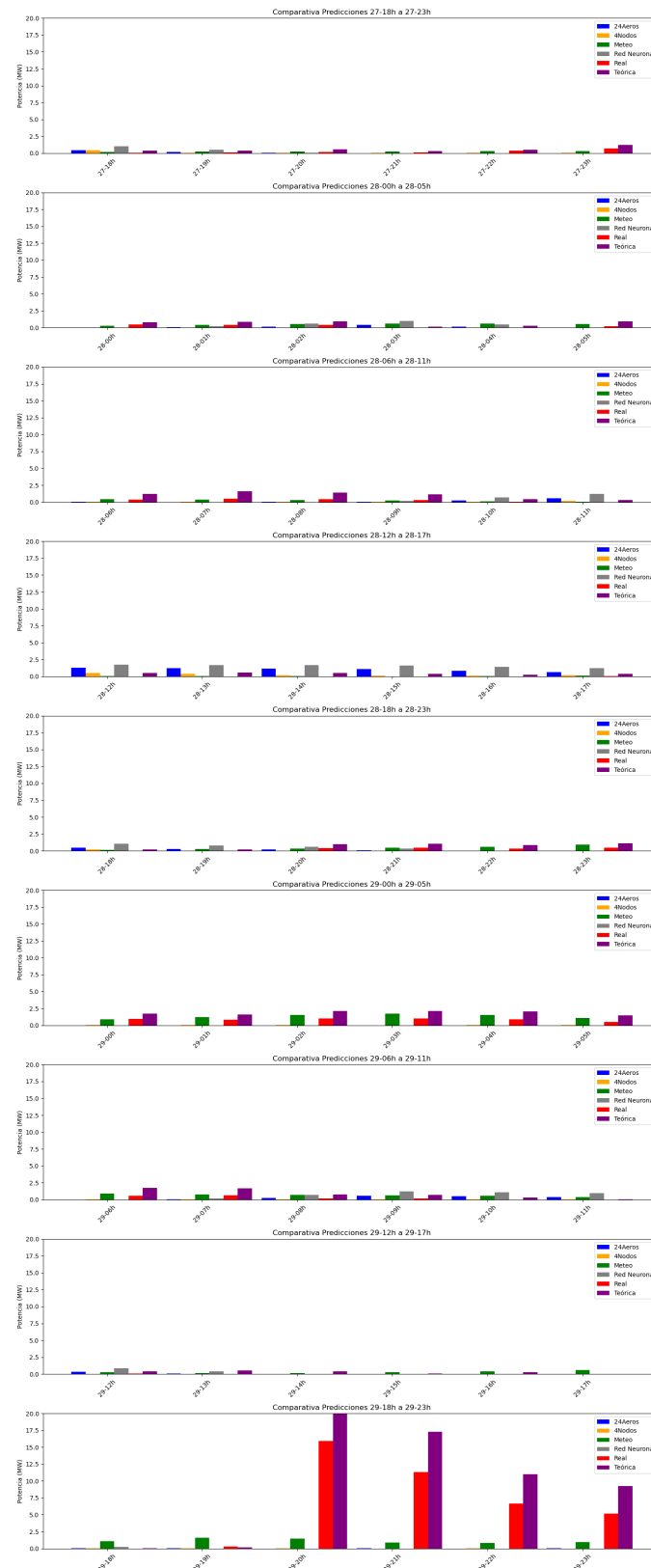


Ilustración 63. Potencia real vs predicha por los diferentes modelos en la semana del 23 al 29 de junio con modelo de red neuronal.

Si se mira desde una escala diaria también se observa una mejora en los días en los que el viento fue reconstruido de manera más acertada.

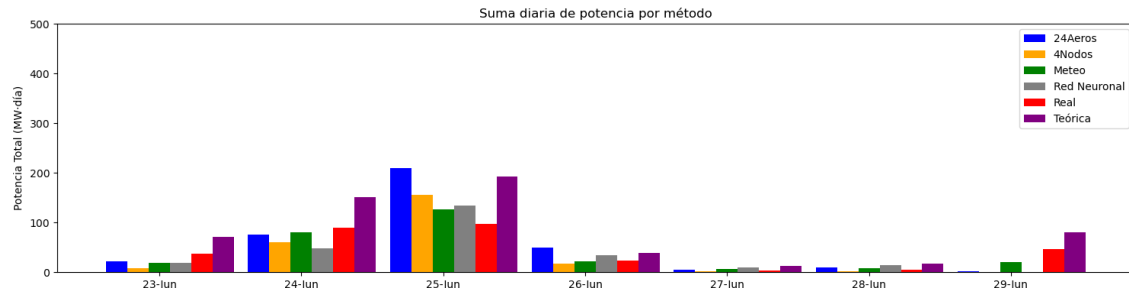


Ilustración 64. Potencia generada vs la predicha de manera diaria con modelo de red neuronal.

