



**Universitat de les
Illes Balears**

Escuela Politécnica Superior

Memoria del Trabajo de Fin de Grado

**Modelo de predicción a corto plazo de la
generación eléctrica en parques eólicos, utilizando
técnicas de Machine-Learning.**

Deva Murti Baer

Grado de Ingeniería Industrial Electrónica y Automática

Año académico 2019-2020

DNI del alumno: X9647408-N

Trabajo tutelado por: Dr. Vicente José Canals Guinand

Departamento de Física, Área de Ingeniería Mecánica

Trabajo supervisado por: Sr. Rafael Muñoz Campos

Responsable de operación del parque eólico Es Milà (Consorti de Residus Urbans i Energia de Menorca)

Se autoriza la Universidad a incluir este trabajo en el Repositorio Institucional para su consulta en acceso abierto y difusión en línea, con fines exclusivamente académicas y de investigación	Autor		Tutor	
	Sí	No	Sí	No
	X		X	

Palabras clave del trabajo:

Predicción, series temporales, energía eólica, redes neuronales artificiales, MATLAB.



AGRADECIMIENTOS

A Vicente José Canals, por la dirección de este TFG y por todo lo que me ha enseñado y transmitido durante su realización. Siempre ha estado dispuesto a dedicarme su tiempo y su conocimiento en la realización de este proyecto, y en la resolución de mis dudas relacionadas con otros ámbitos de mi vida.

A Rafael Muñoz Campos, por sus comentarios y su colaboración en la obtención de los datos meteorológicos históricos del parque eólico de Es Milá.

A mis abuelos y mi madre por su apoyo constante durante estos años, con especial mención a Volker Flache.

A todos los profesores de la Escuela Politécnica Superior de la UIB, por los conocimientos que me han transmitido.

Al BAfög, que ha permitido financiarme durante el grado.

A la existencia de los cafés y las empanadas, por darme la energía necesaria para asistir con "atención" a las clases.



RESUMEN

El gran desarrollo experimentado por las energías renovables y especialmente de la energía eólica, ha dado lugar a una alta penetración de estas energías en los sistemas eléctricos, lo que ha conducido a la necesidad de disponer de herramientas de predicción para facilitar la tarea de gestión y coordinación al operador del sistema eléctrico y a los productores de energía eléctrica. Estas herramientas permiten aprovechar mejor estas fuentes de energía y facilitan su integración en la red eléctrica.

En este trabajo de fin de grado se aborda la elaboración de un modelo de predicción de la energía eléctrica generada en parques eólicos para horizontes cercanos, basado en el uso de redes neuronales artificiales recurrentes. Para ello, se presentan los fundamentos teóricos necesarios y se realiza una etapa de preprocesamiento de las series temporales de las variables más significativas que influyen sobre la producción de la energía eólica.

Los modelos de previsión usados se basan en el uso de redes neuronales recurrentes relativamente simples que incorporan funciones no lineales en su interior, y que tienen la capacidad de operar con información de baja calidad o faltante, permitiendo incluso así obtener previsiones de cierta calidad. Los resultados obtenidos reflejan lo complejo que resulta obtener un modelo de predicción preciso para la energía eólica, y la capacidad de las redes neuronales como ajustadores no lineales universales.

En el marco del presente trabajo se ha desarrollado una herramienta computacional basada sobre MATLAB, encargada de realizar el preprocesamiento de las series temporales y la elaboración de los modelos de predicción propuestos.



ABSTRACT

The great development experienced by renewable energies and especially wind energy, has given rise to a high penetration of these energies in electrical systems, which has led to the need for prediction tools to facilitate the task of management and coordination of the electrical system operator and electric power producers. These tools allow a better use of these energy sources and facilitate their integration in the electricity grid.

In this final degree project, a prediction model of the electrical energy generated in wind farms for close horizons is elaborated, based on the use of recurrent artificial neural networks. For this, the necessary theoretical foundations are presented and a preprocessing stage of the time series of the most significant variables that influence the production of wind energy is carried out.

The forecast models used are based on the use of relatively simple recurrent neural networks that incorporate non-linear functions within them, and that have the capacity to operate with low-quality or missing information, even allowing to obtain forecasts of a certain quality. The results reflect the complexity of obtaining an accurate prediction model for wind energy, and the capacity of neural networks to act as universal nonlinear fitters.

In the framework of this project, a computational tool based on MATLAB has been developed, in charge of preprocessing the time series and for the preparation of the forecast models proposed.

KEYWORDS

Prediction, time series, wind energy, artificial neural networks, MATLAB.



ÍNDICE

Capítulo 1: Introducción	18
1.1 Motivación del trabajo de fin de grado	18
1.2 Objetivos del trabajo de fin de grado	20
1.3 Estructura del documento.....	22
Capítulo 2: La atmósfera y el viento	24
2.1 La atmósfera	24
2.1.1 El viento	25
2.2 Circulación atmosférica general	27
2.3 Circulación atmosférica local: vientos locales	28
2.3.1 Vientos locales en las Islas Baleares.....	29
2.4 La Capa Límite Atmosférica (CLA).....	30
2.4.1 Energía cinética en la CLA.....	30
2.4.2 Perfil de velocidades	32
Capítulo 3: Energía eólica	35
3.1 Evolución histórica de los molinos de viento y la energía eólica	35
3.2 Estado actual de la energía eólica a nivel mundial.....	37
3.3 Energía eólica en España y el mercado eléctrico	40
3.3.1 El mercado eléctrico en España	41
Capítulo 4: Predicción de la energía eólica	44
4.1 Predicciones	44
4.2 La previsión de la energía eólica	46
4.3. Estado del arte de la previsión eólica	47
4.4 Métricas para medir el error en la previsión eólica	53
Capítulo 5. Caso de estudio: parque eólico Es Milá	56
5.1 Parques eólicos.....	56
5.2 Parque eólico Es Milá.....	59
Capítulo 6: Preparación de los datos	66
6.1 Preprocesamiento de datos	66
6.2 Preprocesamiento de los datos del parque eólico de Es Milá	68
6.2.1 Limpieza de los datos	68
6.2.2 Integración de los datos meteorológicos con los datos de generación.....	70



6.2.3 Identificación y eliminación de outliers	73
6.3 Superficie de generación del parque eólico Es Milá.....	78
6.3.1 Curva de potencia de un aerogenerador.....	78
6.3.2 Velocidad y dirección del viento.....	79
6.3.3 Curva de potencia del parque eólico.....	82
6.4 Normalización de los datos	84
6.5 Corte de los datos.....	86
6.5.1 Fraccionamiento de los datos del parque eólico de Es Milá	88
Capítulo 7: Redes neuronales artificiales	91
7.1 Introducción a las redes neuronales artificiales	91
7.2 Evolución histórica de las redes neuronales artificiales	97
7.3 Redes neuronales unidireccionales y recurrentes	99
7.3.1 Redes neuronales unidireccionales	99
7.3.2 Redes neuronales recurrentes	100
7.3.3 Entrenamiento de las redes neuronales MLP	103
Capítulo 8: Modelo de Predicción.....	108
8.1 Introducción	108
8.2 Resultados.....	115
8.2.1 Resultados del entrenamiento	115
8.2.1 Resultados de predicción	120
Capítulo 9: Conclusiones y trabajo futuro.....	139
9.1 Conclusiones	139
9.2 Posibles trabajos futuros	141
Bibliografía	143
ANEXOS	149
ANEXO A: Especificaciones técnicas de los aerogeneradores de Es Milá.....	150
ANEXO B: Códigos MATLAB.....	153
B.1: Preprocesamiento_Fase_1.m	153
B.2: Preprocesamiento_Fase_2.m.....	157
B.3: Regenerador_Automatico_Series_temporales.m	164
B.4: Entrenamiento_Validacion_NARX.m.....	166
B.5: Test_de_NARX_Entrenadas.m.....	172
B.6: Coeficiente_de_Fricción.m	175



ÍNDICE DE FIGURAS

Figura 1: Estructura vertical de la atmósfera terrestre. Fuente:[10]	25
Figura 2: Fuerzas que contribuyen a la velocidad del viento. Fuente: Elaboración propia.	26
Figura 3: Angulo de impacto de la radiación solar sobre el globo terrestre. Fuente:[12]	27
Figura 4: Circulación atmosférica general. Fuente:[13].....	28
Figura 5: Principales corrientes de viento en el Mediterráneo. Fuente:[13].....	29
Figura 6: Viento de Tramontana y Mistral. Fuente:[13]	29
Figura 7: Zonas de la CLA. Fuente: Elaboración propia.....	30
Figura 8: Multipala americana del s. XIX. / Savonius / Darrieux. Fuente:[21]	36
Figura 9: Evolución del tamaño de los aerogeneradores entre 1980-actualidad.Fuente:[22].....	37
Figura 10:Capacidad de las nuevas instalaciones anuales en el periodo 2015-2019. Fuente:[1].....	38
Figura 11: Nueva capacidad instalada por región en 2019. Fuente:[1].....	38
Figura 12: Nueva capacidad instalada en 2019. A) onshore (%) – B) offshore (%) Fuente:[1]	39
Figura 13: Instalación total a nivel mundial 2019. A) onshore (%) – B) offshore (%) Fuente:[1]	39
Figura 14: Evolución cronológica de la potencia de la energía eólica instalada. Fuente:[1]	40
Figura 15: Potencia eólica instalada en España 2002-2018. Fuente:[24].....	40
Figura 16: Potencia eólica de cada comunidad autónoma, respecto a la total (%). Fuente:[24].....	41
Figura 17: Flujo de trabajo de un modelo de predicción. Fuente: Elaboración propia.....	45
Figura 18: Flujo de trabajo del modelo de previsión. Fuente: Elaboración propia.	49
Figura 19: Conexiones eléctricas en un parque eólico. Fuente: Elaboración propia.....	56
Figura 20: Disposición en tresbolillo de aerogeneradores en un parque eólico. Fuente:[6].	59
Figura 21: Localización geográfica del parque eólico de Es Milá. Fuente:[49].....	60
Figura 22: Fotografía del parque eólico Es Milá. Fuente:[49].....	60
Figura 23: Numeración y ubicación de los aerogeneradores en el parque de Es Milá. Fuente:[49]...	61
Figura 24: Mapa eólico de España. Fuente:[50]	62
Figura 25: Ubicación de la estación meteorológica J.Mora. Fuente: Consorci per a la Gestió de Residus Sòlids Urbans.....	65
Figura 26:Etapas de preprocesamiento. Fuente:[51].....	67
Figura 27: Ejemplo de hueco temporal grande. Fuente: Elaboración propia.....	69
Figura 28:Hueco temporal menor a 6 horas. Fuente: Elaboración propia.	69
Figura 29: Eliminación de las velocidades de viento de 0 m/s. Fuente: Elaboración propia.	69
Figura 30: Lectura del archivo con Matlab sin modificar. Fuente: Elaboración propia.	71
Figura 31: Cambio de nomenclatura de las direcciones de viento. Fuente: Elaboración propia.....	71



Figura 32: Huecos temporales. Fuente: Elaboración propia.....	71
Figura 33: Huecos temporales rellenados con NaN. Fuente: Elaboración propia.	72
Figura 34: Fechas y horas unidas. Horas en punto. Fuente: Elaboración propia.	72
Figura 35: Obtención de la media de los valores. Fuente: Elaboración propia.....	72
Figura 36: Frecuencia de las diferentes velocidades de viento. Fuente: Elaboración propia.....	73
Figura 37: Energía generada/velocidad del viento. Fuente: Elaboración propia	74
Figura 38: Velocidad del viento a 45 /30 m. Fuente: Elaboración propia.....	75
Figura 39: Resultado de la aplicación del filtro o. Fuente: Elaboración propia.....	76
Figura 40: Resultado del filtro 1: Fuente: Elaboración propia.....	76
Figura 41: Resultado de aplicar el filtro 2: Fuente: Elaboración propia.....	77
Figura 42: Ejemplo de aplicación del filtro 3. Fuente:[53]	77
Figura 43: Curva de potencia de un aerogenerador. Fuente:[54]	79
Figura 44: Nombre de las direcciones del viento en las Islas Baleares. Fuente:[55].....	80
Figura 45: Rosas de los vientos de las diferentes estaciones del año. Fuente: Elaboración propia. ..	81
Figura 46: Rosa de los vientos global. Fuente: Elaboración propia.	81
Figura 47: Vista superior de la superficie de generación. Fuente: Elaboración propia.	83
Figura 48: Vista superior inclinada de la superficie de generación. Fuente: Elaboración propia.	83
Figura 49: Vista lateral de la superficie de generación. Fuente: Elaboración propia.....	84
Figura 50: Partición de los datos en subconjuntos. Fuente: Elaboración propia.....	87
Figura 51: Flujo de trabajo, usando los tres conjuntos de datos. Fuente: Elaboración propia.....	87
Figura 52: Flujo de trabajo de la fragmentación de los datos de Es Milá. Fuente: Elaboración propia.	88
Figura 53: Ejemplo de la fragmentación de una de las tablas estacionales. Fuente: Elaboración propia.	89
Figura 54: Ejemplo de la regeneración de la señal mediante la media móvil. Fuente: Elaboración propia	89
Figura 55: Clasificación de las series temporales regeneradas según el uso. Fuente: Elaboración propia	90
Figura 56: Neurona artificial. Fuente: Elaboración propia.....	92
Figura 57: Ejemplo del flujo de información en una red neuronal unidireccional multicapa. Fuente: Elaboración propia.....	94
Figura 58: Aprendizaje supervisado. Fuente: Elaboración propia	95
Figura 59: Aprendizaje no supervisado. Fuente: Elaboración propia.....	95
Figura 60: Aprendizaje reforzado. Fuente: Elaboración propia.....	96



Figura 61: Clasificación de las redes neuronales. Fuente: Elaboración propia.	99
Figura 62: NARX en lazo abierto. Fuente: Elaboración propia	102
Figura 63: NARX con lazo cerrado. Fuente: Elaboración propia.	102
Figura 64: Ejemplo de ventana temporal en una NARX. Fuente: Elaboración propia.....	102
Figura 65: Flujo de trabajo de la retropropagación del error. Fuente: Elaboración propia.....	104
Figura 66: Red neuronal MLP. Fuente: Elaboración propia.	104
Figura 67: Errores de salida de la red neuronal. Fuente: Elaboración propia	105
Figura 68: Gradiente descendente. Fuente: Elaboración propia.	106
Figura 69: Deltas de los nodos de la capa oculta. Fuente:[67].....	107
Figura 70: Entradas y salidas del modelo de predicción en lazo abierto. Fuente: Elaboración propia.	108
Figura 71: Entradas y salidas del modelo de predicción en lazo cerrado. Fuente: Elaboración propia.	109
Figura 72: Flujo de trabajo de entrenamiento de las NARX del modelo de predicción. Fuente: Elaboración propia	110
Figura 73: Flujo de trabajo de las NARX durante la fase de prueba. Fuente: Elaboración propia	110
Figura 74: Ejemplo de conexiones entre capas en una NARX. Fuente: Elaboración propia.	113
Figura 75: Predicciones de la generación eléctrica de Es Milá para horizontes de 24 horas. Verano. Fuente: Elaboración propia.	121
Figura 76: Predicciones de la generación eléctrica de Es Milá para horizontes de 48 horas. Verano. Fuente: Elaboración propia.	122
Figura 77: Predicciones de la generación eléctrica de Es Milá para horizontes de 72 horas. Verano. Fuente: Elaboración propia.	124
Figura 78: Predicciones de la generación eléctrica de Es Milá para horizontes de 24 horas. Otoño. Fuente: Elaboración propia.	126
Figura 79: Predicciones de la generación eléctrica de Es Milá para horizontes de 48 horas. Otoño. Fuente: Elaboración propia.	127
Figura 80: Predicciones de la generación eléctrica de Es Milá para horizontes de 72 horas. Otoño. Fuente: Elaboración propia.	129
Figura 81: Predicciones de la generación eléctrica de Es Milá para horizontes de 24 horas. Primavera. Fuente: Elaboración propia.	130
Figura 82: Predicciones de la generación eléctrica de Es Milá para horizontes de 48 horas. Primavera. Fuente: Elaboración propia.	132
Figura 83: Predicciones de la generación eléctrica de Es Milá para horizontes de 72 horas. Primavera. Fuente: Elaboración propia.	133
Figura 84: Predicciones de la generación eléctrica de Es Milá para horizontes de 24 horas. Invierno. Fuente: Elaboración propia.	135



Figura 85: Predicciones de la generación eléctrica de Es Milá para horizontes de 48 horas. Invierno.
Fuente: Elaboración propia.136

Figura 86: Predicciones de la generación eléctrica de Es Milá para horizontes de 72 horas. Invierno.
Fuente: Elaboración propia. 137



ÍNDICE DE TABLAS

Tabla 1: Tabla de clases y longitud de rugosidad. Fuente:[6]	32
Tabla 2: Valores de α en función del tipo de terreno. Fuente: [6].....	34
Tabla 3: Direcciones de viento. Fuente: Elaboración propia.	70
Tabla 4: Direcciones de viento y velocidades de viento predominantes en el parque eólico de Es Milá	80
Tabla 5: Coeficientes de la función de ajuste de la superficie de generación del parque eólico. Fuente: Elaboración propia.....	82
Tabla 6: Normalización de las variables de entrada a la red neuronal. Fuente: Elaboración propia..	86
Tabla 7: Métricas de error de la fase de entrenamiento. Invierno. Fuente: Elaboración propia	116
Tabla 8: Medias de las métricas de error de la fase de entrenamiento. Invierno. Fuente: Elaboración propia	116
Tabla 9: Métricas de error de la fase de entrenamiento. Primavera. Fuente: Elaboración propia....	117
Tabla 10: Medias de las métricas de error de la fase de entrenamiento. Invierno. Fuente: Elaboración propia	117
Tabla 11: Métricas de error de la fase de entrenamiento. Otoño. Fuente: Elaboración propia	118
Tabla 12: Medias de las métricas de error de la fase de entrenamiento. Otoño. Fuente: Elaboración propia	119
Tabla 13: Métricas de error de la fase de entrenamiento. Verano. Fuente: Elaboración propia	119
Tabla 14: Medias de las métricas de error de la fase de entrenamiento. Otoño. Fuente: Elaboración propia	120
Tabla 15: Métricas de error de la fase de predicción para un horizonte de 24 h en verano. Fuente: Elaboración propia	121
Tabla 16: Medias de las métricas de error de la fase de predicción para horizontes de 24 horas en verano. Fuente: Elaboración propia	122
Tabla 17: Métricas de error de la fase de predicción para un horizonte de 48 h en verano. Fuente: Elaboración propia	123
Tabla 18: Medias de las métricas de error de la fase de predicción para horizontes de 48 horas en verano. Fuente: Elaboración propia	123
Tabla 19: Métricas de error de la fase de predicción para un horizonte de 72 h en verano. Fuente: Elaboración propia	124
Tabla 20: Medias de las métricas de error de la fase de predicción para horizontes de 72 horas en verano. Fuente: Elaboración propia	125
Tabla 21: Métricas de error de la fase de predicción para un horizonte de 24 h en otoño. Fuente: Elaboración propia	126



Tabla 22: Medias de las métricas de error de la fase de predicción para horizontes de 24 horas en otoño. Fuente: Elaboración propia.....	127
Tabla 23: Métricas de error de la fase de predicción para un horizonte de 48 h en otoño. Fuente: Elaboración propia.....	128
Tabla 24: Medias de las métricas de error de la fase de predicción para horizontes de 48 horas en otoño. Fuente: Elaboración propia.....	128
Tabla 25: Métricas de error de la fase de predicción para un horizonte de 72 h en otoño. Fuente: Elaboración propia.....	129
Tabla 26: Medias de las métricas de error de la fase de predicción para horizontes de 72 horas en otoño. Fuente: Elaboración propia.....	130
Tabla 27: Métricas de error de la fase de predicción para un horizonte de 24 h en primavera. Fuente: Elaboración propia.....	131
Tabla 28: Medias de las métricas de error de la fase de predicción para horizontes de 24 horas en primavera. Fuente: Elaboración propia.....	131
Tabla 29: Métricas de error de la fase de predicción para un horizonte de 48 h en primavera. Fuente: Elaboración propia.....	132
Tabla 30: Medias de las métricas de error de la fase de predicción para horizontes de 48 horas en primavera. Fuente: Elaboración propia.....	133
Tabla 31: Métricas de error de la fase de predicción para un horizonte de 72 h en primavera. Fuente: Elaboración propia.....	134
Tabla 32: Medias de las métricas de error de la fase de predicción para horizontes de 72 horas en primavera. Fuente: Elaboración propia.....	134
Tabla 33: Métricas de error de la fase de predicción para un horizonte de 24 h en invierno. Fuente: Elaboración propia.....	135
Tabla 34: Medias de las métricas de error de la fase de predicción para horizontes de 24 horas en invierno. Fuente: Elaboración propia.....	136
Tabla 35: Métricas de error de la fase de predicción para un horizonte de 48 h en invierno. Fuente: Elaboración propia.....	137
Tabla 36: Medias de las métricas de error de la fase de predicción para horizontes de 48 horas en invierno. Fuente: Elaboración propia.....	137
Tabla 37: Métricas de error de la fase de predicción para un horizonte de 72 h en invierno. Fuente: Elaboración propia.....	138
Tabla 38: Medias de las métricas de error de la fase de predicción para horizontes de 72 horas en invierno. Fuente: Elaboración propia.....	138





CAPÍTULO 1: INTRODUCCIÓN

En este capítulo se presenta la importancia que tiene disponer de modelos de predicción para la producción eléctrica en parques eólicos, a la vez, que se describen los objetivos del trabajo de fin de grado. Para finalmente realizar un breve resumen del contenido de los diferentes capítulos de este trabajo.

1.1 MOTIVACIÓN DEL TRABAJO DE FIN DE GRADO

La Primera Revolución Industrial, que se inició en la segunda mitad del siglo XVIII en Reino Unido, y que se extendió unas décadas más tarde a gran parte de Europa occidental y América Anglosajona, dio lugar al mayor conjunto de transformaciones económicas, tecnológicas y sociales de la historia de la humanidad desde el Neolítico. Se produce el paso de una economía fundamentalmente agrícola a una economía industrial, hecho que influyó sobre la población, que empieza a experimentar un continuo y rápido crecimiento. A mediados y finales del siglo XIX se inicia la Segunda Revolución Industrial, la cual supone un punto de inflexión en el consumo energético. Se extiende el uso de combustibles fósiles como el petróleo, el gas, y el carbón, pero el crecimiento en el consumo del petróleo es tan rápido que a mediados del siglo XX es la fuente de energía más utilizada. La crisis del petróleo de los años 70 sirve de impulso a las energías renovables, debido al deseo de los países occidentales de buscar alternativas que los libren de la dependencia de los estados petroleros. También influye el auge del movimiento ecologista, que germina en esta época a raíz de varios desastres ambientales que sensibilizan a parte de la sociedad en contra de la contaminación.¹

A raíz del impulso recibido a lo largo de las últimas décadas, las energías renovables han experimentado un crecimiento enorme y a día de hoy satisfacen una parte considerable de la demanda energética mundial. Del conjunto de energías renovables se debe destacar por su madurez tecnológica, la elevada penetración en el sistema eléctrico, y el potencial de crecimiento que todavía tiene, a la energía eólica. Desde el año 2001 al 2019 la potencia de la energía eólica instalada a nivel mundial ha pasado de los 24 GW a los 651 GW. Durante el mismo período, en España ha pasado de 3,4 GW a 24,3 GW convirtiéndose en la principal fuente de energías renovables de este país [1].

Este crecimiento de las energías renovables, y especialmente de la energía eólica, ha permitido reducir la huella de carbono, disminuir la dependencia energética con respecto a terceros, y generar empleo. El aumento de la penetración de la energía eólica ha generado múltiples beneficios ambientales y económicos, pero el uso de este recurso energético eléctrico también presenta una serie de inconvenientes técnicos y ambientales:

¹ Durante los años 50, 60, y 70 ocurrieron varios eventos que avivaron la conciencia medioambiental de parte de la población. Por ejemplo, en 1954, miembros de la tripulación del buque pesquero Daigo Fukuryu fueron expuestos a radiactividad de una prueba de bomba de hidrógeno en el atolón Bikini. En 1969 hubo un vertido en una excavación petrolífera en el Canal de Santa Bárbara de California. Fuente: [68]



- **Inconvenientes técnicos**

- a) Los parques eólicos suelen requerir de largos periodos para su puesta en servicio, generando costes desde el primer instante.
- b) La variabilidad del viento, y consecuentemente la variabilidad de la generación eólica, dificulta la integración en el sistema eléctrico. La generación eléctrica no siempre es proporcional a la demanda, y en el caso de la energía eólica la diferencia puede ser considerable, lo que implica que en ocasiones una parte significativa de la energía generada no puede ser aprovechada debido a que el almacenaje de energía eléctrica a gran escala, esta aun en desarrollo y resulta muy costoso. En el momento que se disponga de sistemas de almacenamiento lo suficientemente eficientes y baratos, estas centrales podrán operar como las convencionales; ajustando la energía inyectada en función de la demanda. Otro problema que puede aparecer está vinculado a la variabilidad del viento, es la desconexión de los aerogeneradores de la red a causa de la aparición de un hueco de tensión.²

- **Factores ambientales**

Para poder instalar un parque eólico en una ubicación dada, este debe superar un estudio de impacto ambiental. Los aerogeneradores son estructuras de gran tamaño que suelen ocupar una extensión considerable de terreno, teniendo un impacto ambiental sobre la población, la flora y la fauna de la región (ruido producido por los aerogeneradores). Por ejemplo, se puede dar el caso que se instalen en zonas en las que intercepten las rutas migratorias de las aves debido a que estas siguen las corrientes de viento.

Si se analiza los inconvenientes expuestos se aprecia como algunos de ellos no son exclusivos de la energía eólica, como por ejemplo la dificultad de almacenaje de la energía, el gran talón de Aquiles de la generación y distribución de la energía eléctrica. El impacto ambiental es inevitable, pero cabe decir que es una característica común entre todas las centrales eléctricas de gran tamaño, y que en el caso de las energías renovables suele ser mucho menor que las grandes centrales convencionales.

La Comisión Europea (2018) presentó su visión estratégica a largo termino para una economía de la Unión Europea próspera, moderna, competitiva y neutra desde el punto de vista del clima para el 2050. Esta estrategia pretende mostrar como Europa puede liderar el camino hacia la neutralidad climática mediante la inversión en soluciones tecnológicas realistas, la capacitación de los ciudadanos y la armonización de las acciones en ámbitos clave como la política industrial, el financiamiento o la investigación, garantizando al mismo tiempo la justicia social para una transición justa; estos objetivos están alineados con los objetivos del Acuerdo de París para contener el cambio climático [2]. Los objetivos de la Unión Europea para el 2030, establecen que se debe reducir como mínimo un 40 % las emisiones de gases de efecto invernadero (respecto el año 1990), que las energías renovables deben proporcionar como mínimo un 32 % de la energía consumida, y se deberá mejorar en un 32,5 % la eficiencia energética en viviendas [3]. En cuanto a los planes nacionales en materia de clima y energía, en España el Plan Nacional Integrado de Energía y Clima 2021-2030 (PNIEC) establece entre otros que para el 2030 el 42 % de la energía eléctrica final utilizada provenga de energías renovables

² Un hueco de tensión es una reducción brusca de tensión de alimentación de una red eléctrica a un valor situado entre el 90 % y el 1% de la tensión declarada, seguida del restablecimiento de la tensión después de un corto período de tiempo. Por convenio un hueco de tensión tiene una durada de entre 10 ms (medio ciclo a 50 Hz) y 1 minuto. Los huecos de tensión se pueden dar en una, dos o tres fases [69].



[4]. En las Islas Baleares la Ley de Cambio Climático y Transición Energética de las Islas Baleares establece que para el 2050 la energía eléctrica provenga al 100 % de energías renovables. Para el 2030 las previsiones son disponer de un 35 % de renovables (85 % en Menorca), un 23 % de reducción del consumo energético, y una reducción del 40 % de las emisiones contaminantes [5].

Teniendo en cuenta que actualmente las energías renovables representan el 22,6 % de la potencia nacional instalada, los plazos de instalación de nuevas centrales eléctricas de energías renovables se van a agilizar, beneficiados por este nuevo marco de la UE. En cuanto a los problemas con los huecos de tensión, las mejoras tecnológicas constantes de los sistemas de control y gestión de los aerogeneradores están reduciendo la magnitud de este problema. Finalmente se llega a la conclusión que el inconveniente principal de la energía eólica es la variabilidad del viento. A fin de poder planificar la generación, se debe intentar predecir de forma precisa el viento que habrá en el futuro. Un problema de predicción consiste en analizar los valores que ha tenido un conjunto de variables en el pasado y el presente con el objetivo de identificar una relación que permita extrapolar los valores que tomará esta variable en el futuro. La importancia de las previsiones radica en poder optimizar la gestión de la generación de manera que los costos de generación sean los mínimos posibles, siempre garantizando la calidad del suministro.

En el sector eléctrico la cuestión que ha suscitado más interés a lo largo de las últimas décadas ha sido conocer la demanda eléctrica que se dará en el corto plazo, pero ante el gran auge de las energías renovables, debido a su mayor penetración, en el sistema eléctrico ha aparecido una nueva cuestión de interés; conocer con antelación la energía eléctrica que se va a obtener a partir de estas fuentes. Dado que la energía eólica es la fuente de energías renovables con más presencia en España y Europa, y en que se han centrado los mayores esfuerzos para crear herramientas de predicción de generación eléctrica fiables que ayuden a integrar la energía eólica en la red. Estas predicciones tienen importantes repercusiones económicas y técnicas referentes a la operación del sistema eléctrico.

1.2 OBJETIVOS DEL TRABAJO DE FIN DE GRADO

El objetivo principal de este Trabajo de Fin de Grado (TFG) es implementar un modelo de predicción de la generación eléctrica de parques eólicos para horizontes cercanos utilizando técnicas de machine-learning.

Generalmente, los datos reales de generación de los parques eólicos no suelen estar disponibles para que la comunidad científica pueda hacer uso de ellos. Por ello, resulta interesante disponer de modelos de predicción que requieran del menor número posible de datos para realizar las previsiones, o incluso que sean capaces de operar con ausencia de algunos de estos datos como es el caso de los modelos que incorporan redes neuronales artificiales, ya que estos funcionan como aproximadores “universales” de relaciones no lineales. A continuación, se presentan los objetivos específicos del presente trabajo:

Objetivo 1: Desarrollo y evaluación de los fundamentos básicos de los procesos atmosféricos.



Adquirir conocimientos básicos sobre los procesos atmosféricos y la variación de la velocidad del viento en la capa límite atmosférica resulta fundamental para entender la evolución histórica de la energía eólica y la variabilidad del viento.

Objetivo 2: Conocer el estado de arte de la energía eólica, del funcionamiento del mercado eléctrico en España, y la presentación de los principales métodos de predicción eólica.

Conocer el despliegue actual de la energía eólica en el mix de generación eléctrica a nivel mundial, permite entender la importancia que tiene disponer de herramientas de predicción eficaces para mejorar el aprovechamiento del recurso eólico. Todos los modelos de predicción comparten algunas características. Por lo tanto, disponer de un conocimiento básico sobre las diferentes metodologías utilizadas, permite decidir qué tipo de modelo de predicción se adapta mejor a la información disponible.

Objetivo 3: Conocer y desarrollar metodologías para el preprocesamiento de series temporales de datos para su posterior uso en la previsión de la generación eólica de parques.

Los datos procedentes de aplicaciones reales suelen presentar una serie de deficiencias. No siempre se tiene disponible una base de datos (u mediante ficheros de texto u hojas de cálculo), y su obtención es parte del proceso. La obtención de la información puede realizarse mediante la colocación de sensores, búsqueda en la web, peticiones, formularios, etc. Una vez se tiene la información que se considera necesaria para resolver el modelo propuesto, se le debe aplicar un preprocesado, es decir se debe preparar para la siguiente fase sin perder información de valor. El objetivo es representar la información en un formato que reduzca el coste de cómputo y optimice los resultados de los algoritmos. Hay muchas técnicas y métodos diferentes que se pueden aplicar. Algunos de ellos son:

- Reducción de la dimensionalidad
- Discreción de variables
- Normalización
- Saneamiento

El preprocesado es una fase crítica, ya que condiciona el resto de los procesos y puede causar la diferencia entre el éxito y el fracaso del modelo. Por ello, esta fase y las que le proceden son tan dependientes que lo habitual es iterar sobre ellas hasta encontrar la combinación que mejor se ajuste.

Objetivo 4: Conocer y desarrollar técnicas de previsión de series temporales mediante el uso de redes neuronales artificiales.

Abordar los fundamentos básicos de las redes neuronales artificiales sobre los cuales se construirá un modelo de predicción, basado sobre redes neuronales recurrentes.



1.3 ESTRUCTURA DEL DOCUMENTO

En este subapartado se presenta de forma resumida la estructura y el contenido de cada una de las partes del trabajo. Éste se estructura alrededor de 9 capítulos, y a fin de guiar al lector a través de los diferentes contenidos, se aporta un breve resumen de cada uno de ellos:

- **Capítulo 1:** En este capítulo se aborda una introducción general al problema a resolver, y los objetivos del trabajo de fin de grado.
- **Capítulo 2:** En este capítulo se introducen fundamentos básicos sobre la atmósfera y se define el concepto del viento realizando una descripción de las fuerzas implicadas en su generación. Posteriormente se describe de forma breve la circulación atmosférica general terrestre, las principales corrientes de aire que afectan al parque eólico de Es Milá, y el flujo de aire en la capa límite de la atmósfera.
- **Capítulo 3:** En este capítulo se realiza una descripción de la evolución histórica del molino de viento hasta el estado de arte actual de la energía eólica a nivel mundial, y nacional. Finalmente se realiza una descripción del funcionamiento del mercado eléctrico español y de una serie de conceptos relacionados.
- **Capítulo 4:** El capítulo se inicia con una introducción del concepto de predicción haciendo hincapié en las cuestiones que se deben responder antes de elegir el modelo de predicción a utilizar. Seguidamente se describen las diferentes metodologías de predicción y se presenta el estado de arte de los modelos de predicción más importantes en función del horizonte. Se presenta el desarrollo matemático de un modelo de predicción físico y de un modelo de predicción estadístico. Finalmente se describen las principales métricas para medir el error en la previsión eólica, y se realiza un desglose de sus ventajas y desventajas.
- **Capítulo 5:** En este capítulo se describen las características óptimas que debe cumplir el emplazamiento de un parque eólico. Para seguidamente abordar la descripción del parque eólico de Es Milá (Menorca), con el objetivo de describir el origen de los datos meteorológicos y de generación que se utilizarán para realizar las predicciones.
- **Capítulo 6:** En este capítulo se describe la metodología seguida para preparar los datos para ser utilizados en el modelo de predicción. Se describen las fases de limpieza, relleno de datos, integración, normalización, corte, y los resultados obtenidos. Gran parte del grueso del trabajo de investigación se encuentra en este capítulo.
- **Capítulo 7:** En este capítulo se presentan los fundamentos teóricos básicos de las redes neuronales, con el objetivo de entender el funcionamiento de un modelo de predicción basado en su uso. También se describe mediante un ejemplo el funcionamiento del algoritmo de aprendizaje de la retropropagación del error.



- **Capítulo 8:** En este capítulo se presentan los modelos utilizados para realizar la predicción de la generación eólica del parque eólico de Es Milá. Se divide en dos partes. En la primera parte se exponen los resultados obtenidos durante la fase de validación del modelo de predicción, y en la segunda se presentan los resultados de predicción.
- **Capítulo 9:** En este capítulo se expone las conclusiones y los posibles trabajos futuros. Se presenta los conocimientos adquiridos durante las diferentes fases de este trabajo y se valora los resultados obtenidos. Finalmente se proponen algunas posibles mejoras futuras del modelo.

CAPÍTULO 2: LA ATMÓSFERA Y EL VIENTO

En este capítulo se realiza una descripción de las capas atmosféricas a partir del concepto de gradiente térmico, para a continuación introducir el concepto de viento y describir de forma breve la circulación atmosférica general en el globo terrestre y algunas características de los vientos locales presentes en Menorca. Seguidamente se describe la capa límite atmosférica y el perfil de velocidades con el objetivo de introducir conceptos fundamentales para comprender la forma con la que se obtiene energía eléctrica a partir de la energía cinética del viento.

2.1 LA ATMÓSFERA

La atmósfera es la capa gaseosa que envuelve a la Tierra. Está constituida por una mezcla de gases juntamente con partículas sólidas y líquidas en suspensión en cantidades y composición variables. La masa total de la atmósfera es aproximadamente de $5,3 \cdot 10^{18} \text{ kg}$. La fuente primaria de energía para los procesos atmosféricos es la radiación solar, de la cual solo una pequeña parte se transforma en energía cinética vinculada al movimiento del aire³.

La temperatura del aire varía con la altura de manera que la atmósfera se puede dividir en capas según el gradiente de temperatura. Se distinguen tres grandes zonas [6]:

a) **Homosfera**: la homosfera es la capa inferior de la atmósfera terrestre clasificada según su composición. Se extiende hasta los 80-100 km de altura aproximadamente y se caracteriza por mantenerse constante la concentración de la mayoría de los gases constituyentes allí presentes [7].

- **Troposfera**: zona inferior de la atmósfera que engloba aproximadamente el 80 % de la masa atmosférica. En esta zona la temperatura disminuye según un gradiente medio de $6,5^\circ\text{C}/\text{km}$ a medida que aumenta la altura. La parte superior de esta región se denomina tropopausa y separa la troposfera de la estratosfera. Su altura varía en función del punto terrestre, siendo de entre 6 y 8 km en los polos y de 16 a 18 km en el ecuador.
- **Estratosfera**: en esta capa se produce una inversión térmica. La temperatura aumenta con la altura, frenando los movimientos verticales del aire. Los principales intercambios energéticos son de tipo radiativo. El límite superior se denomina estratopausa y está a unos 50 km de altura.
- **Mesosfera**: vuelve a disminuir la temperatura con la altura, llegando a sus valores mínimos (alrededor de -90°C) a una altura de 90 km.

³ Se considera que aproximadamente un 2 % de la energía solar que llega al globo terrestre, se transforma en energía cinética del movimiento del aire.

b) Heterosfera: la atmósfera superior o heterósfera es la capa superior de la atmósfera terrestre clasificada según su composición. Se desarrolla sobre la Mesosfera, a partir de los 80 km de altitud aproximadamente. También se le denomina Ionosfera debido a que las capas de nitrógeno, oxígeno y helio se encuentran ionizadas [8].

- **Termosfera**: la temperatura aumenta con la altura llegando a unos 2000 K en la parte superior que se encuentra a una altura de unos 200-500 km. El aire está muy enrarecido.
- **Metasfera**: en alturas superiores a los 500 km el movimiento de las partículas viene condicionado por el campo magnético de la Tierra.

c) Exosfera: la exosfera es la capa de la atmósfera de un planeta o satélite en la que los gases poco a poco se dispersan hasta que la composición es similar a la del espacio exterior. Es la capa menos densa y su ubicación varía en cada astro, en el caso de la Tierra comienza a los 690 km del suelo, y por ejemplo en la Luna se encuentra a nivel del suelo [9]. En la Figura 1, se puede observar la estructura vertical de la atmósfera:

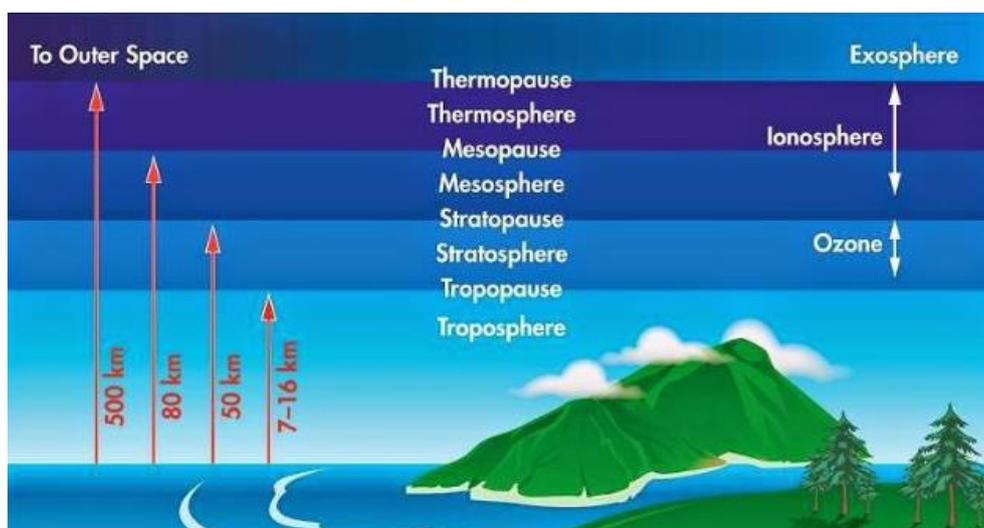


Figura 1: Estructura vertical de la atmósfera terrestre. Fuente:[10]

2.1.1 EL VIENTO

El viento es aire en movimiento [11], [6]. Este movimiento se debe a las diferencias de densidad y presión presentes en la atmósfera, que a su vez se deben a las diferencias de temperatura ocasionadas por la variación de la radiación solar sobre las diferentes partes del globo terrestre. Dos de los parámetros del viento que se deben tener en cuenta siempre a la hora de planificar cualquier aplicación eólica, son la dirección y la velocidad del viento. La rosa de los vientos es una representación conjunta de estas dos magnitudes que permite ver la dirección predominante a simple vista.

Velocidad del viento

La velocidad del viento depende de la acción de varias fuerzas [6]:

- *Fuerza debida al gradiente de presión (F_p)*: fuerza perpendicular que va desde las isóbaras de alta presión a las de baja presión.
- *Fuerza de Coriolis (F_d)*: la rotación de la Tierra sobre su propio eje provoca la desviación de la masa de aire hacia la derecha (hemisferio norte) y hacia la izquierda en el hemisferio sur.
- *Fuerza centrífuga (F_c)*: se debe a la curvatura de las isobaras y actúa en la dirección del radio de curvatura.
- *Fuerza de rozamiento (F_r)*: tiene influencia en las capas bajas de la atmósfera próximas a la superficie terrestre (capa límite). Depende de las características locales de la superficie.

En la Figura 2, se puede observar la representación vectorial de estas fuerzas:

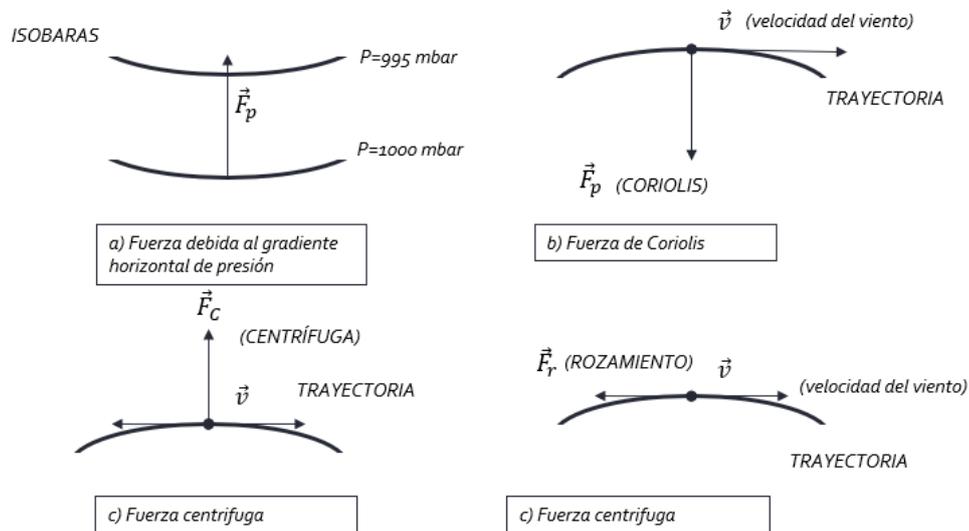


Figura 2: Fuerzas que contribuyen a la velocidad del viento. Fuente: Elaboración propia.

En función de las fuerzas que se consideran, se puede distinguir entre dos tipos de viento, el viento geostrófico y el viento de gradiente. El viento geostrófico es el viento resultante de la acción conjunta de las fuerzas debidas al gradiente de presión y la fuerza de Coriolis. El viento de gradiente es el resultante de las fuerzas debidas al gradiente de presión (F_p), Coriolis (F_d) y la fuerza centrífuga (F_c).

Para alturas superiores a los 1000 m, zona conocida como atmósfera libre, los vientos geostróficos y los de gradiente describen adecuadamente las condiciones de viento reales. A alturas menores, el viento de superficie está muy influenciado por el perfil orográfico, la rugosidad del terreno y los obstáculos. En aplicaciones de energía eólica, interesa estudiar los vientos de superficie que se encuentran en la denominada capa límite de la atmósfera.

2.2 CIRCULACIÓN ATMOSFÉRICA GENERAL

Según se presenta en [6], el movimiento del aire se produce principalmente en la troposfera. Influyen los siguientes factores:

- **Radiación solar:** se distribuye de forma irregular en el globo terrestre, siendo mayor en la zona ecuatorial y menor en los polos.⁴ En la Figura 3, se puede observar los diferentes ángulos de impacto de la radiación solar sobre el globo terrestre:

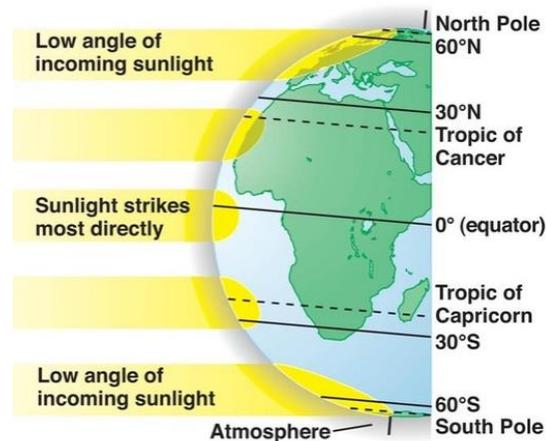


Figura 3: Ángulo de impacto de la radiación solar sobre el globo terrestre. Fuente:[12]

Cabe destacar que el ángulo de impacto varía según la estación del año, debido a la diferente posición relativa de la Tierra respecto al Sol. Los otros factores de influencia son el **movimiento rotacional de la Tierra** que produce el efecto Coriolis y **la orografía local y las diferencias en la presión atmosférica**.

En la zona ecuatorial se origina un movimiento ascendente convectivo de las masas de aire. El aire que se eleva se traslada hacia latitudes superiores, siendo sustituido por aire más fresco de los trópicos, dando lugar a la aparición de los vientos alisios que tienen un carácter muy regular. Este fenómeno da lugar a las denominadas células de Hadley⁵. En los polos tiene lugar un proceso similar dando lugar a las células polares y a las de Ferrel.⁶ En latitudes ecuatoriales se origina una depresión originada por el mayor calentamiento, ejerciendo un efecto de succión de aire hacia ambos hemisferios. Luego los núcleos templados están separados por zonas de altas presiones, en latitudes

⁴ El Ecuador presenta dos periodos de máxima insolación en los equinoccios y dos periodos mínimos en los solsticios.

Desde el Ecuador, todas las latitudes comprendidas entre los Trópicos de Cáncer y de Capricornio poseen dos máximos y dos mínimos, pero un máximo domina en mayor grado a medida que nos acercamos a un Trópico u otro. En el hemisferio Norte el máximo del equinoccio de primavera es menor que el máximo del equinoccio de otoño. Desde el Trópico de Cancer hacia el Círculo Polar Ártico hay un ciclo anual de insolación continuo, con un mínimo en el solsticio de invierno y un máximo en el solsticio de verano. En el hemisferio Sur ocurre lo contrario. En los Polos hay un ciclo de insolación continuo que dura seis meses. En el Polo Norte comienza en el equinoccio de primavera y termina en el equinoccio de otoño, mientras que en el Polo Sur se produce desde el solsticio de verano al solsticio de invierno [70].

⁵ La célula de Hadley recibe este nombre en honor a George Hadley, abogado inglés que, en el siglo XVIII, fue el primero en estudiar por que los vientos alisios en el hemisferio Norte soplan siempre hacia el Oeste sin desviarse al Sur [71].

⁶ La célula de Ferrel se encuentra entre la célula de Hadley y la célula Polar, es decir entre los 30º y 60º. Esta zona también es conocida como Zona de Mezclas [72].

alrededor de los 30°. Son zonas de mucha calma, y pocas precipitaciones. En las zonas templadas los vientos predominantes vienen del oeste, comúnmente conocidos como *Westerlies*. En estas zonas entran masas de aire frío y aire caliente dando lugar a sucesivos frentes cálidos y fríos, generando formaciones nubosas y abundantes precipitaciones. Finalmente, las zonas polares se caracterizan por altas presiones y escasas precipitaciones. Se origina un flujo de aire desde el este hacia zonas de más baja presión, alrededor de los 60°.

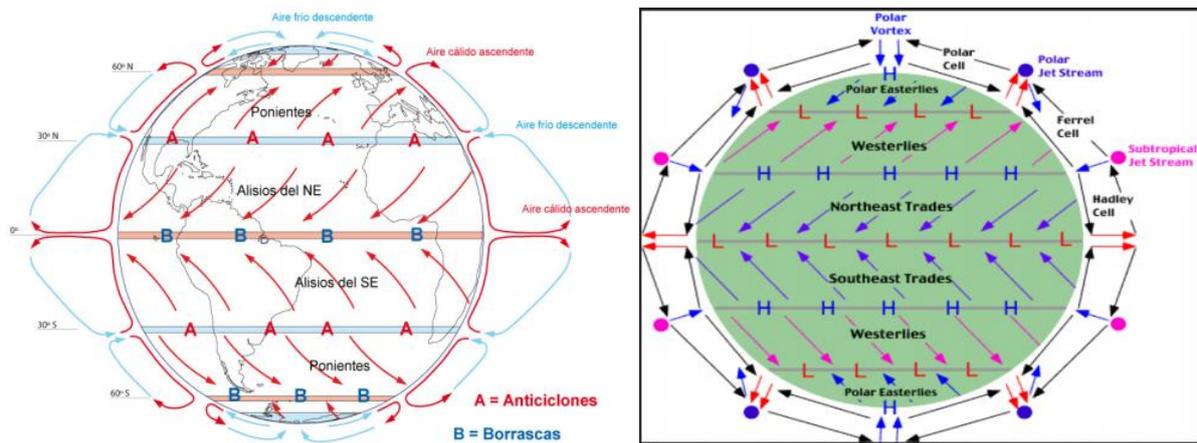


Figura 4: Circulación atmosférica general. Fuente:[13]

2.3 CIRCULACIÓN ATMOSFÉRICA LOCAL: VIENTOS LOCALES

La existencia de topografía compleja, diferencias de rugosidad en el entorno de un emplazamiento, u otros fenómenos locales, pueden originar vientos que se apartan de las leyes generales expuestas para los vientos globales a macro escala [6]. Muchas aplicaciones de la energía eólica son factibles gracias a la existencia de estos fenómenos, como podrían ser las brisas marinas y las brisas de valle:

- *Brisas de mar y brisas de tierra:* durante el día la tierra se calienta más que el mar. Se genera una corriente de aire ascendente. Durante la noche se produce el fenómeno inverso. La tierra se enfría más rápido que el agua del mar, y se genera una corriente descendiente.
- *Brisas de valle y de montaña:* de noche el aire situado a mayor altitud tiende a enfriarse más rápido, descendiendo por la ladera. De día se origina el fenómeno inverso, siendo el aire en contacto con los valles que tiende a ascender.

2.3.1 VIENTOS LOCALES EN LAS ISLAS BALEARES

En el presente trabajo se utilizarán datos meteorológicos del parque eólico Es Milà que se encuentra ubicado en Menorca (Islas Baleares). Por lo tanto, resulta de interés ver algunos detalles sobre los vientos locales presentes en el Mediterráneo. En la Figura 5, se pueden observar algunas de las principales corrientes de aire que circulan en el mar Mediterráneo, y en la Figura 6 se puede ver que los vientos de Tramontana y el Mistral afectan considerablemente a la isla de Menorca.

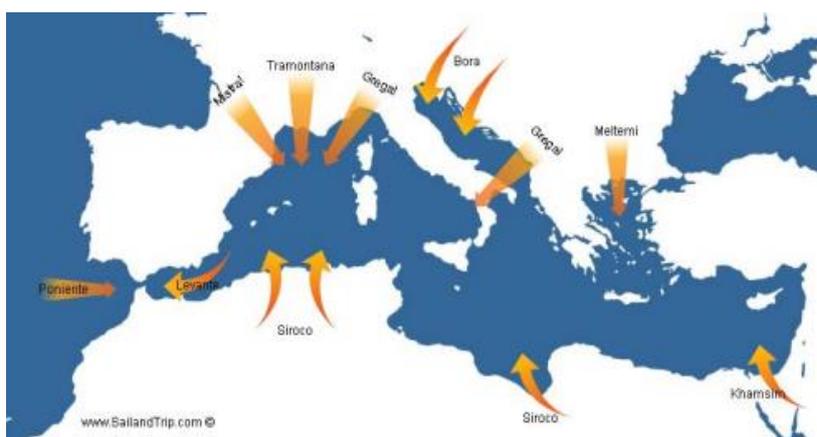


Figura 5: Principales corrientes de viento en el Mediterráneo. Fuente:[13]

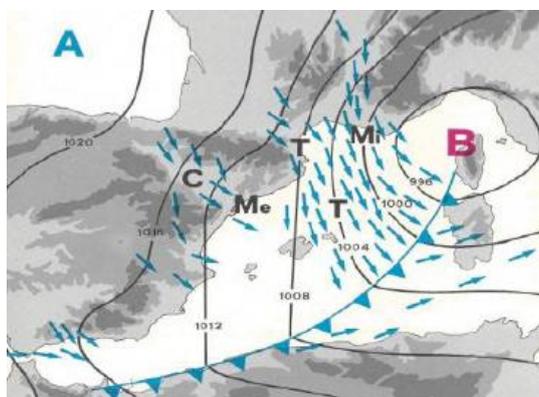


Figura 6: Viento de Tramontana y Mistral. Fuente:[13]

Los vientos de Mistral y de Tramontana se originan debido a un efecto de Venturi⁷. Se produce un aceleramiento de las masas de aire entre los Pirineos, el Macizo Central, y los Alpes, al verse canalizados entre las cadenas de montañas. Suelen ser vientos fuertes que al llegar al mar todavía se aceleran más debido a que disminuye el rozamiento con la superficie. Al tratarse de una isla, la brisa marina también es un fenómeno a tener en cuenta. La existencia de estos vientos juntamente con la orografía de la isla convierte a Menorca en un lugar apto para la instalación de un parque eólico.

⁷ El efecto Venturi consiste en un fenómeno en el que un fluido en movimiento dentro de un conducto cerrado disminuye su presión cuando aumenta la velocidad al pasar por una zona de sección menor [73].

2.4 LA CAPA LÍMITE ATMOSFÉRICA (CLA)

Para poder llevar a cabo una predicción de la generación de energía eléctrica producida por un parque eólico, se tiene que llevar a cabo una predicción de la variación de los campos de velocidades presentes en la Capa Límite Atmosférica (CLA), especialmente para aquellas regiones o alturas incluidas en el área de rotación de las palas del aerogenerador. Según se presenta en [14] y en [6] la CLA se puede dividir en dos zonas :

- *Zona inferior o capa superficial:* parte más cercana a la superficie, llegando a una altura hasta los 100 m. El flujo y la estructura vertical están determinados por las tensiones rasantes a la superficie y el gradiente vertical de temperatura.
- *Zona superior o capa de Ekman:* se encuentra entre los 100-2000 m. En esta zona las variaciones de intensidad y dirección del viento se producen de forma más progresiva. La estructura del flujo también está influenciada por las tensiones rasantes en la superficie, el gradiente vertical de temperatura y la fuerza de Coriolis.

En la Figura 7, se visualiza esta clasificación:

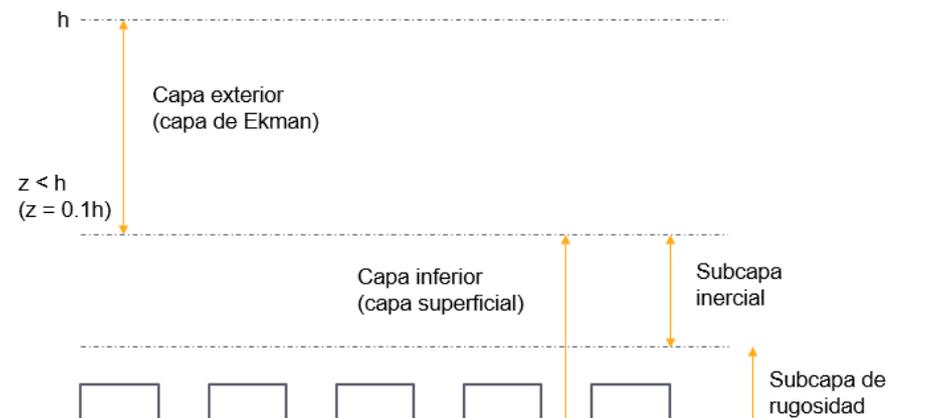


Figura 7: Zonas de la CLA. Fuente: Elaboración propia.

En la actualidad la tendencia es instalar aerogeneradores cada vez más grandes; esto implica aumentar la altura de las torres que en aplicaciones comerciales suelen variar entre 30 y 200 m. Por lo tanto, se puede ver que resulta fundamental disponer de información sobre las características del flujo de aire en la capa límite atmosférica para estudiar.

2.4.1 ENERGÍA CINÉTICA EN LA CLA

Cuando el viento incide sobre el rotor de un aerogenerador tiene una energía cinética asociada ya que cada partícula de aire tiene una cierta velocidad. En cada instante el sistema de control del aerogenerador posiciona el rotor de manera que se maximiza la generación, de forma que se puede



asumir que la componente horizontal del flujo tiene una dirección normal al área de barrido de las palas del aerogenerador [6]. La turbina eólica no aprovecha la energía cinética disponible en su totalidad, ya que el viento aguas abajo del rotor sigue teniendo cierta velocidad y también se producen pérdidas mecánicas y pérdidas debido al rozamiento. En el eje del rotor se obtiene una potencia capaz de accionar una carga, P .

La potencia eólica del viento antes de incidir sobre el rotor se conoce como potencia eólica disponible. Se tiene un diferencial de potencia para cada diferencial de área de la superficie de barrido de las palas:

$$dP = \frac{1}{2} \cdot \rho \cdot v^3 dA \quad (1)$$

Integrando la ecuación (1) se obtiene la ecuación de la potencia eólica disponible:

$$Pd = \int_0^A \frac{1}{2} \cdot \rho \cdot v^3 dA \quad (2)$$

$$Pd = \frac{1}{2} \cdot \rho \cdot v^3 \cdot A \quad (3)$$

Donde:

ρ : densidad del aire al nivel del mar $1,225 \text{ kg/m}^3$.

A : área de barrido por el rotor del aerogenerador (m^2).

v : velocidad del viento medida a la altura del buje del rotor (m/s).

La potencia disponible se relaciona a través de un rendimiento, o coeficiente de potencia C_p con la potencia del rotor:

$$P = C_p \cdot Pd \quad (4)$$

Límite de Betz

El límite de Betz representa el límite máximo para la potencia que se puede extraer de un flujo de viento de velocidad v . De acuerdo con este límite una turbina eólica, independientemente de su diseño puede capturar como máximo el $16/27$ ($59,3 \%$) de la energía cinética del viento, sin tener en cuenta las pérdidas mecánicas ni las pérdidas eléctricas.⁸ Esto significa que el coeficiente de potencia C_p máximo teórico puede ser del $59,3 \%$. Para saber que energía eléctrica produce cada aerogenerador hay que tener en cuenta diferentes parámetros adicionales como el rendimiento de la caja multiplicadora, el rendimiento del generador, etc.

⁸ El nombre de límite de Betz se puso en honor al físico alemán Albert Betz (1885-1968) quién fue el primero en concluir de forma analítica que cualquier turbina eólica no podía superar un rendimiento del $16/27 \%$ [74].

2.4.2 PERFIL DE VELOCIDADES

La velocidad del viento tiende a aumentar con la altura debido a la presencia de turbulencias de origen mecánico y térmico [6]. También juega un papel fundamental las irregularidades de la superficie del suelo y el gradiente vertical de temperatura que genera corrientes convectivas verticales. En el suelo la velocidad del viento es teóricamente nula, y a medida que elevamos la altura, disminuye el rozamiento y aumenta la velocidad. Este efecto se conoce como cizallamiento del viento. La rugosidad del terreno tiene una gran influencia sobre el gradiente de velocidades en las zonas próximas al suelo, y como los aerogeneradores trabajan dentro de esta zona, es un parámetro que tener en cuenta a la hora de implementar cualquier aplicación de aprovechamiento del recurso eólico.

Rugosidad del terreno

A una altura de alrededor de un kilómetro, la superficie terrestre apenas ejerce influencia alguna sobre el viento. Sin embargo, en las capas más bajas de la atmósfera, las velocidades del viento se ven afectadas por la fricción con la superficie, a causa de la rugosidad del terreno, la presencia de obstáculos y la influencia del contorno del terreno. En general, cuanto más pronunciada es la rugosidad del terreno mayor será la ralentización que experimenta el viento. Por ejemplo, la presencia de bosques o ciudades ralentizan mucho el viento, mientras que superficies lisas como pistas de hormigón o superficies de agua ejercen una menor influencia sobre el viento [15]. La rugosidad del terreno se caracteriza a partir de la longitud de rugosidad (z_0), la cual representa la altura respecto al nivel del suelo, hasta que la velocidad del viento es nula. Se puede clasificar en clases en función de su longitud, y la relación entre clase y rugosidad se puede calcular mediante las ecuaciones (5), (6):

$$\text{Si } z_0 \leq 0,03 \text{ m} \rightarrow \text{Clase} = \frac{\ln(5000 \cdot z_0)}{\ln 150} \quad (5)$$

$$\text{Si } z_0 > 0,03 \text{ m} \rightarrow \text{Clase} = \frac{\ln\left(\frac{1000}{9} \cdot z_0\right)}{\ln \frac{10}{3}} \quad (6)$$

Tabla 1: Tabla de clases y longitud de rugosidad. Fuente:[6]

Clase de rugosidad	Longitud de rugosidad (z_0) (m)	Índice de energía	Tipo de Terreno
0	0,0002	100	superficie de agua. Terreno abierto.
0,5	0,0024	73	Pistas de hormigón, césped.
1	0,03	52	Campos abiertos sin cetos. Edificios dispersos. Colinas suaves.
1,5	0,055	45	Campo con algunas casas y arbolado de hasta 8 m situado a 1250 m.
2	0,1	39	Campo con algunas casas y arbolado de hasta 8 m situado a 500 m.
2,5	0,2	31	Campo con muchas casas y arbolado de hasta 8 m situado a 250 m.
3	0,4	24	Pueblos, bosques.
3,5	0,8	18	Ciudades con edificios altos.
4	1,6	13	Grandes ciudades con edificios muy altos.



Para poder deducir la ecuación del perfil de velocidades se realizan las siguientes asunciones:

- Se considera que la atmósfera es neutra, es decir que la variación de temperatura se produce de la misma manera que en un perfil adiabático de temperatura.
- Se define el coeficiente de intercambio turbulento K_m como el producto de una velocidad y una longitud, [14] y [16] :

$$K_m = k \cdot u_* \cdot z \quad (7)$$

Donde:

k = constante de Von Karman ($k=0,4$).

u_* =velocidad de fricción.

z = escala de longitud.

En [14], se presenta la deducción de la siguiente ecuación :

$$\frac{\partial \bar{u}}{\partial z} = \frac{u_*}{k \cdot z} \quad (8)$$

Integrando la ecuación (8) se obtiene el perfil logarítmico de velocidades:

$$\bar{u}(z) = \frac{u_*}{k} \cdot \ln\left(\frac{z}{z_0}\right) \quad (9)$$

Para calcular la velocidad a una determinada altura, conociendo la velocidad a otra altura, la ecuación (9) se puede reformular de la siguiente forma [6]:

$$u' = u \cdot \frac{\ln\left(\frac{z'}{z_0}\right)}{\ln\left(\frac{z}{z_0}\right)} \quad (10)$$

Donde:

u' y u (m/s) : velocidades respectivas a las alturas z' , z (m).

El perfil logarítmico proporciona buenas aproximaciones de la velocidad media del viento en alturas próximas a la superficie, entre 20-100 m. Para alturas superiores es más recomendable utilizar la ley de potencial, o ley de Hellman [17]:

$$\frac{\langle v' \rangle}{\langle v \rangle} = \left(\frac{z'}{z}\right)^\alpha \quad (11)$$

Donde α es un coeficiente que depende de la longitud de rugosidad. En el caso de disponer de datos de velocidad a diferentes alturas se puede obtener despejándolo de la ecuación (11), tal y como se indica en la siguiente expresión:



$$\alpha = \frac{\log_{10}\left(\frac{v'}{v}\right)}{\log_{10}\left(\frac{z'}{z}\right)} \quad (12)$$

Si se aplica la siguiente propiedad de los logaritmos a la ecuación (12), se puede expresar el coeficiente de rugosidad tal como se indica en la ecuación (13):

$$PROPIEDAD \rightarrow \frac{\log_b(x)}{\log_b(a)} = \log_a(x) \rightarrow \alpha = \log_{\frac{z'}{z}}\left(\frac{v'}{v}\right) \quad (13)$$

Existen múltiples variaciones de fórmulas aproximadas para obtener el coeficiente de fricción, pero en el presente trabajo se utilizará la ecuación desarrollada por Abbes & Belhadj [18].

$$\alpha = 0.0910 \cdot \log_{10}Z_o + 0.016 \cdot (\log_{10}Z_o)^2 + 0.24 \quad (14)$$

En la **Tabla 2**, se muestran algunos valores característicos del coeficiente de fricción para diferentes tipos de terreno:

Tabla 2: Valores de α en función del tipo de terreno. Fuente: [6]

Tipo de terreno	α
Terrenos llanos con hielo o hierba	0,08-0,12
Llanos (mar, costas suaves)	0,14
Terrenos poco accidentados	0,13-10,16
Zonas rurales	0,2
Terrenos accidentados y bosques	0,2-0,25
Terrenos muy accidentados y ciudades	0,25-0,4



CAPÍTULO 3: ENERGÍA EÓLICA

En este capítulo se presenta de forma breve la evolución histórica del molino de viento para comprender el origen del aprovechamiento de la energía eólica. Posteriormente, se presenta el estado de arte de la energía eólica a nivel mundial y en España, para que el lector se haga una idea de la importancia de este tipo de energía en el mundo actual. Finalmente, se detalla el funcionamiento del mercado eléctrico y se introduce el concepto de seguridad de suministro, con el objetivo de mostrar la relevancia que tiene el disponer de buenos sistemas de predicción de la energía eólica para mejorar la integración y el aprovechamiento de este recurso en el sistema eléctrico general.

3.1 EVOLUCIÓN HISTÓRICA DE LOS MOLINOS DE VIENTO Y LA ENERGÍA EÓLICA

El origen de los molinos de viento es un tema controvertido y que no ha sido claramente definido. Existen varias tesis en cuanto a los orígenes de estos molinos, y todas ellas cuentan con algún elemento de peso. Tal como se presenta en [19], las principales tesis son:

- a) Las que opinan que el mundo greco-romano ya conocía los fundamentos mecánicos y los principios que regían el movimiento de la rueda expuesta al aire, pero no hay evidencia de su utilización en el ámbito de la molinería.
- b) Las que sostienen que los molinos de viento son de invención oriental, fruto de las culturas árabe y china, y que fueron llevadas al mediterráneo y a Europa por la cultura islámica.
- c) Las que defienden que los molinos de viento fueron inventados en torno al s. XII en la Europa occidental, surgiendo espontáneamente en este continente.

El desarrollo del molino del viento para aplicaciones de energía eólica se puede dividir en una serie de etapas, según se presenta en [20],[6]:

- **Primera etapa:** comprende la evolución del molino de viento desde la edad media, s. XII, hasta el inicio del Renacimiento. Se caracteriza por el uso de los molinos de viento para obtener harina. Esta etapa se caracteriza por una evolución muy lenta debido a las dificultades técnicas.
- **Segunda etapa:** parte desde el Renacimiento (s. XVIII) y termina durante la Revolución Industrial. Aparecen nuevas aplicaciones de los molinos de viento, como por ejemplo el bombeo de agua, y se introducen mejoras en el diseño de las palas y los sistemas de orientación.

- **Tercera etapa:** el Profesor James Blyth usa por primera vez un molino de viento para generar electricidad. A lo largo del s. XIX, en Estados Unidos y Dinamarca aparecen los primeros molinos de viento para la generación eléctrica de mayor tamaño. Se empieza a desarrollar la teoría aerodinámica, y los molinos experimentan una transformación completa de su diseño. Cabe destacar la aparición de la turbina eólica multipala americana, que llegó a convertirse en el modelo de molino de viento más extendido del mundo.

La evolución experimentada a partir del siglo XX se puede dividir en dos períodos:

- **1900-1973:** esta etapa se caracteriza por la instalación extendida de molinos de viento para la generación eléctrica propia (micro eólica). Además, aparecen las primeras turbinas eólicas de eje vertical como el Savonius y el Darrieux, y en 1927 Prandtl y Betz demostraron por primera vez de forma analítica que el rendimiento de los rotores aumentaba con la velocidad de rotación y que en cualquier caso ningún sistema eólico podía aprovechar más del 59,3 % de la energía contenida en el viento. También cabe destacar el desarrollo que experimenta Dinamarca en cuanto a la generación eléctrica mediante la energía eólica. Este desarrollo la convertiría en pionera y en modelo a seguir en los años siguientes. Al final de la II Guerra Mundial, el petróleo presentaba un precio muy bajo, provocando que el interés en el desarrollo de la tecnología eólica, por parte de la mayoría de los países, desapareciera completamente.

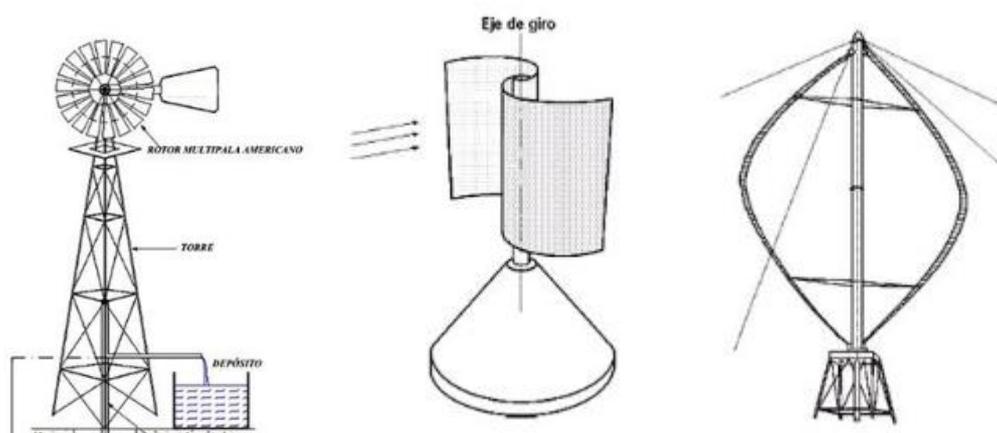


Figura 8: Multipala americana del s. XIX. / Savonius / Darrieux. Fuente:[21]

- **1973-actualidad:** esta etapa se inició con la crisis del petróleo⁹ de los años 70 del siglo pasado. Surge la necesidad de encontrar alternativas energéticas. Durante esta etapa no se aprecian grandes modificaciones en el diseño, pero se produce una evolución tecnológica importante. Se utilizan nuevos materiales más ligeros y resistentes, aparecen los sistemas electrónicos de regulación y control, y aumenta el tamaño y la capacidad de los aerogeneradores.

⁹ La crisis del petróleo fue ocasionada debido a la decisión del bloque árabe de la Organización de Países Exportadores de Petróleo (OPEC), a llevar a cabo un embargo petrolífero a las naciones occidentales, como represalia a los países que apoyaron a Israel en la guerra del Yom Kippur. Fuente:[75]

El espectacular desarrollo experimentado en el aprovechamiento de la energía eólica a lo largo de las últimas décadas ha permitido que la energía eólica sea una de las fuentes de energía más baratas y con una tecnología de aprovechamiento más madura en la actualidad. Se ha consolidado la turbina regulada por paso variable, y el uso de turbinas con rotor de velocidad variable. También se han resuelto problemas de integración en la red eléctrica, como la estabilidad, huecos de tensión, etc.

Las nuevas turbinas se sitúan entre 1 y 3 MW de potencia, y se espera que en un futuro próximo se comercialicen turbinas eólicas de 4 a 6 MW, junto a un gran desarrollo en el sector de la eólica marina (offshore). En la Figura 9, se puede observar la evolución del tamaño de los aerogeneradores a lo largo de las últimas tres décadas:

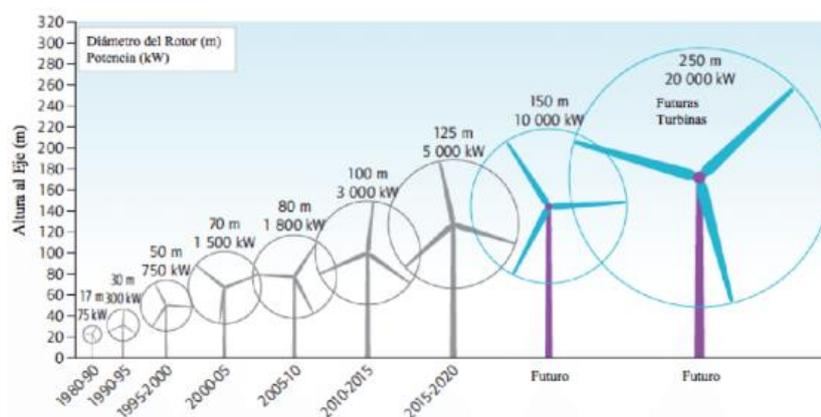


Figura 9: Evolución del tamaño de los aerogeneradores entre 1980-actualidad. Fuente: [22]

3.2 ESTADO ACTUAL DE LA ENERGÍA EÓLICA A NIVEL MUNDIAL

Tal como se presenta en el resumen anual de la energía eólica del *Global Wind Energy Council (GWEC)* [1] a finales del año 2019, la capacidad de producción de energía eléctrica a partir de la energía eólica superó los 651 GW. Esto supone un incremento de un 10 % (60,4 GW de capacidad instalada en 2019) respecto al año 2018. En la Figura 10, se puede visualizar la capacidad de las nuevas instalaciones realizadas a nivel mundial durante el período 2015-2019. Se puede ver que el porcentaje de las nuevas instalaciones de eólica marina aumenta cada año. El coste de realizar una instalación Offshore sigue siendo muy superior a una Onshore debido a las dificultades técnicas que se presentan a la hora de su instalación, el mayor coste de mantenimiento y el mayor coste de evacuación de la energía, pero es el futuro de la energía eólica debido al gran potencial no explotado que todavía existe en este sector y las múltiples ventajas que presenta como la ausencia de obstáculos, la presencia de vientos más constantes y de mayor velocidad, y un menor impacto visual y acústico para la población.



Figura 10: Capacidad de las nuevas instalaciones anuales en el periodo 2015-2019. Fuente:[1]

En la Figura 11, se puede ver la distribución de la nueva potencia eólica instalada por regiones del mundo, durante el 2019.

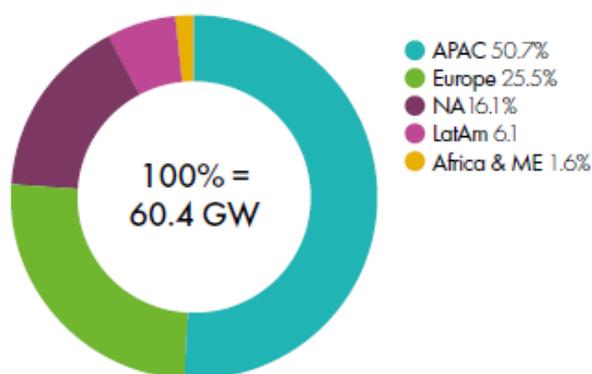


Figura 11: Nueva capacidad instalada por región en 2019. Fuente:[1]

Como se puede observar, la región Asia Pacífica (APAC) se consolida como líder en el desarrollo de la energía eólica, siendo responsable de más del 50,7 % de las nuevas instalaciones a nivel mundial. La APAC incluye gran parte de Asia Oriental, el Sudeste Asiático y Oceanía. Las regiones que menos crecimiento han experimentado han sido América Latina y África, y los países que han instalado una mayor cantidad de capacidad nueva han sido China, Estados Unidos, el Reino Unido, India, y España. A continuación, se presenta la distribución por países de la nueva capacidad instalada (respecto a la total a nivel mundial), y el tipo de instalación:

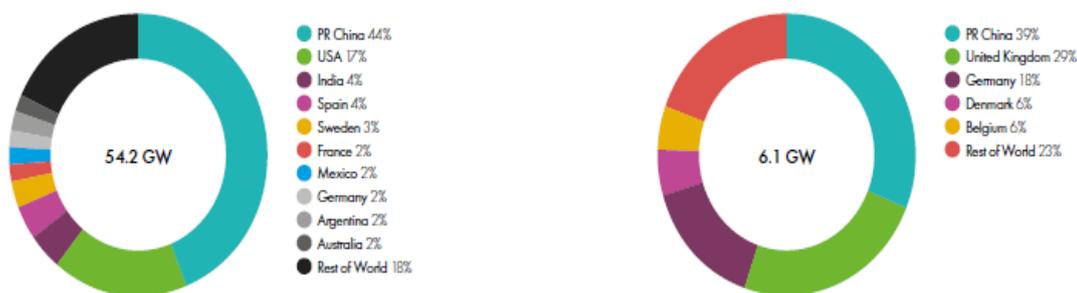


Figura 12: Nueva capacidad instalada en 2019. A) Onshore (%) (parte de la izquierda) – B) Offshore (%) (parte de la derecha); Fuente:[1]

En la Figura 13, se puede observar la distribución por países de la potencia total instalada a nivel mundial según el país y el tipo de eólica:

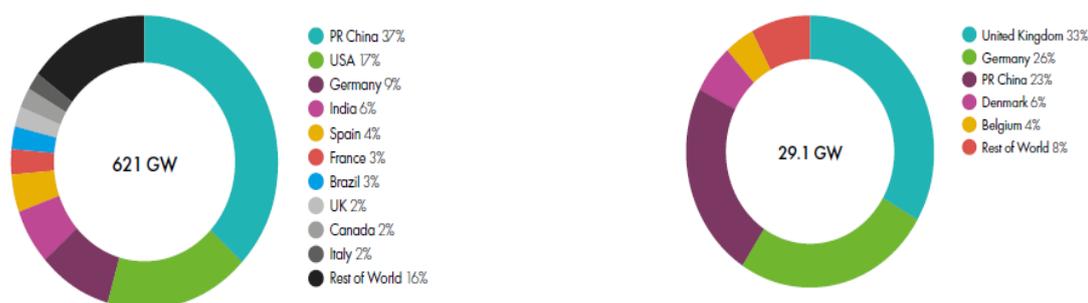


Figura 13: Instalación total a nivel mundial 2019. A) Onshore (%) (parte de la izquierda) – B) Offshore (%) (parte de la derecha); Fuente:[1]

En la Figura 14, se puede observar la evolución cronológica de la capacidad global instalada. En aproximadamente dos décadas se ha multiplicado prácticamente por 30, reflejando el gran desarrollo que ha experimentado el sector de la energía eólica. Esto ha sido posible gracias a que des de 1990 el tamaño de los aerogeneradores y los factores de capacidad se han triplicado, mientras que los costes de generación se han reducido en un 65 %. Se preveía que el año 2020 sería un año de récord, ya que se esperaba que se instalasen más de 76 GW de potencia nueva, pero todavía se desconoce el impacto que tendrá sobre estos números la pandemia ocasionada por el Covid-19.

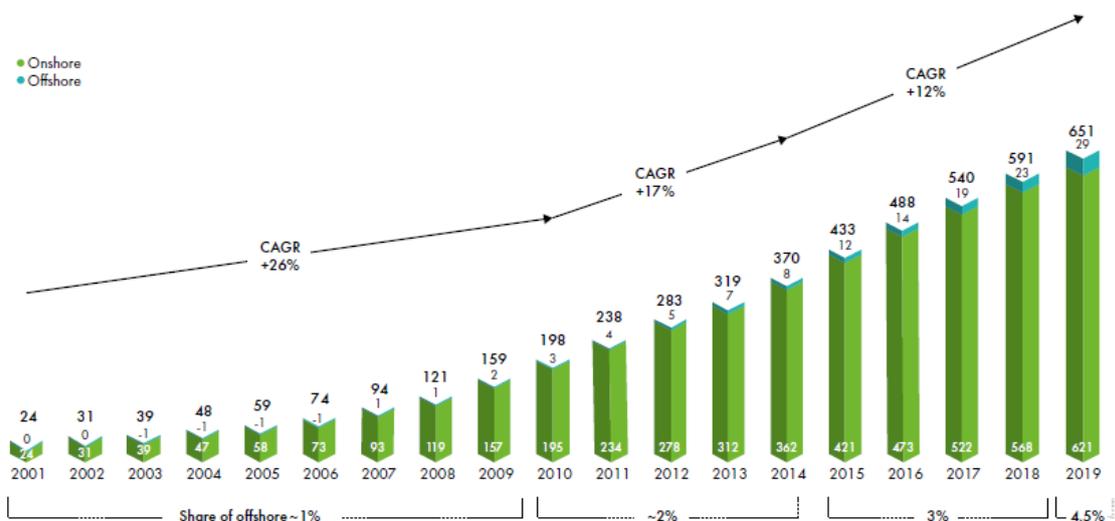


Figura 14: Evolución cronológica de la potencia de la energía eólica instalada. Fuente:[1]

3.3 ENERGÍA EÓLICA EN ESPAÑA Y EL MERCADO ELÉCTRICO

Des de la década del 2000 la energía eólica ha experimentado un crecimiento enorme incentivada por una legislación muy propicia hacia la implementación de las energías renovables, cuando tecnológica y económicamente eran poco rentables. Se incentivaron fuertemente las investigaciones e inversiones en este sector (Real Decreto 661/2007) mediante primas [23]. Para luego, recortar el apoyo de las renovables a raíz de la crisis economía de 2007, mediante sucesivos decretos y cuotas a las renovables. Lo que condujo que de 2013 en adelante no se haya instalado prácticamente nueva potencia eólica en España. En la Figura 15, se puede observar la evolución cronológica de la potencia eólica instalada (en MW) entre el 2002-2018.



Figura 15: Potencia eólica instalada en España 2002-2018. Fuente:[24]

Con una potencia instalada de 23.507 MW a finales del 2018, la generación eólica es la principal fuente de energía renovable en España. En total representa el 22,6 % de la potencia nacional instalada,

siendo la segunda tecnología por detrás del ciclo combinado. En cuanto a las comunidades autónomas, Castilla y León es la región con más potencia eólica instalada con casi un 24 % de toda la potencia nacional, seguida por Castilla-La-Mancha, Galicia y Andalucía [24]. En la Figura 16, se puede ver el desglose por comunidades autónomas a principios del año 2019.

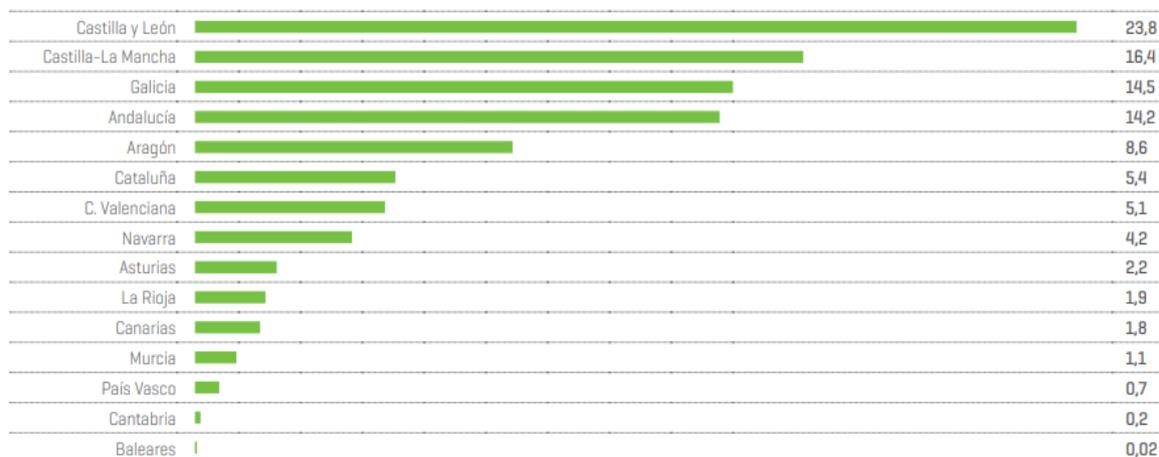


Figura 16: Potencia eólica instalada por comunidad autónoma, respecto al total (%). Fuente:[24]

La gran variabilidad que presenta la generación eólica en la cobertura horaria de la demanda hace que su participación durante el año 2018 oscilase entre valores mínimos del 0,9 % hasta un máximo del 49,2 %. La participación media fue del 23 %. En comparación con el resto de los países europeos, España se mantiene como el segundo país con mayor capacidad eólica instalada por detrás de Alemania, que es claramente el líder con 58 GW instalados, seguida en tercer lugar por Gran Bretaña. En cuanto a la contribución de esta fuente respecto al total de la generación, el líder destacado es Dinamarca con algo menos del 50 % de su producción procedente de esta energía.

3.3.1 EL MERCADO ELÉCTRICO EN ESPAÑA

Antes del 1997 el mercado eléctrico español se caracterizaba por su gran nivel de concentración, y un carácter oligopolista¹⁰, donde entre Endesa e Iberdrola prácticamente abarcaban toda la cuota de mercado. Funcionaba mediante un sistema de intervencionismo estatal llamado Marco Legal Estable, y el sistema de retribución consistía en que el ingreso se dividía entre una parte fija sobre los costes de inversión y otra parte variable sobre los costes de mantenimiento. A partir del 1997 se permite la entrada de nuevos agentes en el mercado eléctrico, como las comercializadoras independientes o las cooperativas de energías renovables. A demás, se establece un sistema de mercado basado en la

¹⁰ Un mercado oligopolista consiste en un mercado dominado por un pequeño número de empresas donde estas tienen la posibilidad de manejar los precios del producto a su antojo mediante una serie de acuerdos recíprocos.



subasta diaria de la energía eléctrica que es gestionada por el *Operador del Mercado de Electricidad (OMEL)* [25].

Se puede distinguir entre el mercado mayorista, el mercado de restricciones técnicas y el mercado minorista:

- En el **mercado mayorista** los generadores y comercializadores negocian el precio de la electricidad mediante un sistema de subasta en el que se estipula un precio y una cantidad de energía eléctrica el día antes de realizar el suministro para cada hora del día siguiente (mercado diario).¹¹ En los mercados intradiarios se realizan los posibles ajustes necesarios en la oferta y la demanda y establecen los precios finales. Finalmente, en los mercados de servicios complementarios se corrigen posibles desequilibrios entre generación y demanda a tiempo real [20].
- Una vez finalizada la subasta en el mercado diario se analizan las restricciones técnicas, es decir si la cantidad de energía ofertada por cada central eléctrica es suficiente o adecuada para el buen funcionamiento del sistema eléctrico. Aparece el **mercado de restricciones técnicas** ya que, debido a la propia estructura del sistema eléctrico, en determinadas zonas geográficas se pueden producir situaciones de exceso o déficit de producción. Es decir, se puede dar el caso que se tenga que incluir centrales de generación que no estaban entre las iniciales unidades oferentes de energía y/o que puedan eliminar otras centrales que sí habían entrado en el programa de casación [26]. En el **mercado minorista** las comercializadoras ponen a disposición de los consumidores domésticos la energía eléctrica, que tienen libertad a la hora de elegir la comercializadora de la cual adquieren la electricidad. La imposibilidad de diferenciar un kWh de otro y el nulo papel que las comercializadoras pueden jugar en la calidad y en la seguridad del suministro eléctrico, hacen que la competencia entre ellas sólo se pueda canalizar a través del nivel de precios [20].

Balance demanda/generación

El operador de un sistema eléctrico debe mantener un equilibrio dinámico continuo entre la energía que se genera y la que se consume ya que todavía no existen formas de almacenar energía eléctrica a gran escala. Este balance mantiene la frecuencia constante, en caso de que haya más generación que demanda la frecuencia baja, y en el caso inverso la frecuencia aumenta. Por lo tanto, se debe planificar con suficiente antelación la energía eléctrica que se va a generar y el tipo de centrales que participarán en su generación. A demás se realiza una reserva rodante de generación para poder mitigar posibles desvíos de última hora [27].

Seguridad de suministro en España

¹¹ La subasta funciona de la siguiente manera: los generadores de electricidad declaran, a través de sus curvas de oferta, los precios mínimos a los que están dispuestos a producir cada kWh, al igual que los demandantes (grandes consumidores, comercializadoras...), que declaran los precios máximos a los que están dispuestos a consumir cada kWh.

El comportamiento maximizador de los agentes les llevará a ofertar sus unidades de producción teniendo en cuenta el impacto de sus pujas sobre la cantidad producida y el precio de mercado, dado el previsible comportamiento de sus rivales. En este cálculo, las empresas se enfrentan a la siguiente disyuntiva: elevar sus pujas por encima de los costes marginales implicará una pérdida de producción, pero que podrá ser al menos en parte, compensada por el incremento resultante en el precio de mercado.

El subastador, conocido como el Operador del Mercado, establece el orden de preferencia económica entre las pujas, casa oferta y demanda, y determina el precio que será pagado por todas las unidades despachadas. El precio depende de la última central de generación de energía eléctrica que entre en el mercado (será la que mayor retribución exija). Este método de determinación de los precios se denomina sistema de precios marginalista. Fuente:[76]



En términos generales el nivel de seguridad de suministro es alto gracias a la gran variedad de fuentes de generación eléctrica que hay a disposición, y el bajo nivel de interconexión con los sistemas eléctricos vecinos. La contribución a la seguridad de suministro de una central no se mide por su potencia instalada, sino por la potencia que puede ofrecer en los momentos de demanda punta con una mínima seguridad o incertidumbre. Esta potencia se denomina potencia firme, que depende de factores como la disponibilidad de la energía primaria utilizada (tipo de combustible, viento, agua, etc.), el estado de mantenimiento de la instalación, y el régimen de explotación.

En función del tipo de tecnología de generación, una misma potencia instalada se puede traducir en diferentes valores de potencia firme [28]:

- *Centrales eólicas, solares (termo solar y fotovoltaica) e hidráulicas (no reguladas)*: aportan poca seguridad de suministro en el corto y medio plazo, ya que ni el viento, ni el sol, ni el caudal de agua están necesariamente disponibles en los momentos de máxima demanda.
- *Centrales nucleares*: la probabilidad de indisponibilidad es muy baja de manera que el nivel de seguridad que aportan es muy elevado.
- *Centrales térmicas convencionales (ciclo combinado, carbón, fuelóleo)*: ofrecen un nivel de seguridad de suministro alto.
- *Centrales hidráulicas regulables*: las que tienen grandes capacidades de embalse y las de bombeo¹², aportan en general una seguridad de suministro elevada, pues tienen incentivos en utilizar el agua en momentos de precios elevados.

Durante las últimas décadas la capacidad instalada de tecnologías de producción a partir de fuentes de energía renovables se ha incrementado enormemente, pero su aportación a la cobertura de la demanda punta no se ha incrementado de forma proporcional. Se debe al carácter intermitente de los recursos naturales necesarios para producir electricidad a partir de las fuentes renovables. Por ejemplo, la energía eólica, que se ha convertido en una de las fuentes de energía más importantes, es una alternativa muy atractiva a las fuentes de energía convencionales pero su integración en el sistema eléctrico conlleva superar una serie de retos debido al bajo nivel de seguridad de suministro que aporta. Un sistema de previsión fiable permite aumentar el nivel de seguridad de suministro del recurso, reduciendo la incertidumbre de generación. En términos económicos reduce los costes de operación del sistema debido a la reducción de reserva necesaria y permite evitar las posibles penalizaciones económicas que se aplican a los agentes, debido a los desvíos producidos en sus compromisos de generación adquiridos en el mercado eléctrico [29].

12 Centrales hidroeléctricas de bombeo: este tipo de central eléctrica cuenta con dos embalses a distinta altura que permiten almacenar el agua en los momentos de menor demanda y aprovecharla para generar energía en las horas de mayor consumo para satisfacer toda la demanda eléctrica.

En las horas valle, generalmente durante la noche en los días laborables y los fines de semana, se usa la energía sobrante —que además en esas horas tiene un coste más bajo en el mercado— para elevar el agua contenida en el embalse situado en el nivel más bajo al depósito superior por medio de una bomba hidráulica que hace subir el agua a través de una tubería forzada y de la galería de conducción. El embalse superior actúa, así, como un depósito de almacenamiento de energía. Fuente: [77]

CAPÍTULO 4: PREDICCIÓN DE LA ENERGÍA EÓLICA

En este capítulo se describe el concepto de predicción y se explica la diferencia entre las predicciones cualitativas y cuantitativas, juntamente con la estructura general del flujo de trabajo de un modelo de predicción cuantitativo. También se muestran las principales preguntas que se debe responder a la hora de diseñar un modelo de predicción, como sería por ejemplo que precisión se requiere, o que coste computacional se puede tolerar, etc. Posteriormente, se realiza una clasificación de los modelos de predicción cuantitativos, en modelos físicos, estadísticos o híbridos, y se presenta el estado del arte actual de la previsión eólica. Con el objetivo de mejorar el entendimiento sobre los modelos físicos y estadísticos se muestra el desarrollo de la estructura de funcionamiento de cada uno de los modelos. Finalmente, se describen las principales métricas que se usan para valorar el resultado de la predicción

4.1 PREDICCIONES

La habilidad de realizar una buena predicción ha sido muy bien valorada a lo largo de la historia. Todas las personas hacen predicciones constantemente, pero el futuro siempre mantiene una cierta incertidumbre, de manera que estas predicciones no siempre son perfectas. El objetivo de una predicción es proporcionar información sobre la evolución temporal de una variable en el futuro, a partir de la información disponible del presente y del pasado. Además, se busca que la información obtenida se ajuste al máximo a lo que sucederá en la realidad. En las últimas décadas se ha producido un gran avance en los sistemas de predicción gracias al incremento en la potencia de cálculo de los ordenadores, que pueden almacenar, analizar y relacionar grandes cantidades de variables con un coste computacional muy reducido.

Clasificación de los métodos de predicción

Los métodos de predicción se pueden clasificar en *cualitativos* y *cuantitativos* [30]:

- ***Cualitativos***: también se conocen como pronósticos subjetivos. No suelen ser reproducibles por terceros, ya que el individuo que realiza el pronóstico no especifica la forma en que ha procesado la información para realizarlo. A pesar de ser un método poco riguroso puede servir como primera solución o aproximación a un problema.
- ***Cuantitativos***: son aquellos métodos que incorporan un modelo matemático o estadístico para realizar la predicción. Una vez establecido el modelo, las predicciones se realizan automáticamente, de manera que son reproducibles por terceros.

Flujo de trabajo de un modelo de predicción cuantitativo

La elaboración de un sistema de predicción suele realizarse en dos etapas. En la primera se construye el modelo de predicción a partir de los datos históricos disponibles y la teoría necesaria. Una vez elaborado el modelo, se tiene que comprobar que este se adecua correctamente a los datos. Podría

incluir variables inadecuadas, no reflejar la relación que se busca, etc. En ese caso se tendría que remodelar el modelo. En la segunda etapa se utiliza el modelo para realizar las predicciones pertinentes. A partir de ese momento se calculan los errores y se realizan correcciones con el objetivo de mejorar el rendimiento [30]. En la Figura 17, se puede visualizar la estructura general del flujo de trabajo de estos modelos.

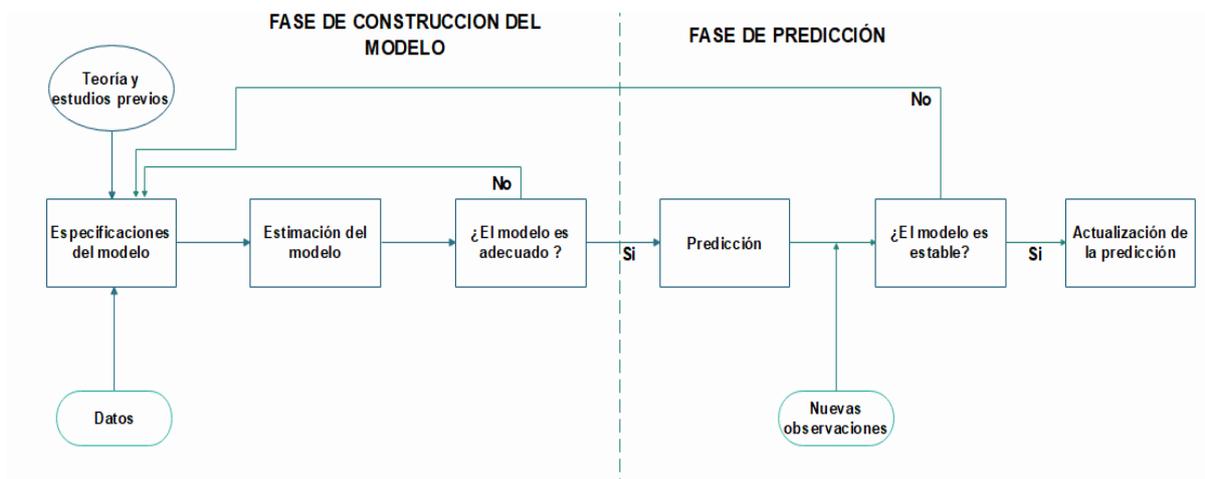


Figura 17: Flujo de trabajo de un modelo de predicción. Fuente: Elaboración propia.

Elección de un modelo de predicción

Entre muchos otros factores, la elección de un modelo de predicción, o unas técnicas concretas, depende de las respuestas a las siguientes preguntas [30]:

- ¿Qué nivel de precisión se necesita?
- ¿Cuál es el horizonte de predicción requerido?
- ¿Cuál puede ser el coste de predicción que se tolera?
- ¿Qué nivel de complejidad se requiere?
- ¿De qué datos se dispone?

Algunas veces, solo se requiere una predicción aproximada, otras, de gran precisión. Normalmente incrementar la precisión implica aumentar considerablemente el coste de adquisición de datos y el coste computacional. Si se puede tolerar una disminución de la precisión, puede resultar más recomendable un modelo menos complejo y más barato. El horizonte de predicción también es un factor esencial, ya que los métodos de predicción para un horizonte cercano o lejano son significativamente diferentes. Finalmente, disponer de buenos datos es fundamental ya que no se puede esperar obtener resultados precisos a partir de una base de datos limitada o incompleta [30]. El auge que han experimentado las fuentes de energía renovables ha planteado un nuevo problema de predicción, el de prever con suficiente antelación la energía que van a generar estas centrales.

4.2 LA PREVISIÓN DE LA ENERGÍA EÓLICA

La energía eólica es uno de los recursos energéticos que requiere un menor coste de inversión para generar energía eléctrica, y se estima que para el año 2025 los costes medios de la electricidad producida podrían disminuir un 35 % para la energía eólica marina, y el 26 % para la energía eólica terrestre en comparación con los precios del 2015 [31]. Esto ha llevado a que haya muchas regiones dónde la penetración de la energía eólica en el sistema eléctrico es tan grande que fluctuaciones en la energía producida, afectan al control, la estabilidad y la gestión del mercado eléctrico. Mediante modelos de previsión se pretende facilitar la gestión de este recurso, y poder planificar la generación basada en fuentes convencionales para poder balancear la demanda y la generación, o incluso para deslastrar la generación si no hay demanda.

Tipos de modelos cuantitativos

Según la metodología utilizada podemos distinguir entre modelos físicos y modelos estadísticos:

- **Físicos:** se basan en la previsión de las condiciones de la capa límite de la atmósfera mediante sistemas numéricos de previsión del tiempo (*Numerical Weather Predictions NWP*) utilizando datos de previsión sobre la temperatura, presión, coeficientes de fricción del suelo y condiciones orográficas. Estos modelos implican un gran costo computacional y requieren resolver y modelar procesos a distintas escalas, lo cual tiene asociado un gran nivel de complejidad, teórico y numérico.
- **Estadísticos:** se basan en el uso de datos históricos de las variables de interés como son la velocidad del viento, la temperatura del aire, o datos de generación. El objetivo es detectar relaciones o patrones entre ellas para poder extrapolar los valores que tomarán estas variables en el futuro. Suelen incluir modelos de caja negra compuestos por redes neuronales y análisis de series temporales.³³ También hay algunos que utilizan modelos de caja gris, los cuales aplican un cierto conocimiento sobre las propiedades del viento para mejorar el resultado.