El desarrollo de este modelo está justificado por momentos como el ocurrido el 25 de junio a las 12:00 horas, instante en el que la velocidad del viento fue predicha con precisión exacta y se evidenció una clara diferencia entre la potencia teórica y la real.

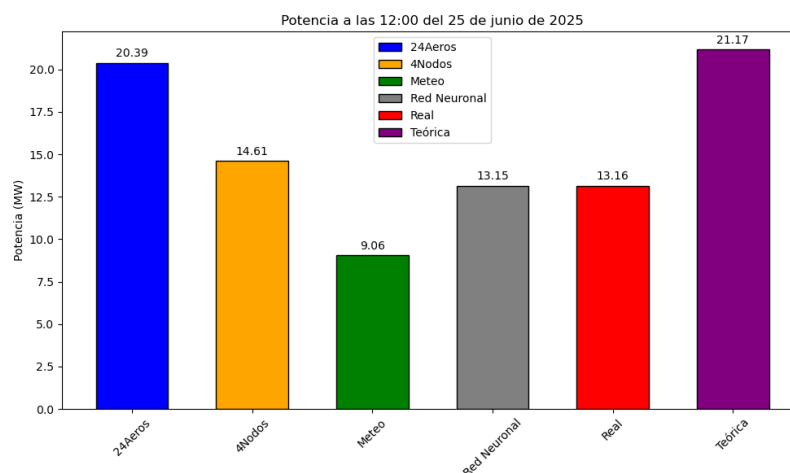


Ilustración 65. Producción del 25 de junio a las 12h.

7. Conclusiones y futuras líneas de investigación

7.1 Conclusiones

El presente estudio se ha centrado en el desarrollo y aplicación de un metamodelo con el objetivo de predecir la velocidad del viento en el parque eólico de “Castríos”, situado en la provincia de Burgos, al norte de España. La investigación ha demostrado la efectividad de combinar técnicas avanzadas de modelización numérica con herramientas de ciencia de datos para abordar los complejos desafíos asociados a la dinámica del viento, fenómeno que depende de múltiples variables atmosféricas como la presión, la temperatura o la humedad.

La elección del parque eólico de Castríos como zona de estudio se fundamentó en la disponibilidad de datos históricos y en su compleja orografía montañosa, que influye directamente en el perfil del viento. Esta configuración lo convierte en un escenario idóneo para evaluar la robustez y precisión del metamodelo.

El trabajo se ha beneficiado de múltiples fuentes de información, desde modelos de topografía de alta resolución hasta registros históricos en diferentes ubicaciones del norte de España. Estos datos han sido fundamentales para validar el metamodelo y analizar el impacto de la rugosidad del terreno sobre el comportamiento del viento.

La metodología empleada ha consistido en la validación y calibración de un metamodelo híbrido a través de técnicas de ciencia de datos como el Análisis de Componentes Principales (PCA), el algoritmo de Máxima Disimilitud (MDA) y las Funciones de Base Radial (RBF).

Los resultados obtenidos han sido prometedores. El metamodelo desarrollado ha demostrado ser una herramienta eficaz tanto para la predicción del viento como para la reconstrucción de series históricas, proporcionando resultados coherentes y precisos en comparación con las observaciones reales.

Más allá de los avances técnicos, este trabajo sienta las bases para la aplicación del metamodelo a escalas mayores, lo que podría contribuir significativamente a la transición energética. La posibilidad de identificar zonas óptimas para la instalación de parques eólicos y predecir su producción energética con hasta 24 horas de antelación permite reducir los costes asociados a las penalizaciones por desvíos en la generación prevista.

Asimismo, el estudio subraya la importancia de contar con bases de datos extensas y de calidad, con registros de variables como el viento, la temperatura o la presión. La validación cruzada entre distintos metamodelos y datos reales resulta esencial para su correcto desarrollo. Estas herramientas no solo son aplicables al viento, sino también a otros fenómenos como el oleaje, y se están extendiendo a ámbitos como la predicción de inundaciones.

En definitiva, este estudio ha puesto de manifiesto la eficacia de los metamodelos de viento tanto en la reconstrucción histórica como en su aplicación futura a la predicción operativa. Los resultados alcanzados en el parque eólico de Castríos refuerzan la necesidad de seguir impulsando el desarrollo de este tipo de tecnologías, fundamentales para una adecuada planificación energética y para avanzar en la transición hacia un modelo energético más limpio y sostenible, en el contexto de la lucha contra el cambio climático.