A partir de una combinación de técnicas procedentes de ambos modelos, surgen los métodos híbridos:

- **MOS (Model Output Statistics):** estos métodos se suelen basar en modelos físicos, pero incorporan un bloque de correcciones a partir de datos estadísticos. Un ejemplo sería un modelo que calcula la velocidad del viento a partir de datos físicos, y luego escala el valor obtenido teniendo en cuenta datos históricos del relieve local del emplazamiento.

³³ Serie temporal: conjunto de valores, correspondientes a una variable, ordenadas según el instante en que fueron tomados, y normalmente tomados en intervalos regulares de tiempo. La predicción del valor futuro de una serie temporal implica determinar los valores que tomará la variable de interés a partir de su valor en el instante presente.



Cada uno de estos modelos tiene sus virtudes, pero en términos generales los físicos son mejores a la hora de realizar una previsión con un horizonte largo, mientras que los estadísticos lo son para previsiones cercanas [32].

Horizontes de predicción

El horizonte de predicción hace referencia al momento del futuro para el cual realizamos la predicción. Se puede utilizar para clasificar las diferentes herramientas de previsión de forma genérica, sin entrar en la metodología utilizada. Una posible opción de clasificación de los horizontes es:

- **Horizonte corto-inmediato (hasta 8 horas de adelanto):** estas predicciones tienen el objetivo de mejorar la seguridad dinámica de la red y la gestión de la regulación secundaria¹⁴.
- **Horizonte corto (de pocas horas a 72h de adelanto):** este tipo de predicciones son las que permiten al operador del mercado eléctrico tomar las medidas necesarias para asegurar la estabilidad del sistema en un futuro. Sirven para organizar el mercado diario y intradiario.
- **Horizonte lejano (múltiples días de adelanto):** este tipo de predicciones son interesantes para la planificación futura de los parques eólicos. Por ejemplo, permiten planificar labores de mantenimiento de los aerogeneradores previendo los días en los que pararlos conllevaran menores pérdidas económicas.

4.3. ESTADO DEL ARTE DE LA PREVISIÓN EÓLICA

En este apartado veremos algunos ejemplos de las herramientas de previsión más exitosas que se han elaborado según el horizonte de previsión.

Horizonte corto-inmediato

Este tipo de previsiones suelen basarse en métodos estadísticos y en el uso de redes neuronales. Un ejemplo de herramienta de este tipo es el WPMS (*Wind Power Management System*). Esta herramienta utiliza redes neuronales artificiales que son entrenadas con una gran cantidad de datos históricos. La ventaja de usar redes neuronales en lugar de otros métodos estadísticos como los modelos ARIMA (modelo Autorregresivo Integrado de Media Móvil), o incluso los modelos NARMA

¹⁴ Regulación primaria y secundaria: son mecanismos de balance que tienen el objetivo de mantener la frecuencia constante en la red eléctrica. La regulación primaria es un control de la frecuencia descentralizado realizado por los generadores conectados a la red con capacidad de prestarla. Determinados generadores son capaces de proveer una regulación primaria significativa por su inercia, midiendo la desviación entre la frecuencia real respecto a la teórica de la red (50 hercios) y oponiéndose a la desviación disminuyendo o aumentando su generación eléctrica durante los primeros 20 segundos, hasta que otros mecanismos de regulación centralizada actúen compensando el exceso o defecto de generación.

La regulación secundaria es un servicio completamente de regulación frecuencia-potencia centralizado que actúa entre los 20 segundos y 15 minutos del desbalance. En este punto los generadores ya no hacen *la guerra por su cuenta* como en la regulación primaria, sino que la consigna de regulación es calculada por un sistema central llamado RCP (Regulación Compartida Peninsular) gestionado por Red Eléctrica de España.



(modelos Autorregresivos No-lineales de Media Móvil) es que pueden modelar mejores problemas no lineales ya que son "aproximadores universales". La red neuronal funciona como un procesador enorme compuesto por unidades muy sencillas. Son capaces de "aprender" e interpolan los resultados incluso cuando las entradas son contradictorias. Identifican las relaciones que hay entre los datos meteorológicos y la velocidad del viento, a partir de datos del pasado. Posteriormente, comparando los resultados de predicción con la realidad se va ajustando la red neuronal hasta encontrar la configuración óptima. Normalmente se necesita como mínimo datos meteorológicos de un año entero para poder detectar las variaciones estacionales. La precisión de la previsión depende en gran medida de la calidad de los datos meteorológicos [33].

Horizonte corto

Se han desarrollado múltiples herramientas para este horizonte. La mayoría de ellas utilizan métodos híbridos, como por ejemplo: el WPPT (The Wind Prediction Tool), Prediktor, Zephyr, Ewind, etc.

Concretamente el WPPT fue desarrollado en Dinamarca para el operador Elsam/Eltra. Puede ser utilizado para realizar predicciones de generación con un adelanto de 36-120 horas. Utiliza modelos no lineales estadísticos que incluyen el modelaje de las curvas de potencia de los parques, y modelos dinámicos de previsión que describen la potencia del viento y las variaciones diurnas. En función de la configuración requiere de:

- Medidas de la generación de los parques eólicos (actualizadas cada 5 min-1 hora).
- Datos de producción de energía de los parques eólicos, con un retraso máximo de unas 3-5 semanas.
- Previsiones meteorológicas de la velocidad del viento en las regiones de funcionamiento.
- Medidas o previsiones locales.

Otro ejemplo que destacar es Prediktor. La idea principal detrás de Prediktor es disponer de un modelo lo más físico posible. Entonces, mediante un método numérico de previsión del tiempo (NWP) obtiene la velocidad del viento en la región de interés. Esta velocidad se escala a la altura del buje de los aerogeneradores mediante un modelo WAsP (*Wind Atlas Analysis and Application Program*), utilizando información sobre la orografía local, obstáculos, longitud de rugosidad del terreno, disposición de los diferentes aerogeneradores y la influencia que tienen entre ellos, etc. Finalmente se utilizan varios filtros MOS para realizar correcciones estadísticas finales, antes de calcular la potencia de generación [34].

Horizonte lejano

Se han creado pocas herramientas de predicción para un horizonte lejano debido a que se requiere de modelos complejos. Se suelen utilizar directamente programas numéricos de predicción del tiempo. Un ejemplo sería la herramienta **Previento**¹⁵. Esta herramienta proporciona predicciones fiables de la velocidad del viento con un adelanto de 10 días y una resolución de 15 minutos. Se basa en un modelo físico que utiliza como entrada los datos meteorológicos que obtiene de un programa

¹⁵ Previento modela la capa límite de la atmósfera teniendo en cuenta la rugosidad y la orografía local de la región de interés. También tiene en cuenta las variaciones diarias del gradiente térmico del aire para actualizar el perfil logarítmico. De esta manera es capaz de calcular la velocidad del viento a la altura del buje del aerogenerador, y posteriormente calcular la potencia de salida a partir de la curva de potencia.[78]

de previsión de tiempo de gran escala llamado LokalModell, del servicio meteorológico alemán. Luego usa las predicciones individuales de algunos parques eólicos para proporcionar predicciones regionales[34].

A continuación, se expone la estructura de funcionamiento de un modelo físico. En la Figura 18, se puede observar el flujo de trabajo del método físico de previsión que se analizará [35]:

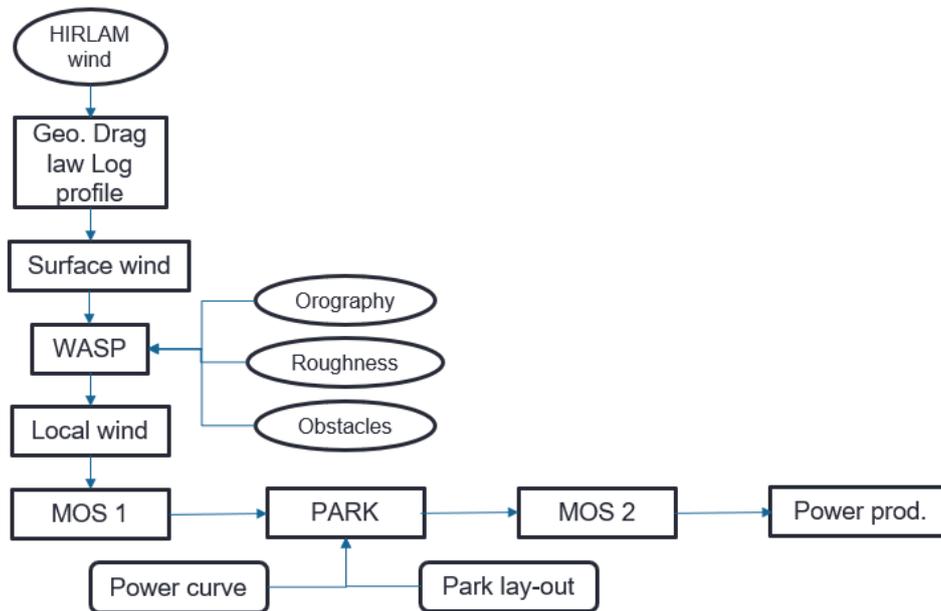


Figura 18: Flujo de trabajo del modelo de previsión. Fuente: Elaboración propia.

Primero se obtienen los datos meteorológicos procedentes del programa numérico de previsión del tiempo, en este caso se usa el HIRLAM (*High Resolution Limited Area Model*).¹⁶Las velocidades de viento que se obtienen representan el viento geostrófico. Para escalar este viento a la superficie se utilizan dos ecuaciones: la ley de resistencia geostrófica (15) y el perfil logarítmico del viento (9) :

$$G = \frac{u_*}{k} \cdot \sqrt{\left[\ln\left(\frac{u_*}{f \cdot z_0}\right) - A \right]^2 + B^2} \quad (15)$$

Donde:

G: módulo del viento geostrófico.

u_* : velocidad de fricción.

k: constante de von Karman.

f: parámetro de Coriolis

¹⁶ HIRLAM es un modelo atmosférico regional. SU resolución horizontal es de 16 km (distancia entre puntos de rejilla), y su alcance temporal de 72 horas



A y B: típicamente se utiliza $A = 1.8$ y $B = 4.5$

Uniendo las ecuaciones (9) , (15) y teniendo en cuenta que la longitud de rugosidad (z_0) es una función que depende de la dirección (θ), se puede definir el viento de superficie (u_s) como una función que depende del viento geostrofico, la dirección y la longitud de rugosidad [32]:

$$u_s = (G, \theta, z) \quad (16)$$

El siguiente paso que se realiza en este modelo es llevar a cabo correcciones locales mediante el WA^{SP}. Este realiza un escalado del viento en función de la dirección. El viento local se podrá expresar mediante el producto del factor de escalado por la velocidad de superficie:

$$w(\theta) \cdot u_s(G, \theta, z) \quad (17)$$

La potencia de cada turbina, para una velocidad de viento determinada, se obtiene a partir de la curva de potencia:

$$pow(w(\theta) \cdot u_s(G, \theta, z)) \quad (18)$$

La potencia global del parque eólico se obtiene de la siguiente manera:

$$P_{tot} = p(\theta) \cdot N \cdot pow(w(\theta) \cdot u_s(G, \theta, z)) \quad (19)$$

Donde:

N: número de turbinas eólicas.

$p(\theta)$: factor que sirve para tener en cuenta el efecto de las estelas de los aerogeneradores.¹⁷

Hasta este instante el método ha sido puramente físico, pero es evidente que hay muchos detalles que no se han considerado. Habitualmente, con el objetivo de mejorar los resultados, el siguiente paso consiste en aplicar dos MOS. Uno para escalar la velocidad del viento, y el otro para aplicar un offset a la generación del parque eólico. Luego se consideraría el efecto de la estela de los aerogeneradores utilizando una herramienta ofimática de simulación de la distribución del parque, para finalmente obtener la previsión de la generación eólica. Si se aplican estos últimos pasos se obtiene un modelo híbrido.

A continuación, con el objetivo de mostrar la estructura de funcionamiento de un modelo de predicción estadístico, se describirá una de las herramientas de previsión de la energía eólica más utilizadas en España, el Sipreólico.

¹⁷ Estela de un aerogenerador: el aerogenerador absorbe energía del viento. Esto provoca que haya una cola de viento turbulento y ralentizado detrás del aerogenerador, que puede afectar al rendimiento de otros aerogeneradores cercanos.



Sipreólico

Es una herramienta de predicción fruto de la colaboración entre la Universidad Carlos III de Madrid y Red Eléctrica de España. Proporciona predicciones a corto plazo, hasta 48 horas de adelanto, de la producción eólica para cada una de las unidades de predicción recogidas.¹⁸ [36] Como datos de entrada utiliza medidas de potencia en tiempo real que consisten en medias cuarto-horarias o veinteminutales de la potencia eólica de los parques. Estos valores se obtienen tras el procesado de las telemidas recibidas cada 12 segundos por el centro de control de Red Eléctrica de España. Actualmente prácticamente la totalidad de la potencia eólica instalada en España proporciona telemidas en tiempo real a Red Eléctrica de España. También utiliza medidas meteorológicas que recibe cuatro veces al día del *Instituto Nacional de Meteorología*, que se generan mediante el programa HIRLAM [37].

Sipreólico combina varios modelos de predicción que se pueden clasificar en paramétricos y no paramétricos.¹⁹ Concretamente utiliza ocho modelos diferentes [37],[38]:

- **Modelo 1:** Persistencia

$$\hat{P}_{t+h} = P_t \quad (20)$$

Donde P_t es la potencia en la hora t .

- **Modelo 2:** Autorregresivo

$$\hat{P}_{t+h} = \alpha_{1t}P_t + \alpha_{2t}P_{t-1} + \dots + \alpha_{qt}P_{t-q+1} \quad (21)$$

Donde los valores de α cambian con el tiempo.

- **Modelo 3:** Autorregresivo y velocidad lineal:

$$\hat{P}_{t+h} = \alpha_{0t} + \alpha_{1t}P_t + \beta_{1t}\hat{v}_{t+h} \quad (22)$$

Donde \hat{v}_{t+h} es la predicción de la velocidad del viento proporcionada por la agencia meteorológica para el período “ $t+h$ ”.

- **Modelo 4:** Autorregresivo con velocidad cuadrática:

$$\hat{P}_{t+h} = \alpha_{0t} + \alpha_{1t}P_t + \beta_{1t}\hat{v}_{t+h} + \beta_{2t}\hat{v}_{t+h}^2 \quad (23)$$

¹⁸ Unidad de producción: este término puede hacer referencia a un parque eólico individual, o a un conjunto de parques eólicos que vierten su potencia en el mismo nodo de la red eléctrica.

¹⁹ **Paramétricos:** estos modelos establecen la relación entre el viento y la potencia mediante modelos basados en polinomios de diferente grado y cuyos coeficientes se estiman de forma recursiva.

No paramétricos: estos modelos no establecen la curva que relaciona la potencia y el viento de antemano, sino que es estimada a partir de la información disponible.



- **Modelo 5:** Autorregresivo con velocidad lineal y dirección del viento:

$$\hat{P}_{t+h} = \alpha_{0t} + \alpha_{1t}P_t + \gamma_{1t}\sin\left(\frac{2\pi\phi_{t+h}}{360}\right) + \gamma_{2t}\cos\left(\frac{2\pi\phi_{t+h}}{360}\right) \quad (24)$$

Donde ϕ_{t+h} es la predicción de la dirección de viento proporcionada por la agencia meteorológica para el periodo "t+h".

- **Modelo 6:** Autorregresivo con velocidad cuadrática y dirección del viento:

$$\begin{aligned} \hat{P}_{t+h} = \alpha_{0t} + \alpha_{1t}P_t + \beta_{1t}\hat{v}_{t+h} + \beta_{2t}\hat{v}_{t+h}^2 + \gamma_{1t}\sin\left(\frac{2\pi\phi_{t+h}}{360}\right) \\ + \gamma_{2t}\cos\left(\frac{2\pi\phi_{t+h}}{360}\right) \end{aligned} \quad (25)$$

- **Modelo 7:** Autorregresivo y predicción no paramétrica basada en la velocidad y dirección del viento:

$$\hat{P}_{t+h} = \alpha_{0t} + \alpha_{1t}P_t + \varphi_{1t}G_{t+h} \quad (26)$$

Donde G_{t+h} es la predicción de potencia obtenida mediante métodos no paramétricos, que utilizan como variables explicativas la velocidad y dirección del viento.

- **Modelo 8:** Predicción no paramétrica basada en velocidad y dirección.

$$\hat{P}_{t+h} = \alpha_{0t} + \varphi_{1t}G_{t+h} \quad (27)$$

Cada vez que se obtienen actualizaciones de la velocidad del viento, o la potencia u otros parámetros que tener en cuenta se actualizan los modelos. La predicción final se obtiene mediante una combinación lineal de las predicciones de los diferentes modelos expuestos previamente. Cabe destacar la velocidad de funcionamiento de Sipleólico, que solo necesita 12 minutos para realizar la predicción para alrededor de 150 parques eólicos.

4.4 MÉTRICAS PARA MEDIR EL ERROR EN LA PREVISIÓN EÓLICA

El error de predicción se define de forma general como la diferencia entre el valor medido y el valor predicho. Por ejemplo, sea P la potencia eólica inyectada a la red para el instante "k", el error de predicción se define como [39]:

$$e(t + k|t) = P(t + k) - \hat{P}(t + k|t) \quad (28)$$

Si se quiere normalizar este error, simplemente se tiene que dividir el resultado de la ecuación (28) por la potencia instalada:

$$e(t + k|t) = \frac{1}{P_{inst}} (P(t + k) - \hat{P}(t + k|t)) \quad (29)$$

Normalizar el error permite obtener unos resultados independientes del tamaño del parque eólico. Cabe destacar que cualquier error de predicción se puede descomponer en una suma de un error sistemático (μ_e) y un error aleatorio (ξ_e):

$$e = \mu_e + \xi_e \quad (30)$$

Los errores sistemáticos se calculan a partir del valor medio de los errores. Se calcula para todo el período de predicción mediante el BIAS. Por otro lado, el valor medio de los errores aleatorios tiene un valor igual a cero [39].

$$BIAS(k) = \hat{\mu}_e(k) = \bar{e}(k) = \frac{1}{N} \cdot \sum_{t=1}^N e(t + k|t) \quad (31)$$

Hay una serie de parámetros que nos sirven para medir la calidad de la predicción [40],[41]:

- **MAE (Mean Absolute Error), o media absoluta del error:** es una medida de la dispersión del error de pronóstico. Consiste en el sumatorio de los errores de pronóstico, dividido entre el número de pronósticos realizados (N).

$$MAE(k) = \frac{1}{N} \cdot \sum_{t=1}^N e(t + k|t) \quad (32)$$

Donde $e(t + k|t)$ viene definido por la ecuación (28).

- **MSE (Mean Squared Error), o media del error cuadrático:** es muy parecido al MAE. La única diferencia es que se realiza el sumatorio de los errores de pronóstico al cuadrado:

$$MSE(k) = \frac{1}{N} \cdot \sum_{t=1}^N (e(t+k|t))^2 \quad (33)$$

- **RMSE (Root Mean Square Error), o raíz cuadrada del valor medio cuadrático del error:**

$$RMSE(k) = \sqrt{MSE} = \sqrt{\frac{1}{N} \cdot \sum_{t=1}^N (e(t+k|t))^2} \quad (34)$$

- **MAPE (Mean Absolute Percentage Error), o media absoluta del porcentaje de error:** se define como la suma del cociente del error de predicción entre el valor objetivo, multiplicado por 100/N, donde N representa la cantidad de muestras que se han realizado:

$$MAPE(k) = \frac{1}{N} \cdot \sum_{t=1}^N \frac{e(t+k|t)}{P(t+k)} \quad (35)$$

Cabe destacar que tanto el error sistemático como el error aleatorio tienen influencia sobre los valores del MAE, MSE, RMSE y MAPE. Una alternativa consiste en calcular el **SDE (Standard Deviation Error)** donde solo contribuyen los errores aleatorios [39]:

$$SDE(k) = \sqrt{\frac{1}{N} \cdot \sum_{t=1}^N ((e(t+k|t)) - \bar{e}(k))^2} \quad (36)$$

También puede resultar interesante calcular el coeficiente de determinación R^2 para cada pronóstico. Representa la habilidad del modelo para explicar las variaciones de los datos. El rango de valores va de 0-1, donde 1 representa un pronóstico perfecto. Existen casos donde puede adquirir valores negativos, por ejemplo, si se trabaja con horizontes temporales grandes:

$$R^2 = \frac{MSE_0(k) - MSE(k)}{MSE_0(k)} \quad (37)$$

Donde $MSE_0(k)$ es la media cuadrática del error para la media global del modelo. No se recomienda usar el R^2 como herramienta principal para la evaluación de sistemas de predicción, y si se utiliza, se tiene que detallar como se ha calculado.

A continuación, se presentan las ventajas y las desventajas de las diferentes métricas definidas previamente [40],[42],[43],[44],[45]:

- **MAE**
 - Tiene como ventaja que todas las diferencias individuales se ponderan por igual en el promedio. Por lo tanto, no es tan sensible a errores grandes o a valores atípicos como el error cuadrático medio.



- La desventaja principal es que puede tomar valores positivos y negativos de manera que formalmente no es diferenciable.
 - Un MAE grande indica que el modelo tiene problemas en ciertas áreas, mientras que un MAE pequeño indica que el modelo realiza buenas predicciones.
- **MSE**
 - Una de las ventajas es que el error nunca es negativo. Una alternativa al MAE sería obtener el valor absoluto del error, pero si se hace de esta manera la función no sería derivable, y disponer de una función derivable posibilita el uso de algoritmos de optimización muy efectivos como por ejemplo el gradiente descendiente.
 - La principal desventaja de esta métrica es que si la predicción no es buena el valor del MSE puede aumentar rápidamente.
 - Otra desventaja es que resulta complicado interpretar si el resultado es bueno o malo, ya que es muy sensible a los outliers.²⁰
 - Puede tomar cualquier valor entre $[0, \infty)$.
 - **RMSE**
 - Una de las desventajas respecto al MSE es que requiere un mayor coste computacional, pero sí que se puede escalar al error original gracias a la raíz. Esto facilita la interpretación.
 - Los outliers le afectan del mismo modo que al MSE.
 - Puede tomar cualquier valor entre $[0, \infty)$.
 - Se puede destacar la utilidad del RMSPE (Root Mean Squared Percent Error), el cual se calcula como el RMSE, pero dividiendo entre el valor de la predicción y multiplicando por 100. Proporciona un porcentaje facilitando la interpretación de la desviación.
 - **MAPE**
 - Es más sencillo de interpretar que el MAE ya que se puede conceptualizar mejor un porcentaje.
 - No le afectan tanto como al MSE o el RMSE la presencia de outliers.
 - Una de las desventajas es que, al dividir entre el valor actual de la predicción y este es cero, no se define el MAPE para ese dato. También puede crecer mucho si el valor actual es muy pequeño.

¿Qué métrica es mejor?

No existe una respuesta clara o única a esta pregunta. Depende de la aplicación y la naturaleza de los datos. Si se quiere dar una mayor importancia a los outliers conviene utilizar el RMSE o el MSE. Si se quiere minimizar el coste computacional es recomendable usar el MSE, y si se quiere mantener una mayor relación con los errores originales el RMSE. Por otro lado, el MAE y el MAPE son más sencillos de interpretar. Si queremos un porcentaje utilizaremos el MAPE, pero tendremos que preocuparnos por los datos que sean ceros, ya que estaríamos cometiendo un error sistemático al incluirlos. A veces resulta más sencillo probarlos a todos y posteriormente decidir cuál es el que mejor se adapta a la aplicación.

²⁰ Outliers: un outlier es una observación anormal y extrema en una muestra estadística o serie temporal de datos que puede afectar potencialmente a la estimación de los parámetros de este.

CAPÍTULO 5. CASO DE ESTUDIO: PARQUE EÓLICO ES MILÁ

En este capítulo se realiza una descripción breve de la instalación eléctrica básica presente en un parque eólico, las características que debe cumplir un emplazamiento óptimo para la instalación de un parque eólico, y la forma más eficiente de distribuir los aerogeneradores sobre el terreno. Posteriormente, se describe el parque eólico Es Milá y el origen de los datos meteorológicos y de generación que se utilizaran en el modelo de predicción desarrollado.

5.1 PARQUES EÓLICOS

Los parques eólicos son instalaciones formadas por un conjunto de aerogeneradores de gran tamaño, que producen energía eléctrica para su inyección a la red de distribución o transporte. Con independencia del tipo de generador eléctrico (síncrono, asíncrono, etc.) la energía eléctrica acostumbra a generarse en trifásica en baja tensión (BT) a 690 V entre fases (Tensión industrial). Esta tensión se eleva a media tensión (MT) entre 20kV y 30kV mediante una celda de transformación BT-MT situada en la base de los aerogeneradores. Posteriormente, pasa a la subestación eléctrica MT/AT donde la tensión se eleva a alta tensión en el intervalo de 132-400kV para su conexión a la red de transporte. En la siguiente imagen se muestra un esquema de la estructura de estas conexiones:

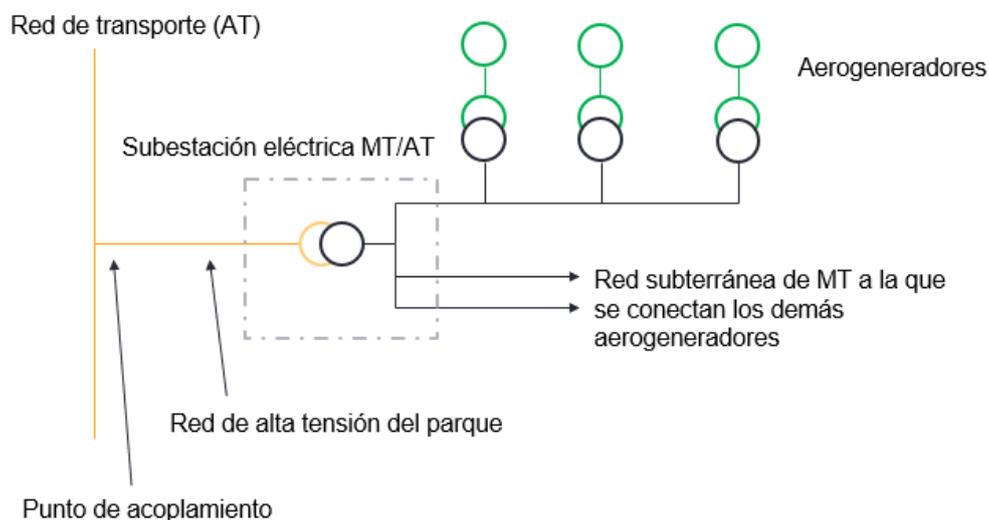


Figura 19: Conexiones eléctricas en un parque eólico. Fuente: Elaboración propia

Aparte de los aerogeneradores, en todos los parques eólicos se puede encontrar los siguientes elementos:

- Generadores eléctricos: están situado en las góndolas de los aerogeneradores, y generan electricidad en BT.



- Instalaciones eléctricas de Baja Tensión de cada aerogenerador, incluyen la aparatada de protección, regulación y control y si procede están equipados con condensadores para la corrección de reactiva.
- Transformadores de baja a media tensión (BT/MT).
- Una red subterránea de media tensión que canaliza la energía producida por los distintos aerogeneradores hasta la subestación transformadora (MT/AT).
- Transformador de MT/AT.
- Equipos de regulación, control, protección y medida de la línea de evacuación de la energía a su interconexión con la red exterior.
- Una línea de evacuación en alta tensión (nodo o punto de conexión con la red).
- Una red de tierras.

Selección de un emplazamiento para parques eólicos

La elección de un emplazamiento para la instalación de un parque eólico se realiza atendiendo a una serie de criterios [6]:

- a) Características eólicas del emplazamiento
- b) Condiciones del terreno.
- c) Proximidad a las redes eléctricas de interconexión.
- d) Impacto medioambiental.

El emplazamiento seleccionado debe aprovechar el recurso eólico disponible de forma óptima, debe transportar la energía eléctrica producida de forma eficiente y ser lo más respetuoso posible con el entorno y el medio ambiente.

a) Características eólicas del emplazamiento

El viento presente en el emplazamiento es el criterio más importante de selección ya que de él dependerá la viabilidad económica del parque. Debe cumplir los siguientes requisitos:

- La velocidad media anual del viento debe ser lo más elevada posible. Las velocidades máximas del viento deben ser admisibles (ausencia de rachas fuertes y frecuentes).
- El viento debe ser poco turbulento.
- Resulta beneficioso si el viento presenta una dirección predominante a lo largo del año.
- La presencia del viento debe ser constante (ausencia de calmas muy duraderas).

En general solo se contemplan emplazamientos cuya velocidad media anual supere los 6 m/s a 10 metros de altura sobre el suelo. Se realiza una caracterización del viento del lugar mediante modelos numéricos de la distribución de velocidades, y se contrastan los resultados con los datos procedentes de estaciones meteorológicas próximas. También se realizan mediciones del viento a diferentes alturas.

b) Condiciones del terreno

El terreno del emplazamiento de los parques eólicos debe cumplir en lo posible las siguientes características:



- El terreno debe presentar la menor rugosidad posible y por lo tanto estar libre de obstáculos ya que estos pueden provocar perturbaciones en el flujo del aire.
- Se deben estudiar los posibles efectos de sombra eólica y estela que pueden ejercer unos aerogeneradores sobre otros, con el objetivo de encontrar el emplazamiento óptimo para cada aerogenerador.
- Debe presentar un nivel de complejidad lo menor posible.
- Debe encontrarse apartado de núcleos habitados para minimizar el impacto ambiental y la contaminación acústica.
- En general, lugares potencialmente atractivos se suelen encontrar en elevaciones suaves, colinas, grandes valles descendentes de cadenas montañosas, o lugares próximos a las costas para aprovechar los efectos de las brisas.

c) Proximidad a las redes eléctricas de interconexión

Resulta interesante que el emplazamiento se encuentre cerca de la red de alta tensión para disminuir los costes de inversión en el sistema de interconexión y se minimicen las pérdidas de transporte. Además, ya que la energía eléctrica todavía no se puede almacenar a gran escala, es un punto a favor del emplazamiento si este se encuentra cerca de un foco de demanda de electricidad.

d) Impacto ambiental

A la hora de seleccionar un emplazamiento se debe considerar el impacto acústico que puede ocasionar sobre poblaciones cercanas, el impacto paisajístico, los efectos sobre la flora y la fauna de la región y también las posibles interferencias con ondas de radio, televisión o telefonía.

Disposición de los aerogeneradores

Es importante colocar los aerogeneradores sobre la superficie del parque eólico de manera que se aproveche de forma óptima el terreno disponible. Por una parte, los aerogeneradores deben separarse entre sí una distancia que impida que las sombras eólicas que generan disminuyan la producción de energía de forma significativa. Por otro lado, conviene que estén próximos entre sí ya que de esta manera se minimizan los costes de la red de media tensión interior del parque, y las pérdidas de energía en dicha red.

En superficies donde se permita la colocación de varias filas de aerogeneradores, estos se suelen colocar en tresbolillo, manteniendo una separación entre ellos de 5 a 10 veces el diámetro del rotor en la dirección del viento dominante, y entre 3 y 5 veces el diámetro del rotor en la dirección perpendicular. Cabe destacar que estos valores solo son orientativos [6].

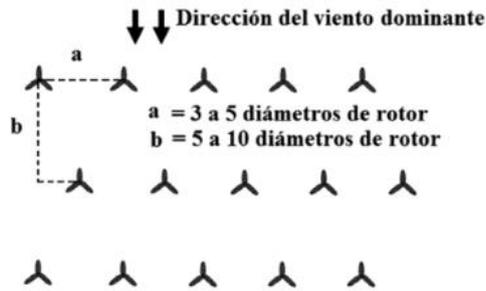


Figura 20: Disposición en tresbolillo de aerogeneradores en un parque eólico. Fuente:[6].

Parques eólicos en España

España es el segundo país europeo con más potencia eólica instalada. En cuanto a número de parques eólicos, consta de 1123 parques eólicos repartidos en 807 municipios diferentes. Las comunidades que suman más parques son [46]:

- Castilla y León: 243 parques
- Galicia: 161 parques
- Andalucía: 153
- Castilla-La Mancha: 139

5.2 PARQUE EÓLICO ES MILÁ

El parque eólico de Es Milá se puso en funcionamiento en el año 2004. Está constituido por cuatro aerogeneradores de 50 m de altura y palas de 27,5 m (59 m de diámetro). La potencia unitaria es de 800 kW y la total del parque de 3,2 MW. Se calculó que la vida útil del parque sería de 20 años. Este parque (es el único parque eólico de las Islas Baleares) produjo el año 2010 el 1,14% de la electricidad consumida en Menorca. En los primeros años de funcionamiento diversos problemas técnicos disminuyeron la capacidad productiva [47].

Emplazamiento

El parque eólico se encuentra ubicado en un paraje cercano al Parque Natural de S'Albufera des Grao, núcleo de la Reserva de la Biosfera de la isla de Menorca, concretamente en la dirección *Carretera Es Grao, Km 2.5, 07701, Mahón*. Las coordenadas de su ubicación son: Latitud (39° 55' 4.5"), Longitud: (4° 15' 33.5"). En la Figura 21, se puede ver la ubicación de Menorca en el archipiélago de las Islas Baleares, y la del parque en la isla de Menorca [48].



Figura 21: Localización geográfica del parque eólico de Es Milá. Fuente:[49]

En la Figura 22, se presenta una fotografía aérea del parque. Se puede ver la distribución en fila de los cuatro aerogeneradores sobre una suave colina, y lo cerca que se encuentra el parque del mar.



Figura 22: Fotografía del parque eólico Es Milá. Fuente:[49]

El parque eólico es gestionado por el '*Consorti per a la Gestió de Residus Sòlids Urbans*' del Consell de Menorca, y se calcula que los cuatro aerogeneradores son capaces de satisfacer la demanda de unas 2.000 viviendas. La compañía GESA Endesa se encarga de distribuir la energía eléctrica generada, la cual se vende al mercado a través de una empresa comercializadora. En la Figura 23 se puede observar un croquis de la distribución de los cuatro aerogeneradores presentes en el parque eólico de Es Milá.



Figura 23: Numeración y ubicación de los aerogeneradores en el parque de Es Milá. Fuente:[49]

Características del emplazamiento

a) Características eólicas

En la Figura 24, se puede observar que la velocidad media del viento a unos 60 m de altura es de unos 7,5 m/s. No son velocidades muy elevadas, pero si nos fijamos en las características de funcionamiento de los aerogeneradores instalados se puede ver que son suficientes para obtener un buen rendimiento. Los estudios iniciales en los que se basó la construcción del parque estimaban que este sería capaz de generar unos 7.040 MWh/año, considerando una disponibilidad del 96 % y unas 2.250 h de viento equivalente al año. Desafortunadamente se ha constatado que la energía generada por el parque se sitúa en torno a los 5.500 MWh/año, un 22 % por debajo de las previsiones iniciales.

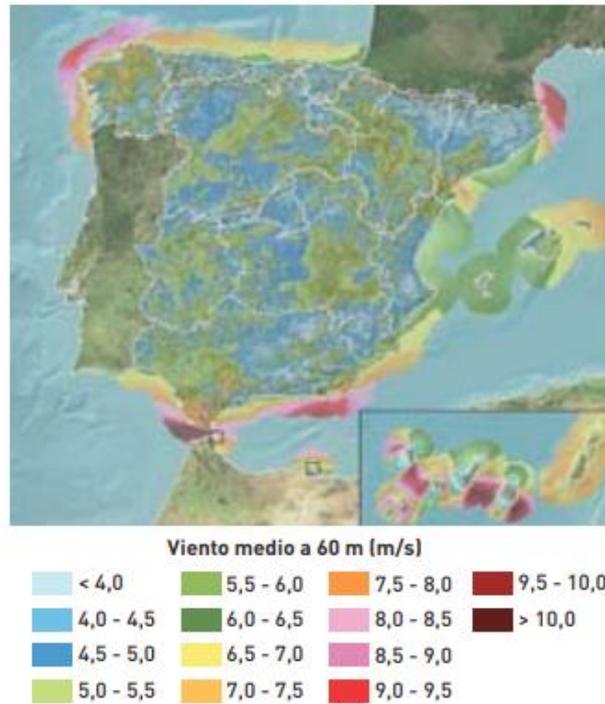


Figura 24: Mapa eólico de España. Fuente:[50]

b) Condiciones del terreno

Observando la Figura 22 se aprecia como los aerogeneradores se encuentran sobre una colina poco pronunciada, y que hay una vegetación débil formada por arbustos y hierba. Por lo tanto, a partir de la **Tabla 2**, el terreno debería presentar un coeficiente de rugosidad con un valor de entre 0,08-0,12, y una longitud de rugosidad de clase 0,5 o clase 1.

Aerogeneradores instalados

Los aerogeneradores instalados fueron fabricados explícitamente para este parque por MADE, actualmente GAMESA. Forman parte de la serie MADE Serie 800, y el nombre del modelo es MADE AE-59. Se trata de un modelo de aerogenerador de eje vertical con una potencia nominal de 800 kW por unidad. Las especificaciones técnicas de mayor relevancia son ²¹:

- Área barrida por el rotor: 2733,97 m²
- Rango de velocidades del rotor: 11,3 a 22,6 rpm
- Altura del buje sobre suelo: 50/60 m.
- Angulo inclinación: 5°
- Velocidad del viento para máximo rendimiento: 11 m/s
- Velocidad de arranque: 3 m/s
- Velocidad de parada de emergencia 25m/s

Para mayor detalle, se puede consultar la ficha técnica en el anexo A (150) del presente trabajo.

²¹ Para más detalle consultar la ficha técnica de los aerogeneradores adjuntada en los anexos.



Las singularidades del sistema eléctrico Balear, formado por varios subsistemas, inciden negativamente en la estabilidad de la red eléctrica de Menorca. Esta inestabilidad de la red, junto a las condiciones climáticas del emplazamiento ha dado lugar a múltiples averías a lo largo de la vida del parque. A esto hay que añadir unos costes de explotación, gestión y mantenimiento sensiblemente superiores a los inicialmente proyectados, basados en los costes medios de los parques presentes en los territorios peninsulares, todo ello ha limitado la disponibilidad del parque eólico.

Problemas y averías del parque

Parámetros de calidad

Durante los primeros 8 años de funcionamiento el aprovechamiento del parque se ha visto afectado por las elevadas exigencias de calidad del suministro que se han establecido para poder inyectar energía eléctrica a la red de distribución de Gesa/Endesa. El gran problema del parque proviene del mal estado de la red de distribución encargada de evacuar la energía (final de línea), que no permitía cumplir con la garantía de estabilidad de frecuencia (± 0.5 Hz) y tensión ($\pm 1\%$) demandada.

Disparo de las protecciones

La inestabilidad de la red de distribución de media tensión a la que vierte la energía el parque eólico ha provocado la desconexión de este en múltiples ocasiones debido al disparo de las protecciones eléctricas. En lugar de una conexión directa a una subestación, el parque vierte la energía a una red aérea de 15kV en antena. Concretamente en el periodo 2004-2014 se acumularon más de 3.283 horas de desconexión de red. Solamente en los últimos años se ha producido el disparo de las protecciones en más de 1.500 ocasiones. Se suelen producir en la madrugada cuando la central térmica de Mahón empieza a aumentar la inyección de energía a la red de distribución.

Humedad ambiental

Como se ha visto en la descripción del emplazamiento del parque, este se encuentra ubicado en una zona con un alto grado de humedad ambiental y componente salina (en forma de aerosol), que combinado con la alta variabilidad de la temperatura diaria propicia la condensación de dicha humedad sobre los delicados componentes de los aerogeneradores. Lo que conlleva que en muchas ocasiones se hayan provocado deflagraciones y arcos voltaicos en los circuitos de potencia de los aerogeneradores, provocando incendios eléctricos y la destrucción de componentes vitales para el funcionamiento del parque.

Averías

A lo largo de la existencia del parque ha habido múltiples averías en los aerogeneradores. Estas averías han provocado que el parque no haya funcionado a plena potencia durante largos períodos de tiempo, debido a la demora en la ejecución de las reparaciones. La principal causa de las demoras en las reparaciones proviene del modelo de aerogenerador instalado, ya que estos están descatalogados y arreglarlos supone un gasto económico muy elevado ya que es complicado encontrar piezas de recambio y se requiere de técnicos especializados no presentes en Menorca. Algunas de las averías de mayor dimensión que han limitado la disponibilidad del parque a lo largo de los últimos años han sido:



- En septiembre del 2016 se confirma una derivación a tierra del alternador correspondiente al aerogenerador 2, y en febrero del 2017 se detecta la misma avería en el aerogenerador 4. Estas averías no fueron arregladas hasta junio del 2018, y tuvieron un coste de 104.193,1 €.
- En 2019 se detectaron problemas en el funcionamiento de la caja multiplicadora del aerogenerador número 2.

Las pérdidas a consecuencia del lucro cesante, el cual se acumula día detrás día cuando un aerogenerador no funciona debido a una avería, se ha estimado en 168 euros/día. Actualmente el parque eólico se encuentra en la etapa final de su vida útil.

Origen de los datos meteorológicos y de generación.

Como se ha mencionado previamente el objetivo de este trabajo es desarrollar una herramienta de predicción basada en un modelo estadístico que incorpora el uso de redes neuronales. Para poder entrenar y testear esta herramienta se requiere de una serie temporal de datos de velocidad de viento, de su dirección y de la generación eléctrica obtenida en cada instante.

- Los datos meteorológicos que se utilizarán proceden de una estación meteorológica denominada J.Mora, ubicada en el parque eólico de Es Milá. Concretamente se encuentra ubicada al suroeste del aerogenerador 4 (ver Figura 23 y Figura 25), en la parte superior de un mástil de 5 m de altura. Cada media hora se transmiten los datos actualizados a través del SCADA hacia la estación de control del parque.²²
- Los datos de generación eléctrica proceden del contador de medida (compañía) que recoge la energía eléctrica vertida a la red, que proporciona un valor para cada hora.

Se dispone de dos tipos de datos:

1. Datos meteorológicos: hojas de cálculo donde se encuentra los valores de variables meteorológicas como la temperatura, presión atmosférica, velocidad y dirección del viento, etc. Los valores proporcionados representan la media de los valores de estas variables para cada media hora.
2. Datos de generación: series temporales anuales de la energía producida por el parque, y la energía consumida. Se dispone de 24 valores por día, es decir uno para cada hora.

Problemas de la estación J.Mora

La estación meteorológica J.Mora se alimenta sobre la una red de distribución en antena, anteriormente descrita, la cual padece de problemas recurrentes debido a la falta de tensión y otros transitorios que provocan el disparo de las protecciones y la desconexión de la red.

²² SCADA (Supervisory Control And Data Acquisition): este concepto hace referencia a una aplicación software diseñada para permitir la comunicación des de un ordenador con dispositivos de campo como microcontroladores, autómatas programables, etc.

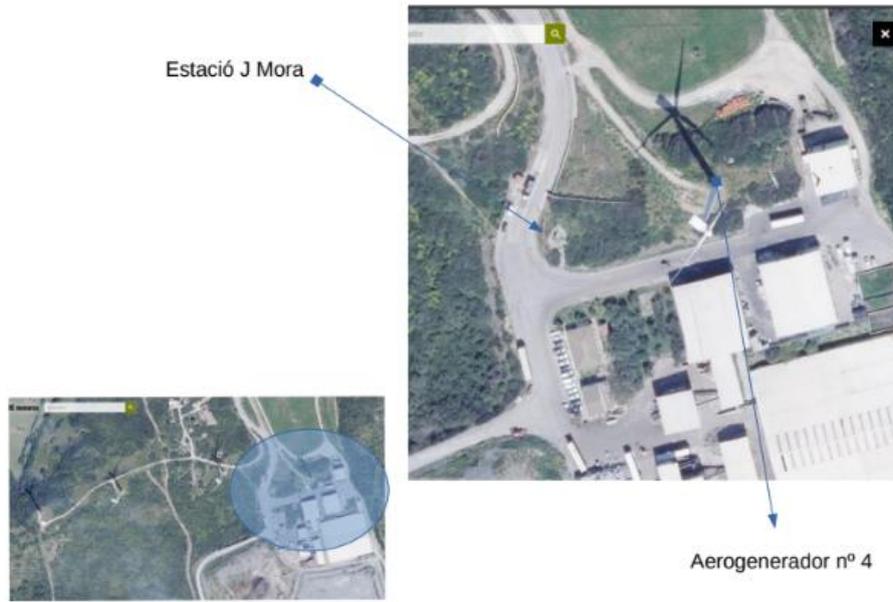


Figura 25: Ubicació de la estació meteorològica J.Mora. Fuente: Consorci per a la Gestió de Residus Sòlids Urbans

CAPÍTULO 6: PREPARACIÓN DE LOS DATOS

En este capítulo se explican los pasos seguidos para preparar los datos de las series temporales para poder utilizarlas a la hora de realizar predicciones con el modelo elaborado. Se explica la importancia del pre-procesamiento de los datos y el tratamiento al que se ha sometido los datos de Es Milá. Los cuales han sido limpiados, se han rellenado los huecos de datos, y se han fusionado en un único archivo. Posteriormente han sido filtrados con el objetivo de eliminar los valores ruidosos (outlayers), y se ha determinado la superficie de generación del parque eólico. Finalmente, se han normalizado y cortado.

6.1 PREPROCESAMIENTO DE DATOS

Cuando se pretende aplicar cualquier metodología de minería de datos o Machine Learning (ML), esta parte de una matriz de variables de entrada a la que aplica una serie de transformaciones con el fin de proporcionar una previsión. El pre-procesamiento de datos es una parte importante de cualquier aplicación de aprendizaje automático ya que disponer de datos de calidad es fundamental para poder obtener buenos resultados. El objetivo de esta fase es buscar y corregir deficiencias que pueden alterar el aprendizaje, como por ejemplo eliminar valores extremos o ruidosos, rellenar huecos de datos, transformar los datos de algunas variables, etc. Hay tres parámetros que definen la calidad de los datos: la precisión, la consistencia y lo completo que están. Desafortunadamente los datos procedentes de la vida real suelen presentar ciertas deficiencias originadas por múltiples causas, como errores de escritura, errores de actualización de los datos, errores en la transmisión de los datos, y una largo etc. Si los datos no son preparados previamente el algoritmo de aprendizaje de la red neuronal no funcionará, y si funciona lo más seguro es que no presente resultados aceptables. Habitualmente el pre-procesamiento incluye alguna, o todas de las siguientes etapas [51]:

- Limpieza de datos.
- Transformación de datos.
- Integración de los datos.
- Relleno de huecos de datos.
- Identificación y eliminación de outliers.
- Normalización de datos.

Limpieza de datos

Esta etapa consiste en filtrar a grosso modo algunos datos erróneos y reducir el tamaño de la base de datos eliminando información irrelevante. Consiste en realizar una primera limpieza de la base de datos que se va a utilizar.

Integración de datos

Esta etapa consiste en la unión de datos procedentes de diferentes bases de datos. No suele ser una tarea sencilla cuando los datos provienen de diferentes fuentes. Por ejemplo, diferentes nombres, o

longitudes diferentes en cuanto a filas y columnas, dificultan la unión de los archivos. Este proceso se debe realizar de forma cuidadosa para evitar redundancias o inconsistencias en la base de datos final. Usualmente se suele disponer de una variable presente en ambos archivos como por ejemplo la fecha o la hora, que facilita la unión ordenada de los diferentes archivos.

Transformación de los datos

En esta etapa se realizan tareas de creación de nuevas variables a partir de los datos existentes, o la modificación de algunas de ellas mediante cambios de formato. También se realiza la normalización, discretización y generalización, pero estos pasos son tratados como pasos independientes. Suelen ser tareas que requieren de la supervisión humana

Rellenos de huecos

Consiste en rellenar aquellas variables que contienen huecos de datos. El origen de un hueco de datos puede deberse a múltiples razones, pero su existencia puede reducir la eficiencia durante el entrenamiento de las redes neuronales o crear complicaciones a la hora de analizar y manejar los datos. Disponer de un valor aproximado al que dio en la realidad es mejor que disponer de un hueco en la serie temporal. Se suelen utilizar métodos de interpolación, medias, medianas, medias móviles,...

Identificación y eliminación de outliers

Esta etapa consiste en identificar valores extremos o imposibles en la base de datos. Estos pueden estar presentes debido a errores humanos, fallos de los sistemas de medida, etc. Una vez identificados se pueden eliminar o corregir aplicando algún método de corrección.

Normalización de los datos

Esta etapa consiste en la transformación de los valores de las variables de interés en otros de equivalentes, rescaldándolos en un rango de valores determinado. El objetivo de esta transformación es facilitar la tarea de predicción ya que se iguala el peso de las variables y se facilita el aprendizaje mediante métodos estadísticos.

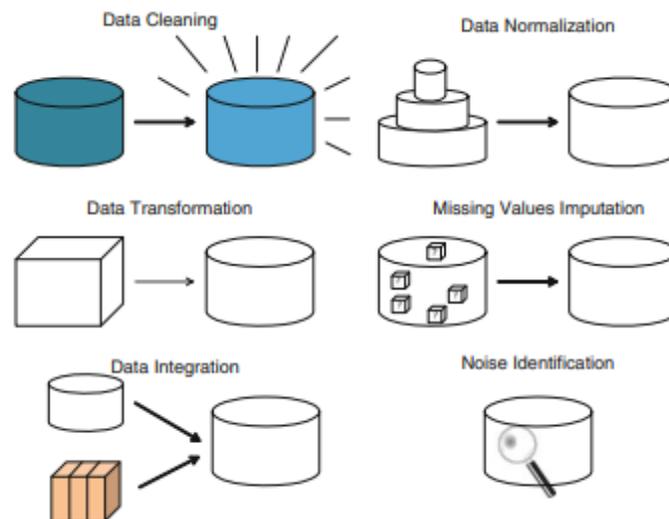


Figura 26: Etapas de pre-procesamiento. Fuente: [51]

6.2 PREPROCESAMIENTO DE LOS DATOS DEL PARQUE EÓLICO DE ES MILÁ

Se dispone de archivos de generación y archivos meteorológicos. El objetivo es unir la información relevante de estos archivos, formando una serie temporal fiable con datos meteorológicos y de generación para el período 2010-2014. Antes de poder unirlos estos deben ser depurados y sometidos a una serie de pretratamientos.

6.2.1 LIMPIEZA DE LOS DATOS

Tal como se ha mencionado en el capítulo 5, se dispone de datos meteorológicos y datos de generación eléctrica, los cuales fueron provistos mediante un conjunto de hojas de cálculo para cada tipo de datos y año, para el período 2008-2014. El objetivo es someter los archivos anuales a las diferentes fases de tratamiento mencionadas, con el objetivo de obtener un archivo final que contenga las series temporales de los datos meteorológicos y de generación para un período de cuatro años. Para limpiar los datos se corregirán de forma manual: los errores de formato, los huecos temporales muy grandes y se omitirá la información irrelevante.

Datos meteorológicos

Se dispone de ficheros de datos meteorológicos de la estación J.Mora para el período 2008-2014. Se trata de hojas de cálculo que recogen la serie temporal anual o mensual de los valores de diferentes variables como la fecha, la temperatura, la humedad, la velocidad y dirección del viento, etc. Se seleccionan los archivos respectivos al período 2010-2014 debido a que estos son los que presentan un menor número de huecos, el mismo formato, es decir recogen las mismas variables, y por lo tanto permiten mantener la relación temporal al tratarse de años consecutivos.

Ficheros de datos:

- 2010-DatosVietao_revisado.ods
- 2011-DatosVietao_revisado.ods
- 2012-DatosVietao_revisado.ods
- 2013-DatosVietao_revisado.ods
- 2014-DatosVietao_revisado.ods

Al abrir los archivos se muestra un mensaje de error de formato debido a que estos fueron escritos mediante Open Office y posteriormente guardados con Excel. Con el objetivo de evitar problemas de formato interno de las variables se ha realizado un copiado especial sin formato de las hojas de cálculo a un libro de Excel nuevo. Observando como en los archivos resultantes han aparecido múltiples huecos datos, para los que no se dispone de información. Se han eliminado todas las filas que forman parte de un hueco temporal consecutivo con una duración superior a 6 horas. Los huecos temporales pequeños han sido tolerados ya que posteriormente serán rellenados.



2806	28/02/2010	03:30	0	0	0	0	0	0	0	0	0	0	0	0
2807	28/02/2010	04:00	0	0	0	0	0	0	0	0	0	0	0	0
2808	28/02/2010	04:30	0	0	0	0	0	0	0	0	0	0	0	0
2809	28/02/2010	05:00	0	0	0	0	0	0	0	0	0	0	0	0
2810	28/02/2010	05:30	0	0	0	0	0	0	0	0	0	0	0	0
2811	28/02/2010	06:00	0	0	0	0	0	0	0	0	0	0	0	0
2812	28/02/2010	06:30	0	0	0	0	0	0	0	0	0	0	0	0
2813	28/02/2010	07:00	0	0	0	0	0	0	0	0	0	0	0	0
2814	28/02/2010	07:30	0	0	0	0	0	0	0	0	0	0	0	0
2815	28/02/2010	08:00	0	0	0	0	0	0	0	0	0	0	0	0
2816	28/02/2010	08:30	0	0	0	0	0	0	0	0	0	0	0	0
2817	28/02/2010	09:00	0	0	0	0	0	0	0	0	0	0	0	0
2818	28/02/2010	09:30	0	0	0	0	0	0	0	0	0	0	0	0
2819	28/02/2010	10:00	0	0	0	0	0	0	0	0	0	0	0	0
2820	28/02/2010	10:30	0	0	0	0	0	0	0	0	0	0	0	0
2821	28/02/2010	11:00	0	0	0	0	0	0	0	0	0	0	0	0
2822	28/02/2010	11:30	0	0	0	0	0	0	0	0	0	0	0	0
2823	28/02/2010	12:00	0	0	0	0	0	0	0	0	0	0	0	0
2824	28/02/2010	12:30	0	0	0	0	0	0	0	0	0	0	0	0
2825	28/02/2010	13:00	0	0	0	0	0	0	0	0	0	0	0	0
2826	28/02/2010	13:30	0	0	0	0	0	0	0	0	0	0	0	0
2827	28/02/2010	14:00	0	0	0	0	0	0	0	0	0	0	0	0

Figura 27: Ejemplo de hueco temporal grande. Fuente: Elaboración propia

En la Figura 27, se puede ver que cuando no se disponía de los datos meteorológicos se han rellenado las filas con ceros. Esto implica que aparte de eliminar todos los huecos temporales grandes se ha tenido que borrar todos los ceros de forma manual. En la Figura 28, se puede ver el aspecto de los huecos temporales que se han tolerado una vez eliminado los ceros y los huecos temporales grandes:

135	03/01/2011	18:00:00	8.5	8.5	8.4	82	5.6	11.3 N	15.29	17.7 NNE	6.7	8.4	6.6	4.7
136	03/01/2011	18:30:00		8.5	8.5			11.3 N	13.68	16.1 NNE				
137	03/01/2011	19:00:00							12.87	NNE				
138	03/01/2011	19:30:00							11.27	NE				
139	03/01/2011	20:00:00							10.46	NNE				
140	03/01/2011	20:30:00							10.46	NNE				
141	03/01/2011	21:00:00							11.27	NE				
142	03/01/2011	21:30:00							8.05	NNE				
143	03/01/2011	22:00:00							7.24	NNE				
144	03/01/2011	22:30:00	8.3	8.4	8.3	82	5.4	1.6 NNE	7.24	4.8 NE	8.3	8.2	8.2	6.3
145	03/01/2011	23:00:00	7.9	8.3	7.9	83	5.2	3.2 NNE	8.05	8 NE	7.9	7.9	7.9	6
146	03/01/2011	23:30:00	8	8.2	7.9	84	5.5	4.8 NNE	6.44	9.7 NE	7.6	8	7.6	5.7
147	04/01/2011	00:00:00	7.7	8.1	7.7	85	5.4	1.6 NNE	4.02	8 NNE	7.7	7.7	7.7	5.8
148	04/01/2011	0:30:00	8.3	8.3	7.7	85	5.9	NNE	4.83	3.2 N	8.3	8.3	8.3	6.4
149	04/01/2011	1:00:00	8.4	8.4	8.3	83	5.7	NNE	4.02	1.6 N	8.4	8.3	8.3	6.4
150	04/01/2011	1:30:00	8.3	8.3	8.3	83	5.6	NNE	4.02	3.2 N	8.3	8.3	8.3	6.4

Figura 28: Hueco temporal menor a 6 horas. Fuente: Elaboración propia.

El siguiente paso de limpieza, ha consistido en eliminar los ceros consecutivos presentes en la columna de la velocidad del viento:

147	04/01/2011	0:00:00	7.7	8.1	7.7	85	5.4	1.6 NNE	4.02	8 NNE	7.7	7.7	7.7	5.8
148	04/01/2011	0:30:00	8.3	8.3	7.7	85	5.9	NNE	4.83	3.2 N	8.3	8.3	8.3	6.4
149	04/01/2011	1:00:00	8.4	8.4	8.3	83	5.7	NNE	4.02	1.6 N	8.4	8.3	8.3	6.4
150	04/01/2011	1:30:00	8.3	8.3	8.3	83	5.6	NNE	4.02	3.2 N	8.3	8.3	8.3	6.4
151	04/01/2011	2:00:00	8.3	8.3	8.2	84	5.8	NNE	4.02	3.2 N	8.3	8.3	8.3	6.4
152	04/01/2011	2:30:00	8.4	8.4	8.3	83	5.7	NNE	4.02	1.6 N	8.4	8.3	8.3	6.4
153	04/01/2011	3:00:00	8.7	8.7	8.4	83	5.9	NNE	4.02	1.6 N	8.7	8.7	8.7	5.8
154	04/01/2011	3:30:00	8.8	8.8	8.7	82	5.9	NNE	4.02	1.6 N	8.8	8.8	8.8	6.1
155	04/01/2011	4:00:00	8.7	8.8	8.7	82	5.8	NNE	4.02	1.6 N	8.7	8.8	8.7	6
156	04/01/2011	4:30:00	8.8	8.9	8.7	83	6	NNE	4.02	1.6 N	8.8	8.9	8.7	6.1
157	04/01/2011	5:00:00	8.5	8.7	8.3	85	6.1	NNE	4.02	1.6 N	8.5	8.7	8.3	6.1
158	04/01/2011	5:30:00	8.8	8.8	8.5	86	6.6	NNE	4.02	1.6 N	8.8	8.8	8.5	6.6
159	04/01/2011	6:00:00	8.4	8.8	8.4	86	6.2	NNE	4.02	1.6 N	8.4	8.8	8.4	6.2
160	04/01/2011	6:30:00	8.8	8.8	8.2	87	6.7	NNE	4.02	1.6 N	8.8	8.8	8.2	6.7
161	04/01/2011	7:00:00	9	9	8.8	88	7.1	NNE	4.02	1.6 N	9	9	8.8	7.1
162	04/01/2011	7:30:00	9.1	9.1	9	87	7	NNE	4.02	1.6 N	9.1	9.1	9	7
163	04/01/2011	8:00:00	9.2	9.2	9.1	87	7.2	NNE	4.02	1.6 N	9.2	9.2	9.1	7.2
164	04/01/2011	8:30:00	9.2	9.3	9.2	85	6.8	NNE	4.02	1.6 N	9.2	9.3	9.2	6.8
165	04/01/2011	9:00:00	8.9	9.2	8.9	84	6.4	NNE	4.02	1.6 N	8.9	9.2	8.9	6.4
166	04/01/2011	9:30:00	9.4	9.4	8.9	81	6.3	NNE	4.02	1.6 N	9.4	9.4	8.9	6.3

Figura 29: Eliminación de las velocidades de viento de 0 m/s. Fuente: Elaboración propia.



Esta primera limpieza ha permitido disponer de cuatro archivos preparados para poder procesarlos de forma automática mediante MATLAB.

Datos de generación

Se dispone de una hoja de cálculo en la que se resume el balance energético del parque eólico durante el período 2006-2016. Este documento contiene las series temporales anuales de la generación energética del parque, de la energía consumida por los aerogeneradores, para cada hora y fecha de dicho periodo; aportando valores de energía horaria [kWh].

Nombre del archivo:

- 2006-2016-Curva horaria anual.ods.

Se repite el mismo proceso que para los datos meteorológicos. Adicionalmente se añade una variable nueva correspondiente al balance de generación de cada instante temporal. Este se calcula restándole a la energía generada la energía consumida por el parque ([AS-AE]). Al final del proceso se dispone de cuatro archivos anuales, aptos para ser procesados con Matlab.

6.2.2 INTEGRACIÓN DE LOS DATOS METEOROLÓGICOS CON LOS DATOS DE GENERACIÓN

El objetivo es unir los datos meteorológicos con los datos de generación de un año en un único archivo. Para unir los archivos y realizar una serie de tareas adicionales se utiliza el script de MATLAB denominado *Preprocesamiento.P1*. Se ha seguido el siguiente proceso:

1. Transformación del formato de las direcciones de viento, de caracteres a su equivalente numérico. Se sigue la nomenclatura expuesta en la Tabla 3:

Tabla 3: Direcciones de viento. Fuente: Elaboración propia.

N	NNE	NE	ENE	E	ESE	SE	SSE	S	SSW	SW	WSW	W	WNW	NW	NNW
0°	22.5°	45°	67.5°	90°	112.5°	135°	157.5°	180°	202.5°	225°	247.5°	270°	292.5°	315°	337.5°

2. Se rellena con "NaN" los huecos de datos.²³ Esto permite utilizar la función *fillmissing(..)* de Matlab para rellenar automáticamente los huecos interpolando. Concretamente se realiza una interpolación cúbica utilizando la función *spline*.²⁴
3. Se fusionan las columnas de fecha y hora. De esta manera se puede obtener el *datevec* de cada una de ellas.²⁵

²³ Un NaN (Not a Number) es una forma de indicarle a MATLAB que el contenido de la celda no es un número real ni complejo. En el caso expuesto sirve para marcar los huecos de datos para que MATLAB sea capaz de identificarlos.

²⁴ El tipo de interpolante a usar depende de las características de los datos que se están ajustando. Se utiliza la interpolación cúbica debido a que es la que se adapta mejor a cambios bruscos en el valor de los datos.

²⁵ *datevec*: convierte un valor temporal en un vector de sus componentes: año, mes, día, hora, minuto, segundo. Permite comparar fechas.

4. Filtrado de las filas por fechas. Se eliminan aquellas que contienen formatos de fecha corruptos, y solo se guarda las fechas con horas en punto.
5. Se dispone de datos meteorológicos para cada media hora, en cambio para la generación solo para cada hora. Se realiza la media de los valores correspondientes a una hora en punto y (hora +30 min) y se asigna el resultado a la hora en punto.

$$Valor(t) = \frac{Valor(t - 30min) + Valor(t)}{2} \quad (38)$$

A continuación, se presenta una serie de figuras en las que se puede ver el aspecto de algunas de las modificaciones expuestas en los puntos anteriores:

- En la Figura 30 se muestra la apariencia de algunas de las columnas y filas de la tabla generada a la hora de leer con MATLAB el archivo de datos meteorológicos del año 2014, antes de aplicar ningún tipo de modificación.

	1 Var1	2 Var2	3 Temp_	4 Temp_1	5 Temp_2	6 Hum_	7 Pto_	8 Vel_	9 Dir_	10 Rec_	11 Vel_1	12 Dir_1
1	'Fecha'	'Hora'	'Ext.'	'Mx.'	'Mjn.'	'Ext.'	'Rocio'	'Viento'	'Viento'	'Viento'	'Mx.'	'Max.'
2	'01/01/2014'	'0:30:00'	'10.9'	'11.2'	'10.9'	'85'	'8.5'	'4.8'	'WSW'	'12.07'	'14.5'	'SSW'
3	'01/01/2014'	'1:00:00'	'10.6'	'10.9'	'10.6'	'86'	'8.4'	'4.8'	'WSW'	'11.27'	'11.3'	'SSW'
4	'01/01/2014'	'1:30:00'	'10.6'	'10.8'	'10.6'	'85'	'8.2'	'8'	'WSW'	'12.07'	'17.7'	'SSW'
5	'01/01/2014'	'2:00:00'	'10.9'	'11'	'10.6'	'82'	'7.9'	'6.4'	'WSW'	'12.87'	'14.5'	'SSW'
6	'01/01/2014'	'2:30:00'	'11.2'	'11.3'	'10.9'	'83'	'8.4'	'4.8'	'WSW'	'13.68'	'11.3'	'SSW'

Figura 30: Lectura del archivo con Matlab sin modificar. Fuente: Elaboración propia.

- En la Figura 31, se puede apreciar el cambio de caracteres a números de las direcciones de viento.

	9 Dir_	15
1	'Viento'	247.5000
2	'WSW'	247.5000
3	'WSW'	247.5000
4	'WSW'	247.5000
5	'WSW'	236.2500
6	'WSW'	225

Figura 31: Cambio de nomenclatura de las direcciones de viento. Fuente: Elaboración propia.

- En la Figura 32, se muestra la presencia de los huecos temporales, y en la Figura 33 se puede apreciar como estos han sido rellenados con "NaN", para su posterior relleno.

348	'08/01/2014'	'5:30:00'	'12.6'	'12.9'	'12.5'	'93'	'11.5'	'8'	'S'		'16.1'	'N'
349	'08/01/2014'	'6:00:00'	'12.6'	'12.6'	'12.4'	'93'	'11.5'	'8'	'S'		'16.1'	
350	'08/01/2014'	'6:30:00'	'12.4'	'12.6'	'12.3'	'93'	'11.3'	'4.8'	'S'		'11.3'	
351	'08/01/2014'	'7:00:00'	'12.3'	'12.4'	'12.3'	'93'	'11.2'	'4.8'	'S'		'9.7'	
352	'08/01/2014'	'7:30:00'	'12.9'	'12.9'	'12.3'	'93'	'11.8'	'6.4'	'SSE'		'14.5'	
353	'08/01/2014'	'8:00:00'	'12.9'	'12.9'	'12.9'	'93'	'11.8'	'4.8'	'SSE'		'8'	

Figura 32: Huecos temporales. Fuente: Elaboración propia.

347	12.6000	12.9000	12.5000	93	11.5000	8	NaN	16.1000	12	12.7000	12.1000	10.4000
348	12.6000	12.6000	12.4000	93	11.5000	8	NaN	16.1000	12.1000	12.7000	12.2000	10.5000
349	12.4000	12.6000	12.3000	93	11.3000	4.8000	NaN	11.3000	12.4000	12.5000	12.5000	10.8000
350	12.3000	12.4000	12.3000	93	11.2000	4.8000	NaN	9.7000	12.3000	12.4000	12.4000	10.8000
351	12.9000	12.9000	12.3000	93	11.8000	6.4000	NaN	14.5000	12.7000	13	12.8000	11.1000
352	12.9000	12.9000	12.9000	93	11.8000	4.8000	NaN	8	12.9000	13.1000	13.1000	11.4000

Figura 33: Huecos temporales rellenados con "NaN". Fuente: Elaboración propia.

- En la columna 1 de la Figura 34 se puede ver el aspecto de las fechas una vez unidas con las horas, y la asignación de la media de los valores a la hora en punto.

	1	2	3	4	5	6	7	8	9	10	11	12
1	'01/01/2014 01:00:00'	10.7500	11.0500	10.7500	85.5000	8.4500	4.8000	11.6700	12.9000	10.6500	10.7500	10.6500
2	'01/01/2014 02:00:00'	10.7500	10.9000	10.6000	83.5000	8.0500	7.2000	12.4700	16.1000	10.0500	10.7000	10
3	'01/01/2014 03:00:00'	11	11.2500	10.8500	84	8.4000	4.8000	13.6800	12.9000	10.9500	11	10.9000
4	'01/01/2014 04:00:00'	10.9000	11.1500	10.7500	87.5000	8.9000	6.4000	12.4700	13.7000	10.4500	10.9500	10.5000
5	'01/01/2014 05:00:00'	10.2000	10.7000	10.1500	88.5000	8.3500	7.2000	12.0700	19.3000	9.4000	10.3000	9.5000
6	'01/01/2014 06:00:00'	10.3500	10.4500	10.0500	88.5000	8.5500	8	11.2650	15.3000	9.4000	10.4500	9.5000

Figura 34: Fechas y horas unidas. Horas en punto. Fuente: Elaboración propia.

- En la Figura 35, se presenta un ejemplo del resultado del método para la obtención de las medias. Por ejemplo, el valor de la fila 1 y la columna 1 (10,75) surge de hacer la media de los valores de la primera columna y las filas 1 y 2: $(10,9 + 10,6) / 2 = 10,75$.

	1	2		1	2
1	10.7500	11.0500	1	10.9000	11.2000
2	10.7500	10.9000	2	10.6000	10.9000
3	11	11.2500	3	10.6000	10.8000
4	10.9000	11.1500	4	10.9000	11
5	10.2000	10.7000	5	11.2000	11.3000
6	10.3500	10.4500	6	10.8000	11.2000

Figura 35: Obtención de la media de los valores. Fuente: Elaboración propia.

Para los datos de generación se realizan las correcciones expuestas en los puntos 2,3 y 4. Posteriormente se procede a unir los dos archivos procesados. Se busca para cada fecha de generación si están disponibles los datos meteorológicos. Si los hay, se carga la fila correspondiente de ambos archivos. De esta manera se obtiene una serie temporal anual filtrada y completa de datos meteorológicos y de generación. Se repite el proceso para los archivos de cada año, que finalmente se fusionan en un único archivo denominado:

- out_2010_2011_2012_2013_2014_Meteo_Gen.csv

6.2.3 IDENTIFICACIÓN Y ELIMINACIÓN DE OUTLIERS

Llegado a este punto del pre-procesamiento se dispone de un archivo con una cantidad considerable de datos, pero que aún contiene errores que deberán de ser filtrados. La presencia de estos errores en los datos se debe a:

- Errores de medida.
- Errores humanos a la hora de recolectar los datos en las hojas de cálculo.
- Mal funcionamiento o falta de calibrado de los sensores de la estación meteorológica. Por ejemplo, el anemómetro puede indicar una velocidad, pero la velocidad del viento ser otra.
- Desconexiones de red o errores de transferencia de la información.
- Relleno de valores inapropiados durante el proceso de interpolación.
- Deslastrado de la red.²⁶

Antes de iniciar la fase de filtrado se analiza los valores de algunas variables de interés, como la velocidad del viento y la generación del parque. Si se observa la Figura 36 y la Figura 37, se puede ver que hay valores de velocidad negativos. Estos se han generado al interpolar de forma cúbica mediante la función de Matlab *spline*, a la hora de rellenar los huecos. La función spline selecciona el valor que mejor se ajusta a los datos, independientemente de si es positivo o negativo.

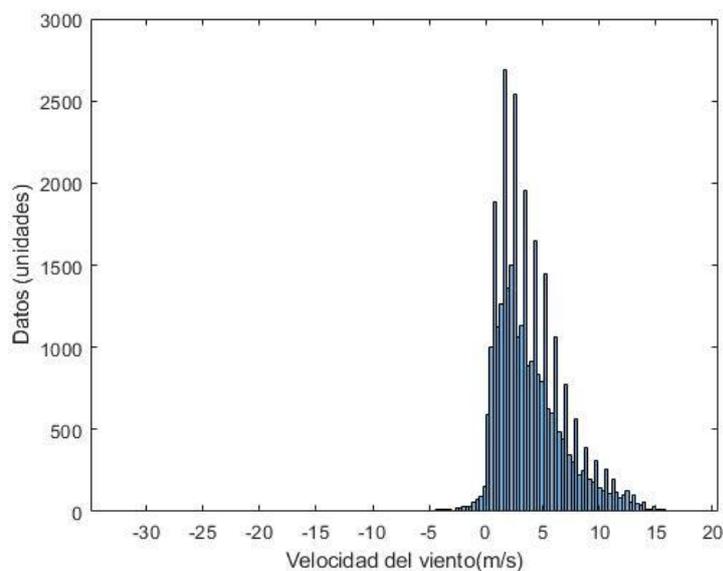


Figura 36: Frecuencia de las diferentes velocidades de viento. Fuente: Elaboración propia

²⁶ Deslastrado de red: cuando REE (Red Eléctrica de España) detecta que hay un exceso de generación eléctrica debido a una falta de demanda, desconecta instalaciones de más de 100 kW con el objetivo de asegurar la estabilidad de la red eléctrica.

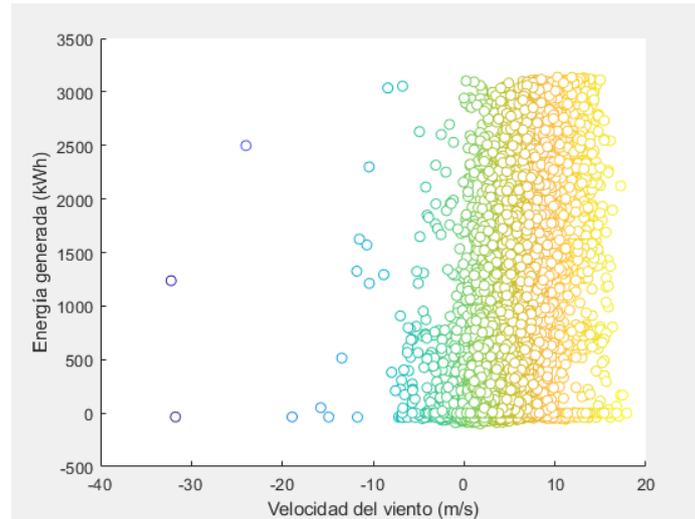


Figura 37: Energía generada/velocidad del viento. Fuente: Elaboración propia

En la Figura 37 se puede ver a simple vista que hay una serie de errores. Por ejemplo, hay datos de generación con velocidades negativas y generaciones elevadas con velocidades bajas. Estos datos deben de ser filtrados.

Fase de filtrado

La fase de filtrado se compone de los siguientes filtros:

- *Filtro 0*: este filtro tiene el objetivo de eliminar las filas que contienen velocidades de viento negativas.
- *Filtro 1*: este filtro tiene el objetivo de eliminar las filas de datos cuando hay generación eléctrica pero la velocidad del viento no llega a la mínima de operación: 3 m/s.
- *Filtro 2*: este filtro tiene el objetivo de eliminar las filas cuando hay generación y la velocidad del viento es superior a la velocidad de parada de seguridad: 25 m/s.
- *Filtro 3*: este filtro tiene el objetivo de eliminar las filas de datos cuando hay velocidades superiores a la velocidad del viento que provoca la generación máxima y hay una generación menor a la máxima (consultar ecuación (39)). En estos casos el sistema de control ha limitado la generación debido a que Red Eléctrica de España ha llevado a cabo un deslastrado de la red, o alguno de los aerogeneradores no ha estado operativo.
- *Filtro 4*: este filtro tiene el objetivo de eliminar las filas de datos cuando se tiene direcciones de viento que no están comprendidas en el rango $[0^\circ, 360^\circ]$.

$$FILTRO\ 3 \rightarrow v_i > \text{rated speed} \ \& \ P_i < (\text{rated capacity} - 100kW) \quad (39)$$

Donde:

v_i : velocidad del viento.

rated speed: velocidad del viento de potencia nominal.

P_i : potencia de salida.

rated capacity: potencia máxima que puede producir la turbina.

Antes de realizar la fase de filtrado se tiene que calcular la velocidad del viento a la altura del buje del aerogenerador. El primer paso que se tiene que realizar para calcular la velocidad a esa altura es determinar el coeficiente de rugosidad, para lo que se utilizará la ecuación de Hellman (11).

Cálculo del coeficiente de rugosidad

El objetivo es calcular el coeficiente de rugosidad real del terreno. Para ello se utilizan datos de velocidad a dos alturas diferentes (a 35 m i 45 m), del año 2007 de Es Milá. Se ha elaborado un script en MATLAB denominado *Coeficiente_de_Friccion.m* que tiene el objetivo de leer estos datos, representar las velocidades una respecto a la otra, y ajustarle una función no lineal que permita obtener el coeficiente de rugosidad. Se ajusta una función matemática a un modelo basado en la ecuación (13), y se obtiene una α prácticamente igual a cero. A priori, esto indicaría que la relación entre las velocidades es 1:1, es decir que son la misma velocidad. Debido a la incoherencia del resultado obtenido se vuelve a revisar el procedimiento seguido y se llega a la conclusión que los datos utilizados contienen múltiples valores erróneos o incoherentes y que por lo tanto no se pueden utilizar para obtener el coeficiente α de forma experimental. En la Figura 38, se puede ver el resultado del ajuste realizado:

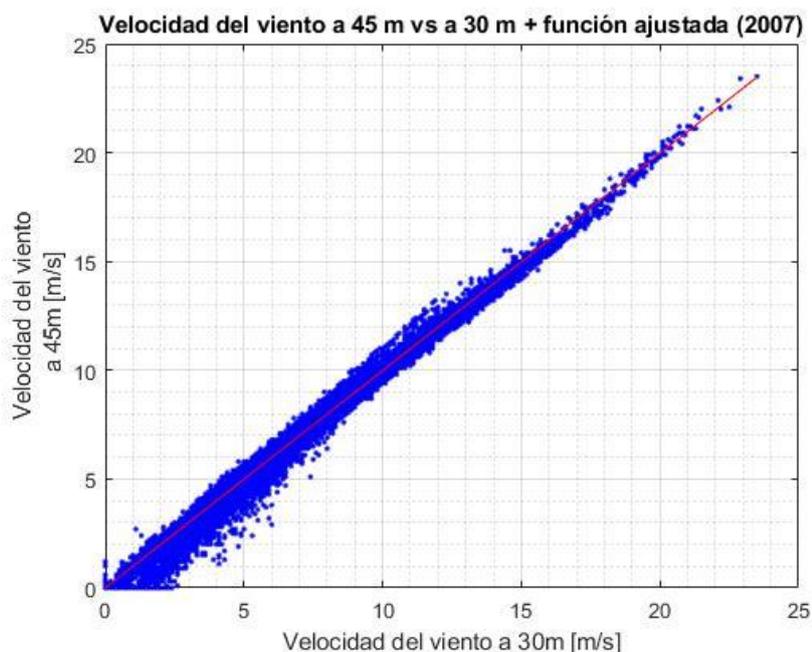


Figura 38: Velocidad del viento a 45/30 m. Fuente: Elaboración propia.

Al no disponer de datos útiles para obtener el coeficiente de fricción real del terreno del parque de Es Milá, se calculará de forma aproximada utilizando la ecuación (14), [52]. Basándose en la orografía del parque eólico de Es Milá se considera que la longitud de rugosidad a utilizar debe ser de clase 1. La clase 1 incluye el intervalo de 0,009-0,04 m. Utilizando $z_0=0,009$ se obtiene un coeficiente de rugosidad de 0,121:

$$\begin{aligned} \alpha &= 0.0910 \cdot \log_{10} 0.009 + 0.016 \cdot (\log_{10} 0.009)^2 + 0.24 \\ &= 0.120798 = 0,121 \end{aligned} \quad (40)$$

Consultando la **Tabla 2**, se aprecia como este valor corresponde a un terreno llano con hierba. El terreno del parque de Es Milá no es una explanada, pero es una colina muy suave, de manera que se considera que el resultado obtenido es válido. Se calcula las velocidades de viento a la altura del buje y se inicia el filtrado de los datos.

Resultados de la fase de filtrado

- **Filtro 0:** en la Figura 39 se ha representado los datos de generación respecto a la velocidad del viento. Comparando esta figura con la Figura 37 se puede ver que se diferencian en que se ha eliminado los datos de generación relacionados a velocidades negativas.

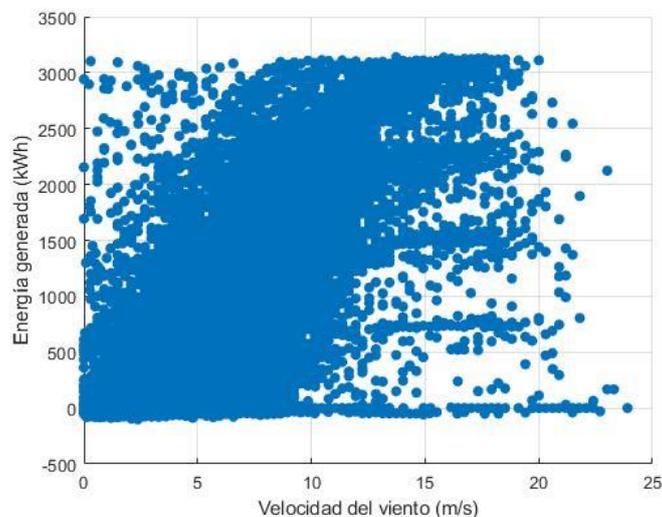


Figura 39: Resultado de la aplicación del filtro 0. Fuente: Elaboración propia.

- **Filtro 1:** en la Figura 40 se ha representado los datos de generación respecto a la velocidad del viento. Comparando esta figura con la Figura 39 se puede ver que se ha eliminado los datos cuando se disponía de velocidades inferiores a la velocidad mínima de funcionamiento de los aerogeneradores (3 m/s).

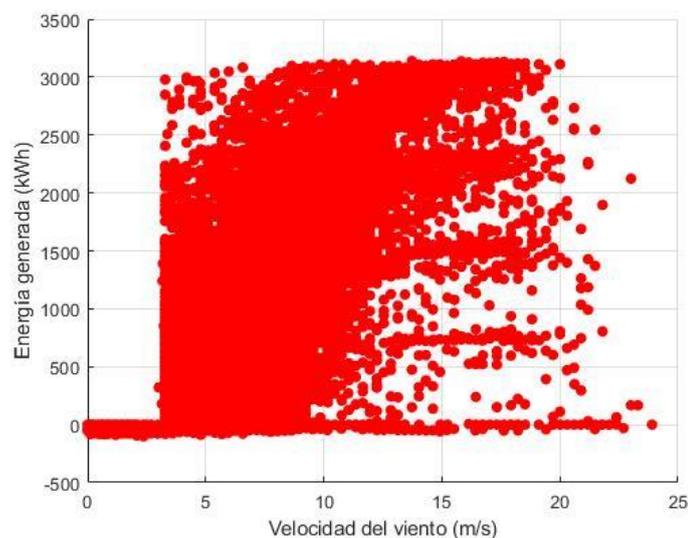


Figura 40: Resultado del filtro 1. Fuente: Elaboración propia.

- **Filtro 2:** comparando la Figura 41 con la Figura 40, se puede apreciar como el filtro 2 no ha detectado datos incoherentes. Esto se debe a que en ningún momento se han dado velocidades de viento superiores a 25 m/s (90 km/h).

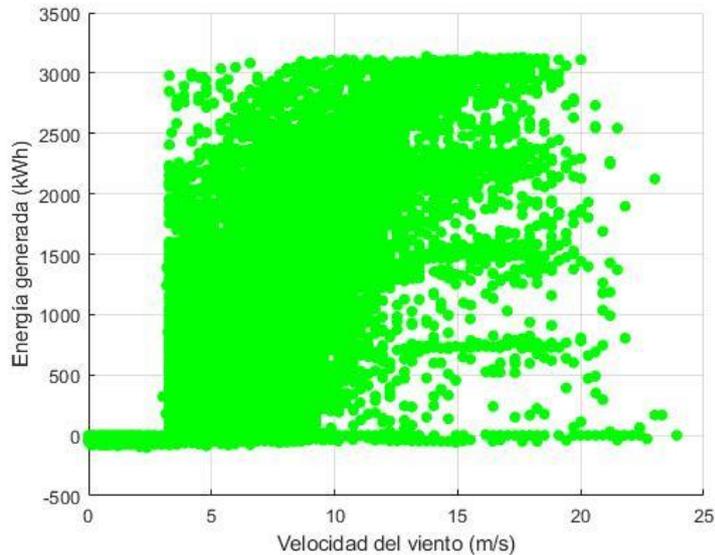


Figura 41: Resultado de aplicar el filtro 2: Fuente: Elaboración propia.

- **Filtro 3:** el filtro 3 pretende eliminar los datos cuando se dispone de una generación eléctrica inferior a la nominal, con velocidades de viento que deberían dar lugar a una generación nominal. En la Figura 42 se puede apreciar la gran cantidad de datos que suele eliminar este filtro:

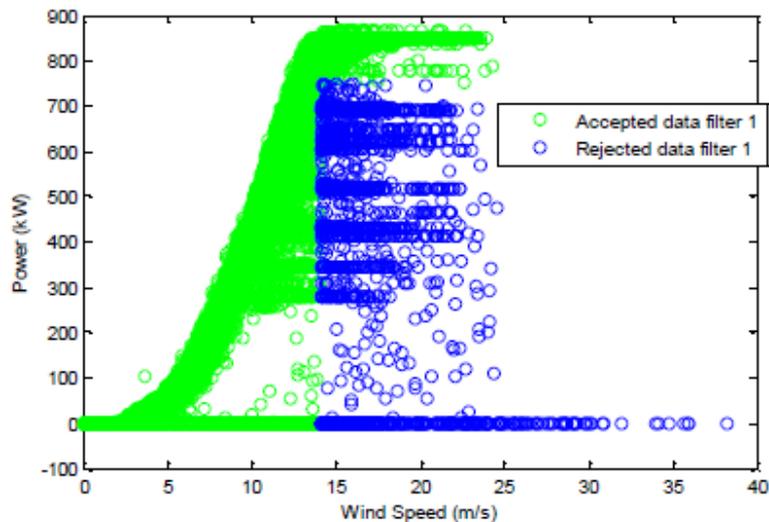


Figura 42: Ejemplo de aplicación del filtro 3. Fuente:[53]

Cabe destacar que en el presente caso de estudio no se ha aplicado el filtro 3, debido a que no se podría aplicar de forma adecuada ya que no se dispone de la información necesaria.



- Se dispone de datos de la energía total producida en cada instante, pero se desconoce qué porcentaje de esta generación corresponde a cada aerogenerador.
- A partir de la velocidad del viento se puede calcular la potencia eólica disponible, posteriormente se puede calcular la potencia mecánica de rotación del eje del rotor, luego la potencia de accionamiento, la eléctrica y finalmente la transformada. No se dispone de información sobre el coeficiente de potencia de los aerogeneradores, del rendimiento de la caja multiplicadora, del rendimiento eléctrico, ni del rendimiento del transformador.
- De los datos de generación se deduce que no siempre han estado funcionando todos los aerogeneradores, pero no se especifica cuantos han estado funcionando en cada instante.
- **Filtro 4:** se elimina una fila de datos que contenía un valor incoherente de dirección de la velocidad del viento.

Conclusiones sobre la fase de filtrado

Se ha constatado la importancia de la fase de filtrado, debido a la gran cantidad de datos incoherentes presentes. El no haber podido aplicar el filtro 3 se ha limitado la eficacia de la fase de filtrado ya que la base de datos seguramente sigue conteniendo una gran cantidad de valores ruidosos, que no reflejan lo sucedido en la realidad. Este hecho tendrá consecuencias negativas sobre la calidad de los resultados de predicción (limitando la bondad de esta).

6.3 SUPERFICIE DE GENERACIÓN DEL PARQUE EÓLICO ES MILÁ

Los fabricantes de turbinas eólicas publican las curvas de potencia para sus turbinas que son utilizadas para planificar y estimar la producción total eólica de un parque que incorpora estos aerogeneradores. Disponer de una función parecida para todo el parque eólico puede resultar muy útil ya que permite predecir de forma aproximada la energía total que se va a producir, dada una determinada velocidad de viento. El objetivo de este apartado es obtener una función que se ajuste a la superficie de generación del parque, y que a partir de una velocidad y una dirección de viento proporcione un valor aproximado de la generación eléctrica del parque.

6.3.1 CURVA DE POTENCIA DE UN AEROGENERADOR

La producción de un aerogenerador puede ser caracterizada mediante una curva de potencia, que es una representación gráfica de la potencia eléctrica producida en función de la velocidad de viento. A partir de esta curva se puede predecir la energía que va a producir un determinado aerogenerador, sin disponer de un conocimiento detallado sobre el funcionamiento interno de este.

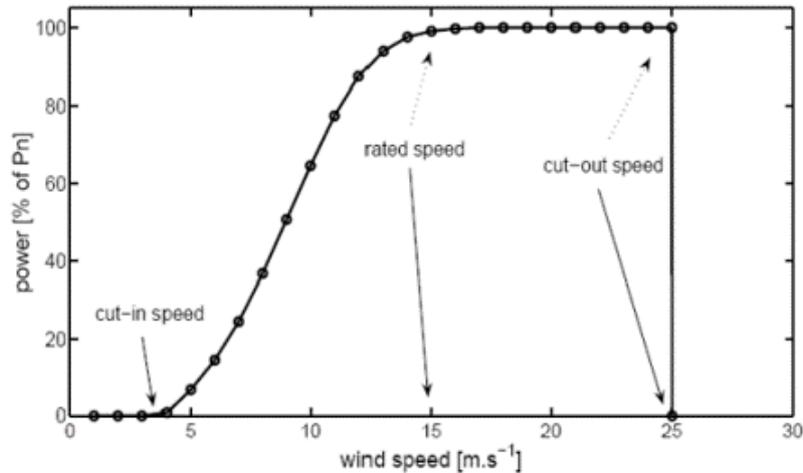


Figura 43: Curva de potencia de un aerogenerador. Fuente:[54]

Observando la figura anterior se puede ver que la curva de potencia presenta tres puntos clave:

- Velocidad de arranque (cut-in speed): indica la velocidad mínima a partir de la cual el aerogenerador empieza a producir energía.
- Velocidad de funcionamiento nominal (rated speed): velocidad a partir de la cual el aerogenerador produce la potencia máxima.
- Velocidad de corte (cut-off speed): velocidad a partir de la cual se para el aerogenerador para evitar averías.

Los parques eólicos suelen diseñarse mayormente con el mismo modelo de aerogeneradores, por lo tanto, la curva de potencia del parque debe seguir una forma similar a la curva de potencia de los aerogeneradores. De todas maneras, hay que tener en cuenta que esta curva también dependerá de otros factores como las características del terreno, la dirección del viento, el efecto de las estelas de los aerogeneradores, posibles diferencias de altura entre los aerogeneradores, etc.

6.3.2 VELOCIDAD Y DIRECCIÓN DEL VIENTO

Hay múltiples factores que influyen en la producción energética de un parque eólico, pero la velocidad del viento y la dirección de donde viene son los factores clave. Se procederá a obtener una función matemática que se ajuste a una curva de generación en 3 dimensiones formada por la velocidad y dirección del viento, y la generación del parque. Previamente se analizarán las velocidades y las direcciones de viento características del parque de Es Milá, con el objetivo de disponer de información que facilite la tarea de ajuste de la superficie de generación.

Rosas del viento según la estación del año

Cuando se habla de la dirección del viento, se hace referencia a las coordenadas desde las que sopla, no hacia las que se dirige. Para conocer esas coordenadas desde las que sopla el viento, se utiliza la rosa de los vientos. La rosa de los vientos es un diagrama en el que se marcan los puntos cardinales

(N, S, E, O, NE, SE, SO, NO...) y las velocidades del viento. El nombre varía según la dirección. En las Islas Baleares, se distingue a los vientos con los siguientes nombres [55]:

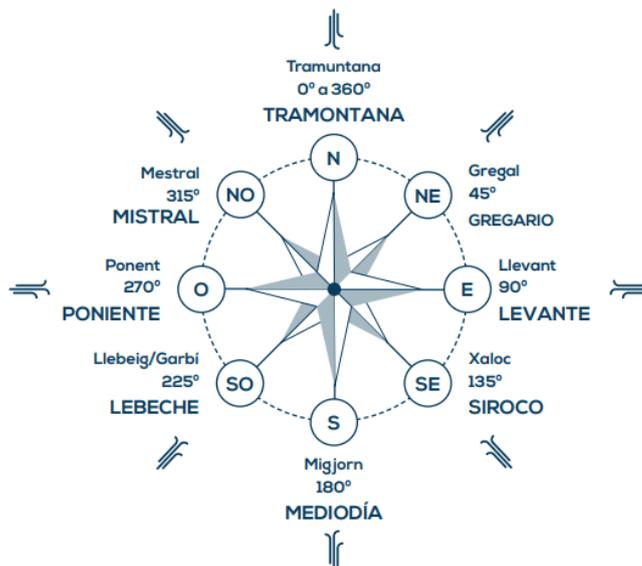


Figura 44: Nombre de las direcciones del viento en las Islas Baleares. Fuente:[55]

La velocidad del viento y la dirección predominante tienden a variar con la estación del año. Para la clasificación de los datos del parque de Es Milá, se ha seguido la nomenclatura de estaciones del hemisferio norte del globo terrestre. Cabe destacar que esta nomenclatura tiende a experimentar pequeñas variaciones a lo largo de los años, pero con el objetivo de facilitar la tarea de clasificación de los datos se ha mantenido constante.

- Primavera: 20/03 - 20/06
- Verano: 21/06 – 22/09
- Otoño: 23/09 – 21/12
- Invierno: 22/12 -19/03

En la Figura 45 se puede ver la representación de las rosas de los vientos para cada estación, y en la Figura 46 la rosa de los vientos obtenida para todo el conjunto de datos. Observando las rosas del viento se puede ver que hay ligeras diferencias en cuanto a las direcciones del viento predominantes a lo largo de las estaciones, pero la tendencia global es la presencia de vientos procedentes del norte, norte-este y del sudoeste. Se confirma la influencia de los vientos locales procedentes de Tramontana y Mistral (consultar apartado 2.3.1 Vientos locales en las Islas Baleares). En la Tabla 4, se puede ver un desglose de las direcciones de viento y velocidades, en función de la época del año.

Tabla 4: Direcciones de viento y velocidades de viento predominantes en el parque eólico de Es Milá

Periodo	Direcciones predominantes	Rango de velocidades predominante
Primavera	SW > N > (NNE, NE)	[3,11] m/s
Verano	NNE > NE	[3,11] m/s
Otoño	NNE > N > SW > NE	[11,25] m/s, [3,11] m/s
Invierno	N > SW > NNE	[11,25] m/s
Global	NNE > N > SW > NE	[11,25] m/s, [3,11] m/s

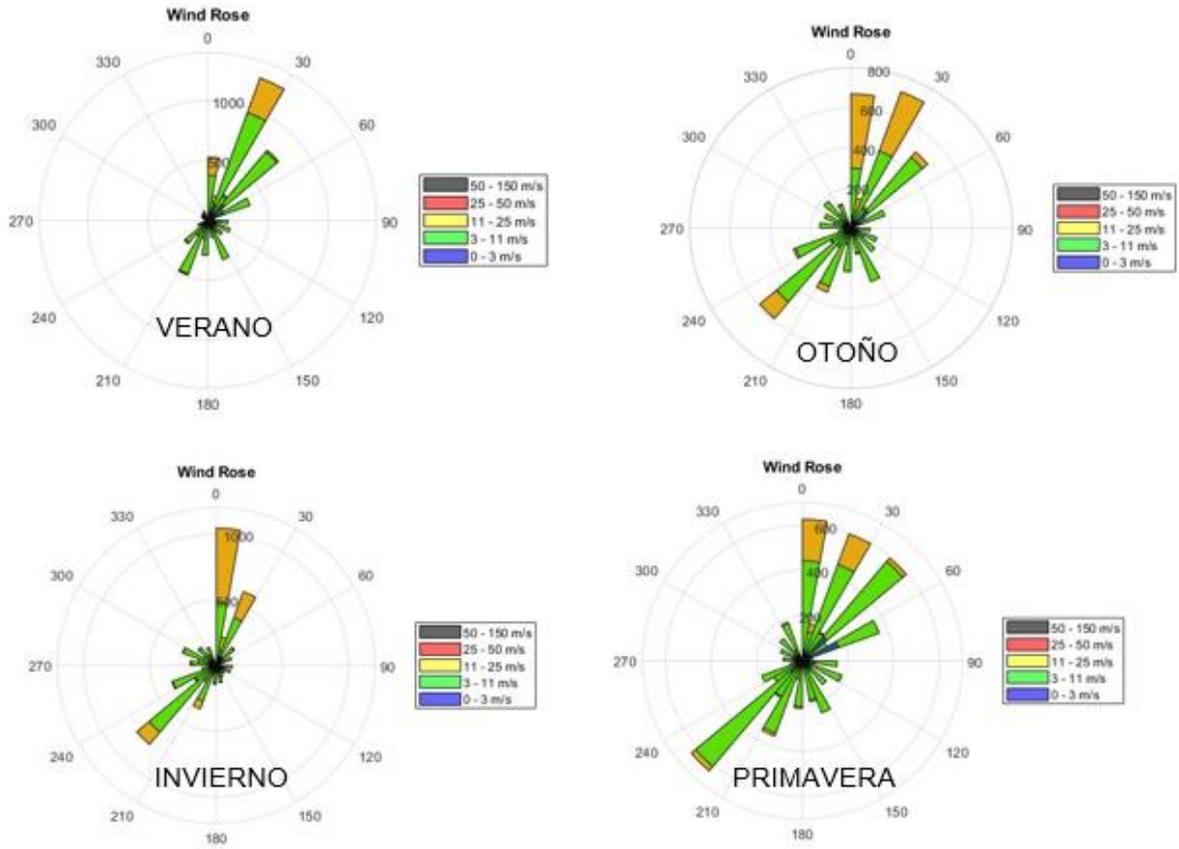


Figura 45: Rosas de los vientos de las diferentes estaciones del año. Fuente: Elaboración propia.

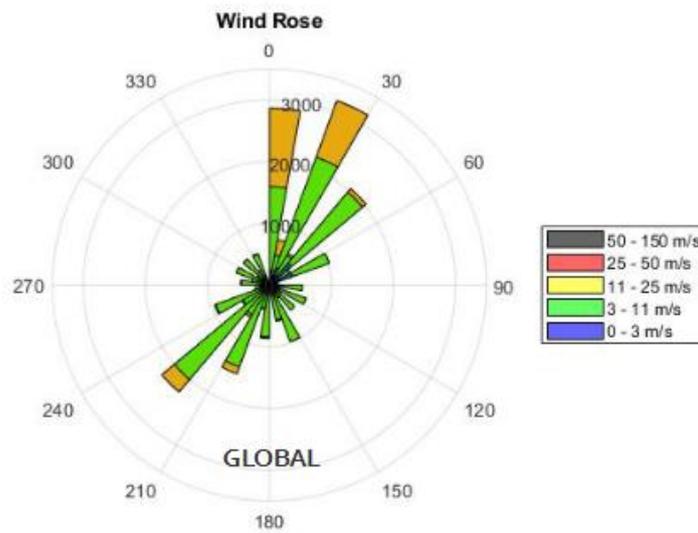


Figura 46: Rosa de los vientos global. Fuente: Elaboración propia.

6.3.3 CURVA DE POTENCIA DEL PARQUE EÓLICO

Concretamente se pretende obtener una función que dependa de la velocidad y la dirección del viento, y proporcione como salida la generación eléctrica del parque:

$$E_{\text{parque}} = f(\theta, v) \quad (41)$$

Donde:

θ es la dirección del viento.

v : velocidad del viento.

Esta función modela el parque como una caja negra ya que las velocidades de viento, la dirección, y la generación eléctrica obtenida en cada instante, toman en cuenta en conjunto todas las variables que influyen sobre el funcionamiento del parque.

Ajuste de la función

cftool (Curve fitting tool): se utiliza esta herramienta de Matlab para realizar dicho ajuste. Se llega a la conclusión que la superficie de generación que mejor se adapta es cuando se dispone de una función polinómica de grado 5, que utiliza el LAR (Least Absolute Residuals) para modelar la forma de la curva. Este método minimiza las diferencias absolutas entre los residuos, y es especialmente útil debido a que le afecta muy poco la presencia de outliers. Una vez finalizado el ajuste, se obtiene la siguiente función:

$$\begin{aligned} f(x,y) = & p_{00} + p_{10} * x + p_{01} * y + p_{20} * x^2 + p_{11} * x * y + p_{02} * y^2 + p_{30} * x^3 + p_{21} * x^2 * y \\ & + p_{12} * x * y^2 + p_{03} * y^3 + p_{40} * x^4 + p_{31} * x^3 * y + p_{22} * x^2 * y^2 \\ & + p_{13} * x * y^3 + p_{04} * y^4 + p_{50} * x^5 + p_{41} * x^4 * y + p_{32} * x^3 * y^2 \\ & + p_{23} * x^2 * y^3 + p_{14} * x * y^4 + p_{05} * y^5 \end{aligned}$$

Donde los coeficientes obtenidos, son:

Tabla 5: Coeficientes de la función de ajuste de la superficie de generación del parque eólico. Fuente: Elaboración propia.