7.2 Futuras líneas de investigación

Existen distintas líneas de trabajo que pueden abordarse en el futuro con el fin de mejorar la precisión, aplicabilidad y utilidad operativa del metamodelo de viento desarrollado en este estudio:

1- Incorporar nuevas variables predictoras

Una línea prioritaria es la inclusión de nuevas variables atmosféricas como la temperatura, la humedad relativa o la estabilidad atmosférica. Estas variables condicionan el comportamiento del viento a escala local y su integración en el modelo puede mejorar la capacidad predictiva, especialmente en situaciones meteorológicas complejas.

2- Aumentar la resolución espacial en el entorno del parque

El desarrollo del metamodelo a una resolución de 100 metros en la zona del parque eólico permitiría una caracterización más precisa del viento en cada aerogenerador, mejorando así las estimaciones individuales de producción. Este enfoque resulta especialmente relevante en emplazamientos de orografía accidentada como Castríos, donde pequeñas variaciones topográficas generan grandes diferencias en el viento.

3- Ampliar el conjunto de datos de entrenamiento

Contar con una base de datos más amplia y representativa, que incluya eventos extremos de viento y datos de diferentes estaciones del año, contribuiría a aumentar la robustez del metamodelo. Esta ampliación permitiría capturar con mayor fidelidad la variabilidad estacional y mejorar la respuesta del modelo ante situaciones fuera del rango habitual.

4- Automatizar el proceso de predicción y estimación de potencia

Una mejora clave sería el diseño de un sistema automatizado que integre los datos de entrada meteorológicos, la ejecución del metamodelo y la transformación de viento a potencia mediante las curvas de los aerogeneradores. Esta automatización permitiría el uso operativo en tiempo

real, reduciendo errores humanos y facilitando la planificación energética diaria.

5- Desarrollar un modelo de potencia ajustado considerando múltiples variables

Una línea de trabajo prioritaria es el desarrollo de un modelo de potencia que integre no solo la velocidad del viento predicha, sino también otros parámetros relevantes que afectan la producción real del parque eólico. Este enfoque multidimensional permitiría un ajuste más preciso del modelo a los valores reales observados, reduciendo las desviaciones entre la energía estimada y la generada, y mejorando así la precisión y fiabilidad del sistema bajo condiciones operativas reales.

6- Aplicación del modelo en otras regiones

La metodología desarrollada puede ser extrapolable a otros emplazamientos de interés con características topográficas distintas. Su implementación en diferentes contextos permitiría evaluar su generalización y crear un marco predictivo común para la planificación eólica regional.

7- Estimar el impacto de errores de predicción en la planificación energética

Una aplicación futura relevante es cuantificar cómo los errores en la predicción de viento afectan a la programación y a los costes del sistema eléctrico, especialmente en lo que respecta a las penalizaciones por desvíos. Esto permitiría dimensionar el valor económico de contar con un metamodelo más preciso y apoyar decisiones estratégicas en la operación de parques eólicos.

Abordar las líneas de investigación propuestas permitirá consolidar los avances logrados en este estudio y continuar mejorando la precisión, escalabilidad y aplicabilidad operativa del metamodelo de viento en entornos de orografía compleja. La metodología desarrollada para el parque eólico de Castrís supone un paso significativo hacia una predicción más fiable del viento a corto plazo y una comprensión más detallada del comportamiento del recurso eólico a alta resolución espacial y temporal. Estos avances no solo fortalecen la planificación energética local y reducen los costes derivados de desviaciones en la predicción, sino que además contribuyen a los objetivos globales de integración de energías renovables, resiliencia del sistema y mitigación del cambio climático mediante una implantación más informada y eficiente de parques eólicos.

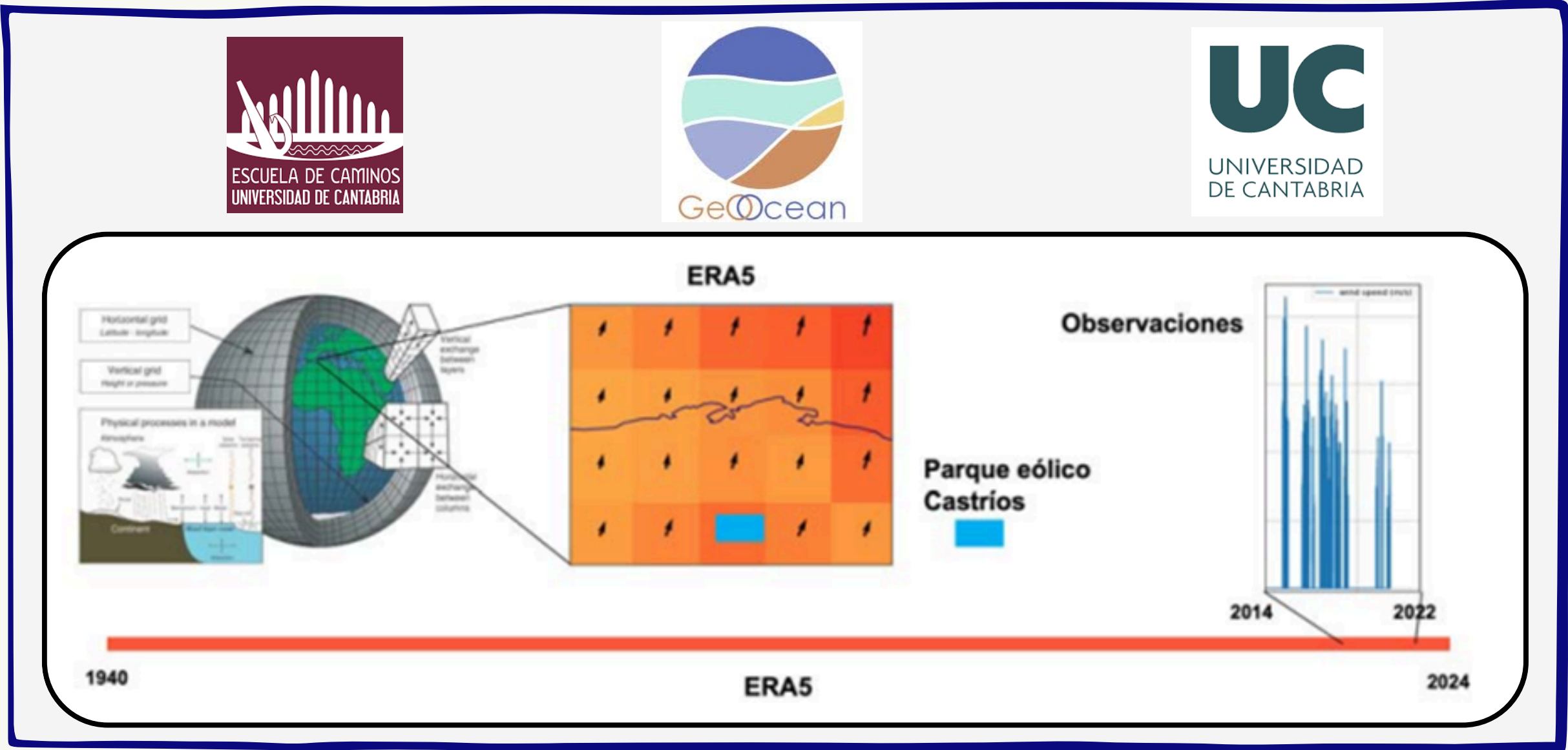
8. Bibliografía

- [1] Red Eléctrica, “La producción renovable crece en España un 10,3% en 2024 y alcanza sus mayores registros | Red Eléctrica,” Informes del sistema eléctrico español 2024. Accessed: Apr. 11, 2025. [Online]. Available: <https://www.ree.es/es/sala-de-prensa/actualidad/nota-de-prensa/2025/03/la-produccion-renovable-crece-en-Espana-un-10-3-por-ciento-2024-alcanza-mayores-registros>
- [2] OMIE, “Mercado de electricidad ,” El mercado de electricidad está estructurado en un mercado diario, un mercado intradiario de subastas y un mercado intradiario continuo. Accessed: Apr. 11, 2025. [Online]. Available: <https://www.omie.es/es/mercado-de-electricidad>
- [3] L. A. Fernández Jiménez, “Modelos avanzados para la predicción a corto plazo de la producción eléctrica en parques eólicos,” 2007.
- [4] Parque Eólico Los Castríos, de 26,4 MW, en el término municipal de Espinosa de los Monteros, “Proyecto Constructivo”, Burgos, 2003.
- [5] A. Ricondo, L. Cagigal, B. Pérez-Díaz, and F. J. Méndez, “HySwash: A hybrid model for nearshore wave processes,” *Ocean Engineering*, vol. 291, 2024, doi: 10.1016/j.oceaneng.2023.116419.
- [6] Luis Alfredo Fernández Jiménez, “13747,” La Rioja, 2007.
- [7] “Base de datos de energía eólica.” Accessed: Apr. 11, 2025. [Online]. Available: https://www.thewindpower.net/index_es.php
- [8] P. Camus, F. J. Mendez, R. Medina, and A. S. Cofiño, “Analysis of clustering and selection algorithms for the study of multivariate wave climate,” 2011. doi: 10.1016/j.coastaleng.2011.02.003.
- [9] “GeoOcean/BlueMath.” Accessed: Apr. 14, 2025. [Online]. Available: <https://github.com/GeoOcean/BlueMath?tab=readme-ov-file>
- [10] P. Camus, F. J. Mendez, R. Medina, and A. S. Cofiño, “Analysis of clustering and selection algorithms for the study of multivariate wave climate,” 2011. doi: 10.1016/j.coastaleng.2011.02.003.
- [11] P. Camus, F. J. Mendez, R. Medina, A. Tomas, and C. Izaguirre, “High resolution downscaled ocean waves (DOW) reanalysis in coastal areas,” *Coastal Engineering*, vol. 72, 2013, doi: 10.1016/j.coastaleng.2012.09.002.
- [12] “ERA5 hourly data on single levels from 1940 to present.” Accessed: Apr. 11, 2025. [Online]. Available: <https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=overview>
- [13] H. Hersbach *et al.*, “The ERA5 global reanalysis,” *Quarterly Journal of the Royal Meteorological Society*, vol. 146, no. 730, pp. 1999–2049, Jul. 2020, doi: 10.1002/QJ.3803.
- [14] B. Bell *et al.*, “The ERA5 global reanalysis: Preliminary extension to 1950,” *Quarterly Journal of the Royal Meteorological Society*, vol. 147, no. 741, pp. 4186–4227, Oct. 2021, doi: 10.1002/QJ.4174.
- [15] MeteoCantabria, “MeteoCantabria Vigía Virgen del Mar.” Accessed: May 02, 2025. [Online]. Available: <https://www.meteocantabria.es/meteocantabria/boya/boya-red-vigia>