$p_{00} = 102$ (73.34, 130.6)	$p_{21} = 0.01658$ (0.01431, 0.01885)	$p_{04} = -0.5538$ (-0.5836, -0.5239)
$p_{10} = -11.16$ (-12.16, -10.16)	$p_{12} = 0.05771$ (0.01517, 0.1003)	$p_{50} = -1.126e-08$ (-1.164e-08, -1.087e-08)
$p_{01} = 8.996$ (-11.54, 29.53)	$p_{03} = 8.547$ (7.951, 9.142)	$p_{41} = 2.404e-07$ (2.291e-07, 2.517e-07)
$p_{20} = 0.2525$ (0.2374, 0.2677)	$p_{40} = 8.518e-06$ (8.189e-06, 8.847e-06)	$p_{32} = -6.295e-07$ (-8.319e-07, -4.271e-07)
$p_{11} = -0.1601$ (-0.4685, 0.1483)	$p_{31} = -0.0001202$ (-0.0001283, -0.000112)	$p_{23} = -1.435e-05$ (-1.888e-05, -9.831e-06)

$p_{02} = -26.19 (-31.5, -20.87)$	$p_{22} = 0.0001234 (-3.531e-05, 0.000282)$	$p_{14} = 0.0004149 (0.0003701, 0.0004598)$
$p_{30} = -0.002266 (-0.002371, -0.002162)$	$p_{13} = -0.007862 (-0.01024, -0.005481)$	$p_{05} = 0.009857 (0.009309, 0.01041)$

La función obtenida se denomina *fitobject*. Esta proporciona un valor de generación a partir de un dato de velocidad y un dato de dirección del viento. Se añade una columna nueva al archivo de datos filtrado con los valores de generación obtenidos a partir de esta función. En las siguientes imágenes se puede ver el aspecto de la superficie de generación que se ha ajustado a los datos:

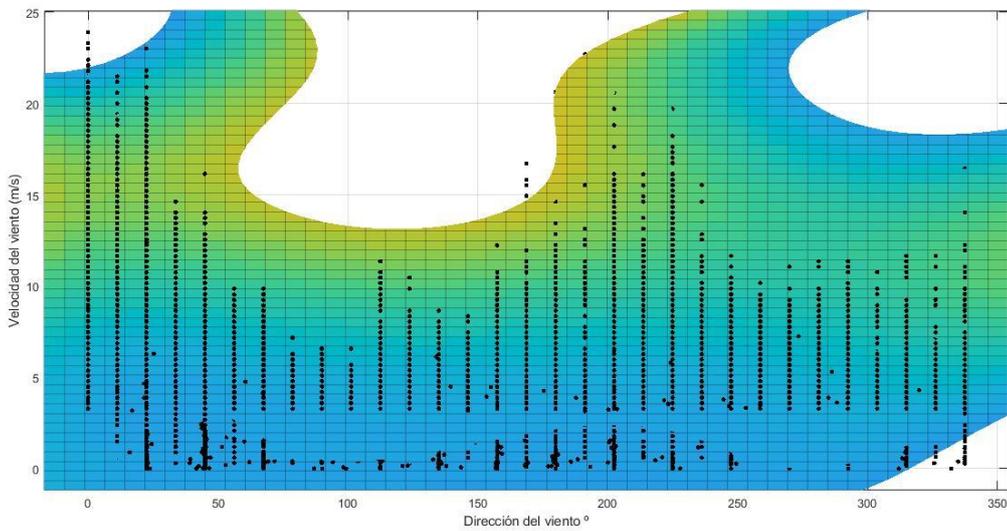


Figura 47: Vista superior de la superficie de generación. Fuente: Elaboración propia.

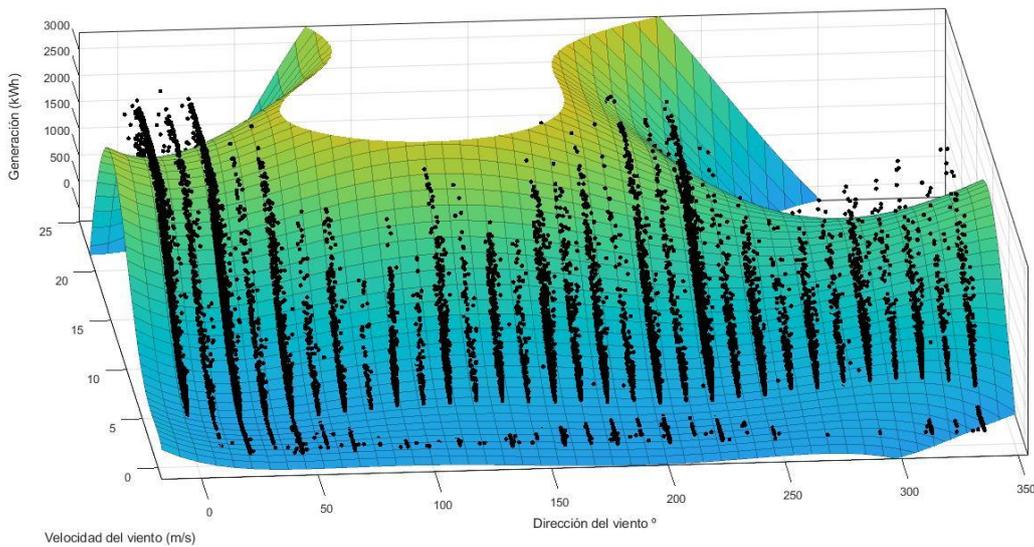


Figura 48: Vista superior inclinada de la superficie de generación. Fuente: Elaboración propia.

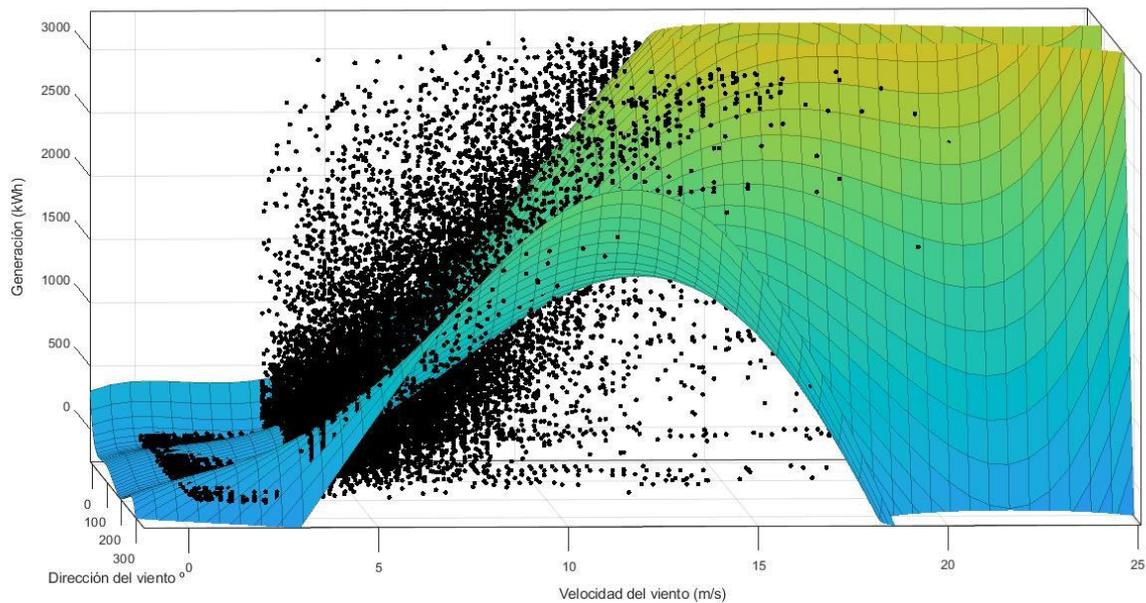


Figura 49: Vista lateral de la superficie de generación. Fuente: Elaboración propia.

Se puede ver que la superficie de generación se adapta a las direcciones de viento predominantes. Entre $[0,50]^\circ$ que corresponde a los vientos procedentes del N, NNE, N hay el primer pico de generación. Luego hay otro pico de generación entre $[200,250]^\circ$ que corresponde a los vientos procedentes del SSW, SW. También se puede ver que la forma que tiene la superficie de generación (vista lateral) es muy parecida a la de la curva de potencia de un aerogenerador. Cabe destacar que si hubiese sido posible realizar una mejor fase de filtrado se habría obtenido una función que se adaptase mejor a la realidad.

Valor de esta función

Disponer de esta función permite determinar de forma aproximada que energía va a producir el parque eólico. Es una variable explicativa que servirá como entrada a la red neuronal y que aporta mucha información sobre el parque eólico sin tener que realizar un modelado complejo.

6.4 NORMALIZACIÓN DE LOS DATOS

Las variables de entrada a un modelo de predicción pueden tener un origen muy diverso, de manera que pueden presentar rangos y características muy distintas. Por ejemplo, una variable de entrada puede disponer de datos comprendidos en un rango muy grande. La red neuronal podría haber sido entrenada con datos procedentes de la zona inferior del rango, y luego recibir valores de entrada procedentes de la zona alta. Es muy probable que no sea capaz de predecir correctamente. Además, la normalización de los datos tiene el objetivo de conseguir que el modelo de predicción de un trato interno similar a todas las variables de entrada. Es una buena opción cuando se dispone de información sobre los máximos y mínimos de los datos, y estos están distribuidos de forma uniforme



a lo largo del rango. La existencia de valores ruidosos puede provocar que la mayoría de los datos sean comprimidos en una parte muy pequeña de la escala [56].

Técnicas de normalización

La técnica de normalización más sencilla consiste en escalar todas las variables entre los valores [0,1]. Esta técnica se utiliza especialmente cuando se trabaja con funciones de activación/transferencia sigmoideas en la red neuronal. Por ejemplo, el valor normalizado de la variable v , para la muestra j de un total de N muestras positivas, se podría calcular de la siguiente manera:

$$v_{jnormalizado} = \frac{v_j}{\max(v_k)_{k=1\dots N}} \quad (42)$$

Donde:

$\max(v_k)$: valor máximo de la variable v para las N muestras.

Normalización mín-max

Este método consiste en escalar los valores de la variable dentro del rango [0,1] o [-1,1] en función de si la variable puede tomar solo valores positivos o también negativos. A continuación, se presenta la ecuación general de este método:

$$v_{jnormalizado} = \frac{v_j - \min(v_k)_{k=1\dots N}}{\max(v_k)_{k=1\dots N} - \min(v_k)_{k=1\dots N}} \quad (43)$$

Donde:

$\max(v_k)$: valor máximo de la variable v para las N muestras.

$\min(v_k)$: valor mínimo de la variable v para las N muestras.

Si se quiere escalar el valor entre un rango arbitrario [a,b], la ecuación anterior se reformula de la siguiente manera [57]:

$$v_{jnormalizado} = a + \frac{(b - a) \cdot (v_j - \min(v_k)_{k=1\dots N})}{\max(v_k)_{k=1\dots N} - \min(v_k)_{k=1\dots N}} \quad (44)$$

Donde:

a: límite inferior del rango.

b: límite superior del rango.

Normalización de los datos del parque eólico de Es Milá

Se va a realizar el escalado de las variables que posteriormente se utilizaran como entradas a la red neuronal. Concretamente, se va a realizar la normalización de los valores de las siguientes variables:



- *Velocidad del viento a la altura del buje.*
- *Dirección del viento.*
- *Generación eléctrica real del parque.*
- *Generación eléctrica obtenida a partir del ajuste de la superficie del parque (ver 6.3.3 Curva de potencia del parque eólico).*

Se utilizará la ecuación (44) para todas las variables. Por lo tanto, el primer paso realizado ha sido obtener los valores mínimos y máximos de cada variable. En la Tabla 6, se puede observar los resultados obtenidos con Matlab:

Tabla 6: Normalización de las variables de entrada a la red neuronal. Fuente: Elaboración propia.

Variable	Mínimo	Máximo	Rango de normalización	Ecuación
<i>Velocidad del viento</i>	0 m/s	23,9 m/s	[0,1]	$Dato = 0 + ((1-0) * (dato - Min_Vel)) / (Max_Vel - Min_Vel);$
<i>Dirección del viento</i>	0°	337, 5°	[0,1]	$Dato = 0 + ((1-0) * (dato - Min_Dir)) / (Max_Dir - Min_Dir);$
<i>Generación eléctrica real.</i>	-101 kWh	3138 kWh	[-1,1]	$Dato = -1 + ((1-(-1)) * (dato - Min_Gen)) / (Max_Gen - Min_Gen);$
<i>Generación eléctrica ajustada</i>	-1153 kWh	1718 kWh	[-1,1]	$Dato = -1 + ((1-(-1)) * (dato - Min_Gen)) / (Max_Gen - Min_Gen);$

6.5 CORTE DE LOS DATOS

Fraccionamiento de los datos

El objetivo de esta fase de pre-procesamiento de los datos es disponer de una base de datos que nos permita entrenar, evaluar y testear la red neuronal. Una de las características más importantes de las redes neuronales es su capacidad de generalización, y por ello es muy importante diseñar una red neuronal que sea capaz de predecir a partir de nuevos datos. Si se utilizan todos los datos para entrenar y validar el modelo generado por una red neuronal no se podrá saber cómo va a responder ante la entrada de nuevos datos. Pueden ocurrir dos tipos de problemas [51]:

- *Ajuste deficiente (underfitting):* sucede cuando el modelo se ajusta mal a los datos y hay errores grandes a la hora de entrenar y testear los datos.
- *Sobre ajuste (overfitting):* sucede cuando se ha ajustado demasiado el modelo a los datos de entrenamiento. Presenta una precisión muy elevada a la hora de tratar con los datos conocidos, y una baja precisión al tratar con datos nuevos. También puede aparecer en caso de disponer datos muy ruidosos ya que el modelo podría ajustarse de forma errónea a una región de los datos compuesta por outliers.

Con el objetivo de evitar estos problemas se cortan los datos. Una parte se utilizará para entrenar el modelo, otra para validarlo, y finalmente otra para testearlo. Cada uno de los trozos de los datos está compuesto por un conjunto de datos que conforman las variables de entrada, y otros datos que

conforman la salida a predecir (targets)[58]. A continuación, se describe cada uno de estos conjuntos de datos:

- Conjunto de entrenamiento: mediante los datos de entrenamiento se ajustan los parámetros de la red neuronal, como por ejemplo los pesos que interconectan las neuronas de diferentes capas. Para ello se utiliza algún método de entrenamiento supervisado, que intenta predecir la salida a partir de los datos de entrada. En función de la salida predicha, y su comparación con los *targets*, se van ajustando los parámetros. Habitualmente representa alrededor del 70 % de los datos
- Conjunto de validación: cuando se cree que se ha obtenido un modelo ajustado, se utilizan estos datos nuevos para realizar una nueva previsión. Los datos de validación permiten detectar insuficiencias en el modelo e ir modificando su estructura (cantidad de neuronas en la capa oculta, cantidad de capas ocultas...). Habitualmente representa un 10 % de los datos.
- Conjunto de testeo: estos datos se utilizan para realizar una evaluación final del modelo. Habitualmente representa un 20% de los datos.

En la Figura 50 se puede ver la dimensión habitual de cada subconjunto, respecto al total de los datos.

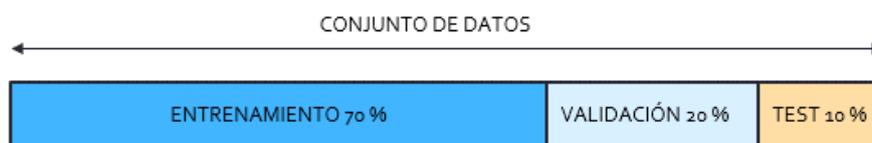


Figura 50: Partición de los datos en subconjuntos. Fuente: Elaboración propia.

En la Figura 51 se puede ver como se suelen utilizar los diferentes conjuntos de datos en el flujo de trabajo de un modelo de predicción.

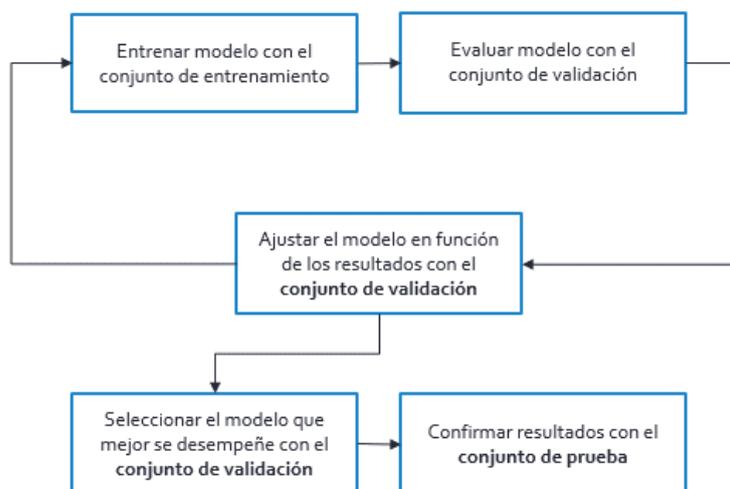


Figura 51: Flujo de trabajo, usando los tres conjuntos de datos. Fuente: Elaboración propia.

6.5.1 FRACCIONAMIENTO DE LOS DATOS DEL PARQUE EÓLICO DE ES MILÁ

El proceso de fragmentación al que se ha sometido los datos de Es Milá difiere del expuesto en el apartado anterior, debido a que las series temporales contienen una cantidad muy grande de huecos temporales, y si se quiere obtener resultados de predicción de cierta calidad se debe regenerar la señal temporal. A continuación, se expone el proceso seguido para obtener archivos aptos para el entrenamiento, el testeo y la validación de la red neuronal:

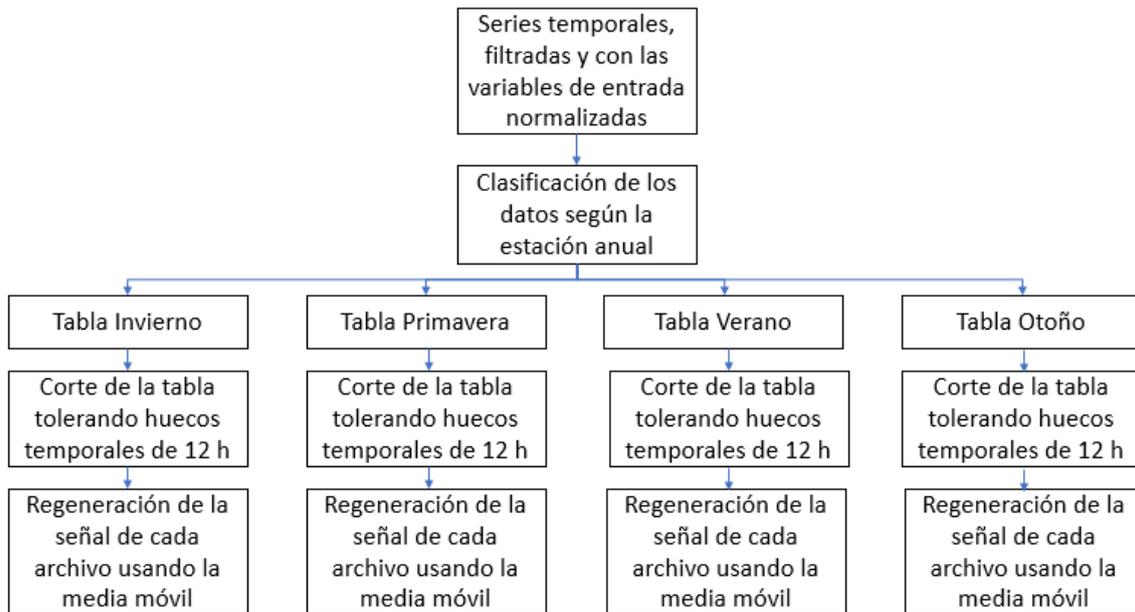


Figura 52: Flujo de trabajo de la fragmentación de los datos de Es Milá. Fuente: Elaboración propia.

Clasificación de los datos según la estación del año

La clasificación de los datos según la estación anual permite mejorar la capacidad de predicción debido a que se incorpora información estacional al modelo de predicción. El objetivo es disponer de cuatro redes neuronales recurrentes NARX que han sido entrenadas y validadas con los datos de una de las estaciones del año. Posteriormente a la hora de realizar las predicciones se consulta a que estación pertenece el periodo que se quiere prever y se selecciona la red neuronal correspondiente para realizar la previsión.

Corte de las tablas de datos estacionales con tolerancias de huecos de 12 horas

Se ha generado un script de MATLAB para cortar las tablas de las estaciones en diferentes archivos cuando detecta huecos temporales mayores a 12 horas consecutivas. De esta forma se obtiene archivos relativamente grandes, y al regenerar la señal como mucho se "regeneran" datos para periodos como máximo de 12 horas consecutivas.

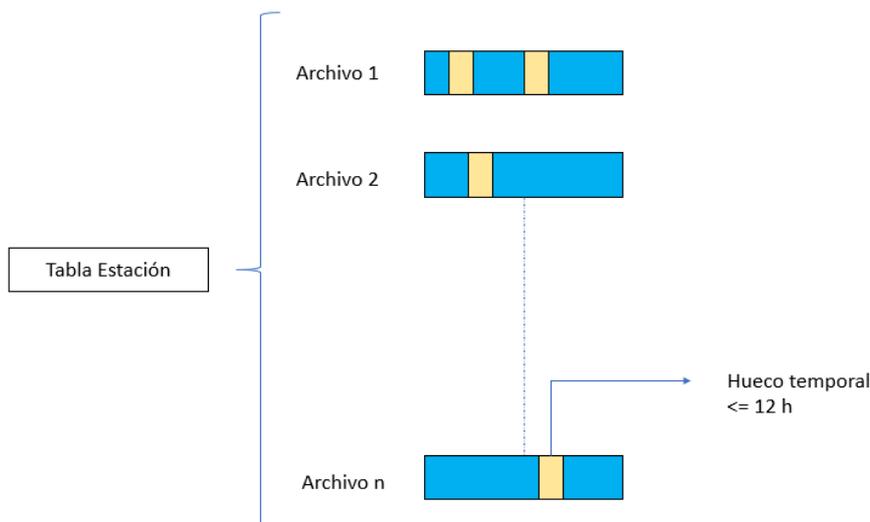


Figura 53: Ejemplo de la fragmentación de una de las tablas estacionales. Fuente: Elaboración propia.

Regeneración de las señales

Una vez finalizada la fragmentación de las tablas de las estaciones se dispone de múltiples archivos que contienen múltiples huecos temporales. Para ello se ha generado un script en MATLAB encargado de leer cada uno de estos archivos, identificar el tamaño de los huecos y reconstruir las señales temporales. El programa primeramente se encarga de identificar la cantidad de filas faltantes, para luego reconstruir la matriz de fechas y rellenar los huecos de valores de las demás variables con "NaN". Posteriormente substituye los NaN por valores numéricos que obtiene realizando una media móvil de los últimos 16 valores disponibles. A continuación, se presenta un ejemplo del aspecto de la serie temporal de uno de los archivos antes de ser regenerada y posteriormente a ser reconstruida:

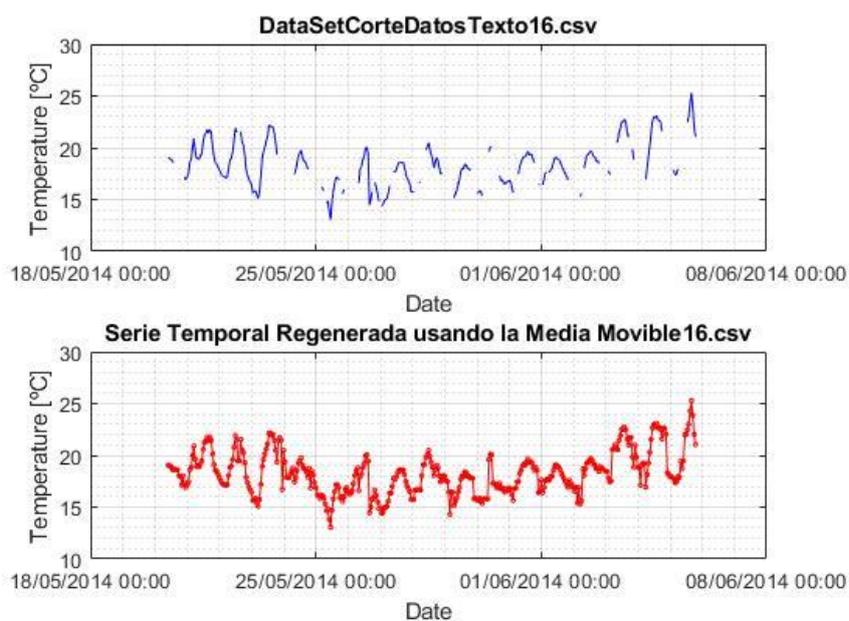


Figura 54: Ejemplo de la regeneración de la señal mediante la media móvil. Fuente: Elaboración propia

Series temporales regeneradas

Se han seleccionado los archivos con las señales regeneradas de mayor tamaño de cada estación. De las cuales, una parte se destinará a entrenar las redes neuronales, una parte a validar el entrenamiento y optimizar la configuración de la red, y finalmente la última parte se destinará a realizar las pruebas de predicción. Se utilizará aproximadamente el 70 % de los archivos para entrenar, un 15% para la validación y un 15 % para las pruebas. Adicionalmente, con el objetivo de mejorar la capacidad de predicción, se recorta los archivos regenerados de manera que todos empiezan a la misma hora, concretamente a las 00:00:00. De esta manera se asegura que todos los archivos son procesados de la misma manera ya que disponen de un punto de inicio temporal común



Figura 55: Clasificación de las series temporales regeneradas según el uso. Fuente: Elaboración propia

CAPÍTULO 7: REDES NEURONALES ARTIFICIALES

En el presente capítulo se realiza una introducción a las redes neuronales. Se aborda su funcionamiento, las diferentes arquitecturas y fortalezas. A continuación, se realiza un pequeño desglose de la evolución histórica por la que han pasado las redes neuronales, especificando los avances más relevantes que se han producido a lo largo de las últimas décadas. Posteriormente, se presentan las principales características de las redes neuronales unidireccionales y las redes neuronales recurrentes, haciendo especial énfasis en las NARX (Nonlinear Autoregressive Exogeneous) debido a que se utilizará este tipo de red a la hora de hacer las predicciones de generación eólica mediante el Neural Network Toolbox de Matlab. Finalmente, se explica mediante un ejemplo el funcionamiento del algoritmo de aprendizaje basado en la retropropagación del error.

7.1 INTRODUCCIÓN A LAS REDES NEURONALES ARTIFICIALES

Las redes neuronales artificiales (Artificial Neural Networks, ANN) se inspiran en los sistemas neuronales biológicos. El cerebro (animal) está compuesto por una multitud de neuronas que reciben señales en forma de pulsos eléctricos de otras neuronas. El conjunto de pulsos eléctricos que recibe una neurona genera una respuesta en esta que es transmitida a todas las neuronas interconectadas a ella. Las redes neuronales artificiales pretenden emular este comportamiento. También se componen por elementos simples de procesamiento denominados neuronas, que se conectan entre sí mediante conexiones que tienen un valor numérico asociado llamado peso. Cada una de estas neuronas es capaz de recibir información de otras neuronas o entradas externas y proporcionar una respuesta. Esta respuesta es el resultado de una función de activación (generalmente no lineal) cuya entrada es una combinación lineal de los valores de entrada, concretamente una suma ponderada [59]:

$$y_i = f_i \left(\sum_{j=1}^n w_{ij} x_j - \theta_i \right) \quad (45)$$

Donde:

y_i : valor de salida de la neurona i .

x_j : valor de la entrada j .

w_{ij} : peso de la conexión entre las neuronas i y j .

θ_i : umbral o sesgo (bias) de la neurona.

f_i : función de activación.

Las funciones de activación pueden ser lineales o no lineales, pero en la mayoría de las aplicaciones son de tipo no lineal (sigmoidales, gaussianas, lineales a tramos, tangentes hiperbólicas, función de Heaviside, etc.) ya que esta no linealidad convierte a las redes neuronales en “aproximadores universales” y allí radica gran parte de sus capacidades. A continuación, se describe la neurona artificial y el origen de la ecuación (45) de una forma más detallada.

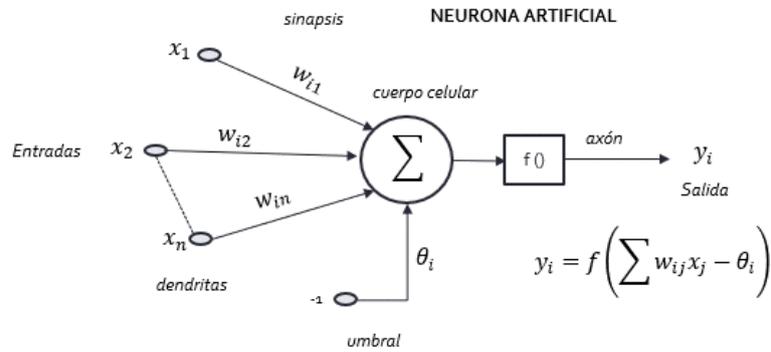


Figura 56: Neurona artificial. Fuente: Elaboración propia

Tal como se presenta en la Figura 56, una neurona artificial consta de [60]:

- Un nodo que recibe un conjunto de entradas x_j y unos pesos sinápticos w_{ij} , con $j=1, \dots, n$.
- Una regla de propagación h_1 , definida a partir del conjunto de entradas y pesos:

$$h_1(x_1, \dots, x_n, w_{i1}, \dots, w_{in}) \rightarrow \text{Lineal} = \sum_{j=1}^n w_{ij}x_j \quad (46)$$

- Un umbral θ_i , que representa un parámetro de off-set que se acostumbra a restar al potencial post-sináptico:

$$h_1(x_1, \dots, x_n, w_{i1}, \dots, w_{in}) \rightarrow \text{Lineal} = \sum_{j=1}^n w_{ij}x_j - \theta_i \quad (47)$$

- Una función de activación, (f) la cual representa la salida de la neurona. Si se le denomina y_i se obtiene el resultado presentado en la ecuación (46):

$$y_i = f_i(h_i) = f_i\left(\sum_{j=1}^n w_{ij}x_j - \theta_i\right) \quad (48)$$



Perceptron

En este caso la función de activación es una función escalonada:

$$y_i = \begin{cases} 1 & \text{si } \sum_{j=1}^n w_{ij}x_j \geq \theta_i \\ 0 & \text{si } \sum_{j=1}^n w_{ij}x_j < \theta_i \end{cases} \quad (49)$$

Neurona sigmoideal

Las funciones sigmoideas más utilizadas son la función logística o Log-sigmoidea y la función tangente hiperbólica sigmoidea [60]:

a) Log-sigmoidea:

$$\text{Forma general} \rightarrow g(a) = \frac{1}{1 + e^{-a}}$$

$$\text{En el ejemplo} \rightarrow y_i = \frac{1}{1 + e^{-(\sum_{j=1}^n w_{ij}x_j - \theta_i)}} \quad (50)$$

b) Tangente hiperbólica sigmoidea:

$$\text{Forma general} \rightarrow g(a) = \frac{e^{(a)} - e^{-(a)}}{e^{(a)} + e^{-(a)}}$$

$$\text{En el ejemplo} \rightarrow y_i = \frac{e^{(\sum_{j=1}^n w_{ij}x_j - \theta_i)} - e^{-(\sum_{j=1}^n w_{ij}x_j - \theta_i)}}{e^{(\sum_{j=1}^n w_{ij}x_j - \theta_i)} + e^{-(\sum_{j=1}^n w_{ij}x_j - \theta_i)}} \quad (51)$$

Ambas funciones son continuas y diferenciables, un requisito indispensable en algunas reglas de aprendizaje como por ejemplo la regla de retro-propagación (backpropagation). Existen muchas otras funciones de activación, como el escalón simétrico, la función lineal a tramos, la función gaussiana, etc. La elección de la función de activación depende de la naturaleza de los datos y de la regla de aprendizaje que se quiere utilizar.

Arquitecturas de interconexión de neuronas artificiales

La estructura de una red neuronal depende de la colocación de las neuronas y en la forma en que estas están interconectadas. Las neuronas se agrupan en capas, y en función del número de capas se pueden clasificar en redes monocapa o multicapa.

- *Monocapa*: están formadas por una capa de entrada y una capa de salida.
- *Multicapa*: estas redes incorporan una o varias capas intermedias denominadas capas ocultas.

Una capa de entrada, también denominada sensorial, está compuesta por neuronas que reciben datos o señales procedentes del entorno. Una capa de salida se compone de neuronas que proporcionan la respuesta de la red neuronal. Una capa oculta no tiene conexión directa con el entorno, su función es la de proporcionar grados de libertad a la red neuronal gracias a las cuales es capaz de representar mejor determinadas características del entorno que trata de modelar. En función de la forma de interconexión de las capas, las redes neuronales se pueden dividir en redes unidireccionales (feed-forward) y redes recurrentes (feed-back):

- *Unidireccionales*: la propagación de la señal se produce desde la entrada hacia la salida, no hay ningún tipo de realimentación.
- *Recurrentes*: estas redes se caracterizan por disponer de lazos de realimentación que pueden darse sobre una misma neurona, entre neuronas de una misma capa, o neuronas de capas distintas.

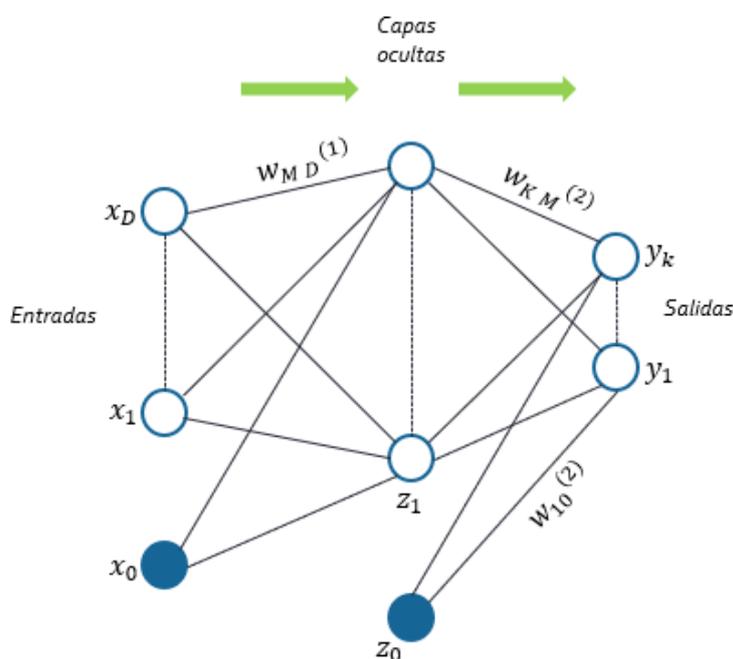


Figura 57: Ejemplo del flujo de información en una red neuronal unidireccional multicapa. Fuente: Elaboración propia.

Capacidades de las redes neuronales

La estructura de las redes neuronales se asemeja a una estructura de procesamiento distribuida en paralelo, que es capaz de resolver problemas no lineales de forma óptima, y con un coste computacional relativamente pequeño. Una de las características más importantes es que son capaces de “aprender”. Este aprendizaje consiste en el ajuste de los pesos de las conexiones entre neuronas, con el objetivo de determinar las relaciones existentes entre los pares de vectores de entrada y salida. Las redes neuronales pueden aprender de diferentes maneras [61]:

- *Aprendizaje supervisado*: en este tipo de aprendizaje la red neuronal dispone del valor de los vectores de entrada y del vector de valores de la salida objetivo. Durante el entrenamiento la red neuronal va ajustando los pesos de manera que se minimice la diferencia entre el valor de salida proporcionado y el valor de la salida objetivo.

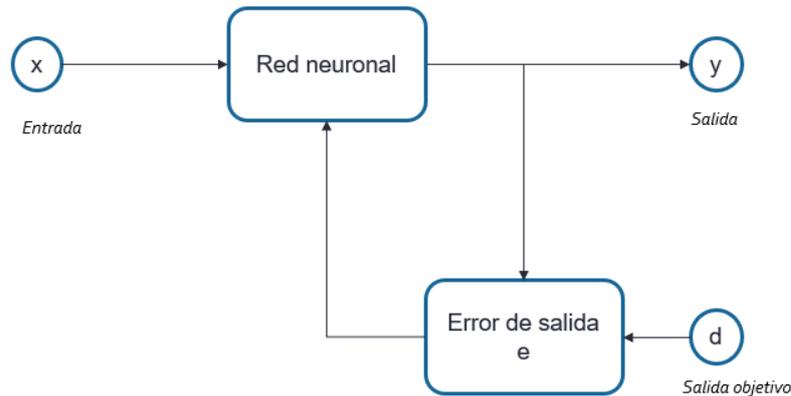


Figura 58: Aprendizaje supervisado. Fuente: Elaboración propia

Las principales estrategias de aprendizaje del aprendizaje supervisado son la corrección del error y el aprendizaje estocástico.

- a) Corrección del error: se ajustan los pesos de las conexiones de la red en función de la diferencia entre los valores deseados y los obtenidos en la salida. Algunos ejemplos son la regla de aprendizaje del Perceptron, la regla delta o error cuadrático mínimo (LMS, usado en modelos neuronales como el Adaline y el Madaline), y la regla delta generalizada o retropropagación del error.
- b) Aprendizaje estocástico: se realizan ajustes aleatorios de los pesos de conexión, y se evalúa el efecto a partir de los valores objetivos mediante una distribución de probabilidad.
- Aprendizaje no supervisado: en este tipo de aprendizaje la red neuronal no recibe un vector de valores de salida objetivo. Intenta encontrar un patrón que le permita clasificar los valores de entrada.

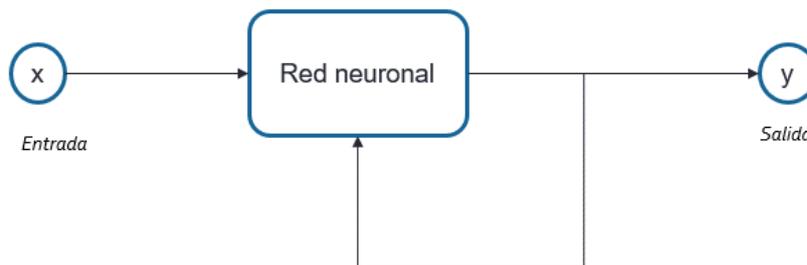


Figura 59: Aprendizaje no supervisado. Fuente: Elaboración propia

La principal estrategia de aprendizaje no supervisado es el aprendizaje hebbiano, que consiste en el ajuste de los pesos de conexión en función de la correlación de los valores de las neuronas conectadas. Es un aprendizaje competitivo, las neuronas “luchan” para activarse.

- Aprendizaje reforzado: en este tipo de aprendizaje la red neuronal no recibe un vector de salida objetivo, pero sí que recibe información sobre si la salida proporcionada es la esperada o no.

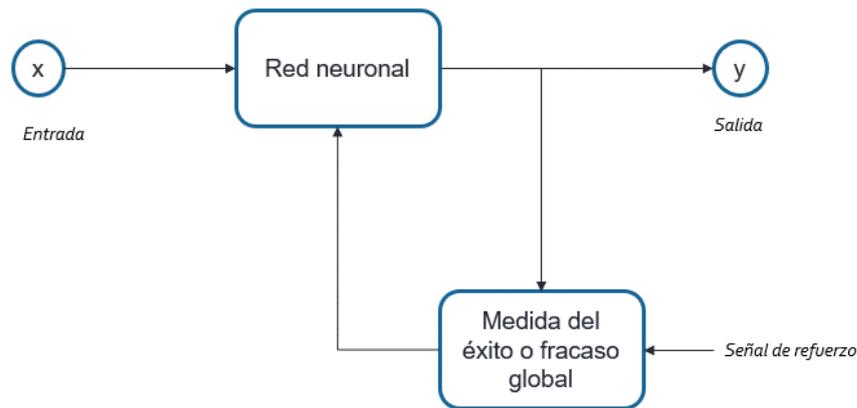


Figura 6o: Aprendizaje reforzado. Fuente: Elaboración propia.

A parte de la capacidad de aprendizaje, las redes neuronales presentan una elevada tolerancia a fallos. En la computación tradicional la pérdida de un fragmento pequeño de información puede acarrear comúnmente la inutilización del sistema. Se entiende por ello que las redes neuronales pueden reconocer patrones de información con ruido, distorsión o incompletos, y además pueden seguir trabajando, aunque se destruya parte de la red (con cierta degradación). La explicación de este fenómeno se encuentra en que mientras la computación tradicional almacena la información en espacios únicos, localizados y direccionables, las redes neuronales lo hacen de forma distribuida y con un alto grado de redundancia [62].

Otra característica a resaltar es la capacidad de adaptabilidad que tienen frente a cambios del sistema. Esta característica las hace muy útiles para realizar operaciones en tiempo real, pero hay que destacar que esta es limitada ya que a medida que aumenta la adaptabilidad también se incrementa la inestabilidad frente a variaciones en las entradas.

Como desventajas se puede mencionar que las redes neuronales suelen requerir de gran cantidad de datos para que su fase de entrenamiento sea satisfactoria. Además, durante la fase de entrenamiento existe la posibilidad de que el entrenamiento se detenga en un mínimo local en lugar del mínimo global.

7.2 EVOLUCIÓN HISTÓRICA DE LAS REDES NEURONALES ARTIFICIALES

El concepto de red neuronal artificial puede hacer pensar que se está hablando de una disciplina de aparición reciente, pero en realidad sus bases se instauraron antes del advenimiento de la era de los computadores. A continuación, se presentará de forma breve la evolución de la teoría de la inteligencia artificial y de los modelos de redes neuronales más importantes a lo largo de las últimas décadas [63].

- Años 40

1943: el neurofisiólogo Warren McCulloch y el matemático Walter Pitts presentan el primer modelo simplificado de neurona artificial. Este modelo abrió el camino para la investigación de redes neuronales, dividida en dos segmentos: uno enfocado en procesos biológicos en el cerebro y el otro en aplicaciones de redes neuronales artificiales.

1949: el psicólogo Donald Hebb escribió su obra titulada "*The Organization of Behaviour*". En ella indica que la conectividad entre las diferentes neuronas del cerebro cambia conforme se van adquiriendo nuevos conocimientos. La regla de Hebb puede interpretarse como un modo de aprendizaje no supervisado en el que las conexiones entre dos neuronas son fortalecidas si estas se activan.

- Años 50

1956: los investigadores Rochester, Holland, Habait y Duda que trabajaban para IBM verifican por primera vez mediante simulaciones numéricas una teoría neuronal basada en los postulados de Hebb.

1957: Franck Rosenblatt desarrolló y simuló el primer modelo de red neuronal, al que denominó "*Perceptron*". Este se caracterizaba por el aprendizaje basado en el ensayo y el error, y representa una de las primeras implementaciones de un sistema conexionista. Posteriormente se demostrarían las limitaciones del *Perceptron*.

1959: Bernard Widrow y Marcian Hoff, de la Universidad de Stanford (USA) desarrollaron un dispositivo electrónico denominado ADALINE (*ADaptive Linear Element*) que tenía la función de reconocer patrones binarios. Posteriormente desarrollaron MADALINE, que fue la primera red neuronal aplicada a un problema del mundo real. Cabe destacar que estos sistemas todavía se utilizan comercialmente hoy en día.

- Años 60

1969: cabe destacar los trabajos de Marvin Minsky y Seymour Papert que evidenciaron las limitaciones del *Perceptron* de una capa, concluyendo que este solo permite resolver problemas linealmente separables. Demostraron matemáticamente la imposibilidad de implementar la función lógica XOR de dos entradas mediante un *Perceptron*. Esta demostración desalentó el interés de la comunidad científica por las redes neuronales artificiales, y juntamente con una visión crítica de la población sobre la existencia de "maquinas pensantes" se inicia el período conocido como Invierno de la Inteligencia Artificial que duraría la década de los 80 del siglo pasado.



- Años 70

1974: Paul Werbos de la Universidad de Harvard (USA) presentó el método "*Back-propagation*" para el aprendizaje de redes neuronales artificiales

1975: Kunihiko Fukushima (Japón) presentó la primera red neuronal multicapa de la historia a la cuál denominó '*Cognitron*', utilizando como base una extensión del concepto de los perceptrones.

- Años 80

1981: Cristopher Von der Malsburg del Instituto Max Planck (Alemania) presentó un nuevo tipo de red neuronal basada en un modelo de neuronas más realistas a las que denomina "*Spiking Neurons*" o neuronas pulsantes.

1982: John Hopfield del Instituto Tecnológico de California (USA), enfocó el uso de las redes neuronales en el desarrollo de dispositivos útiles. Sus estudios fueron clave para el resurgimiento de las redes neuronales. El mismo año Teuvo Kohonen presenta los mapas auto-organizativos (SOM) "*Self Organizative Maps*". Los SOM son redes neuronales que utilizan una función de vecindad para conservar las propiedades topológicas del espacio de entrada.

1986: David E. Rumelhard, Geoffrey E. Hinton y Ronald J. Williams desarrollan el algoritmo de aprendizaje de retro propagación "*back-propagation*" para redes neuronales multicapa. A partir de este momento se produce un aumento exponencial de los trabajos científicos relacionados con las redes neuronales.

1988: D.S. Broomhead y D. Lowe presentan el diseño de una red neuronal de tres capas donde se usan neuronas con un elemento no lineal consistente en una función de base radial (RBF) en la capa oculta.

1988, 1989: Desarrollo de las redes neuronales de retardo temporal (TDNN) por parte de K. Lang y G.E. Hinton. Estas redes neuronales son un tipo de red multicapa cuyas neuronas de salida son replicadas a lo largo del tiempo, es decir, las salidas de una capa se almacenan en varios instantes de tiempo tras los cuales sirven para alimentar la siguiente capa.

- Años 90

En los años 90 se inició un fuerte interés por el estudio de las máquinas de soporte vectorial (SOV), las cuales se basan en un algoritmo de aprendizaje supervisado desarrollado por Vladimir Vapnik. También destacan los trabajos realizados por R.J. Williams y D. Zisper en la consecución de algoritmos de aprendizaje que permitieran la implementación de redes neuronales recurrentes "*Fedback Networks*".

- Del año 2000 en adelante

Aparece el concepto de aprendizaje profundo, o "*Deep Learning*". El 2009 se demuestra que con un conjunto de datos suficientemente grandes las redes neuronales no precisan de pre-entrenamiento. El 2012, algoritmos de reconocimiento de patrones artificiales alcanzan un desempeño de nivel humano en determinadas tareas. Se apunta al aprendizaje profundo como la principal tecnología para la creación de sistemas artificiales inteligentes.

7.3 REDES NEURONALES UNIDIRECCIONALES Y RECURRENTE

Existe una gran variedad de modelos de redes neuronales que se pueden clasificar des de una triple óptica: en función de la forma de aprendizaje²⁷, en función de la arquitectura y en función del área de aplicación. En la Figura 61, se muestran estas tres formas de clasificar las redes neuronales:

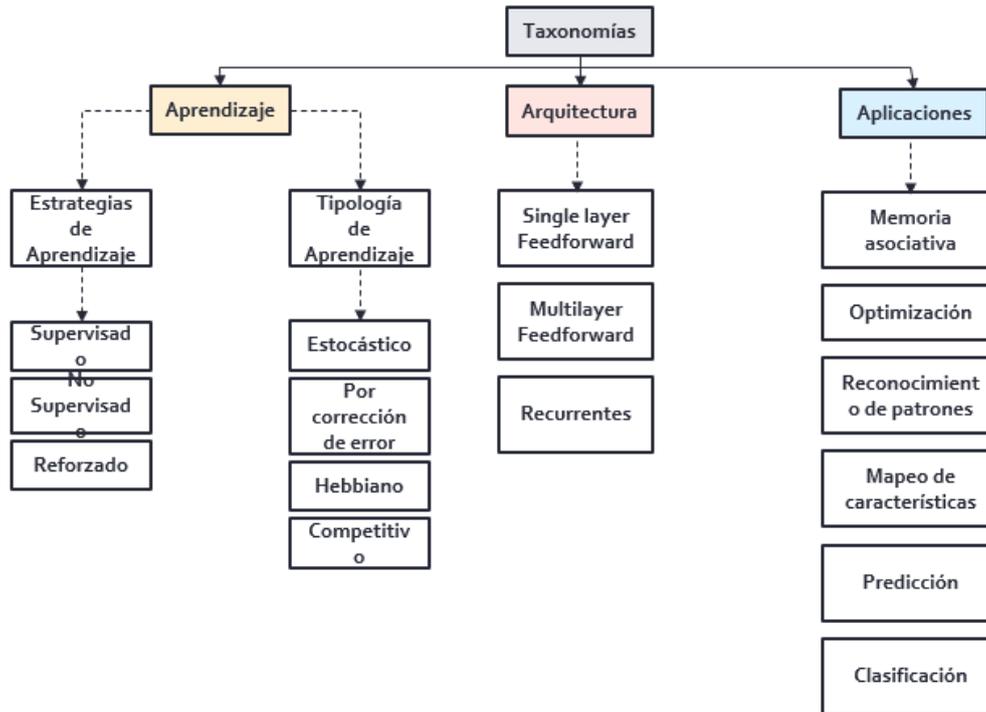


Figura 61: Clasificación de las redes neuronales. Fuente: Elaboración propia.

En el presente trabajo se partirá de la clasificación de las redes neuronales según la arquitectura, con el objetivo de describir las redes neuronales unidireccionales y las recurrentes.

7.3.1 REDES NEURONALES UNIDIRECCIONALES

Las redes unidireccionales son redes que únicamente disponen de conexiones hacia delante (Feed-Forward). Las neuronas de una capa reciben señales de las neuronas de la capa anterior y así sucesivamente hasta llegar a la capa de salida. Los modelos más conocidos, entre otros, son: el perceptrón simple, el multilayer perceptrón (MLP), las MLP modulares, etc.

²⁷ El concepto de aprendizaje hace referencia a la forma en que se produce el ajuste de los pesos de conexión entre las neuronas.



Perceptrones simples

Son redes neuronales que no disponen de capa oculta. Están constituidas por un vector de entrada y un vector de salida objetivo. La relación de estos vectores se aprende mediante la regla de aprendizaje, "*perceptron learning rule*". Estas redes son muy útiles a la hora de realizar tareas de clasificación cuando los datos de entrada son linealmente separables [60].

Perceptrones multicapa

Este tipo de red neuronal también es conocida como MLP (*Multilayer Perceptron*). Se compone de forma muy parecida al perceptrón simple, pero incluye una o varias capas ocultas. Es capaz de aproximar con un error muy pequeño cualquier función continua (lineal o no). Se suele utilizar en tareas de clasificación o para aproximar funciones. Cuanto mayor sea el número de capas ocultas, mayor capacidad de ajuste dispondrá, pero se necesitará más tiempo para entrenar la red y aumenta el riesgo de sobre ajuste. El método de entrenamiento más utilizado es el de la retropropagación del error (backpropagation). Posteriormente, se presentará un ejemplo de funcionamiento de esta regla de aprendizaje [60].

Muchas aplicaciones requieren que las redes neuronales respondan a secuencias temporales, es decir a series temporales. El algoritmo de aprendizaje de la retropropagación del error permite convertir una red MLP en una red TDNN (time-delay neural network) utilizando entradas con retardos. Aparte de las TDNN, existen otras variaciones de las MLP como por ejemplo las FFBP o las MLP modulares. Las FFBP se estructuran igual que las MLP, pero permiten conexiones unidireccionales entre neuronas de capas no consecutivas. Esto les da la ventaja de que necesitan menos tiempo de entrenamiento. Las MLP modulares se componen por varias MLP conectadas en paralelo que comparten la capa de entrada y la capa de salida.