- [16] "IFS Documentation CY48R1 - Part I: Observations | ECMWF." Accessed: Jun. 30, 2025. [Online]. Available: <https://www.ecmwf.int/en/elibrary/81367-ifs-documentation-cy48r1-part-i-observations>
- [17] F. Rabier, H. Järvinen, E. Klinker, J. F. Mahfouf, and A. Simmons, "The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics," *Quarterly Journal of the Royal Meteorological Society*, vol. 126, no. 564, 2000, doi: 10.1002/qj.49712656415.
- [18] L. , Breiman and A. Cutler, "Setting up, using, and understanding random forest V4.0," *Mach Learn*, vol. 45, pp. 1–33, 2003.
- [19] L. Vanneschi and S. Silva, "Artificial Neural Networks," in *Natural Computing Series*, 2023. doi: 10.1007/978-3-031-17922-8_7.

DESARROLLO DE UN METAMODELO DE VIENTO DE ALTA RESOLUCIÓN: APLICACIONES DE PREDICCIÓN DE CORTO PLAZO Y DE ESTIMACIÓN DE RECURSO EÓLICO

Trabajo de Fin de Master - Universidad de Cantabria

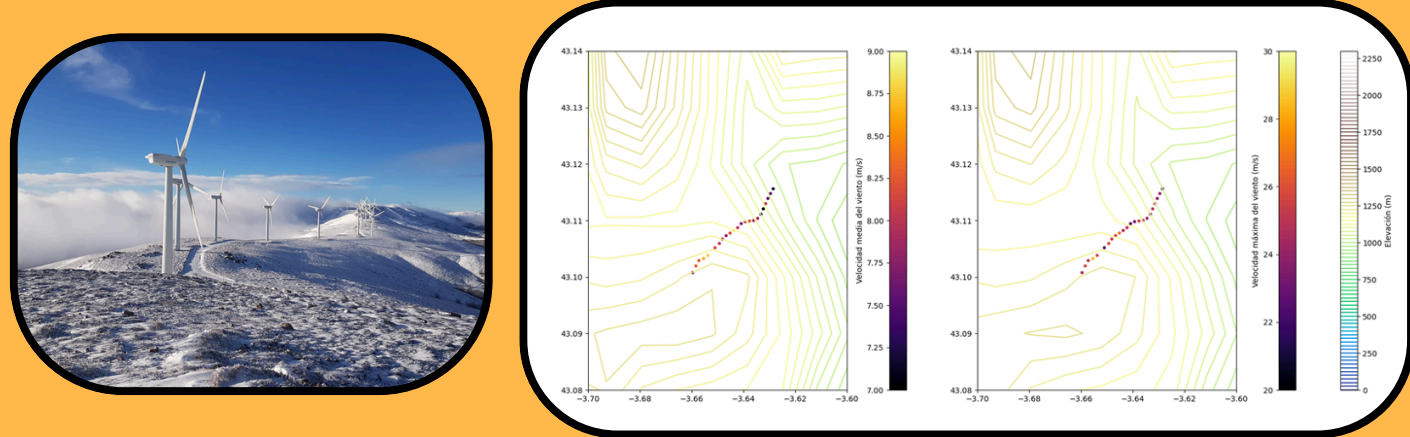


01. INTRODUCCIÓN

La integración de energías renovables exige herramientas precisas para predecir la producción. Los modelos meteorológicos actuales no resuelven bien la escala de parque eólico, especialmente en terrenos complejos. Este TFM plantea un metamodelo (MDA + PCA + RBF) para predecir viento y potencia a alta resolución y baja carga computacional.