7.3.2 REDES NEURONALES RECURRENTE

Las redes neuronales recurrentes se caracterizan por tener conexiones entre neuronas que no siguen la dirección de capa entrada hacia capa de salida. Estas conexiones se pueden ver como una especie de retroalimentación entre neuronas dentro de las capas. Actualmente poseen un número importante de aplicaciones, como por ejemplo servir de memorias asociativas, realizar tareas de clasificación de patrones, o en modelos de predicción. Se pueden clasificar en redes de primer orden y de segundo orden. A su vez las redes de primer orden pueden dividirse en parcialmente recurrentes y redes totalmente recurrentes [64].

Rede neuronales recurrentes de primer orden

Disponen de información sobre la salida objetivo. Las parcialmente recurrentes pueden memorizar secuencias a partir del pasado. Su estructura se asemeja a una red neuronal MLP a la que se han añadido lazos de realimentación. Se puede destacar dos modelos, la red secuencial de Jordan y la red recurrente simple de Elman:



- *Elman*: parten de la estructura de una MLP, pero incluyen una capa oculta adicional denominada capa de contexto. Esta capa tiene como entradas sus propias salidas y las salidas de las neuronas de la capa oculta, y su función es extraer información temporal de los datos.
- *Jordan*: este tipo de red tiene prácticamente la misma estructura que las redes Elman, con la diferencia de que las entradas de la capa de contexto no proceden de la capa oculta sino de la capa de salida.

Son redes neuronales recurrentes completas aquellas en las que cada una de las neuronas de una capa dispone de un lazo de realimentación hacia ella misma. Se pueden destacar los modelos RTRL (Real-Time Recurrent Learning) y las TDRB (Time-Dependent Recurrent Backpropagation).

Redes neuronales recurrentes de segundo orden

Estas no disponen de información explícita sobre la salida objetivo. Son muy utilizadas en procesos de aprendizaje de lenguajes

Memoria

Dado que la salida de una neurona recurrente en un instante de tiempo determinado es una función de las entradas de los instantes de tiempo anteriores se podría decir que estas tienen en cierta forma memoria. Esta "memoria" es la que convierte a las redes neuronales recurrentes en una herramienta muy útil a la hora de tratar con series temporales, ya que al "recordar" la información relevante sobre la entrada que recibieron anteriormente, disponen de más información que les permite ser más precisas prediciendo que otros modelos de predicción. También cabe destacar que son modelos no lineales hecho que les permite representar complejas relaciones no lineales.

NARX

En el presente trabajo las predicciones de generación eólica se realizarán utilizando un tipo de red neuronal recurrente dinámica denominada NARX (Nonlinear Autoregressive with Exogenous Inputs model). Es un tipo de red neuronal con una gran capacidad para modelar sistemas no lineales y especialmente series temporales. Se ha demostrado que el algoritmo de aprendizaje basado en el gradiente descendiente funciona mejor con las NARX. También se ha visto que el entrenamiento es más rápido y que este tipo de red neuronal tiene una mayor capacidad para identificar dependencias a largo plazo que las redes neuronales convencionales. Al disponer de entradas retardadas dispuestas en forma de cascada la propagación de la información no tiene que pasar por una no linealidad en cada iteración, y por lo tanto el gradiente no se modifica siempre durante la diferenciación [65], [66].

Arquitectura de las NARX

Habitualmente se distingue entre dos arquitecturas diferentes:

- *Lazo abierto*: el valor futuro de la serie temporal $y(t+1)$ es predicho a partir de los valores presentes de la entrada $u(t)$, de los valores pasados de $u(t)$ y de los valores reales pasados de $y(t)$.

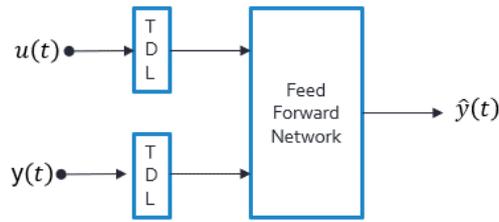


Figura 62: NARX en lazo abierto. Fuente: Elaboración propia

Este tipo de arquitectura presenta dos ventajas. La primera es que al usar valores reales como entrada de la red se obtienen resultados más precisos. La segunda es que tiene una estructura muy similar a las redes MLP de manera que se pueden utilizar los mismos algoritmos de aprendizaje.

- **Lazo cerrado:** el valor futuro de la serie temporal $y(t+1)$ es predicho a partir de los valores presentes y pasados de $u(t)$ y de los valores pasados predichos de $y(t)$.

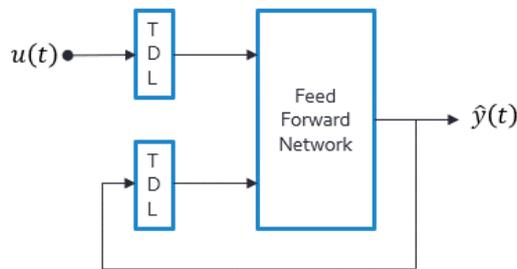


Figura 63: NARX con lazo cerrado. Fuente: Elaboración propia.

Se suele utilizar la arquitectura en lazo abierto para la fase de entrenamiento y la arquitectura en lazo cerrado para realizar las predicciones una vez finalizado el entrenamiento. En la Figura 64, se puede ver un ejemplo de una NARX en lazo cerrado que recibe como entradas los tres valores temporales de la variable $u(t)$, el bias o umbral (b), y tres valores temporales de la salida $y(t)$

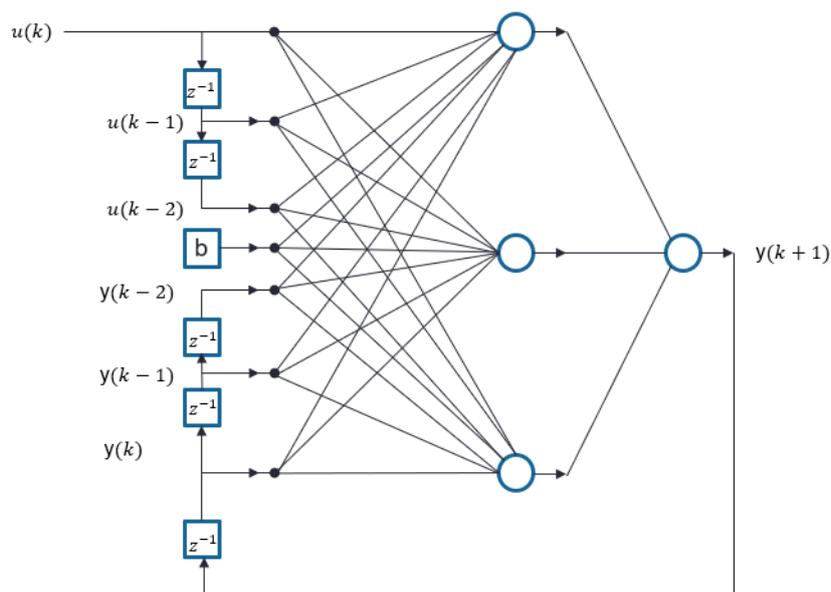


Figura 64: Ejemplo de ventana temporal en una NARX. Fuente: Elaboración propia.

7.3.3 ENTRENAMIENTO DE LAS REDES NEURONALES MLP

Un sistema neuronal se compone por un o unos modelos determinados de neurona y una arquitectura de red en particular. A su vez, existen una serie de factores que influyen sobre el proceso de aprendizaje, como:

- Los pesos iniciales que normalmente se inicializan de forma aleatoria.
- La constante de aprendizaje.
- La función de coste a utilizada.
- La técnica de optimización utilizada (método del gradiente descendiente, método de Newton...).
- El número de capas y de neuronas presentes en las capas ocultas.

Una vez definidos estos parámetros y la estructura del modelo, se procede a entrenar la red. Una de las formas más utilizadas es utilizar una regla de aprendizaje que tiene el objetivo de minimizar una función de error o coste determinada. El proceso de aprendizaje suele ser iterativo, actualizándose los pesos sinápticos sucesivamente hasta que la red alcanza el rendimiento deseado o se ha ejecutado un número de iteraciones prefijado. Con el objetivo de sobrepasar las limitaciones de las redes neuronales de una capa, aparecieron las redes neuronales multicapa. Fue una evolución muy lenta debido a que las reglas de aprendizaje que se habían desarrollado anteriormente no funcionaban para la capa oculta debido a que se disponía de información sobre el error de salida (diferencia entre la salida objetivo y la salida predicha) pero no se sabía cómo calcular el error en la capa oculta ya que no había una salida objetivo-definida para esta capa. En 1986, la introducción de la regla de aprendizaje basada en la retropropagación del error (backpropagation rule) proporcionó una forma sistemática para determinar el error en la capa oculta, permitiendo incrementar la profundidad de las redes neuronales, y abriendo todo un abanico de aplicaciones que anteriormente no estaban disponibles para este tipo de metodologías [67].

Retro-propagación del error (backpropagation)

Dada la importancia de este algoritmo de entrenamiento en el proceso de la evolución de las redes neuronales, se va a realizar una breve explicación de su forma de actuar mediante un ejemplo de aplicación en una red MLP. A continuación, se presenta el flujo de trabajo de este algoritmo:

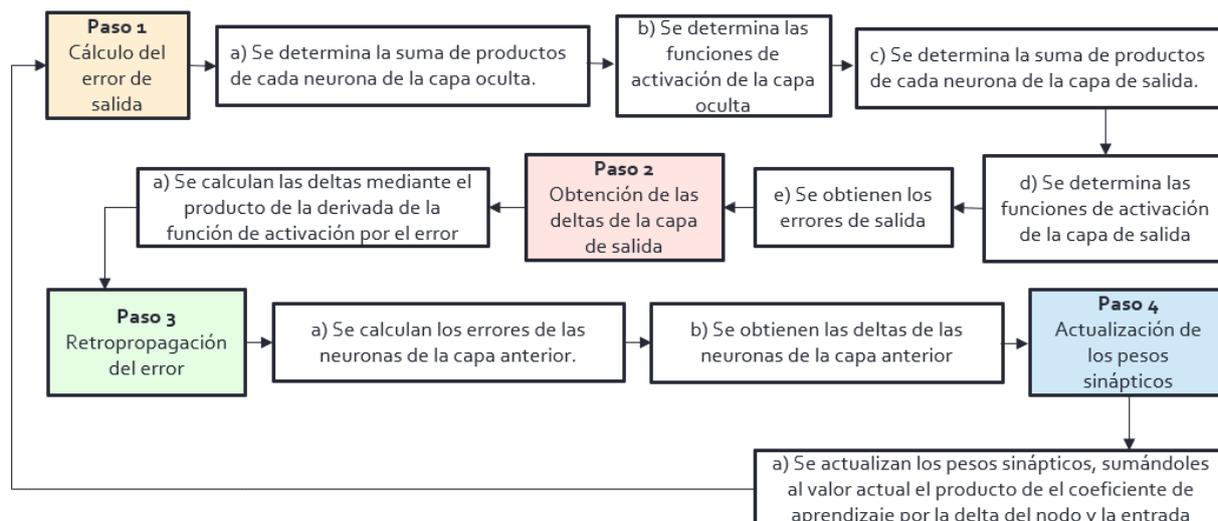


Figura 65: Flujo de trabajo de la retro-propagación del error. Fuente: Elaboración propia

Para el ejemplo de aplicación se parte de una red neuronal con dos nodos de entrada, dos nodos en la capa oculta y dos nodos de salida. Se omiten los umbrales por conveniencia [67]:

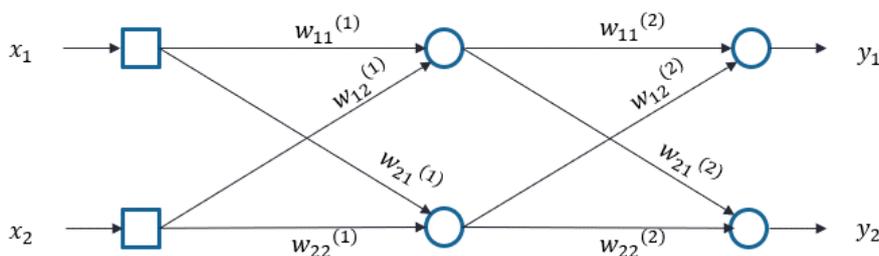


Figura 66: Red neuronal MLP. Fuente: Elaboración propia.

Paso 1: determinación del error de salida.

- a) Cada neurona recibe una "señal" por parte de las neuronas de la capa anterior conectadas a ella. Estas señales se componen por un valor numérico que se genera a partir del producto de la salida de la neurona por el peso sináptico de la conexión. A partir de la suma de las señales que recibe la neurona, esta determina la "salida" o respuesta que proporcionará. Por lo tanto, se puede definir una operación de matrices que permite obtener la suma de productos que recibe cada neurona:

$$\begin{pmatrix} v_1^{(1)} \\ v_2^{(1)} \end{pmatrix} = \begin{pmatrix} w_{11}^{(1)} & w_{12}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad (52)$$

- b) Obtención de las funciones de salida de la capa oculta. La salida de cada una de las neuronas se obtiene aplicándole una función de activación a la suma de señales que ha recibido:

$$\begin{aligned} y_1^{(1)} &= \varphi(v_1^{(1)}) \\ y_2^{(1)} &= \varphi(v_2^{(1)}) \end{aligned} \quad (53)$$

Donde:

φ : función de activación de las neuronas de la capa oculta.

- c) Determinación de la suma de productos de los nodos de salida. Se repite el mismo proceso mencionado en el apartado a, pero esta vez con diferentes entradas y otros pesos sinápticos:

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} w_{11}^{(2)} & w_{12}^{(2)} \\ w_{21}^{(2)} & w_{22}^{(2)} \end{pmatrix} \cdot \begin{pmatrix} y_1^{(1)} \\ y_2^{(1)} \end{pmatrix} \quad (54)$$

- d) Cálculo de la salida de la red neuronal. Se repite el proceso mencionado en el apartado b, pero con diferentes funciones de activación, y otras sumas de productos:

$$\begin{aligned} y_1 &= \varphi(v_1) \\ y_2 &= \varphi(v_2) \end{aligned} \quad (55)$$

Donde:

φ : función de activación de las neuronas de la capa de salida.

- e) Obtención de los errores de salida. Estos se definen como la diferencia entre la salida y la salida objetivo:

$$\begin{aligned} e_1 &= d_1 - y_1 \\ e_2 &= d_2 - y_2 \end{aligned} \quad (56)$$

Donde:

d_i : salidas objetivo (target).

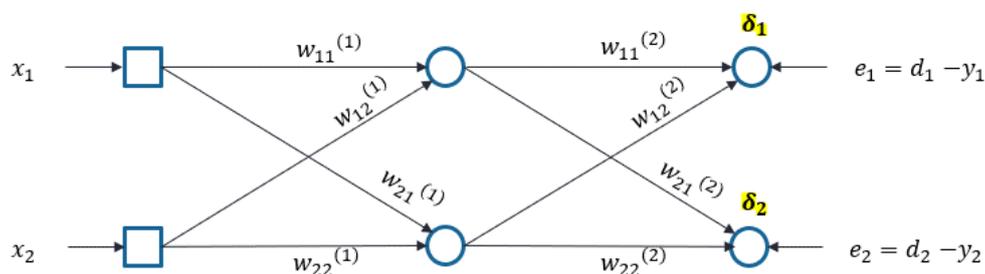


Figura 67: Errores de salida de la red neuronal. Fuente: Elaboración propia

Una vez obtenidos los errores de salida, se tiene que actualizar los pesos sinápticos de manera que el error se minimice a lo largo de las iteraciones. El objetivo es encontrar el mínimo global de la función de error mediante la técnica del gradiente descendiente. Esta técnica consiste en actualizar los pesos en función de la variación que ha experimentado el error. La derivada de la función de error respecto

a los pesos proporciona información mediante su pendiente de la manera en que influye sobre el error el cambio de los valores de los pesos. Si se obtiene una derivada negativa se tiene que incrementar el valor de los pesos, y si es positiva, se tiene que disminuir.

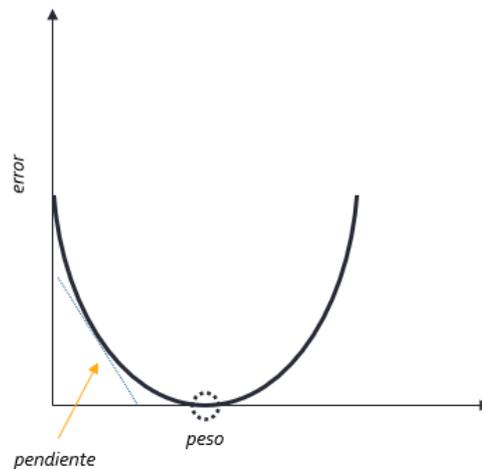


Figura 68: Gradiente descendente. Fuente: Elaboración propia.

Paso 2: cálculo de los deltas para la capa de salida

- a) Se calculan para cada una de las neuronas de salida, mediante el producto de la derivada de la salida por el error total que se ha producido en ese nodo:

$$\begin{aligned} \delta_1 &= \varphi'(v_1) \cdot e_1 \\ \delta_2 &= \varphi'(v_2) \cdot e_2 \end{aligned} \quad (57)$$

Paso 3: se calcula la delta para todos los nodos de todas las capas mediante la retro-propagación del error.

- a) Se calcula el error de cada neurona de la capa anterior como la suma de la multiplicación de los pesos sinápticos por la delta correspondiente. De esta manera se obtiene el error como la suma de las contribuciones de cada peso. Siguiendo con el ejemplo anterior, los errores de las neuronas de la capa oculta serían:

$$\begin{aligned} e_1^{(1)} &= w_{11}^{(2)} \cdot \delta_1 + w_{21}^{(2)} \cdot \delta_2 \\ e_2^{(1)} &= w_{12}^{(2)} \cdot \delta_1 + w_{22}^{(2)} \cdot \delta_2 \end{aligned} \quad (58)$$

A partir de los errores de los nodos de la capa interna se pueden calcular los deltas correspondientes. Se repite el mismo proceso mencionado en el *paso 2*, pero con los errores y las salidas de la capa oculta:

$$\begin{aligned} \delta_1^{(1)} &= \varphi'(v_1^{(1)}) \cdot e_1^{(1)} \\ \delta_2^{(1)} &= \varphi'(v_2^{(1)}) \cdot e_2^{(1)} \end{aligned} \quad (59)$$

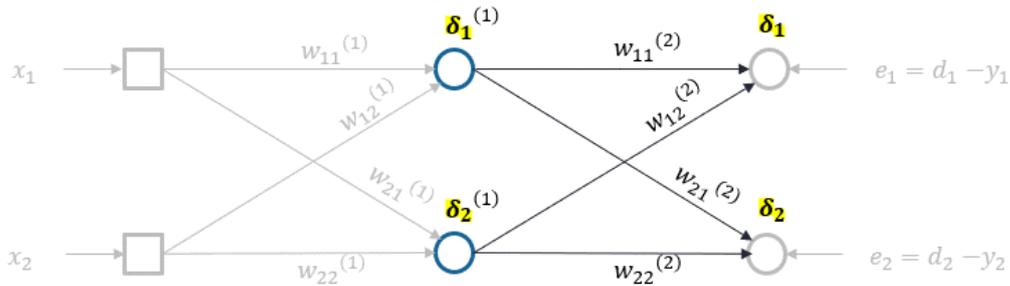


Figura 69: Deltas de los nodos de la capa oculta. Fuente:[67].

En resumen, el error de los nodos de la capa oculta se calcula como la suma de productos de los pesos y la delta, hacia la izquierda, y los deltas se calculan como el producto de la derivada de la función de activación por el error. En caso de haber más capas ocultas, simplemente se repetirá el proceso para cada capa oculta.

Paso 4: se actualizan los pesos sinápticos de todas las conexiones.

$$w_{ij} = w_{ij} + \alpha \cdot \delta_i \cdot x_i \quad (60)$$

Donde:

α : es el coeficiente de aprendizaje (limita la velocidad de entrenamiento con el objetivo de evitar el sobre ajuste).

δ_i : delta del nodo.

x_i : entrada del nodo.

Volviendo al ejemplo expuesto, el peso sináptico w_{21} , es decir el que hay entre la primera neurona de la capa oculta y la segunda neurona de salida, se actualizaría de la siguiente manera:

$$w_{21}^{(2)} = w_{21}^{(2)} + \alpha \cdot \delta_2 \cdot y_1^{(1)} \quad (61)$$

Donde:

$y_1^{(1)}$: es la salida de la primera neurona de la capa oculta.

Los demás pesos se actualizarían de la misma manera. Este proceso se repite en cada iteración hasta que se ha minimizado el error, o se ha llegado al número máximo de iteraciones establecido.

CAPÍTULO 8: MODELO DE PREDICCIÓN

En este capítulo se presenta la estructura del modelo de predicción elaborado, y el flujo de trabajo que se ha seguido para entrenarlo y testarlo. Posteriormente se describe como se definen los elementos básicos de una red neuronal usando la Neural Network Toolbox de MATLAB, y finalmente se presentan los resultados que se ha obtenido durante la fase de entrenamiento del modelo, y de las pruebas de predicción que se han realizado.

8.1 INTRODUCCIÓN

Una vez finalizada la tarea de regeneración de las series temporales se dispone de datos preparados para ser usados por el modelo de predicción. Al disponer de datos clasificados según la estación del año, el objetivo es disponer de una red neuronal para cada una de ellas. Estas redes neuronales recurrentes serán entrenadas mediante datos de una estación, y mediante un algoritmo de optimización se seleccionará su mejor configuración durante la fase de validación. Una vez finalizado el entrenamiento, al realizar las predicciones se mirará a que estación pertenece el período para el que se quiere predecir la generación eléctrica del parque, y se cargará la red neuronal correspondiente a esa estación. Durante la fase de entrenamiento se usará la *Neural Network Toolbox* de MATLAB para definir una red neuronal recurrente del tipo NARX con una arquitectura en lazo abierto y tres entradas independientes:

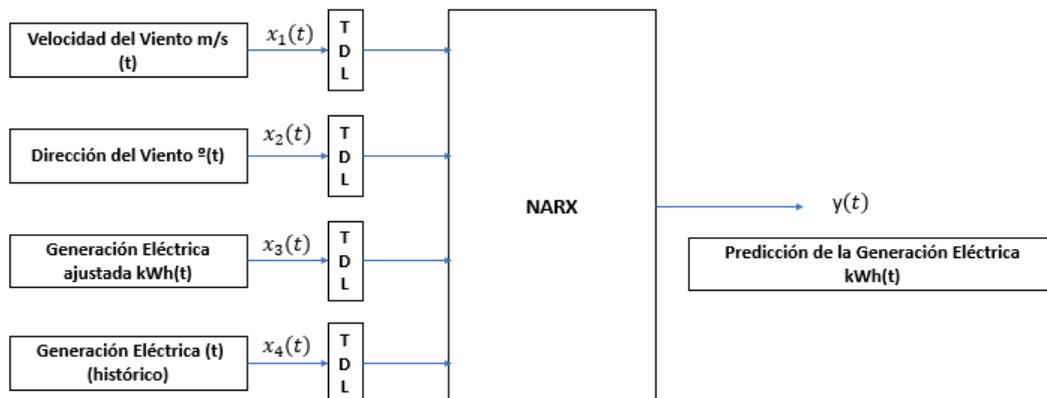


Figura 70: Entradas y salidas del modelo de predicción en lazo abierto. Fuente: Elaboración propia.

Durante la fase de prueba se cierra el lazo de la red:

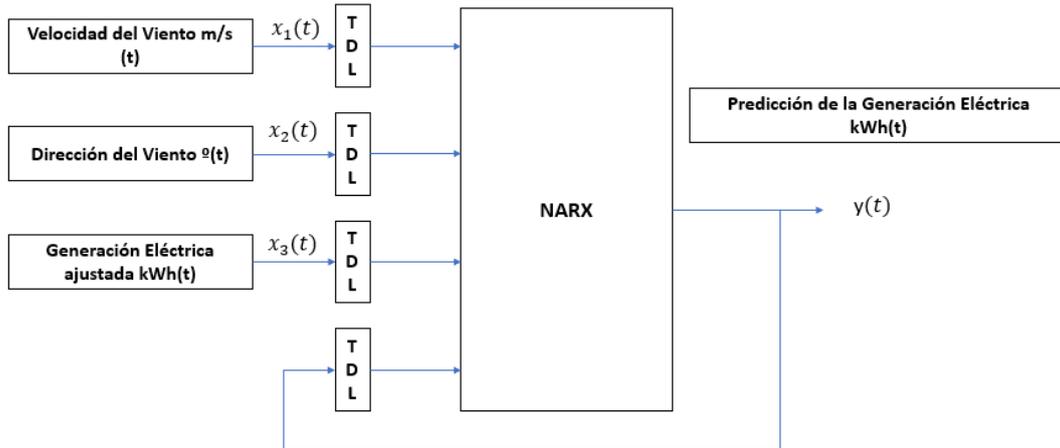


Figura 71: Entradas y salidas del modelo de predicción en lazo cerrado. Fuente: Elaboración propia.

A continuación, se expone el flujo de trabajo seguido para entrenar las NARX de cada estación, y para realizar las predicciones.

Flujo de trabajo del entrenamiento de las NARX

Una vez definidos los parámetros de la red neuronal y preparados los datos de entrada y las salidas objetivo se inicia el proceso de entrenamiento. Para entrenar la red se usa el algoritmo de aprendizaje de Levenberg-Marquardt que se basa en el ajuste de mínimos cuadrados de una función. Suele encontrar de forma más rápida el mínimo global que el algoritmo del gradiente descendente, de manera que se puede considerar como una versión mejorada de la retropropagación del error.

Se configura los entrenamientos de tal manera que estos finalizan cuando la red neuronal consigue minimizar el gradiente por debajo de un valor de 0,0003, y se ha fijado un número máximo de 1.000 de epochs. Una vez finalizada la fase de entrenamiento se realiza la predicción para los datos de validación con un horizonte de 24 horas, y se calculan los valores del: RMSE para cada una de las predicciones, y se calcula el RMSE medio, cuyo valor se compara con el mejor RMSE medio que se había obtenido hasta el momento. Si el valor nuevo es menor ($RMSE(t) < RMSE(t-1)$), se guarda la configuración de la red, ya que significa que la nueva configuración de la red realiza mejores predicciones que la estructura guardada anteriormente. Este proceso se repite múltiples veces de manera que una vez finalizado el proceso se dispone de la mejor combinación de pesos posibles.

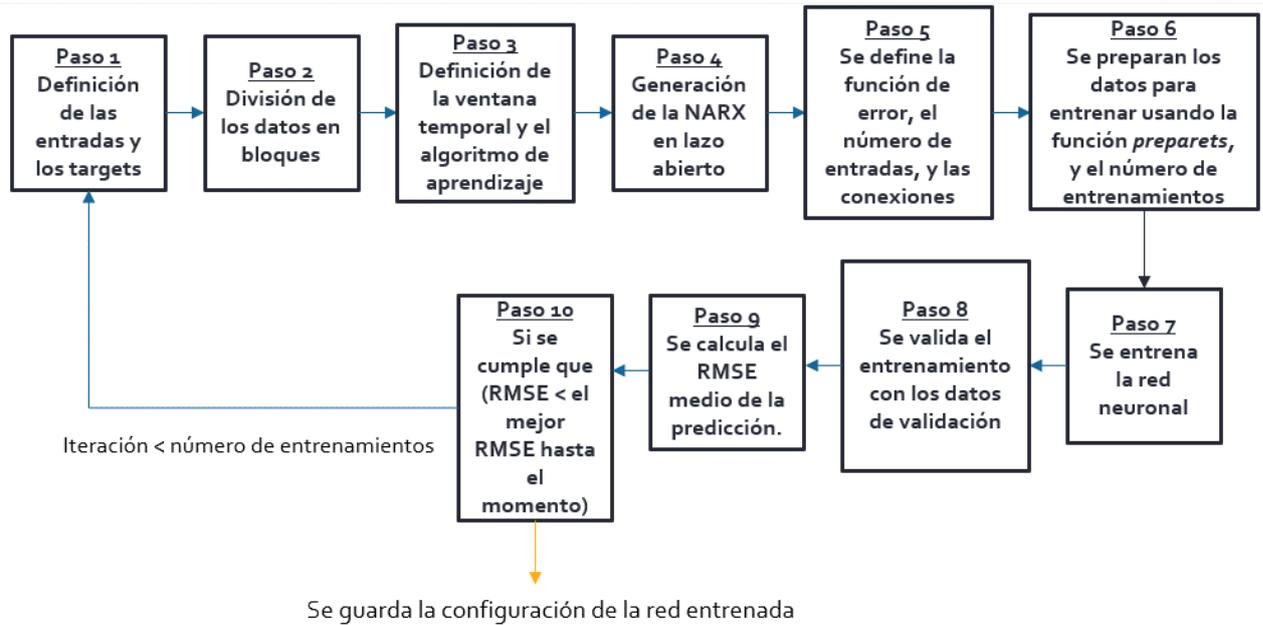


Figura 72: Flujo de trabajo de entrenamiento de las NARX del modelo de predicción. Fuente: Elaboración propia

Flujo de trabajo de las predicciones

Una vez obtenido los datos de las series temporales del archivo que se usará para realizar las pruebas de predicción, se determina a que estación del año pertenece y se carga la red neuronal entrenada para la estación correspondiente, y se cierra el lazo. Se definen los valores de las variables de entrada y los targets, y se preparan para la predicción según el período de previsión establecido. A continuación, se calcula las métricas de error y se representa gráficamente los resultados de la predicción.

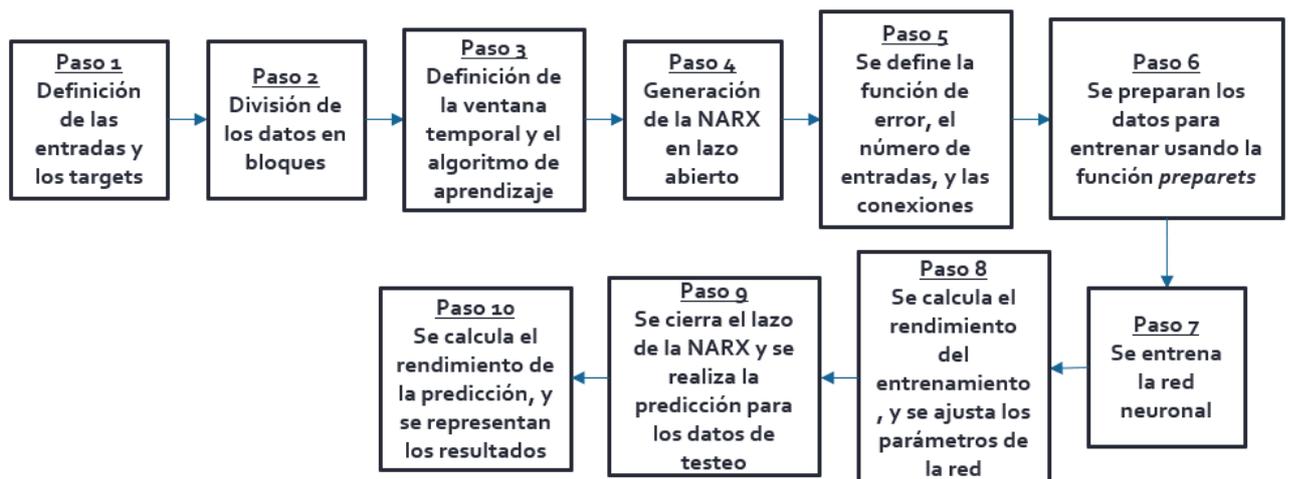


Figura 73: Flujo de trabajo de las NARX durante la fase de prueba. Fuente: Elaboración propia



Definición de una NARX usando la Neural Network Toolbox de MATLAB

La Neural Network Toolbox de MATLAB permite crear de forma relativamente sencilla una red neuronal con gran capacidad computacional. El problema es que resulta complicado encontrar bibliografía útil sobre su uso, y por eso a continuación se va a mostrar mediante ejemplos como se usan algunas funciones y se definen algunos de los parámetros más importantes.

Paso 1: Tanto las variables de entrada como las variables de salida se tienen que definir en formato cell, y deben ser matrices de filas. Si se quiere entrenar la red neuronal con datos de diferentes archivos, se puede utilizar la función de MATLAB "catsamples". Esta realiza una combinación de las variables de diferentes archivos, a fin de que el entrenamiento se realice para el conjunto de series temporales no consecutivas en el tiempo destinadas al entrenamiento de la red neuronal.

Ejemplo:

```
X_Input_train = catsamples(x1,x2,x3,x4,x5,'pad');  
T_target_train = catsamples(t1,t2,t3,t4,t5,'pad');
```

Donde:

x1,x2,x3,x4,x5: son datos de las mismas variables de entrada pero de diferentes archivos.

t1,t2,t3,t4,t5: son datos de las mismas variables de salidas objetivo pero de diferentes archivos.

Paso 2: Al definir una NARX se debe especificar el tamaño de la ventana temporal de retardos de las variables, el tamaño de la ventana de retardos de la salida realimentada, el nombre de neuronas presentes en la capa oculta y la función de entrenamiento:

Ejemplo:

```
inputDelays = 1:36;  
feedbackDelays = 1:36;  
hiddenLayerSize = 20;  
trainFcn = 'trainlm';
```

Donde:

inputDelays: tamaño de la ventana temporal de las variables de entrada. En el ejemplo expuesto se almacenan los últimos 36 valores en una especie de "shift register" que funciona como una memoria.

feedbackDelays: tamaño de la ventana temporal de la salida realimentada

hiddenLayerSize: número de neuronas que hay en la capa oculta.

trainFcn: define el algoritmo de aprendizaje. 'trainlm' hace referencia al algoritmo de Levenberg-Marquardt.



Paso 4: Para entrenar la NARX se define la red en lazo abierto:

Ejemplo:

```
net = narxnet(inputDelays,feedbackDelays,hiddenLayerSize, 'open',trainFcn);
```

Paso 5: Usualmente se define una gran cantidad de parámetros adicionales, como la función de error, las entradas, las conexiones entre capas, etc. A continuación, se presenta como se definen algunos de estos elementos:

- Se fija la función de error que se va a utilizar para valorar el rendimiento de la predicción:

Ejemplo:

```
net.performFcn = 'mse';
```

- Se especifica el número de entradas de la red neuronal. La red generada dispondrá de 6 entradas:

Ejemplo:

```
net.numInputs=6;  
net.inputs{1}.name='x1';  
net.inputs{2}.name='y';  
net.inputs{3}.name='x3';  
net.inputs{4}.name='x4';  
net.inputs{5}.name='x5';  
net.inputs{6}.name='x6';
```

Se utiliza la función `net.inputConnect` para definir las conexiones entre las capas:

Ejemplo:

```
net.inputConnect=logical([1 1 0 1 1 1; 0 0 1 0 0 1]);
```

En el ejemplo se ha generado una matriz de dos filas y cuatro columnas. Cada columna hace referencia a las conexiones de una entrada, y cada fila con que capa se conecta. Un "1" significa que hay conexión, es decir que se define un peso sináptico, y un "0" que no. En el ejemplo, la $x_1(t)$ (primera columna de la matriz) solo se conecta a la capa oculta, y $y(t)$ también. Luego $x_3(t)$ solo se conecta a la capa de salida, $x_4(t)$ y $x_5(t)$ se conectan a la capa oculta, y $x_6(t)$ se conecta simultáneamente a la capa oculta y a la capa de salida. En Figura 74 se puede visualizar las conexiones que genera la matriz de ejemplo:

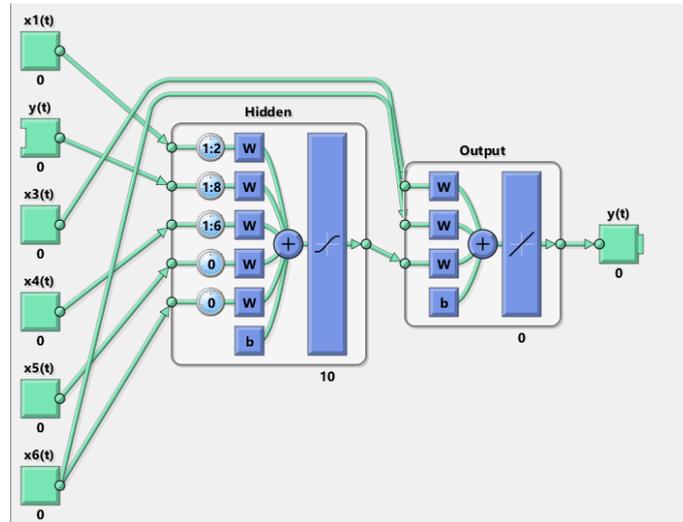


Figura 74: Ejemplo de conexiones entre capas en una NARX. Fuente: Elaboración propia.

Paso 8: La Neural Network Toolbox de Matlab dispone de una función que prepara los datos de forma automática, `preparets`. Esta función rellena con datos las ventanas temporales definidas, redefine las entradas y salidas desplazadas, etc. Es una función muy útil que ahorra mucho trabajo al programador

Ejemplo:

```
[Xs,Xi,Ai,Ts] = preparets(net, X_input_train, {}, T_target_train);
```

Donde:

Xs: contiene los datos de entrada, menos los que se utilizan para rellenar la ventana temporal.

Xi: contiene los valores de entrada de la ventana temporal. En este caso se trata de un vector de 4 filas de 36 valores.

Ai: en el caso expuesto se encuentra vacío. Se utiliza cuando se define una NARX sin realimentación.

Ts: contiene los valores de salida objetivo desplazados.

Paso 9: Para entrenar la red se define el número de iteraciones o generaciones (epochs) que se quiere reentrenar sobre el conjunto de datos de entrada, el gradiente mínimo objetivo, etc..

Ejemplo:

```
rng('default')
net.divideFcn='divideblock';
[neto, tro, Yo, Eo, Xof, Aof] = train(net,Xs,Ts,Xi,Ai);
view(neto);
```

Donde:

Net.divideFcn: esta función divide los datos de entrenamiento en una serie de partes. Luego analiza cada parte con un orden aleatorio. El objetivo es reducir al máximo el peligro del



overfitting. "Divideblock" divide los datos en bloques, de esta manera no se pierde el orden temporal dentro de las diferentes partes generadas.

neto: contiene la estructura de la red neuronal definida.

Yo: contiene los valores de predicción.

Eo: contiene el error de predicción (salida objetivo-salida predicha)

Paso 10: Se calcula el rendimiento del entrenamiento. La función "perform" de Matlab permite calcular el MSE a partir de la salida de la red neuronal entrenada y la salida objetivo.

Ejemplo:

```
[Y,Xf,Af] = neto(Xs,Xi,Ai);  
perf = perform(neto,Ts,Y);
```

Paso 12: Una vez entrenada la red, se cierra el lazo y se realiza la predicción para datos nuevos:

Ejemplo:

```
disp('Forecast task...');  
[netc,Xic,Aic] =  
closeloop(neto,Xf,Af);  
view(netc)  
Y2 = netc(X_Input_test,Xic,Aic);  
e2 = gsubtract(T_target_test,Y2);
```

Paso 13: Para poder calcular el rendimiento de la predicción se suele calcular alguna de las métricas de error más comunes como el MSE, RMSE, RMSPE, MAE, MAPE...

Ejemplo:

```
to_test = cell2mat(T_target_test);  
MSEooo2 = mean(var(to_test',1))  
MSE_Eo2 = mse(e2)  
NMSEo2 = mse(e2)/MSEooo2  
R2o2 = 1 - NMSEo2
```

En caso de duda se recomienda consultar los scripts de entrenamiento (por ejemplo, *Entrenamiento_NARX_catesamples_Primavera*) y testeo (*Test_de NARX_Entrenadas*) generados a lo largo de la realización de este trabajo de fin de grado. A continuación, se presentarán los resultados de la fase de entrenamiento y de la fase de testeo.

8.2 RESULTADOS

En este apartado se muestran los resultados obtenidos con el modelo de predicción elaborado, los cuales se han separado en dos apartados. En el primer apartado se presentan los resultados de la fase de validación del entrenamiento obtenidos para cada una de las redes de cada estación. En el segundo apartado se presentan los resultados de predicción que se ha obtenido, para los ficheros seleccionados para el test de la metodología.

8.2.1 RESULTADOS DEL ENTRENAMIENTO

El modelo de predicción está compuesto por cuatro redes NARX donde cada una es entrenada con datos de una estación del año. A continuación, se exponen las métricas de error que se han obtenido durante la fase de obtención de la configuración óptima de la red neuronal para cada estación. Se ha utilizado el RMSE y el RMSPE como métrica de la calidad de los resultados. Dado que en el caso expuesto el RMSE indica la cantidad de energía [kWh] media que se ha desviado la predicción del valor real de generación eléctrica para el horizonte de predicción, y el RMSPE su porcentaje. Hay que tener en cuenta que el RMSPE es sensible a la presencia de valores ruidosos y en algunos casos presenta valores muy grandes. En estos casos no es posible una correcta interpretación de dicho valor y por lo tanto no se aportará.

INVIERNO

Para entrenar la red neuronal se han utilizado los cuatro archivos de datos históricos de mayor tamaño que se ha obtenido, correspondientes a los siguientes períodos:

- 02/01/2010 – 03/02/2010
- 29/12/2010 – 10/03/2011
- 27/12/2012 – 05/03/2013
- 14/01/2014 – 19/03/2014

Para la fase de validación se utilizan datos históricos para un período de dos semanas. A continuación, se presentan las métricas de error que se ha obtenido para las predicciones con un horizonte de 24 horas utilizando la configuración optimizada de la NARX. Observando Tabla 7 se puede ver que el RMSE mínimo ha sido de 361 kWh y el máximo de 1341 kWh. Esto significa que el error acumulado medio mínimo que ha habido en 24 horas ha sido de 361 kWh, con un RMSPE de un 10.2 %; estableciendo una media un 10 % de diferencia entre la predicción y la generación real. Por otro lado, se puede ver que hay algunas predicciones muy malas, por ejemplo, para el 09/02/2010 se ha obtenido un RMSE de 1361 kWh con un RMSPE de un 263 %. Esto indica que el error acumulado para las 24 horas ha sido de 1361 kWh y que se ha predicho con una desviación de más del 200 %. Seguramente corresponde a un día en el que en algún instante el parque estuvo parado, o la velocidad del viento no alcanzó la mínima de funcionamiento.

Tabla 7: Métricas de error de la fase de entrenamiento. Invierno. Fuente: Elaboración propia

Período de predicción	MSE	NMSE	R ²	MAPE (%)	MAE (kWh)	RMSE (kWh)	RMSPE (%)
07/02/2010 0:00:00 - 07/02/2010 23:00:00	1256629.5	1.7	-0.7	61.2	972.2	1121.0	8.4
08/02/2010 0:00:00 - 08/02/2010 23:00:00	130676.6	0.3	0.7	80.6	297.8	361.5	10.2
09/02/2010 0:00:00 - 09/02/2010 23:00:00	644043.3	1.3	-0.3	644.3	556.6	802.5	217.2
10/02/2010 0:00:00 - 10/02/2010 23:00:00	1800168.0	2.0	-1.0	1754.7	1091.5	1341.7	263.2
11/02/2010 0:00:00 - 11/02/2010 23:00:00	1064842.3	1.2	-0.2	84.1	828.6	1031.9	11.5
12/02/2010 0:00:00 - 12/02/2010 23:00:00	1212114.2	1.2	-0.2	207.9	872.1	1101.0	52.1
13/02/2010 0:00:00 - 13/02/2010 23:00:00	544289.1	0.9	0.1	669.9	601.7	737.8	99.3
14/02/2010 0:00:00 - 14/02/2010 23:00:00	223620.3	0.4	0.6	64.8	397.2	472.9	12.1
15/02/2010 0:00:00 - 15/02/2010 23:00:00	968587.4	0.7	0.3	71.9	664.4	984.2	7.9
16/02/2010 0:00:00 - 16/02/2010 23:00:00	849643.8	4.3	-3.3	134.3	841.6	921.8	18.2
17/02/2010 0:00:00 - 17/02/2010 23:00:00	218359.1	0.5	0.5	85.2	358.6	467.3	16.4
18/02/2010 0:00:00 - 18/02/2010 23:00:00	477075.2	1.2	-0.2	196.3	545.7	690.7	49.2
19/02/2010 0:00:00 - 19/02/2010 23:00:00	1308715.6	1.8	-0.8	57256401319702200.0	949.1	1144.0	14067606693706900.0

Medias de las métricas de error

Observando la Tabla 7 y la Tabla 8 se aprecia lo sensible que son las métricas de error porcentuales como el MAPE y el RMSPE a la presencia de valores ruidosos o a los instantes en los que el valor predicho predicción es muy cercano a cero, ya que para calcular estas métricas se divide entre el valor predicho. Observando el RMSE se aprecia que se ha obtenido un error de desviación acumulado de alrededor de 860 kWh.

Tabla 8: Medias de las métricas de error de la fase de entrenamiento. Invierno. Fuente: Elaboración propia

MSE	NMSE	R ²	MAPE (kWh)	MAE	RMSE (kWh)	RMSPE (%)
822981.9	1.4	-0.4	4404338563054330.0	690.6	859.9	1082123591823670.0

PRIMAVERA

Para entrenar la red neuronal se utilizan los cinco archivos de datos históricos de mayor tamaño disponibles, correspondientes a los siguientes períodos:

- 21/03/2014 - 05/05/2014
- 30/04/2013 - 29/05/2013
- 20/03/2011 - 18/04/2011
- 05/04/2010 - 30/04/2010
- 08/05/2010 - 20/06/2010

Para la fase de validación se utilizan datos históricos para un período de dos semanas. A continuación, se presentan las métricas de error que se ha obtenido para las predicciones con un horizonte de 24 horas utilizando la configuración optimizada de la NARX. Observando la Tabla 9 se puede ver que aparentemente los resultados de predicción son muy buenos ya que el RMSE más grande es de 736 kWh de error acumulado en 24 horas junto a un RMSPE del 5 %. También se dan algunos días en los que la predicción ha sido mala como para el día 11/06 en el cual se tiene un RMSPE de un 188 %. Al disponer de un RMSE pequeño seguramente están vinculados una generación eléctrica muy pequeña.

Tabla 9: Métricas de error de la fase de entrenamiento. Primavera. Fuente: Elaboración propia

Período de predicción	MSE	NMSE	R ²	MAPE (%)	MAE (kWh)	RMSE (kWh)	RMSPE (%)
01/06/2011 0:00:00 - 01/06/2011 23:00:00	215437.9	4.9	-3.9	47.5	391.7	464.2	5.3
02/06/2011 0:00:00 - 02/06/2011 23:00:00	103344.6	2.4	-1.4	39.6	279.3	321.5	4.5
03/06/2011 0:00:00 - 03/06/2011 23:00:00	541554.9	6.5	-5.5	46.1	627.6	735.9	5.2
04/06/2011 0:00:00 - 04/06/2011 23:00:00	502726.2	8.4	-7.4	5587349266966730.0	682.9	709.0	924126509064943.0
05/06/2011 0:00:00 - 05/06/2011 23:00:00	33563.3	3.7	-2.7	224.3	110.6	183.2	39.4
06/06/2011 0:00:00 - 06/06/2011 23:00:00	341515.6	0.9	0.1	50.5	419.1	584.4	7.9
07/06/2011 0:00:00 - 07/06/2011 23:00:00	303091.0	1.2	-0.2	7159138478682150.0	428.7	550.5	3507247254139230.0
08/06/2011 0:00:00 - 08/06/2011 23:00:00	166028.0	4.1	-3.1	313.1	301.4	407.5	73.5
09/06/2011 0:00:00 - 09/06/2011 23:00:00	72045.7	3.4	-2.4	359.2	187.7	268.4	56.2
10/06/2011 0:00:00 - 10/06/2011 23:00:00	64603.2	1.1	-0.1	129.9	198.6	254.2	17.2
11/06/2011 0:00:00 - 11/06/2011 23:00:00	78730.7	1.7	-0.7	746.0	220.2	280.6	188.2
12/06/2011 0:00:00 - 12/06/2011 23:00:00	208303.2	2.0	-1.0	55.5	375.7	456.4	7.2
13/06/2011 0:00:00 - 13/06/2011 23:00:00	33250.3	0.7	0.3	60.5	147.0	182.3	8.3

Medias de las métricas de error

Observando la Tabla 10 se puede ver que tanto el MAPE como el RMSPE no se pueden interpretar debido a la presencia de outliers que corrompe el resultado de las medias. El RMSE medio es de 415 kWh, menos de la mitad del obtenido para la estación invierno. Esto se debe a que se dispone de una mayor cantidad de datos y de mejor calidad para esta estación del año.

Tabla 10: Medias de las métricas de error de la fase de entrenamiento. Invierno. Fuente: Elaboración propia

MSE	NMSE	R ²	MAPE (kWh)	MAE	RMSE (kWh)	RMSPE (%)
204938.0	3.2	-2.2	4848663934488580.0	336.2	415.2	980654795753006.0

OTOÑO

Para entrenar la red neuronal se utilizan los cinco archivos de datos históricos de mayor tamaño disponibles, correspondientes a los siguientes períodos

- 14/11/2013 - 17/12/2013
- 25/11/2012 - 12/12/2012
- 26/09/2011 - 17/10/2011
- 04/10/2010 - 22/10/2010
- 06/12/2010 - 21/12/2010

Para la fase de validación se utilizan datos históricos para un período de dos semanas. A continuación, se presentan las métricas de error que se ha obtenido para las predicciones con un horizonte de 24 horas utilizando la configuración optimizada de la NARX. Comparando los resultados obtenidos para esta estación con los obtenidos para la estación de primavera se puede ver que los resultados son bastante peores. La razón es que se dispone de una menor cantidad de datos para esta estación del año. El peor RMSE obtenido alcanza prácticamente los 2000 kWh, valor bastante grande. También se puede ver que hay una mayor cantidad de valores ruidosos en las columnas de los MAPE y los RMSPE.