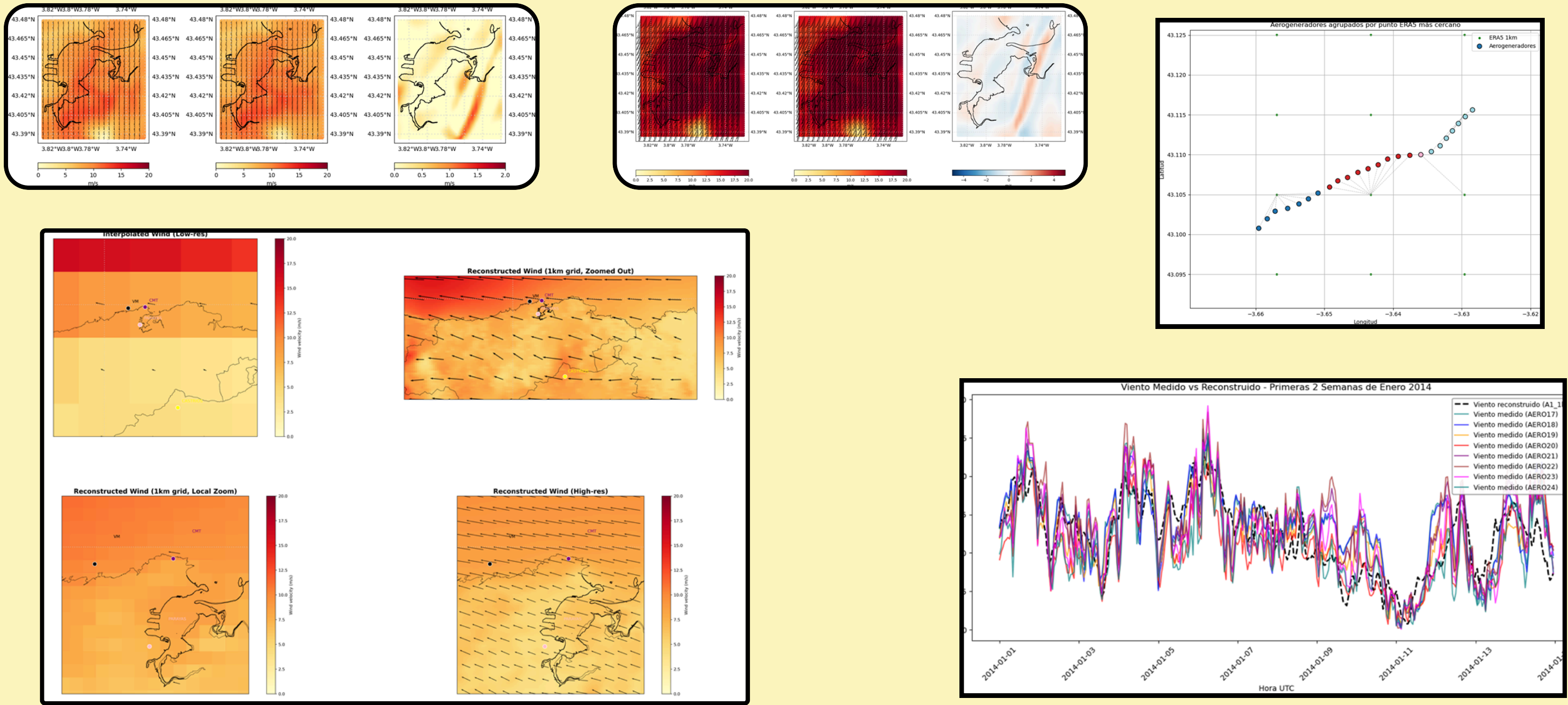
02. SITIO DE ESTUDIO

El parque eólico Castríos (Burgos), a 1.350 m de altitud, presenta una orografía compleja que dificulta la predicción del viento. Cuenta con 24 aerogeneradores de 1100 kW. Las diferencias topográficas condicionan la generación.



05. METAMODELO

El metamodelo se calibra con casos representativos simulados en alta resolución (WRF) con resolución de 1 km y 100m. Se valida en estaciones (AGL, VM, CMT, Parayas) y luego en Castríos. El modelo reproduce correctamente el viento diario y horario, especialmente en zonas elevadas. Se observa mejor desempeño que ERA5, con menor error en la estimación de potencia.



03. MÉTODOS Y HERRAMIENTAS

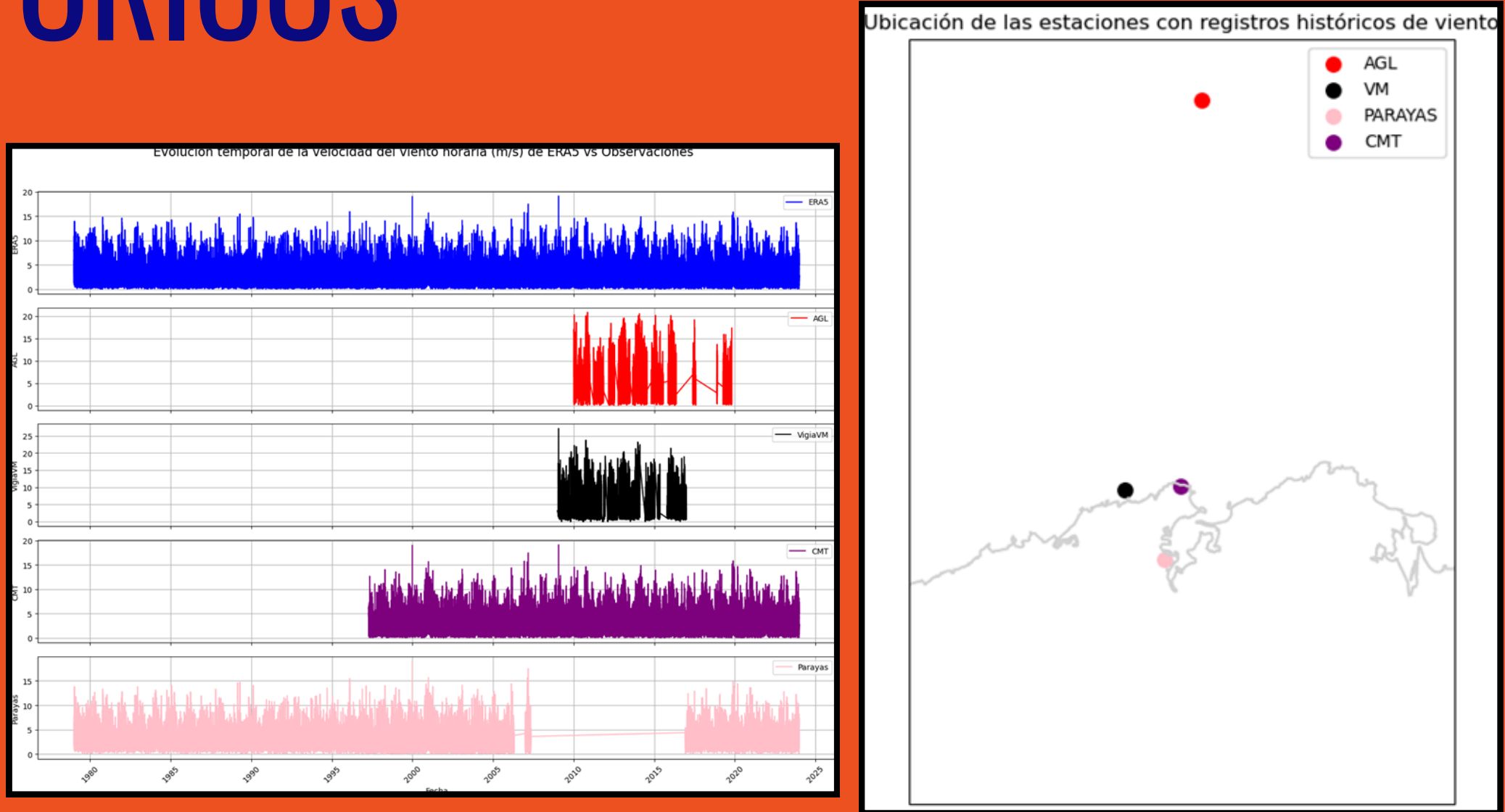
El desarrollo del metamodelo se apoya en un enfoque híbrido basado en tres componentes clave:

- PCA (Principal Component Analysis):** Reducción de la dimensionalidad de los campos de viento simulados a alta resolución, conservando las estructuras espaciales dominantes.
- MDA (Maximum Dissimilarity Algorithm):** Selección de días representativos a partir de datos de viento de baja resolución (ERA5), maximizando la diversidad del conjunto de entrenamiento.
- RBF (Radial Basis Functions):** Interpolación multivariable empleada para reconstruir las series horarias de viento sobre todo el dominio.

El código de modelado ha sido desarrollado en Python, inspirándose en herramientas disponibles en la biblioteca BlueMath-tk, promovida por el grupo GeoOcean de la Universidad de Cantabria. Se ha contado además con el clúster Altamira (SSC, IFCA), que ha permitido ejecutar simulaciones numéricas complejas y procesos de entrenamiento del modelo.