Tabla 11: Métricas de error de la fase de entrenamiento. Otoño. Fuente: Elaboración propia

Período de predicción	MSE	NMS E	R ²	MAPE (%)	MAE (kWh)	RMSE (kWh)	RMSPE (%)
29/10/2013 0:00:00 - 29/10/2013 23:00:00	3715172.5	47.8	-	112.2	1396.1	1927.5	14.3
30/10/2010 0:00:00 - 30/10/2010 23:00:00	2067146.6	11.4	10.4	168.2	1223.9	1437.8	21.2
31/10/2010 0:00:00 - 31/10/2010 23:00:00	1418145.9	7.0	-6.0	3798826741289220	976.0	1190.9	1393227820759560
01/11/2010 0:00:00 - 01/11/2010 23:00:00	2980318.3	10.7	-9.7	8746476380156470	1147.2	1726.4	1450519142423370
02/11/2010 0:00:00 - 02/11/2010 23:00:00	456964.2	12.1	11.1	304.5	561.7	676.0	66.7
03/11/2013 0:00:00 - 03/11/2013 23:00:00	1542958.9	8.5	-7.5	80.7	1102.3	1242.2	8.8
04/11/2013 0:00:00 - 04/11/2013 23:00:00	1816806.8	3.6	-2.6	60.2	1128.2	1347.9	7.0
05/11/2013 0:00:00 - 05/11/2013 23:00:00	629446.9	1.8	-0.8	28.9	670.4	793.4	3.6
06/11/2013 0:00:00 - 06/11/2013 23:00:00	5397360.5	5.8	-4.8	1850844573423080	1739.3	2323.2	641151367616336.0
07/11/2013 0:00:00 - 07/11/2013 23:00:00	2828387.6	8.2	-7.2	323.5	1315.2	1681.8	102.7
08/11/2013 0:00:00 - 08/11/2013 23:00:00	623395.0	3.3	-2.3	143.0	674.0	789.6	18.5
09/11/2013 0:00:00 - 09/11/2013 23:00:00	426244.8	0.8	0.2	4627111433557710	502.2	652.9	1013749323299770
10/11/2013 0:00:00 - 10/11/2013 23:00:00	103409.9	3.7	-2.7	153.9	292.7	321.6	20.7

Observando la tabla de las medias de las métricas se puede observar que el RMSE medio es de 1240 kWh. No es un valor pequeño, pero en el supuesto de una generación a potencia nominal del parque resultaría en un error medio de unos 50 kWh, para cada hora.

Tabla 12: Medias de las métricas de error de la fase de entrenamiento. Otoño. Fuente: Elaboración propia

MSE	NMSE	R ²	MAPE (kWh)	MAE	RMSE (kWh)	RMSPE (%)
1846596.7	9.6	-8.6	10148537478572200.0	979.2	1239.3	2314797717134290.0

VERANO

Para entrenar la red neuronal se utilizan los cuatro archivos de datos históricos de mayor tamaño disponibles, correspondientes a los siguientes períodos:

- 04/08/2013 - 22/09/2013
- 29/06/2011 - 05/08/2011
- 21/06/2010 - 22/09/2010
- 22/06/2013 - 25/07/2013

Para la fase de validación se utilizan datos históricos para un período de dos semanas. A continuación, se presentan las métricas de error que se ha obtenido para las predicciones con un horizonte de 24 horas utilizando la configuración optimizada de la NARX. Observando la tabla de los resultados de las métricas de error se puede ver que para los días 26/08 y 27/08 la previsión no fue muy buena, sobre todo la de día 27 ya que el RMSE es bastante elevado y el RMSPE también. Por otro lado, la previsión de día 28/08 fue muy buena con un RMSE de 170,7 kWh de error acumulado en 24 horas, con una variación porcentual de un 11.3 %. En términos generales se puede decir que la precisión de la previsión es bastante buena para 1/3 de los casos.

Tabla 13: Métricas de error de la fase de entrenamiento. Verano. Fuente: Elaboración propia

Período de predicción	MSE	NMSE	R ²	MAPE (%)	MAE (kWh)	RMSE (kWh)	RMSPE (%)
25/08/2011 00:00:00 - 25/08/2011 23:00:00	577928.5	1.2	-0.2	94.6	528.8	760.2	12.7
26/08/2011 0:00:00 - 26/08/2011 23:00:00	1978524.6	3.3	-2.3	99.2	1129.9	1406.6	12.2
27/08/2011 0:00:00 - 27/08/2011 23:00:00	7270779.8	979.1	978.1	5196.3	2412.9	2696.4	708.1
28/08/2011 0:00:00 - 28/08/2011 23:00:00	29139.9	0.7	0.3	90.1	145.5	170.7	11.3
29/08/2011 0:00:00 - 29/08/2011 23:00:00	144330.7	3.6	-2.6	381.5	343.1	379.9	76.4
30/08/2011 0:00:00 - 30/08/2011 23:00:00	55509.8	1.2	-0.2	50.1	181.7	235.6	11.2
31/08/2011 0:00:00 - 31/08/2011 23:00:00	65514.3	2.7	-1.7	446.9	231.8	256.0	52.9
01/09/2011 0:00:00 - 01/09/2011 23:00:00	12941.3	6.7	-5.7	1143.7	100.5	113.8	244.8
02/09/2011 0:00:00 - 02/09/2011 23:00:00	162570.7	1.6	-0.6	531.2	268.7	403.2	83.3
03/09/2011 0:00:00 - 03/09/2011 23:00:00	94919.5	1.2	-0.2	429.1	245.5	308.1	102.1
04/09/2011 0:00:00 - 04/09/2011 23:00:00	70328.1	1.8	-0.8	325.6	220.5	265.2	46.0
05/09/2011 0:00:00 - 05/09/2011 23:00:00	120451.4	3.9	-2.9	659.3	306.5	347.1	133.3
06/09/2011 0:00:00 - 06/09/2011 23:00:00	43533.5	6.6	-5.6	3399849494709530.0	178.3	208.6	1665579292859400.0



Observando la tabla de las medias de las métricas se puede observar que el RMSE medio es de 580.9 kWh. Es un valor relativamente pequeño, pero si se tiene en cuenta que el RMSPE de la mayoría de los días supera el 50 %, se ve que los resultados tampoco presentan gran calidad.

Tabla 14: Medias de las métricas de error de la fase de entrenamiento. Otoño. Fuente: Elaboración propia

MSE	NMSE	R ²	MAPE (kWh)	MAE	RMSE (kWh)	RMSPE (%)
817420.9	78.0	-77.0	261526884209152.0	484.1	580.9	128121484066223.0

Una vez analizado los resultados de validación de las NARX de todas las redes observa que se ha obtenido los mejores resultados de validación para la estación de primavera, que es la estación para la cual se dispone de archivos de datos de mayor tamaño. Queda reflejada la necesidad de disponer de una gran cantidad de datos históricos para poder entrenar adecuadamente una red neuronal. En términos generales se puede afirmar que se ha obtenido resultados aptos, pero que la presencia de valores ruidosos ha dificultado la interpretación de los resultados medios ya que tanto el MAPE como el RMSPE han resultado alterados.

8.2.1 RESULTADOS DE PREDICCIÓN

Una vez finalizada la etapa de entrenamiento de las NARX de cada estación se procede a realizar una serie de predicciones con las redes neuronales entrenadas para cada estación del año, para horizontes temporales de 24 h, 48 h y 72 h utilizando series temporales de test (nuevas), es decir que no se han utilizado ni para entrenar ni para validar los modelos, con el objetivo de ver cómo reacciona el modelo de predicción al someterle a datos no entrenados. Para aquellas estaciones para las que se tiene un archivo de prueba lo suficientemente grande se realiza una previsión para un período de 14 días. Se expone los resultados obtenidos para cada estación de la siguiente manera:

- Se representa la generación eléctrica real que se ha dado en todo el rango temporal y se muestra la curva de predicción a partir de una fecha de inicio (tercer día del archivo de prueba, es decir a partir de la hora 48 del archivo de prueba). La longitud de la curva de predicción depende del horizonte de predicción.
- Se muestran los resultados de las métricas de error para todo el rango de predicción.
- Se muestran las medias de las métricas de error.

RESULTADOS DE PREDICCIÓN: VERANO

Se realiza la predicción de la generación eléctrica del parque eólico de Es Milá para un período comprendido entre el 27/08/2012 00:00:00 – 10/09/2012.23:00:00 para horizontes de 24 horas, 48 h y 72 h. A continuación, se expone los resultados obtenidos para cada horizonte.

PREDICIONES PARA UN HORIZONTE DE 24 HORAS

GRÁFICOS

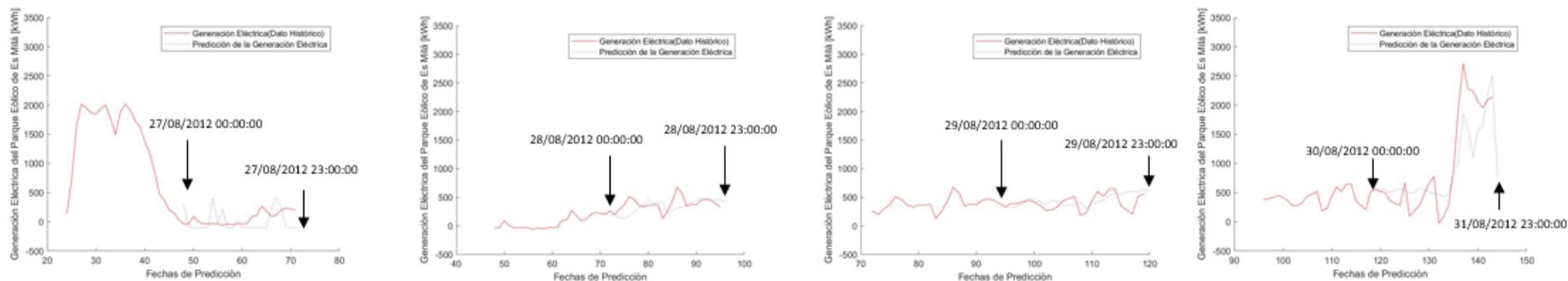


Figura 75: Predicciones de la generación eléctrica de Es Milá para horizontes de 24 horas. Verano. Fuente: Elaboración propia.

MÉTRICAS DE ERROR

Tabla 15: Métricas de error de la fase de predicción para un horizonte de 24 h en verano. Fuente: Elaboración propia

Período de predicción	MSE	NMSE	R ²	MAPE (%)	MAE (kWh)	RMSE (kWh)	RMSPE (%)
27/08/2012 00:00:00 - 27/08/2012 23:00:00	51899.6	4.0	-3.0	651.2	187.7	227.8	207.3
28/08/2012 0:00:00 - 28/08/2012 23:00:00	23365.4	1.9	-0.9	31.6	109.8	152.9	5.2
29/08/2012 0:00:00 - 29/08/2012 23:00:00	21176.5	1.3	-0.3	36.2	117.4	145.5	5.5
30/08/2012 0:00:00 - 30/08/2012 23:00:00	282934.6	0.4	0.6	148.0	391.9	531.9	41.6
31/08/2012 0:00:00 - 31/08/2012 23:00:00	2680965.6	6.4	-5.4	59.5	1308.7	1637.4	7.8
01/09/2012 0:00:00 - 01/09/2012 23:00:00	3412183.7	3.0	-2.0	695.0	1541.3	1847.2	211.7
02/09/2012 0:00:00 - 02/09/2012 23:00:00	3436137.3	22.0	-21.0	54.0	1530.8	1853.7	6.4
03/09/2012 0:00:00 - 03/09/2012 23:00:00	3847011.3	28.2	-27.2	59.8	1608.1	1961.4	7.3
04/09/2012 0:00:00 - 04/09/2012 23:00:00	2838145.1	2.9	-1.9	207.4	1503.4	1684.7	47.2
05/09/2012 0:00:00 - 05/09/2012 23:00:00	7212148.3	74.7	-73.7	1694.4	2340.7	2685.5	322.5
06/09/2012 0:00:00 - 06/09/2012 23:00:00	353073.4	4.9	-3.9	349.6	478.8	594.2	106.8
07/09/2012 0:00:00 - 07/09/2012 23:00:00	102450.9	3.2	-2.2	376.1	230.4	320.1	59.7
08/09/2012 0:00:00 - 08/09/2012 23:00:00	53652.6	498.6	-497.6	478.1	172.2	231.6	62.5

Tabla 16: Medias de las métricas de error de la fase de predicción para horizontes de 24 horas en verano. Fuente: Elaboración propia

MSE	NMSE	R ²	MAPE (kWh)	MAE	RMSE (kWh)	RMSPE (%)
1870395.7	50.1	-49.1	372.4	886.2	1067.2	84.0

Observando la Figura 75 se concluye que las curvas de predicciones se han ajustado bastante bien a la generación real que se ha dado. Aparentemente la predicción para las horas del día 30/08 ha sido más imprecisa que las demás, consultando la Tabla 15 se observa que efectivamente hay un mayor error RMSE que en las predicciones de los días anteriores, junto a un RMSPE elevado. Observando la primera predicción se observa que esta presenta un RMSE acumulado relativamente bajo (227.8 kWh) pero con un RMSPE muy elevado. Esto indica que se trata de un período en el que hubo una generación eléctrica muy baja, hecho que se confirma observando el primer gráfico. Observando los resultados medios se puede ver que hay un RMSE de 1067 kWh i un RMSPE del 84 %, valores elevados, vinculados a los malos resultados obtenidos en algunos días puntuales donde las previsiones han sido muy malas. En términos generales se puede decir que se ha obtenido predicciones relativamente buenas.

PREDICIONES PARA UN HORIZONTE DE 48 HORAS

GRÁFICOS

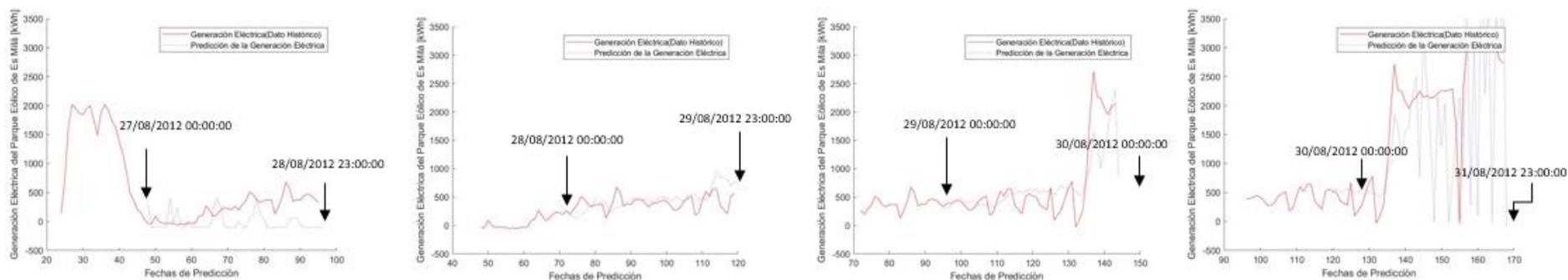


Figura 76: Predicciones de la generación eléctrica de Es Milá para horizontes de 48 horas. Verano. Fuente: Elaboración propia.



MÉTRICAS DE ERROR

Tabla 17: Métricas de error de la fase de predicción para un horizonte de 48 h en verano. Fuente: Elaboración propia

Período de predicción	MSE	NMSE	R ²	MAPE (%)	MAE (kWh)	RMSE (kWh)	RMSPE (%)
27/08/2012 00:00:00 - 28/08/2012 23:00:00	121520.4	3.2	-2.2	385.7	298.5	348.6	148.3
28/08/2012 00:00:00 - 29/08/2012 23:00:00	42519.7	2.8	-1.8	45.2	146.8	206.2	7.6
29/08/2012 00:00:00 - 30/08/2012 23:00:00	178494.6	0.3	0.7	111.7	277.6	422.5	37.9
30/08/2012 00:00:00 - 31/08/2012 23:00:00	1449589.7	1.3	-0.3	207.2	845.4	1204.0	81.9
31/08/2012 00:00:00 - 01/09/2012 23:00:00	2936558.8	3.7	-2.7	276.4	1450.0	1713.6	102.4
01/09/2012 00:00:00 - 02/09/2012 23:00:00	4486397.2	6.2	-5.2	392.8	1863.7	2118.1	151.4
02/09/2012 00:00:00 - 03/09/2012 23:00:00	4131146.3	27.9	-26.9	64.4	1782.0	2032.5	7.2
03/09/2012 00:00:00 - 04/09/2012 23:00:00	3210401.9	3.6	-2.6	84.6	1406.0	1791.8	11.3
04/09/2012 00:00:00 - 05/09/2012 23:00:00	1828187.9	2.1	-1.1	308.4	1099.2	1352.1	92.7
05/09/2012 00:00:00 - 06/09/2012 23:00:00	8282575.9	95.9	-94.9	1939.6	2484.4	2877.9	396.7
06/09/2012 00:00:00 - 07/09/2012 23:00:00	3223187.9	42.3	-41.3	2142.9	1286.9	1792.5	389.5
07/09/2012 00:00:00 - 08/09/2012 23:00:00	54707.1	2.5	-1.5	289.6	150.2	233.9	45.2
08/09/2012 00:00:00 - 09/09/2012 23:00:00	51418.8	4.0	-3.0	361.2	176.2	226.8	49.9

Tabla 18: Medias de las métricas de error de la fase de predicción para horizontes de 48 horas en verano. Fuente: Elaboración propia

MSE	NMSE	R ²	MAPE (kWh)	MAE	RMSE (kWh)	RMSPE (%)
2306669.7	15.1	-14.1	508.4	1020.5	1255.4	117.1

Observando los gráficos presentados, a primera vista parece que las mejores predicciones son las realizadas para los períodos 28/08/2012 00:00:00 al 29/08/2012 23:00:00 y del 29/08/2012 00:00:00 al 30/08/2012 23:00:00. Consultando las métricas de error se aprecia como efectivamente presentan un error porcentual RMSPE bastante bajo y un RMSE relativamente pequeño. Las curvas de predicción expuestas en el primer y último gráfico no se ajustan tan bien a la generación eléctrica real que ha habido, hecho que se confirma con un RMSPE que supera el 148 % en el primer caso y alcanza el 90 % en el segundo. Fijándose en los demás resultados de las métricas de error se aprecia que únicamente se ha obtenido predicciones con un RMSPE inferior al 30 % para tres períodos, hecho que ha disparado el valor de las medias. Se dispone de algunas predicciones bastante buenas y otras muy poco precisas.

PREDICIONES PARA UN HORIZONTE DE 72 HORAS

GRÁFICOS

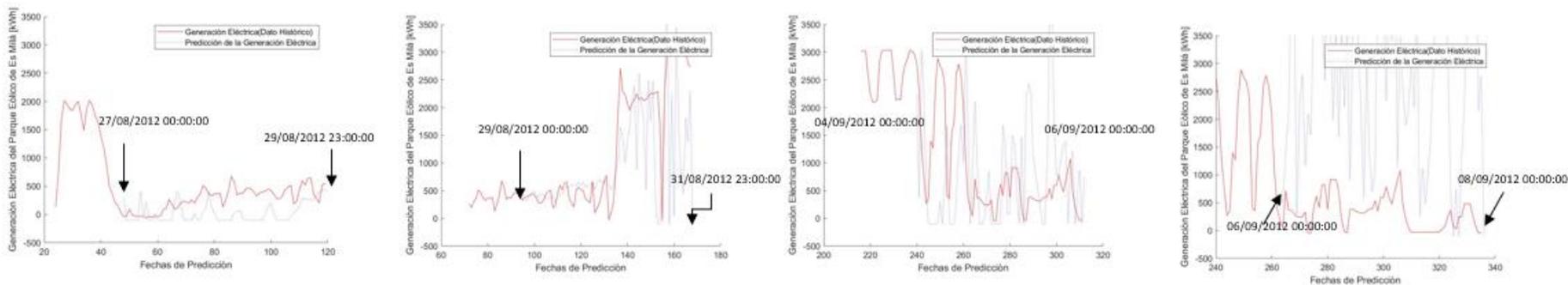


Figura 77: Predicciones de la generación eléctrica de Es Milá para horizontes de 72 horas. Verano. Fuente: Elaboración propia.

MÉTRICAS DE ERROR

Tabla 19: Métricas de error de la fase de predicción para un horizonte de 72 h en verano. Fuente: Elaboración propia

Período de predicción	MSE	NMSE	R ²	MAPE (%)	MAE (kWh)	RMSE (kWh)	RMSPE (%)
27/08/2012 00:00:00 - 29/08/2012 23:00:00	119013.2	3.0	-2.0	283.9	299.0	345.0	121.6
28/08/2012 00:00:00 - 30/08/2012 23:00:00	127058.5	0.3	0.7	83.6	236.5	356.5	27.2
29/08/2012 00:00:00 - 31/08/2012 23:00:00	999819.2	0.9	0.1	131.0	622.0	999.9	46.3
30/08/2012 00:00:00 - 01/09/2012 23:00:00	2271779.1	2.0	-1.0	200.6	1146.0	1507.2	70.5
31/08/2012 00:00:00 - 02/09/2012 23:00:00	3173639.1	5.0	-4.0	204.2	1490.6	1781.5	84.0
01/09/2012 00:00:00 - 03/09/2012 23:00:00	4267527.5	7.8	-6.8	283.6	1795.9	2065.8	124.1
02/09/2012 00:00:00 - 04/09/2012 23:00:00	3758012.4	5.2	-4.2	91.3	1648.0	1938.6	17.8
03/09/2012 00:00:00 - 05/09/2012 23:00:00	2278804.7	1.8	-0.8	100.8	1099.3	1509.6	13.4
04/09/2012 00:00:00 - 06/09/2012 23:00:00	1690144.6	2.5	-1.5	358.9	1004.0	1300.1	92.6
05/09/2012 00:00:00 - 07/09/2012 23:00:00	8307219.7	94.1	-93.1	2918.0	2429.7	2882.2	538.6
06/09/2012 00:00:00 - 08/09/2012 23:00:00	7828221.1	109.0	-108.0	5208.6	2173.3	2797.9	761.0
07/09/2012 00:00:00 - 09/09/2012 23:00:00	228082.6	10.6	-9.6	426.1	301.9	477.6	74.8
08/09/2012 00:00:00 - 10/09/2012 23:00:00	119013.2	3.0	-2.0	283.9	299.0	345.0	121.6



Tabla 20: Medias de las métricas de error de la fase de predicción para horizontes de 72 horas en verano. Fuente: Elaboración propia

MSE	NMSE	R ²	MAPE (kWh)	MAE	RMSE (kWh)	RMSPE (%)
2920776.8	20.2	-19.2	857.5	1187.2	1496.8	164.3

Observando los gráficos presentados, a primera vista parece que las mejores predicciones son las realizadas para los períodos 27/08/2012 00:00:00 al 29/08/2012 23:00:00 y del 29/08/2012 00:00:00 al 31/08/2012 23:00:00. Consultando las métricas de error se aprecia que la predicción para el período del 29/08/2012 00:00:00 -31/08/2012 23:00:00 presenta un RMSPE inferior al 50 % y un RMSE de unos 1000 kWh, mientras que para el período del 27/08/2012 00:00:00 - 29/08/2012 23:00:00 se ha obtenido un RMSPE del 130 %, pero con un RMSE de 345 kWh, valores que hacen pensar que se trata de un período en el que la generación eléctrica del parque ha sido baja (y que el valor predicho ha sido cercano a cero). Los dos gráficos de más de la derecha representan ejemplos de predicciones imprecisas. A simple vista la predicción para el período 04/09/2012 0:00:00 - 06/09/2012 23:00:00 no parece tan mala, pero presenta un RMSPE del 90 %. La del período 06/09/2012 0:00:00 - 08/09/2012 23:00:00 presenta un RMSPE de más del 700 % de manera que de media los valores predichos son 7 veces, mayores o menores que los reales. Si se observa los valores de las métricas medias se puede ver que el RMSPE es del 164 % con un RMSE de 1496 kWh. El RMSE no es muy elevado teniendo en cuenta que se trata del valor acumulado en 72 horas, pero el RMSPE sí que es grande.

RESULTADOS DE PREDICCIÓN: OTOÑO

Se realiza la predicción de la generación eléctrica del parque eólico de Es Milá para un período comprendido entre el 11/12/2011 00:00:00-20/12/2011 23:00:00 para horizontes de 24 horas, 48 h y 72 h. A continuación, se expone los resultados obtenidos para cada horizonte.

PREDICIONES PARA UN HORIZONTE DE 24 HORAS

GRÁFICOS

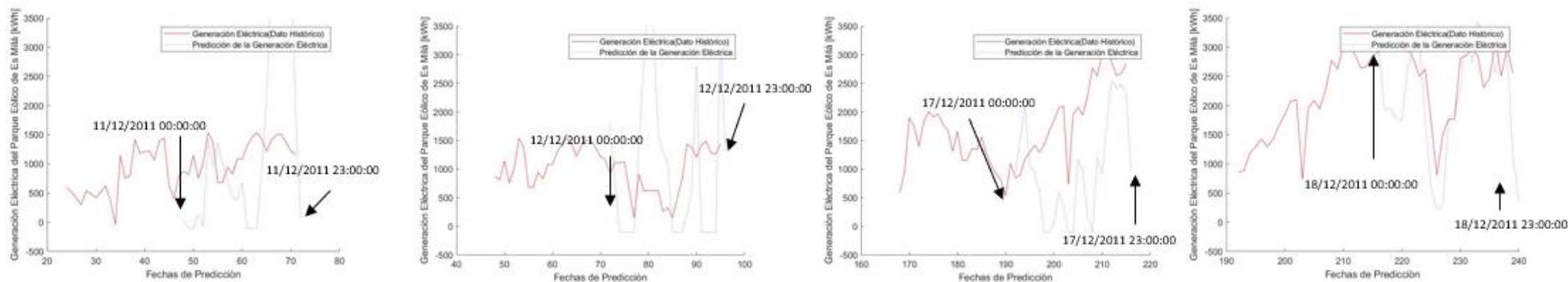


Figura 78: Predicciones de la generación eléctrica de Es Milá para horizontes de 24 horas. Otoño. Fuente: Elaboración propia.

MÉTRICAS DE ERROR

Tabla 21: Métricas de error de la fase de predicción para un horizonte de 24 h en otoño. Fuente: Elaboración propia

Período de predicción	MSE	NMSE	R ²	MAPE (%)	MAE (kWh)	RMSE (kWh)	RMSPE (%)
11/12/2011 00:00:00 - 11/12/2011 23:00:00	3594498.5	46.6	-45.6	111.4	1373.7	1895.9	14.1
12/12/2011 0:00:00 - 12/12/2011 23:00:00	1988529.1	10.2	-9.2	162.3	1187.7	1410.2	20.8
13/12/2011 0:00:00 - 13/12/2011 23:00:00	1450710.0	5.5	-4.5	49610723971984800.0	996.8	1204.5	15149983441190000.0
14/12/2011 0:00:00 - 14/12/2011 23:00:00	2875843.5	10.7	-9.7	83966173249502100.0	1125.6	1695.8	14212127044306800.0
15/12/2011 0:00:00 - 15/12/2011 23:00:00	819800.6	21.4	-20.4	313.4	662.7	905.4	66.2
16/12/2011 0:00:00 - 16/12/2011 23:00:00	1517568.9	8.2	-7.2	82.0	1096.4	1231.9	8.9
17/12/2011 0:00:00 - 17/12/2011 23:00:00	1897766.5	3.6	-2.6	60.5	1161.5	1377.6	7.0
18/12/2011 0:00:00 - 18/12/2011 23:00:00	906261.1	2.6	-1.6	31.3	753.4	952.0	4.0
19/12/2011 0:00:00 - 19/12/2011 23:00:00	5188349.4	5.1	-4.1	1776810790486990.0	1686.3	2277.8	628197479419049.0
20/12/2011 0:00:00 - 20/12/2011 23:00:00	3594498.5	46.6	-45.6	111.4	1373.7	1895.9	14.1

Tabla 22: Medias de las métricas de error de la fase de predicción para horizontes de 24 horas en otoño. Fuente: Elaboración propia

MSE	NMSE	R ²	MAPE (kWh)	MAE	RMSE (kWh)	RMSPE (%)
2248814.2	12.7	-11.7	15039300890219400.0	1116.0	1439.0	3332256440546230.0

Observando la Figura 78 se puede llegar a la conclusión que las dos primeras predicciones no han sido demasiado precisas debido a que la curva de predicción y la curva de la generación real son bastante diferentes, sin embargo si se analiza los RMSE y RMSPE se puede ver que solo presentan un error del 14 y 20 % respectivamente, y aunque los RMSE son grandes se puede considerar que la calidad de las predicciones es buena. En cuanto a las predicciones expuestas para los días 17 y 18 de diciembre del 2011, hay que destacar la del 17 ya que presenta un RMSPE de solo un 7 %. La predicción de día 19 es buena al solo presentar una desviación de un 4 % con un RMSE acumulado de 952 kWh. Si se observa las métricas medias se aprecia cómo tanto el RMSPE y el MAPE divergen en algún punto y por lo tanto no se pueden interpretar correctamente, y en cuanto al RMSE presenta un valor medio de 1439 kWh, un valor relativamente elevado.

PREDICIONES PARA UN HORIZONTE DE 48 HORAS

GRÁFICOS

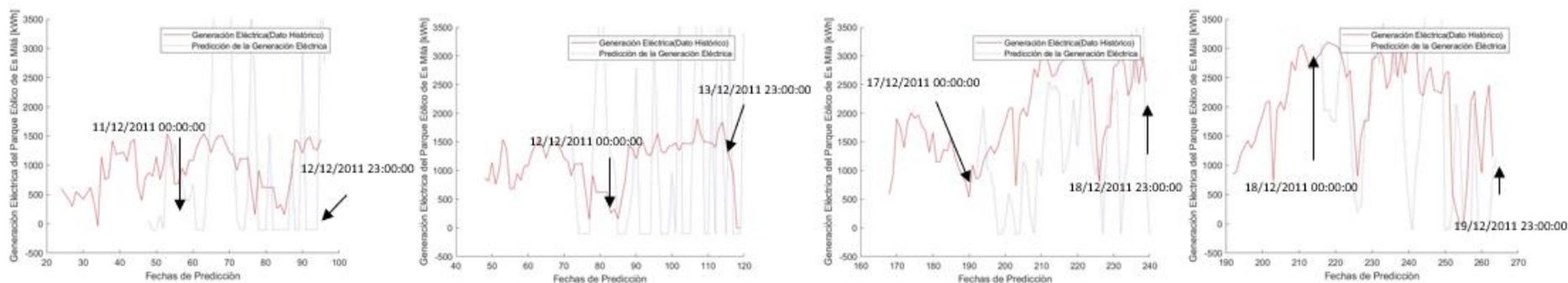


Figura 79: Predicciones de la generación eléctrica de Es Milá para horizontes de 48 horas. Otoño. Fuente: Elaboración propia.



MÉTRICAS DE ERROR

Tabla 23: Métricas de error de la fase de predicción para un horizonte de 48 h en otoño. Fuente: Elaboración propia

Período de predicción	MSE	NMSE	R ²	MAPE (%)	MAE (kWh)	RMSE (kWh)	RMSPE (%)
11/12/2011 00:00:00 - 12/12/2011 23:00:00	2811786.9	18.5	-17.5	147.9	1294.8	1676.8	25.3
12/12/2011 0:00:00 - 13/12/2011 23:00:00	2977365.9	11.2	-10.2	16186324822593500.0	1415.1	1725.5	10705260150994400.0
13/12/2011 0:00:00 - 14/12/2011 23:00:00	1244201.4	2.7	-1.7	57488369498839600.0	899.4	1115.4	14648520053541400.0
14/12/2011 0:00:00 - 15/12/2011 23:00:00	2117851.8	13.6	-12.6	42839884310970700.0	940.4	1455.3	10151519317362000.0
15/12/2011 0:00:00 - 16/12/2011 23:00:00	1711537.7	4.8	-3.8	210.5	1058.2	1308.3	47.9
16/12/2011 0:00:00 - 17/12/2011 23:00:00	2215882.2	4.8	-3.8	76.5	1273.5	1488.6	8.4
17/12/2011 0:00:00 - 18/12/2011 23:00:00	2147050.3	4.2	-3.2	55.2	1210.8	1465.3	6.5
18/12/2011 0:00:00 - 19/12/2011 23:00:00	2808988.2	3.3	-2.3	7507196806450000.0	1266.4	1676.0	4947933682653120.0
19/12/2011 0:00:00 - 20/12/2011 23:00:00	3888678.8	5.3	-4.3	906536117595459.0	1530.6	1972.0	448712485299320.0

Tabla 24: Medias de las métricas de error de la fase de predicción para horizontes de 48 horas en otoño. Fuente: Elaboración propia

MSE	NMSE	R ²	MAPE (kWh)	MAE	RMSE (kWh)	RMSPE (%)
2435927.0	7.6	-6.6	13880923506272200.0	1209.9	1542.6	4544660632205590.0

Si se compara los dos primeros gráficos presentados en la Figura 79 con los expuestos en la Figura 78 se llega a la conclusión que durante esos períodos hay datos incoherentes que provocan que la predicción diverja y no proporcione resultados correctos. Por otro lado, las predicciones para los períodos del 17-18 de diciembre y del 18-19 siguen siendo muy buenas a pesar de que se haya aumentado el horizonte de predicción.

PREDICIONES PARA UN HORIZONTE DE 72 HORAS

GRÁFICOS

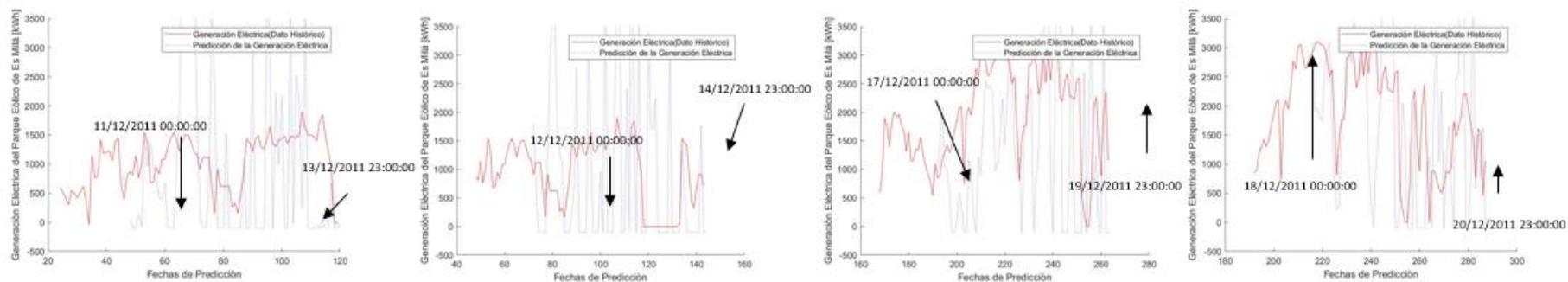


Figura 80: Predicciones de la generación eléctrica de Es Milá para horizontes de 72 horas. Otoño. Fuente: Elaboración propia.

MÉTRICAS DE ERROR

Tabla 25: Métricas de error de la fase de predicción para un horizonte de 72 h en otoño. Fuente: Elaboración propia

Período de predicción	MSE	NMSE	R ²	MAPE (%)	MAE (kWh)	RMSE (kWh)	RMSPE (%)
11/12/2011 00:00:00 - 12/12/2011 23:00:00	2493441.8	12.2	-11.2	668612255096435.0	1263.6	1579.1	371195487326049.0
12/12/2011 0:00:00 - 13/12/2011 23:00:00	2398149.2	6.4	-5.4	40024192654514600.0	1201.9	1548.6	14756618792837700.0
13/12/2011 0:00:00 - 14/12/2011 23:00:00	2377402.3	6.2	-5.2	38588083636208000.0	1091.9	1541.9	12001357142544100.0
14/12/2011 0:00:00 - 15/12/2011 23:00:00	2545165.0	6.6	-5.6	28755538784076200.0	1157.1	1595.4	8317018266814120.0
15/12/2011 0:00:00 - 16/12/2011 23:00:00	2698701.2	3.8	-2.8	173.9	1347.4	1642.8	39.9
16/12/2011 0:00:00 - 17/12/2011 23:00:00	3267981.9	5.6	-4.6	79.0	1547.9	1807.8	8.6
17/12/2011 0:00:00 - 18/12/2011 23:00:00	2671017.1	3.6	-2.6	6246282527728720.0	1373.0	1634.3	5083526883489900.0
18/12/2011 0:00:00 - 19/12/2011 23:00:00	2802287.0	3.3	-2.3	5039077308439080.0	1295.2	1674.0	4053782841276670.0
19/12/2011 0:00:00 - 20/12/2011 23:00:00	3323065.2	4.5	-3.5	608496846057325.0	1427.4	1822.9	367624768284615.0

Tabla 26: Medias de las métricas de error de la fase de predicción para horizontes de 72 horas en otoño. Fuente: Elaboración propia

MSE	NMSE	R ²	MAPE (kWh)	MAE	RMSE (kWh)	RMSPE (%)
2730801.2	5.8	-4.8	13325587112457800.0	1300.6	1649.6	4994569353619230.0

Al realizarse la predicción utilizando prácticamente los mismos datos problemáticos que en las predicciones para horizontes de 24 horas y 48 horas se obtiene resultados incluso peores al haber aumentado el horizonte de predicción. Los resultados obtenidos demuestran la necesidad de disponer de datos de calidad para poder obtener resultados adecuados en la previsión de series temporales.

RESULTADOS DE PREDICCIÓN: PRIMAVERA

Se realiza la predicción de la generación eléctrica del parque eólico de Es Milá para un período comprendido entre el 28/03/2013 00:00:00 - 11/04/2013 23:00:00 para horizontes de 24 horas, 48 h y 72 h. A continuación, se expone los resultados obtenidos para cada horizonte.

PREDICIONES PARA UN HORIZONTE DE 24 HORAS

GRÁFICOS

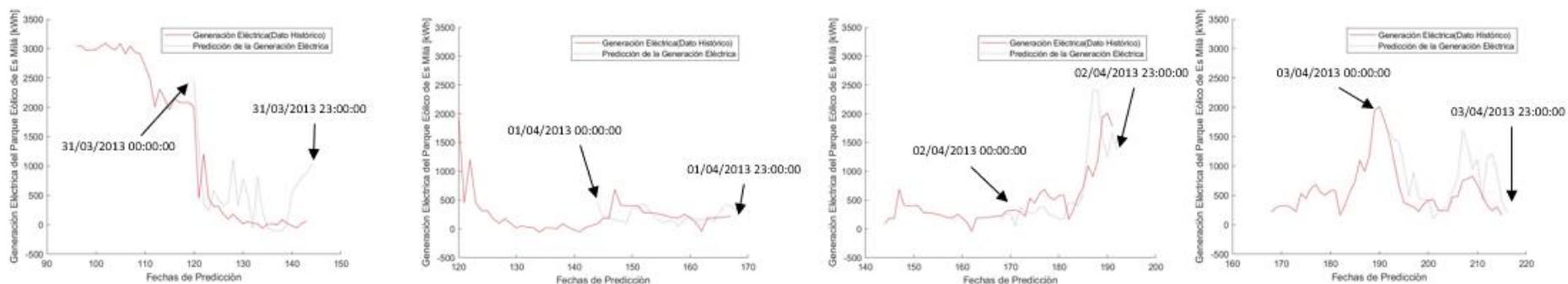


Figura 81: Predicciones de la generación eléctrica de Es Milá para horizontes de 24 horas. Primavera. Fuente: Elaboración propia.



MÉTRICAS DE ERROR

Tabla 27: Métricas de error de la fase de predicción para un horizonte de 24 h en primavera. Fuente: Elaboración propia

Período de predicción	MSE	NMSE	R ²	MAPE (%)	MAE (kWh)	RMSE (kWh)	RMSPE (%)
28/03/2013 00:00:00 - 28/03/2013 23:00:00	154611.9	0.7	0.3	2558313438061550.0	323.6	393.2	1279156719030610.0
29/03/2013 0:00:00 - 29/03/2013 23:00:00	2983468.9	1.7	-0.7	198662919033275000.0	1429.2	1727.3	32408955573975900.0
30/03/2013 0:00:00 - 30/03/2013 23:00:00	5334021.6	28.4	-27.4	84.8	2167.5	2309.6	9.0
31/03/2013 0:00:00 - 31/03/2013 23:00:00	301146.2	1.5	-0.5	296135131748622.0	443.4	548.8	148067565873833.0
01/04/2013 0:00:00 - 01/04/2013 23:00:00	31781.9	1.8	-0.8	74.3	128.4	178.3	14.0
02/04/2013 0:00:00 - 02/04/2013 23:00:00	227028.0	0.8	0.2	50.5	328.3	476.5	6.6
03/04/2013 0:00:00 - 03/04/2013 23:00:00	199015.7	2.0	-1.0	87.8	339.2	446.1	12.6
04/04/2013 0:00:00 - 04/04/2013 23:00:00	103843.9	4.6	-3.6	128.7	245.2	322.2	19.0
05/04/2013 0:00:00 - 05/04/2013 23:00:00	135536.2	4.4	-3.4	526.9	283.5	368.2	85.3
06/04/2013 0:00:00 - 06/04/2013 23:00:00	31034.8	1.9	-0.9	354.6	131.6	176.2	89.4
07/04/2013 0:00:00 - 07/04/2013 23:00:00	81231.6	22.1	-21.1	2578459423477030.0	242.3	285.0	1289229711738020.0
08/04/2013 0:00:00 - 08/04/2013 23:00:00	7493.2	21.9	-20.9	15410755502600800.0	77.8	86.6	1834980968341970.0
09/04/2013 0:00:00 - 09/04/2013 23:00:00	88521.1	1.1	-0.1	124686974275235.0	235.7	297.5	62343487137547.0

Tabla 28: Medias de las métricas de error de la fase de predicción para horizontes de 24 horas en primavera. Fuente: Elaboración propia

MSE	NMSE	R ²	MAPE (kWh)	MAE	RMSE (kWh)	RMSPE (%)
744518.1	7.1	-6.1	16894713038726100.0	490.5	585.8	2847902617392170.0

Observando la Figura 81 se aprecia como las curvas de predicción se ajustan muy bien a la generación real que hubo en el parque eólico de Es Milá en los diferentes períodos expuestos. Analizando las métricas de error de la tabla 27 se puede ver que los RMSPE obtenidos para las predicciones de los períodos expuestos son muy buenos, excepto para las predicciones del día 31 de marzo. Si se analizan los resultados obtenidos en la tabla anterior, prácticamente la mitad de las predicciones presentan valores enormes del RMSPE. Estos se deben a la existencia de algún instante de predicción en el cual el valor predicho es muy cercano a cero, y al dividir entre este valor se incrementa no linealmente el RMSPE. Considerando los valores medios obtenidos, se concluye que a pesar de disponer de estos valores de RMSPE ruidosos, el modelo ha realizado predicciones relativamente buenas teniendo en cuenta que el error acumulado medio RMSE es de unos 585,8 kWh y que en más del 40 % de las predicciones se ha obtenido un RMSPE inferior al 30 %.

PREDICIONES PARA UN HORIZONTE DE 48 HORAS

GRÁFICOS

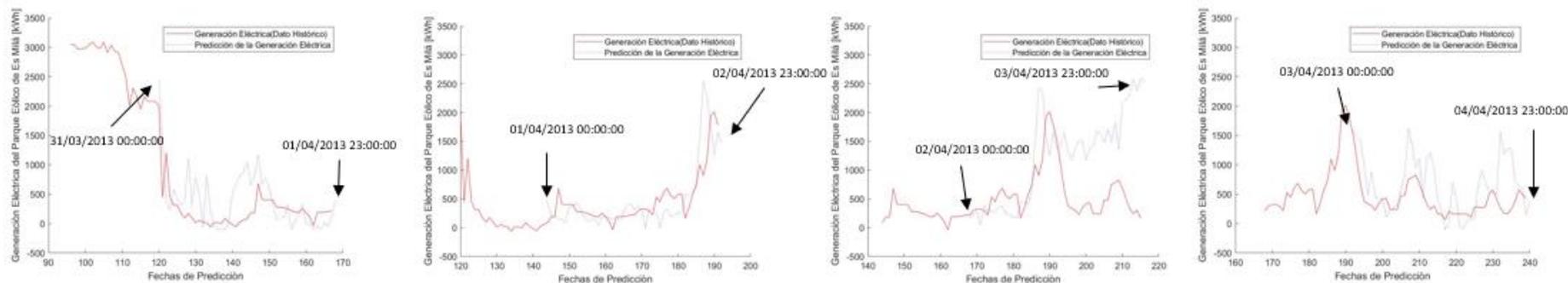


Figura 82: Predicciones de la generación eléctrica de Es Milá para horizontes de 48 horas. Primavera. Fuente: Elaboración propia.

MÉTRICAS DE ERROR

Tabla 29: Métricas de error de la fase de predicción para un horizonte de 48 h en primavera. Fuente: Elaboración propia

Período de predicción	MSE	NMSE	R ²	MAPE (%)	MAE (kWh)	RMSE (kWh)	RMSPE (%)
28/03/2013 00:00:00 - 29/03/2013 23:00:00	1262432.4	1.2	-0.2	91524590346246800.0	816.0	1123.6	21106000409383600.0
29/03/2013 0:00:00 - 30/03/2013 23:00:00	4780182.4	3.1	-2.1	101358632159834000.0	1965.3	2186.4	23149253981411400.0
30/03/2013 0:00:00 - 31/03/2013 23:00:00	3640938.1	2.2	-1.2	1252868837985300.0	1628.3	1908.1	877008186588605.0
31/03/2013 0:00:00 - 01/04/2013 23:00:00	187972.8	1.7	-0.7	151089352933019.0	334.5	433.6	105762547052738.0
01/04/2013 0:00:00 - 02/04/2013 23:00:00	137841.6	0.6	0.4	66.0	233.7	371.3	11.4
02/04/2013 0:00:00 - 03/04/2013 23:00:00	1104201.1	5.7	-4.7	239.1	804.7	1050.8	40.7
03/04/2013 0:00:00 - 04/04/2013 23:00:00	233899.3	3.3	-2.3	129.9	355.6	483.6	19.8
04/04/2013 0:00:00 - 05/04/2013 23:00:00	210212.2	7.3	-6.3	460.9	373.9	458.5	87.4
05/04/2013 0:00:00 - 06/04/2013 23:00:00	110027.1	3.7	-2.7	617.3	250.5	331.7	131.1
06/04/2013 0:00:00 - 07/04/2013 23:00:00	147585.9	14.1	-13.1	3702789864264190.0	287.6	384.2	2591952904984000.0
07/04/2013 0:00:00 - 08/04/2013 23:00:00	145330.3	69.5	-68.5	40502219210501500.0	344.6	381.2	6311342948422560.0
08/04/2013 0:00:00 - 09/04/2013 23:00:00	63246.2	1.0	0.0	7862630358469830.0	159.4	251.5	1310700691672830.0
09/04/2013 0:00:00 - 10/04/2013 23:00:00	100209.9	1.7	-0.7	63615803202535.2	265.5	316.6	44531062241105.0

Tabla 30: Medias de las métricas de error de la fase de predicción para horizontes de 48 horas en primavera. Fuente: Elaboración propia

MSE	NMSE	R ²	MAPE (kWh)	MAE	RMSE (kWh)	RMSPE (%)
2435927.0	7.6	-6.6	13880923506272200.0	1209.9	1542.6	4544660632205590.0

Analizando los resultados de predicción obtenidos se observa que al aumentar el horizonte de predicción a dos días aumenta tanto el error acumulado RMSE como el RMSPE. Hay que destacar la variación que ha experimentado el RMSPE cuando se realiza la predicción empezando el día 02/04. Para un horizonte de 24 horas se obtiene una variación del 9 %, un resultado bueno. En cambio, para un horizonte de 48 horas el RMSPE es del 40 %, hecho que nos indica que el modelo de predicción ha tenido problemas para predecir la generación eléctrica del parque para el día 3 en el cual hay una disminución abrupta de la energía producida, reflejando que el parque debe haber tenido algún problema de funcionamiento o ha dejado de soplar el viento. En términos generales las predicciones obtenidas presentan cierta calidad teniendo en cuenta que solo hay un error acumulado medio de 1542 kWh en 48 horas de producción.

PREDICIONES PARA UN HORIZONTE DE 72 HORAS

GRÁFICOS

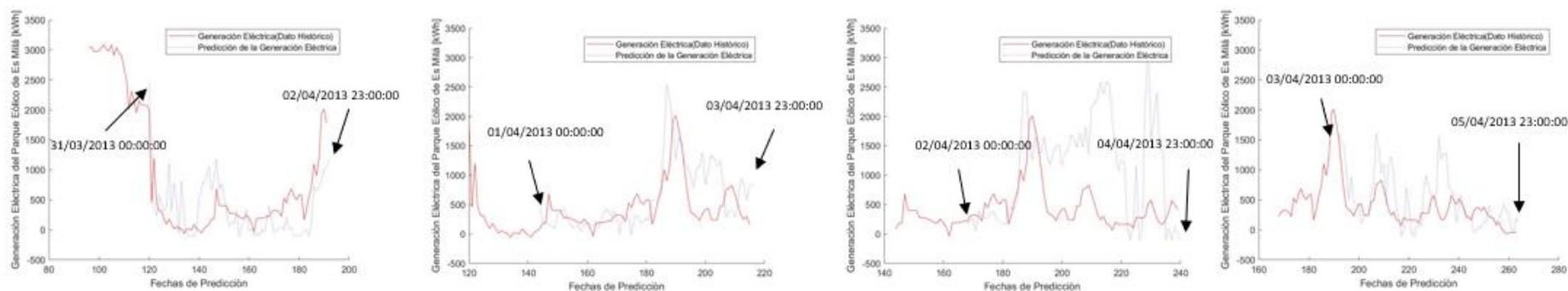


Figura 83: Predicciones de la generación eléctrica de Es Milá para horizontes de 72 horas. Primavera. Fuente: Elaboración propia.