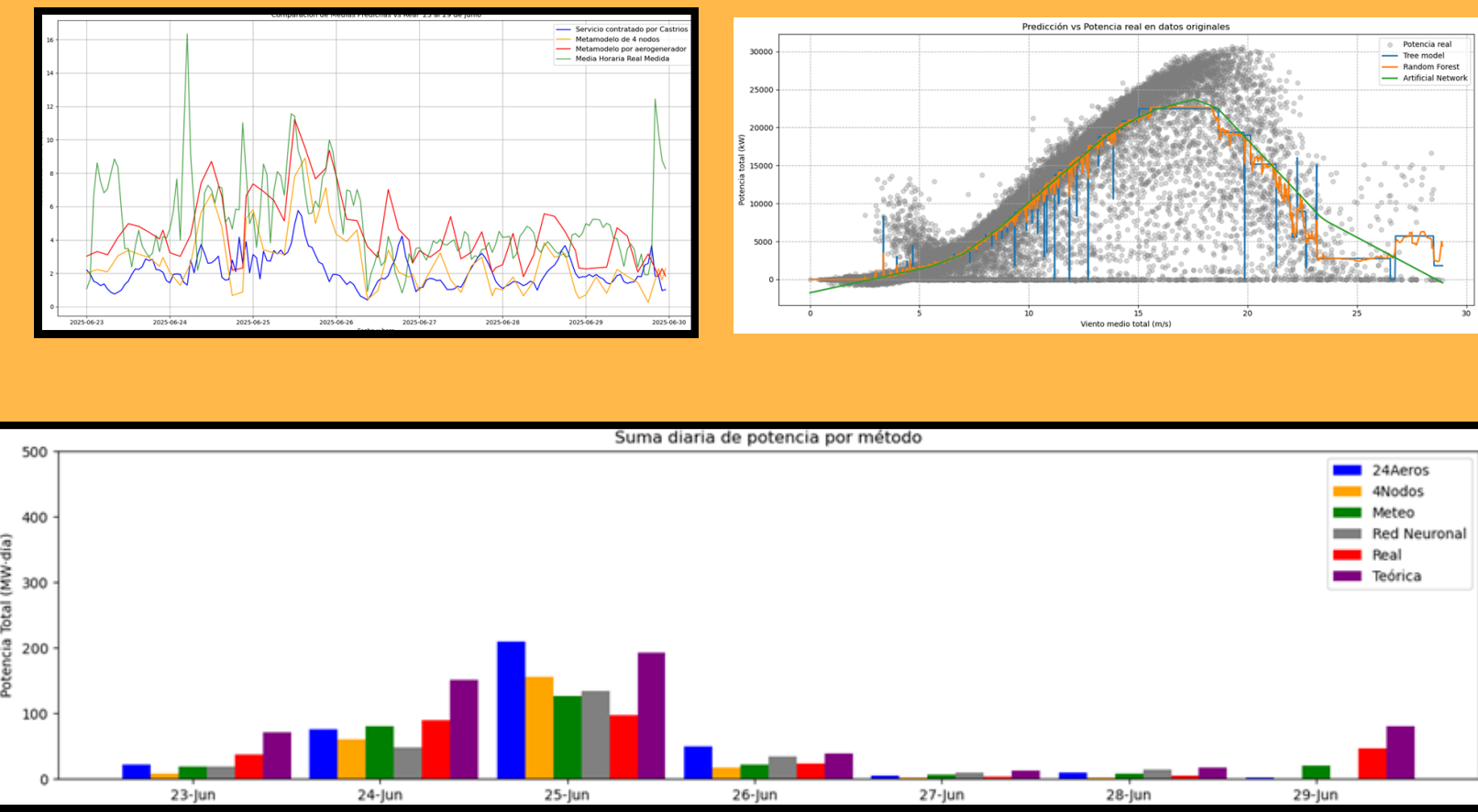
04. DATOS HISTÓRICOS

Se utilizan datos de ERA5, predicciones IFS del ECMWF, observaciones de estaciones en Santander y los aerogeneradores del parque eólico Castríos.



06. PREDICCIÓN

Una vez calibrado y validado, el metamodelo se aplica para predicciones operativas 24, 48 y 72 h vista. Las predicciones se transforman en potencia estimada para cada aerogenerador. Las reconstrucciones son fiables incluso en eventos extremos.



07. CONCLUSIÓN Y FUTURAS LÍNEAS DE INVESTIGACIÓN

Conclusiones:

- Se ha desarrollado un metamodelo eficiente y preciso para la predicción de viento en zonas complejas, reduciendo la necesidad de simulaciones numéricas completas.
- El modelo mejora significativamente la estimación del recurso eólico respecto a productos como ERA5, especialmente en terrenos montañosos.
- Ha demostrado ser útil no solo para predicción, sino también para la reconstrucción de series históricas incompletas.
- Puede aplicarse operativamente para predecir producción energética a escala de aerogenerador.

Entre las **futuras líneas de investigación** se propone:

- Incorporar nuevas variables predictoras (dirección del viento, temperatura, humedad).
- Aplicar el modelo a otras regiones con diferente topografía.
- Aumentar la resolución a 100 metros en la zona del parque.
- Automatizar el proceso de predicción y generación de curvas de potencia horarias.