MÉTRICAS DE ERROR

Tabla 31: Métricas de error de la fase de predicción para un horizonte de 72 h en primavera. Fuente: Elaboración propia

Período de predicción	MSE	NMSE	R ²	MAPE (%)	MAE (kWh)	RMSE (kWh)	RMSPE (%)
28/03/2013 00:00:00 - 30/03/2013 23:00:00	2779597.1	2.0	-1.0	61434314068028700.0	1316.6	1667.2	17291893504453800.0
29/03/2013 0:00:00 - 31/03/2013 23:00:00	3730877.7	2.2	-1.2	68451554533323400.0	1614.2	1931.5	18969241816905600.0
30/03/2013 0:00:00 - 01/03/2013 23:00:00	3096956.8	2.2	-1.2	840966754264268.0	1474.8	1759.8	718522309811099.0
31/03/2013 0:00:00 - 02/03/2013 23:00:00	226828.1	1.0	0.0	101416141009861.0	387.4	476.3	86649988862060.0
01/04/2013 0:00:00 - 03/04/2013 23:00:00	234187.6	1.4	-0.4	108.4	352.3	483.9	16.8
02/04/2013 0:00:00 - 04/04/2013 23:00:00	1192328.7	7.5	-6.5	306.3	849.8	1091.9	52.9
03/04/2013 0:00:00 - 05/04/2013 23:00:00	173638.3	2.6	-1.6	162.9	296.9	416.7	29.4
04/04/2013 0:00:00 - 06/04/2013 23:00:00	170393.3	5.1	-4.1	536.9	323.8	412.8	109.2
05/04/2013 0:00:00 - 07/04/2013 23:00:00	212281.0	9.1	-8.1	2833874037603910.0	363.0	460.7	2421263039103240.0
06/04/2013 0:00:00 - 08/04/2013 23:00:00	270462.0	36.9	-35.9	44072336126772600.0	391.2	520.1	9301665969973090.0
07/04/2013 0:00:00 - 09/04/2013 23:00:00	213308.9	4.6	-3.6	27186421113898500.0	400.6	461.9	5170807733221100.0
08/04/2013 0:00:00 - 10/04/2013 23:00:00	140589.5	3.0	-2.0	5277655994042320.0	265.7	375.0	1073841388089050.0
09/04/2013 0:00:00 - 11/04/2013 23:00:00	79144.4	1.7	-0.7	42701018588272.7	227.3	281.3	36483766273928.6

Tabla 32: Medias de las métricas de error de la fase de predicción para horizontes de 72 horas en primavera. Fuente: Elaboración propia

MSE	NMSE	R ²	MAPE (kWh)	MAE	RMSE (kWh)	RMSPE (%)
963122.6	6.1	-5.1	16172403060579500.0	635.7	795.3	4236182270514860.0

Observando los resultados se puede ver que los RMSPE han aumentado y como se ha acentuado los problemas que ha tenido el modelo de predicción para prever la generación producida a partir de la caída brusca de generación de día 3. En términos generales, la calidad de la predicción es relativamente buena todavía, pero se ha reflejado con claridad como aumenta el error cuando el horizonte de predicción se hace mayor.

RESULTADOS DE PREDICCIÓN: INVIERNO

Se realiza la predicción de la generación eléctrica del parque eólico de Es Milá para un período comprendido entre el 24/12/2011 00:00:00 - 29/12/2011 23:00:00 para horizontes de 24 horas, 48 h y 72 h. Hay que destacar que se dispone de una cantidad muy pequeña de datos para la estación de invierno, de manera que el período de predicción será más corto que para las demás estaciones. A continuación, se expone los resultados obtenidos para cada horizonte de previsión.

PREDICCIÓNES PARA UN HORIZONTE DE 24 HORAS

GRÁFICOS

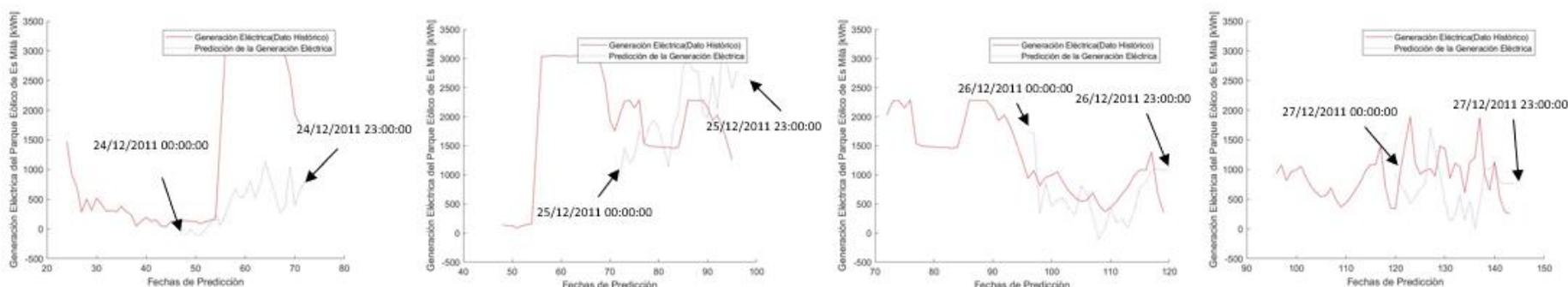


Figura 84: Predicciones de la generación eléctrica de Es Milá para horizontes de 24 horas. Invierno. Fuente: Elaboración propia.

MÉTRICAS DE ERROR

Tabla 33: Métricas de error de la fase de predicción para un horizonte de 24 h en invierno. Fuente: Elaboración propia

Período de predicción	MSE	NMSE	R ²	MAPE (%)	MAE (kWh)	RMSE (kWh)	RMSPE (%)
24/12/2011 00:00:00 - 24/12/2011 23:00:00	3505335.5	2.3	-1.3	90.2	1584.7	1872.3	9.9
25/12/2011 0:00:00 - 25/12/2011 23:00:00	632133.4	4.1	-3.1	39.9	652.4	795.1	5.6
26/12/2011 0:00:00 - 26/12/2011 23:00:00	218515.8	3.0	-2.0	64.1	417.4	467.5	8.2
27/12/2011 0:00:00 - 27/12/2011 23:00:00	475636.5	2.6	-1.6	69.3	582.0	689.7	8.4

Tabla 34: Medias de las métricas de error de la fase de predicción para horizontes de 24 horas en invierno. Fuente: Elaboración propia

MSE	NMSE	R ²	MAPE (kWh)	MAE	RMSE (kWh)	RMSPE (%)
1207905.3	3.0	-2.0	65.9	809.2	956.1	8.0

Observando la Figura 84 se observa como las curvas de predicción para los días 24 y 25 no se ajustan prácticamente nada a la generación real, sobre todo la de día 24 que presenta un RMSE acumulado de 1872 kWh, es decir un error de unos 80 kWh por hora. A pesar de ello, el RMSPE indica que la desviación media es de tan solo el 10 %. Las curvas de predicción para los días 26 y 27 se ajustan mucho mejor, y consecuentemente se tiene un RMSE y un RMSPE muy inferior. En términos generales se ha obtenido un RMSPE del 8 % con un error acumulado medio de 956 kWh. Son los mejores resultados que se han obtenido para horizontes de 24 horas.

PREDICIONES PARA UN HORIZONTE DE 48 HORAS

GRÁFICOS

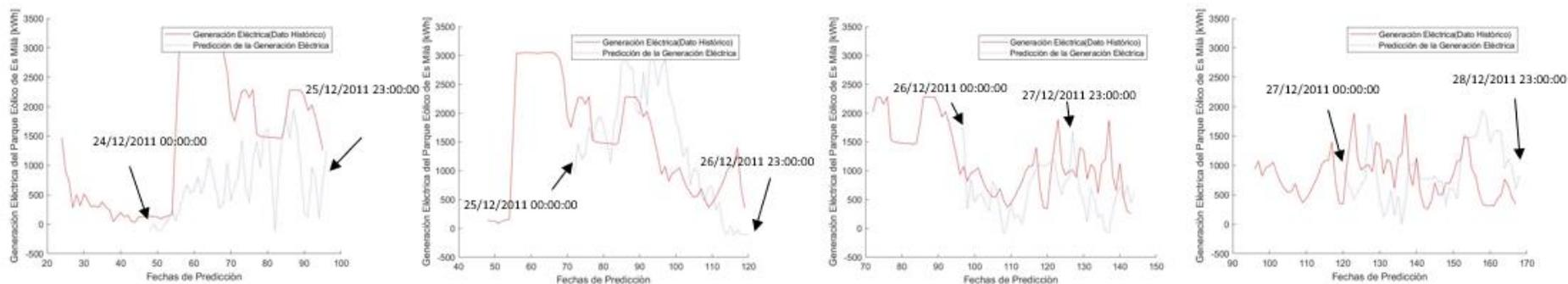


Figura 85: Predicciones de la generación eléctrica de Es Milá para horizontes de 48 horas. Invierno. Fuente: Elaboración propia.

MÉTRICAS DE ERROR

Tabla 35: Métricas de error de la fase de predicción para un horizonte de 48 h en invierno. Fuente: Elaboración propia

Período de predicción	MSE	NMSE	R ²	MAPE (%)	MAE (kWh)	RMSE (kWh)	RMSPE (%)
24/12/2011 00:00:00 - 25/12/2011 23:00:00	2325151.6	2.6	-1.6	68.2	1224.5	1524.8	8.0
25/12/2011 0:00:00 - 26/12/2011 23:00:00	675141.6	1.7	-0.7	63.5	673.7	821.7	8.0
26/12/2011 0:00:00 - 27/12/2011 23:00:00	346777.3	2.6	-1.6	59.1	475.3	588.9	7.4
27/12/2011 0:00:00 - 28/12/2011 23:00:00	520335.7	3.0	-2.0	103.0	582.7	721.3	15.4

Observando los resultados se puede ver a simple vista que las predicciones realizadas son bastante buenas. Si se comparan con los obtenidos para el horizonte de 24 horas se puede decir que se ha obtenido casi mejores resultados para este horizonte, aunque sea más lejano. Los RMSE y los RMSPE se mantienen muy bajos, aunque en términos generales aumenta levemente el RMSPE y disminuye el RMSE.

Tabla 36: Medias de las métricas de error de la fase de predicción para horizontes de 48 horas en invierno. Fuente: Elaboración propia

MSE	NMSE	R ²	MAPE (kWh)	MAE	RMSE (kWh)	RMSPE (%)
966851.5	2.5	-1.5	73.4	739.0	914.2	9.7

PREDICIONES PARA UN HORIZONTE DE 72 HORAS

GRÁFICOS

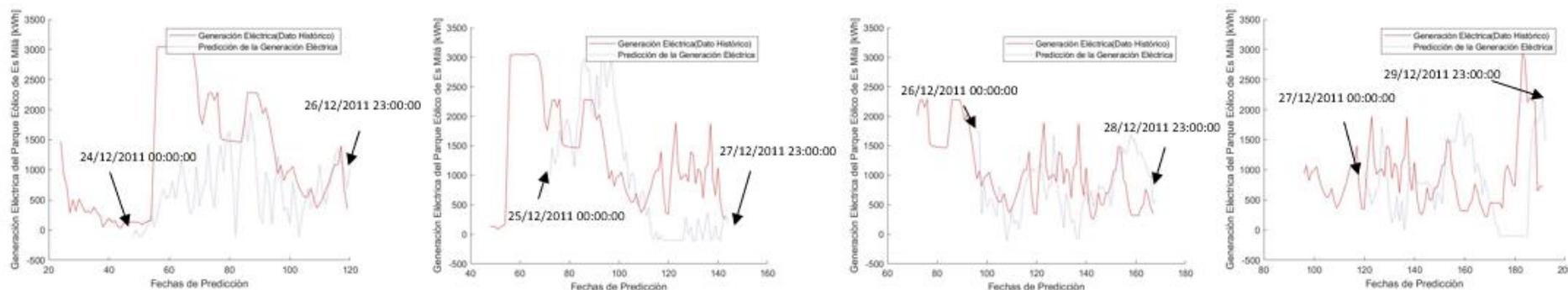


Figura 86: Predicciones de la generación eléctrica de Es Milá para horizontes de 72 horas. Invierno. Fuente: Elaboración propia.



MÉTRICAS DE ERROR

Tabla 37: Métricas de error de la fase de predicción para un horizonte de 72 h en invierno. Fuente: Elaboración propia

Período de predicción	MSE	NMSE	R ²	MAPE (%)	MAE (kWh)	RMSE (kWh)	RMSPE (%)
24/12/2011 00:00:00 - 26/12/2011 23:00:00	1616176.3	1.8	-0.8	62.9	932.2	1271.3	7.7
25/12/2011 0:00:00 - 27/12/2011 23:00:00	836488.6	2.4	-1.4	72.6	769.7	914.6	8.5
26/12/2011 0:00:00 - 28/12/2011 23:00:00	371226.7	2.7	-1.7	78.5	483.0	609.3	11.8
27/12/2011 0:00:00 - 29/12/2011 23:00:00	941585.1	2.6	-1.6	102.1	740.9	970.4	14.2

Tabla 38: Medias de las métricas de error de la fase de predicción para horizontes de 72 horas en invierno. Fuente: Elaboración propia

MSE	NMSE	R ²	MAPE (kWh)	MAE	RMSE (kWh)	RMSPE (%)
941369.2	2.4	-1.4	79.0	731.4	941.4	10.5

Observando los resultados de predicción obtenidos para un horizonte de 72 horas se puede afirmar que los resultados siguen siendo muy buenos. El error medio porcentual máximo que se ha obtenido es de un 14,2 %, prácticamente la mitad del que se suele obtener en aplicaciones de energía eólica (valores del 30% son habituales). El error medio acumulado máximo de energía ha sido de 1271.3 kWh, un valor pequeño teniendo en cuenta la duración del horizonte de predicción.



CAPÍTULO 9: CONCLUSIONES Y TRABAJO FUTURO

En este capítulo se analiza los conocimientos que se han adquirido a lo largo de la realización del proyecto, a la vez que se evalúa el grado de consecución de los objetivos inicialmente establecidos para el proyecto. Finalmente se exponen algunas propuestas para posibles trabajos futuros relacionados con el presente TFG.

9.1 CONCLUSIONES

Durante el abordaje de este proyecto se ha puesto en escena la complejidad del desarrollo de una aplicación de predicción para la energía eólica, y la gran cantidad de tiempo que requiere. La primera fase de este trabajo ha consistido en adquirir los conocimientos teóricos necesarios para desarrollar el modelo de predicción. El primer paso realizado ha consistido en obtener conocimientos básicos sobre la energía eólica, por ejemplo, cómo se obtiene este tipo de energía, la historia del aprovechamiento de la energía eólica, el concepto del viento y las causas de su variabilidad, los tipos de aerogeneradores que hay, etc. A continuación, se ha estudiado el estado del arte de los modelos de predicción para la energía eólica para hacerse una idea del tipo de modelo que podría adaptarse mejor a las series temporales de datos que se tenía a disposición. Una vez decidido que el modelo de predicción a elaborar se basaría en el uso de redes neuronales recurrentes se ha tenido que adquirir conocimientos sobre las bases del funcionamiento de estas estructuras como, por ejemplo, entender el concepto de neurona artificial, las diferentes arquitecturas de redes que existen, las posibles técnicas de aprendizaje y algoritmos de entrenamiento que pueden usar, etc. Finalmente, se ha analizado diferentes metodologías de pre-procesamiento y ciencia de datos con el objetivo de establecer un guion aproximado del tratamiento de los datos a realizar, para conseguir los mejores resultados de predicción posibles, a la vez que estos se realizaban de forma eficiente en términos computacionales.

Una vez adquirida la base teórica necesaria se inicia la etapa de pre-procesamiento de las series temporales de datos del parque eólico. Durante la realización de esa etapa se ha podido observar lo complejo que resulta generar una base de datos de calidad a partir de archivos históricos de datos que presentan errores de formato, huecos temporales, diferentes tamaños, errores de medida; y todo ello sin disponer de una información detallada sobre el origen de los datos. Cada fase del pre-procesamiento ha presentado una serie de dificultades, agravadas por la baja calidad de los datos. Se ha rellenado los huecos temporales, unido y transformado diferentes hojas de cálculo y se ha realizado una fase de filtrado de los datos con el objetivo de eliminar valores ruidosos. Una vez filtrados los datos estos se han clasificado según la estación anual y al detectar la gran cantidad de huecos temporales se decidió regenerar las series temporales. Esta etapa ha ocupado la mayor parte del TFG.



Finalizada la fase de pre-procesamiento se ha iniciado la etapa de elección o diseño del modelo de predicción. Al tener conocimiento de la baja calidad de los datos se ha decidido elaborar un modelo de predicción basado en redes neuronales recurrentes del tipo NARX, debido a la capacidad de este tipo de redes de proporcionar resultados relativamente buenos a partir de datos ruidosos. Con el objetivo de incorporar las relaciones estacionales de los datos, se ha generado un modelo basado en cuatro redes neuronales, dónde cada una ha sido entrenada con datos de una estación diferente. Posteriormente, dependiendo de la estación a la que pertenece el período que se quiere predecir se utiliza una u otra NARX entrenada para realizar la predicción.

Como se refleja en el apartado de resultados se aprecia como la variabilidad del viento, y la baja calidad de los datos provoca que en función de los períodos de predicción y la respuesta que tuvo el parque, se obtenga predicciones de buena calidad o per el contrario de baja calidad. También se ha mostrado como se incrementan los errores a medida que aumenta el horizonte de predicción, así como la necesidad de disponer de una gran cantidad de datos para la fase de entrenamiento. Se ha hecho constancia de este hecho al analizar los resultados de entrenamiento, y percatarse que la red neuronal que ha sido entrenada con los datos de primavera ha proporcionado mejores resultados, al disponer de una mayor cantidad de datos. Los peores resultados se han obtenido al realizar predicciones para períodos de otoño, estación para la cual se disponía de una menor cantidad de datos para entrenar a la red neuronal.

En términos generales, los resultados de predicción obtenidos son poco alentadores al obtener en muchos casos desviaciones de más del 50%, para horizontes de previsión mayores a 24 horas, cuando los resultados de las aplicaciones comerciales obtienen de media unos resultados con desviaciones máximas del 30%. Ahora bien, si analizamos los resultados de la metodología desarrollada para el horizonte de previsión más lejano analizado (72h), peor caso posible, esta obtiene resultados relativamente buenos para las estaciones de verano (RMSPE del 164% y RMSE de 1496 kWh) e invierno (RMSPE del 14,2% y RMSE de 1271,3 kWh), mientras que para las estaciones de transición (primavera y otoño) los resultados obtenidos presentan gran número de resultados espurios, debido fundamentalmente a la calidad de los datos que se disponen para este periodo y a la gran variabilidad meteorológica, que dificultan en gran medida poder dispone de valores medios para comparar los resultados con las otras estaciones del año.

Probablemente, con datos históricos de mayor calidad y con una mayor información sobre el origen de los datos, e incorporando una serie de mejoras seguramente se podría obtener resultados cercanos a la desviación típica del 30 % que obtienen algunos de los modelos de predicción de la energía eólica que se utilizan hoy en día comercialmente. Pese a este punto, se puede considerar que se ha cumplido el objetivo principal inicialmente establecido para este TFG; al obtener algunos resultados bastante aceptables pese a la calidad de la información de la que se ha dispuesto.

En cuanto a los objetivos específicos cabe remarcar que se han cumplido satisfactoriamente. Se han asimilado los fundamentos teóricos que rigen los procesos atmosféricos, se ha investigado sobre el estado del arte de la energía eólica a nivel global y nacional, y se ha analizado los modelos de predicción para la energía eólica más importantes que hay en la actualidad. A su vez, se ha adquirido una buena base teórica de las metodologías de pre-procesamiento de datos (data science), al haber dedicado gran parte del tiempo de realización del TFG a corregir las múltiples deficiencias que presentaban los datos aportados para el parque eólico de Es Milá. Finalmente, antes de desarrollar el modelo de predicción, se ha programado algunas redes neuronales simples con el objetivo observar



la forma en que funcionan y ver de primera mano la forma en que estas “aprenden”, y se ha analizado múltiples modelos de predicción basados en el uso de redes neuronales. Hay que destacar que aparte del modelo de predicción generado usando las funciones del Neural Network Toolbox (NNT) de MATLAB también se ha empezado la programación de una librería propia. Este trabajo opcional no se ha finalizado por varias razones, siendo la principal causa la dificultad de preparar los datos de entrada al tener que entrenar la red neuronal con múltiples archivos. El Neural Network Toolbox de MATLAB permite usar funciones como *preparets*, o *catsamples* que de forma simple permite entrenar una red neuronal con diferentes archivos, y con tamaños diferentes. Generar las librerías para replicar estas funciones requiere de una gran cantidad de tiempo, de manera que al final se ha decidido aplazar esta tarea para posibles trabajos futuros.

El desarrollo de un proyecto de esta índole requiere dedicar gran cantidad de tiempo a abordar cada una de las etapas, desde el aprendizaje teórico previo, pasando por la aplicación de las técnicas de pre-procesamiento de datos, hasta llegar a la obtención de un modelo de predicción de la generación eólica. Los modelos de predicción para la energía eólica todavía tienen un gran margen de mejora y teniendo en cuenta la relevancia que tiene este tipo de energía a nivel mundial, resulta de gran interés seguir investigando en esta dirección en el futuro para ampliar los conocimientos adquiridos, para poder ser competitivo algún día en este mercado.

9.2 POSIBLES TRABAJOS FUTUROS

1. **Mejora del modelo de predicción:** la principal limitación que presenta el modelo de predicción generado es la baja calidad de los datos. En caso de haber dispuesto de información sobre los aerogeneradores que estaban en funcionamiento en cada momento y la potencia instantánea producida por cada uno de ellos, se podría haber aplicado el filtro 3 (6.2.3 Identificación y eliminación de outliers) y consecuentemente se habría obtenido datos de mejor calidad. Por lo tanto, se trataría de obtener esta información y realizar de nuevo la etapa de pre-procesamiento tomándolos en cuenta.
2. **Mejora del modelo de predicción:** otra mejora que se podría aplicar es disgregar la generación del parque eólico a partir de la geometría del parque y la generación de cada aerogenerador teniendo en cuenta el efecto de la estela de los aerogeneradores. Se trataría de generar un “grid” sobre el cual se representarían las coordenadas geográficas de los aerogeneradores y su orientación predominante. A partir de aquí se podría proyectar la estela de forma dinámica y usando la pseudoinversa se podría ajustar los valores de la velocidad del viento a la altura del buje de los aerogeneradores que proporcionan la generación agregada del parque. Para poder realizar este bloque se requeriría de información adicional.
3. **Librería:** la programación de una librería propia que incorpora todas las funciones de preparación de datos usadas de la NNT de MATLAB y el modelo de predicción basado en el uso de redes neuronales recurrentes NARX, permitiría saber en todo momento que está realizando el programa y detectar posibles errores. La NNT permite crear redes neuronales



muy potentes de forma relativamente sencilla, pero funciona como una caja negra. Solo permite modificar ciertos parámetros y se debe usar funciones predefinidas donde el programador solo puede modificar las entradas, y, por lo tanto, es complejo conocer exactamente cómo operan a bajo nivel estas funciones.

BIBLIOGRAFIA

- [1] J.;Zha. F. Lee, "GWEC[GLOBAL WIND REPORT 2019," [Online]. Available: <https://gwec.net/global-wind-report-2019/>.
- [2] "La Comisión Europea aboga por una Europa climáticamente neutra para 2050 - Actualidad RETEMA." <https://www.retema.es/noticia/la-comision-europea-aboga-por-una-europa-climaticamente-neutra-para-2050-L2QOv> (accessed Jul. 09, 2020).
- [3] "Objetivos de reducción de emisiones de gases de efecto invernadero." <https://www.miteco.gob.es/es/cambio-climatico/temas/mitigacion-politicas-y-medidas/objetivos.aspx> (accessed Jul. 09, 2020).
- [4] "El Gobierno envía a la Comisión Europea el Plan Nacional Integrado de Energía y Clima (PNIEC) 2021-2030." [https://www.miteco.gob.es/es/prensa/ultimas-noticias/el-gobierno-envia-a-la-comision-europea-el-plan-nacional-integrado-de-energia-y-clima-\(pniec\)-2021-2030/tcm:30-508439](https://www.miteco.gob.es/es/prensa/ultimas-noticias/el-gobierno-envia-a-la-comision-europea-el-plan-nacional-integrado-de-energia-y-clima-(pniec)-2021-2030/tcm:30-508439) (accessed Jul. 09, 2020).
- [5] "La Ley balear de Cambio Climático y Transición Energética, en 3 claves: Ethic." <https://ethic.es/2019/02/ley-balear-coches-diesel/> (accessed Jul. 09, 2020).
- [6] M. Villarubia, *Ingeniería de la Energía Eólica*. Marcombo Ediciones Técnicas, 2011.
- [7] "Homosfera - Wikipedia, la enciclopedia libre." <https://es.wikipedia.org/wiki/Homosfera> (accessed Jul. 09, 2020).
- [8] "Heterosfera - Wikipedia, la enciclopedia libre." <https://es.wikipedia.org/wiki/Heterosfera> (accessed Jul. 09, 2020).
- [9] "Exosfera - Wikipedia, la enciclopedia libre." <https://es.wikipedia.org/wiki/Exosfera> (accessed Jul. 09, 2020).
- [10] "Structure of Earths Atmosphere ~ EduGeneral." <https://edugeneral.org/blog/geography/structure-of-earths-atmosphere/> (accessed Jul. 09, 2020).
- [11] "Viento." <https://es.wikipedia.org/wiki/Viento>.
- [12] "¿Por qué las regiones que reciben rayos inclinados del sol tienen climas más fríos? - Quora." <https://es.quora.com/Por-qué-las-regiones-que-reciben-rayos-inclinados-del-sol-tienen-climas-más-fríos> (accessed Jul. 09, 2020).
- [13] E. R. UIB, "The scales of wind," 2020. https://ad.uib.es/estudis1920/pluginfile.php/247260/mod_resource/content/1/vent.pdf.
- [14] S. H. Derbyshire, "Atmospheric boundary layer flows—their structure and measurement. by J. C. Kaimal and J. J. Finnigan. Oxford University press. 1994. pp. 289. price £50.00 (hardback). isbn 0 19 506239 6," *Q. J. R. Meteorol. Soc.*, 1995, doi: 10.1002/qj.49712152512.
- [15] "Rugosidad y cizallamiento del viento." <http://drømstørre.dk/wp-content/wind/miller/windpowerweb/es/tour/wres/shear.htm> (accessed Jul. 15, 2020).
- [16] "Chapter 5: Wind power 5.1. The wind resource: from General Circulation to local winds."



- [17] "Log wind profile - Wikipedia." https://en.wikipedia.org/wiki/Log_wind_profile (accessed Jun. 10, 2020).
- [18] M. Abbes and J. Belhadj, "Development of a methodology for wind energy estimation and wind park design," *J. Renew. Sustain. Energy*, vol. 6, no. 5, p. 053103, Sep. 2014, doi: 10.1063/1.4895919.
- [19] "Origen y expansión de los molinos de viento en España." http://ve.scielo.org/scielo.php?script=sci_arttext&pid=So378-18442005000600004 (accessed Jun. 11, 2020).
- [20] "History of wind power - Wikipedia." https://en.wikipedia.org/wiki/History_of_wind_power (accessed Jun. 11, 2020).
- [21] "Mecanismos para el aprovechamiento eólico: las máquinas eólicas (página 2) - Monografias.com." <https://www.monografias.com/trabajos-pdf2/mecanismos-aprovechamiento-eolico-maquinas/mecanismos-aprovechamiento-eolico-maquinas2.shtml> (accessed Jun. 11, 2020).
- [22] "23 Evolución del tamaño de las turbinas eólicas desde 1980 y prospectiva | Download Scientific Diagram." https://www.researchgate.net/figure/Figura-2-23-Evolucion-del-tamano-de-las-turbinas-eolicas-desde-1980-y-prospectiva_fig26_312304898 (accessed Jun. 11, 2020).
- [23] "Energía eólica en España - Wikipedia, la enciclopedia libre." https://es.wikipedia.org/wiki/Energía_eólica_en_España (accessed Jun. 12, 2020).
- [24] RED ELÉCTRICA DE ESPAÑA, "Las energías renovables en el sistema eléctrico español," pp. 17–26, 2019, [Online]. Available: https://www.ree.es/sites/default/files/11_PUBLICACIONES/Documentos/Renovables-2018.pdf.
- [25] C. E. Marín and R. G. Marín, "La energía eólica en la producción de electricidad en España," *Rev. Geogr. Norte Gd.*, no. 51, pp. 115–136, May 2012, doi: 10.4067/s0718-34022012000100007.
- [26] "De mercados mayoristas y sanciones por infracciones - CNMC Blog." <https://blog.cnmc.es/2019/08/28/de-mercados-mayoristas-y-sanciones-por-infracciones/> (accessed Jun. 12, 2020).
- [27] "Mercado eléctrico (III): Los servicios de ajuste del sistema eléctrico peninsular | AEQ." <https://www.aeqenergia.com/blog/mercado-electrico-iii-los-servicios-de-ajuste-del-sistema-electrico-peninsular> (accessed Jun. 16, 2020).
- [28] "2.2. La seguridad de suministro en España: sistema eléctrico | Energía y Sociedad." <http://www.energiaysociedad.es/manenergia/2-2-la-seguridad-de-suministro-en-espana-sistema-electrico/> (accessed Jun. 13, 2020).
- [29] G. W. Chang, H. J. Lu, Y. R. Chang, and Y. D. Lee, "An improved neural network-based approach for short-term wind speed and power forecast," *Renew. Energy*, vol. 105, pp. 301–311, May 2017, doi: 10.1016/j.renene.2016.12.071.
- [30] "Statistical Methods for Forecasting - Bovas Abraham, Johannes Ledolter - Google Libros." <https://books.google.es/books?hl=es&lr=&id=WIPxdb2P8sAC&oi=fnd&pg=PP2&dq=Bovas+Abraham,+Johannes+Ledolter,+Statistical+Methods+for+Forecasting+2005+Wiley->



- Interscience%3B+2+ed&ots=auMxcSzi0f&sig=jsl2Kuof1uBDjbBmns6jrZCMh6s#v=onepage&q&f=false (accessed Jun. 13, 2020).
- [31] "El coste medio global de fotovoltaica y eólica será de entre 5 y 6 centavos por kWh en 2025 – El Periodico de la Energía | El Periodico de la Energía con información diaria sobre energía eléctrica, eólica, renovable, petróleo y gas, mercados y legislación energética." <https://elperiodicodelaenergia.com/el-coste-medio-global-de-fotovoltaica-y-eolica-sera-de-entre-5-y-6-centavos-por-kwh-en-2025/> (accessed Jul. 10, 2020).
- [32] L. Landberg, "A mathematical look at a physical power prediction model," *Wind Energy*, vol. 1, no. 1, pp. 23–28, Sep. 1998, doi: 10.1002/(SICI)1099-1824(199809)1:1<23::AID-WEg>3.0.CO;2-9.
- [33] M. Cali, Ü;Lange, B.;Bobschinski, J.;Kurt, "(PDF) Artificial neural network based wind power forecasting using a multi-model approach." https://www.researchgate.net/publication/242601478_Artificial_neural_network_based_wind_power_forecasting_using_a_multi-model_approach (accessed Jun. 15, 2020).
- [34] X. Wang, Xiaochen;Guo, Peng; Huang, "A Review of Wind Power Forecasting Models - ScienceDirect." <https://www.sciencedirect.com/science/article/pii/S1876610211019291> (accessed Jun. 15, 2020).
- [35] L. Landberg, "A Mathematical Look at a Physical Power Prediction Model."
- [36] "Electric Power Research Trends - Michael C. Schmidt - Google Llibres." https://books.google.es/books?id=qDr449uMORMC&pg=PA113&lpg=PA113&dq=Sipreolico+spanish&source=bl&ots=zakDag6P_K&sig=ACfU3U1pdxwbVsdgPXEgW82LXtE6OQi36A&hl=ca&sa=X&ved=2ahUKEwj_gq_K6obqAhXlXlUKHXZZCKMQ6AEwAXoECA5QAQ#v=onepage&q=Sipreolico spanish&f=false (accessed Jun. 16, 2020).
- [37] F. S. J.Usaola, O.Ravelo, I.Sanchez, C.Velasco, J.Dominguez, M.Lobo, G.González, "SIPREOLICO, A Wind Power Prediction Tool for the Spanish Peninsular Power System Operation," *Glob. Wind. Conf. Paris, Fr.*, 2002.
- [38] F. S. I.Sánchez, J.Usaola, O.Ravelo, C.Velasco, J.Domínguez, M.Lobo, G.González, "SIPREOLICO, A Wind Power Prediction System Based on a Flexible Combination of Dynamic Models. Application to the Spanish Power System," *Proc. World Wind Energy Conf. Exhib. Berlin, Germany*, 2004.
- [39] T. Madsen,H;Pinson,P;Kariniotakis,G.;Nielsen,H.;Nielsen, "A protocol for standarising the performance if a short-term wind power prediction," 2004.
- [40] "Tutorial: Understanding Linear Regression and Regression Error Metrics." <https://www.dataquest.io/blog/understanding-regression-error-metrics/> (accessed Jul. 01, 2020).
- [41] "Forecast KPI: RMSE, MAE, MAPE & Bias - Analytics Vidhya - Medium." <https://medium.com/analytics-vidhya/forecast-kpi-rmse-mae-mape-bias-cdc5703d242d> (accessed Jul. 01, 2020).
- [42] "Tutorial: Understanding Linear Regression and Regression Error Metrics." <https://www.dataquest.io/blog/understanding-regression-error-metrics/> (accessed Jun. 17, 2020).



- [43] "Forecast KPI: RMSE, MAE, MAPE & Bias - Analytics Vidhya - Medium." <https://medium.com/analytics-vidhya/forecast-kpi-rmse-mae-mape-bias-cdc5703d242d> (accessed Jun. 17, 2020).
- [44] "Aprendizaje automatico y las Metricas de regresión - sitiobigdata.com." <https://sitiobigdata.com/2018/08/27/machine-learning-metricas-regresion-mse/#> (accessed Jun. 17, 2020).
- [45] "What does RMSE really mean? - Towards Data Science." <https://towardsdatascience.com/what-does-rmse-really-mean-806b65f2e48e> (accessed Jun. 17, 2020).
- [46] "Parques Eólicos en España. ¿Donde encontrar? - Avatar Energía, blog de energías renovables." <https://avatarenergia.com/parques-eolicos-en-espana/> (accessed Jun. 19, 2020).
- [47] G. de les I. B. C. d'Economia i C. D. G. d'Indústria i Energia, "Energías Renovables y Eficiencia Energética en las Islas Baleares: Estrategias y líneas de Actuación," 2015.
- [48] "Es Milá (España) - Parques eólicos - Acceso en línea - The Wind Power." https://www.thewindpower.net/windfarm_es_695_es-mila.php (accessed Jun. 19, 2020).
- [49] "ENERGIA EÒLICA. VISITA AL PARC EÒLIC ES MILÀ."
- [50] "Análisis del recurso. Atlas eólico de España | IDAE." <https://www.idae.es/publicaciones/analisis-del-recurso-atlas-eolico-de-espana> (accessed Jun. 19, 2020).
- [51] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*, Intelligen. Springer.
- [52] M. Abbes and J. Belhadj, "Development of a methodology for wind energy estimation and wind park design," *J. Renew. Sustain. Energy*, vol. 6, no. 5, p. 053103, Sep. 2014, doi: 10.1063/1.4895919.
- [53] F. Fernandes, D.;Baptista, D;Jardim, B, Figueira, A;Morgado-Dias, "LOIRAL WIND PARK MODELLING USING ARTIFICIAL NEURAL NETWORKS," *10ª Port. Conf. Autom. Control*, 2012.
- [54] "Estrategias de programación diaria y operación horaria en el mercado de la electricidad de un parque eólico dotado de un sistema acumulador de energía." <http://bibing.us.es/proyectos/abreproy/11513/fichero/PDF+Dividido+por+apartados%252F6.+Predicción+de+potencia+eólica.pdf+>.
- [55] "EL VIENTO Programa educativo sobre el Mediterráneo y su litoral 3 M E D C L I C / L I B R O D E L A L U M N O."
- [56] "Normalization." <https://developers.google.com/machine-learning/data-prep/transform/normalization> (accessed Jun. 25, 2020).
- [57] P. Juszczak, D. M. J. Tax, and R. P. W. Dui, "Feature scaling in support vector data descriptions," *Proc. 8th Annu. Conf. Adv. Sch. Comput. Imaging*, pp. 25–30, 2002.
- [58] "Training, validation, and test sets - Wikipedia." https://en.wikipedia.org/wiki/Training,_validation,_and_test_sets (accessed Jun. 24, 2020).
- [59] M. Jordan, J. Kleinberg, and B. Schölkopf, "Pattern Recognition and Machine Learning."



- [60] A. Larrañaga, P.; Inza, I.;Moujahid, "Tema 8. Redes Neuronales. Universidad del País Vasco-Euskal Herriko Unibertsitatea," .
- [61] "CAPÍTULO 2 HERRAMIENTAS DE EXTRACCIÓN DE INFORMACIÓN: REDES NEURONALES." Accessed: Jun. 27, 2020. [Online]. Available: https://www.tdx.cat/bitstream/handle/10803/1455/3.CAPITULO_2.pdf?sequence=5&isAllowed=y.
- [62] G. Sotolongo, ; Guzmán, and M. Victoria, "Aplicaciones de las redes neuronales. El caso de la Bibliometría," 2001.
- [63] J. M. Flórez López, R.;Fernández Fernández, *Las Redes Neuronales Artificiales. Fundamentos teoricos y aplicaciones practicas*, Netbiblo. netbiblio, 2008.
- [64] J. Torres, "Deep Learning, Introducción práctica con Keras (Segunda parte)," WHAT THIS., 2019.
- [65] E. Diaconescu, "The use of NARX Neural Networks to predict Chaotic Time Series," vol. 3, no. 3, 2008.
- [66] C. L. Lin, T.;Horne, G.B;;Tiño,P.;Giles, "Learning Long-Term Dependencies in NARX Recurrent Neural Networks," *IEE Trans. NEURAL NETWORKS*, vol. 7, 1996.
- [67] P. Kim, *MATLAB Deep Learning*, Apress. Berkeley, CA: Apress, 2017.
- [68] "Ecologismo - Wikipedia, la enciclopedia libre." <https://es.wikipedia.org/wiki/Ecologismo> (accessed Jul. 09, 2020).
- [69] "Hueco de tensión - Wikipedia, la enciclopedia libre." https://es.wikipedia.org/wiki/Hueco_de_tensión (accessed Jul. 09, 2020).
- [70] "2.2. Variación latitudinal y estacional de la insolación | Temas de Ecología." https://fjferrier.webs.ull.es/Apuntes3/Leccion02/22_variacion_latitudinal_y_estacional_de_la_insolacion.html (accessed Jun. 12, 2020).
- [71] "Célula de Hadley - Wikipedia, la enciclopedia libre." https://es.wikipedia.org/wiki/Célula_de_Hadley (accessed Jun. 18, 2020).
- [72] "Célula de Ferrel - Wikipedia, la enciclopedia libre." https://es.wikipedia.org/wiki/Célula_de_Ferrel (accessed Jun. 18, 2020).
- [73] "Efecto Venturi - Wikipedia, la enciclopedia libre." https://es.wikipedia.org/wiki/Efecto_Venturi (accessed Jun. 12, 2020).
- [74] "Albert Betz - Wikipedia, la enciclopedia libre." https://es.wikipedia.org/wiki/Albert_Betz (accessed Jun. 12, 2020).
- [75] "Crisis del petróleo de 1973 | Economipedia." <https://economipedia.com/definiciones/crisis-del-petroleo-1973.html> (accessed Jun. 12, 2020).
- [76] N. Fabra, *El Funcionamiento del Mercado Eléctrico Español bajo la Ley del Sector Eléctrico*, Thomson:Ci., vol. Del Monopo. 2006.
- [77] "¿Sabes para qué sirven las centrales hidroeléctricas de bombeo? - Iberdrola." <https://www.iberdrola.com/medio-ambiente/central-hidroelectrica-bombeo> (accessed Jun.



13, 2020).

- [78] "(PDF) Previento - A Wind Power Prediction System with an Innovative Upscaling Algorithm."
https://www.researchgate.net/publication/250448566_Previento_-_A_Wind_Power_Prediction_System_with_an_Innovative_Upscaling_Algorithm (accessed Jun. 16, 2020).



ANEXOS

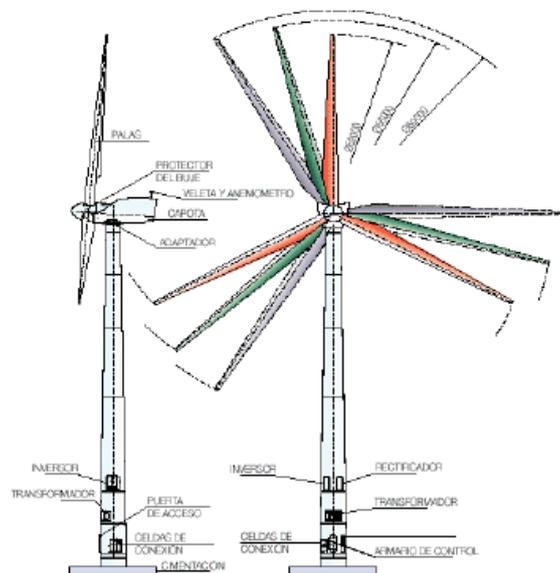
ANEXO A: ESPECIFICACIONES TÉCNICAS DE LOS AEROGENERADORES DE ES MILÁ

En el presente anexo se expone la ficha técnica de los aerogeneradores que se encuentran instalados en el parque eólico de Es Milá. Esta información se ha obtenido gracias a Rafael Campos Muñoz, Ingeniero Técnico Industrial que trabaja para el "Consorti de Residus i Energia de Menorca".



MADE Serie 800 Aerogeneradores de paso y velocidad variables 800 kW

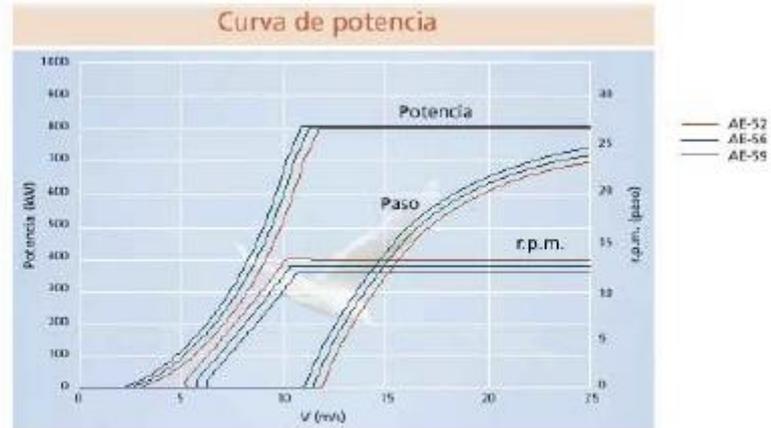
Innovaciones Tecnológicas	que aportan	Ventajas Adicionales
Alto Rendimiento		
Generador síncrono		Alto rendimiento a bajas cargas
Velocidad variable total		Adaptación a la velocidad del viento en todo el rango
Control por lógica difusa		Optimización de producción
Paso variable		Adaptación del perfil de pala
Gran superficie de barrido		Alta producción específica
Gama de rotores		Optimización en todos los regímenes de velocidad del viento
Gran Fiabilidad		
Electrónica de IGBT		Mayor robustez que IGBT
Un solo dispositivo para toda la corriente		Sin averías por falta de sincronismo
Generador sin escobillas		IP 54 real y no mantenimiento
Mando independiente de paso en cada pala		Total seguridad de operación
Componentes de última tecnología		Fiabilidad a largo plazo



Para mejorar el funcionamiento de este aerogenerador, MADE Tecnologías renovables S.A. se reserva el derecho de modificar cualquiera de estas características sin previo aviso. La capacidad de MADE para adaptarse a la demanda del mercado está respaldada por la acreditación de su sistema de "Garantía de Calidad" certificado por la Norma ISO 9001 de GERMANISCHER LLOYD, una de las más prestigiosas instituciones europeas en este campo.

Características Técnicas

		AE-52	AE-56	AE-59
Rotor	Potencia nominal	800 kW	800 kW	800 kW
	Diámetro del rotor	52 m	56 m	59 m
	Control de potencia	Por cambio de paso y velocidad 100% variable	Por cambio de paso y velocidad 100% variable	Por cambio de paso y velocidad 100% variable
	Sistema de orientación	Activo a barlovento	Activo a barlovento	Activo a barlovento
	Área barrida por el rotor	2123,7 m ²	2463 m ²	2733,97 m ²
	Número de palas	3	3	3
	Tipo de pala	LM 25,1 P	LM 27,1 P	LM 28,6 P
	Rango de velocidades del rotor	12,8 a 25,71 rpm	11,9 a 23,8 rpm	11,3 a 22,6 rpm
	Altura de buje sobre suelo	50/60 m	50/60 m	50/60 m
	Ángulo de inclinación	5°	5°	5°
Multiplicador	Tipo	Ejes paralelos, planetario	Planetario	Planetario
	Relación de multiplicación	1:58,34	1:63,02	1:66,37
Generador	Tipo de generador	Síncrono trifásico, de cuatro polos	Síncrono trifásico, de cuatro polos	Síncrono trifásico, de cuatro polos
	Tensión de alimentación	1.000 V ± 5%	1.000 V ± 5%	1.000 V ± 5%
	Aislamiento	Clase H, uso F	Clase H, uso F	Clase H, uso F
	Protección	IP 54	IP 54	IP 54
Convertidor de frecuencia	Topología	Rectificador de diodos, chopper elevador, inversor de IGCT 'S	Rectificador de diodos, chopper elevador, inversor de IGCT 'S	Rectificador de diodos, chopper elevador, inversor de IGCT 'S
	Tensión de alimentación	1000 V	1000 V	1000 V
	Frecuencia de red	50 Hz ± 2%	50 Hz ± 2%	50 Hz ± 2%
	Gestión de coseno de φ	Máxima eficiencia a todas las cargas	Máxima eficiencia a todas las cargas	Máxima eficiencia a todas las cargas
Sistema de freno	Freno principal	Cambio de paso de las palas	Cambio de paso de las palas	Cambio de paso de las palas
	Freno de seguridad	Freno de disco, pinza hidráulica en el eje rápido	Freno de disco, pinza hidráulica en el eje rápido	Freno de disco, pinza hidráulica en el eje rápido
Sistema de orientación	Tipo	Moto-reductores eléctricos, con etapas reductoras y planetaria	Moto-reductores eléctricos, con etapas reductoras y planetaria	Moto-reductores eléctricos, con etapas reductoras y planetaria
	Tipo de pinzas de freno	Pinzas de freno hidráulicas	Pinzas de freno hidráulicas	Pinzas de freno hidráulicas
Torre	Tipo	Tronco-cónica de acero soldado	Tronco-cónica de acero soldado	Tronco-cónica de acero soldado
Condiciones de operación	Clase de viento	I según IEC 61400-1	II según IEC 61400-1	III según IEC 61400-1
	Velocidad de arranque	3,5 m/s	3,3 m/s	3 m/s
	Velocidad de parada	25 m/s	25 m/s	25 m/s
	Temperatura ambiente de funcionamiento	-10°C ÷ 40°C	-10°C ÷ 40°C	-10°C ÷ 40°C
Estimación de pesos	Rotor	15.500 kg	15.200 kg	15.700 kg
	Góndola	28.200 kg	29.000 kg	29.500 kg



Tecnología punta al servicio de su inversión

ANEXO B: CÓDIGOS MATLAB

En el presente anexo se expone los diferentes códigos de MATLAB desarrollados a lo largo de este trabajo de.

B.1: PREPROCESAMIENTO_FASE_1.M

Este script de MATLAB se utiliza para realizar la primera fase de preparación de los datos. Se realizan transformaciones de variables, se rellena los huecos temporales interpolando, se filtra fechas erróneas, y se realiza los tratamientos necesarios para la unión de los archivos de generación con los archivos de datos meteorológicos.

```

1  %
2  %
3  %
4  % Trabajo de fin de grado: INGENIERÍA ELECTRÓNICA INDUSTRIAL Y AUTOMÁTICA
5  % Universidad de las Islas Baleares (UIB)
6  % Deva Murti Baer
7  % 2019/2020
8  %
9  %
10 %
11 % DESCRIPCIÓN DE LA FUNCIÓN DEL SCRIPT
12 % -----
13 %
14 % DATOS METEOROLÓGICOS
15 % -----
16 % - Lectura de los archivos anuales de datos meteorológicos.
17 % - Conversión de la dirección del viento de caracteres a grados.
18 % - Relleno de los huecos de las variables interpolando.
19 % - Filtrado de las filas de las series temporales que contienen fechas
20 % erróneas.
21 % - Reducción de las filas a la mitad, realizando la media de los valores de
22 % dos filas adyacentes.
23 %
24 %
25 % DATOS DE GENERACIÓN
26 % -----
27 % - Lectura de los archivos anuales de datos meteorológicos.
28 % - Conversión de la dirección del viento de caracteres a grados.
29 % - Relleno de los huecos de datos realizando una interpolación cúbica.
30 % - Filtrado de las filas de las series temporales que contienen fechas
31 % erróneas (errores de formato, o fechas incoherentes)
32 %
33 %
34 % INTEGRACIÓN DE LOS DATOS EN UN ARCHIVO
35 % -----
36 % - Se realiza la unión de los datos meteorológicos con los de generación
37 % en un único archivo.
38 %
39 %
40 %
41 %
42 % -----
43 % Inicialización del espacio de memoria.
44 % -----
45 close all;
46 clear all;
47 clc;
48 t_start = cputime;
49 date_start = now;
50 % -----
51 % Mensaje de bienvenida
52 % -----
53 disp('-----');
54 disp('- Universidad de las Islas Baleares');
55 disp('- Trabajo de fin de grado: Deva Murti Baer (2019-2020)');
56 disp(str_txt);
57 disp('-----');
58 disp(' ');
59
60 %
61 %

```

```

62 %----- DATOS METEOROLÓGICOS
63 %-----
64 %+++++
65 %-----
66 %
67 % Directorio donde se hallan los datos
68 %-----
69 %
70 str_dir_input_data = pwd; % Se lee el directorio donde está ubicado el .m
71 str_hist_dir = strcat(str_dir_input_data, '\DatosMeteo'); % Se obtiene la dirección de la carpeta que
72 % contiene las hojas de cálculo con los datos.
73 str_out_dir = strcat(str_dir_input_data, '\out'); % Se genera un directorio de salida
74 %-----
75 %
76 % Se define el tipo de archivo al que se quiere acceder
77 %-----
78 file_type = '*.csv';
79 list_dir_hist = dir(fullfile(str_hist_dir, file_type)); % Se obtiene una lista de todos los excels
80 % incluidos en la carpeta
81 names_hist = {list_dir_hist.name}; % Se obtiene una lista de todas las hojas de
82 % cálculo que tienen la extensión .csv y se encuentran en la carpeta
83 names_hist_size = (list_dir_hist.bytes); % Se obtiene el tamaño en bytes de cada una de las
84 % hojas de cálculo
85 num_excels = length(names_hist); % Se obtiene la cantidad de archivos que cumplen
86 % los requisitos
87 cd(str_hist_dir); % Se cambia el directorio de acción
88 %-----
89 %
90 % Se guarda el contenido del archivo seleccionado en una variable
91 %-----
92 a = readtable(char(names_hist(:,3)), 'Delimiter',';'); % Se obtiene la tabla de datos en formato double
93 % en caso que toda la columna sea numérica o en formato cell en caso contrario
94 %-----
95 %
96 % Se guarda el contenido de interés de la tabla en un conjunto de variables
97 %-----
98 fecha_table = table2array(a(2:end,1)); % Se obtiene, las fechas, horas, y los datos
99 meteorológicos
100 hora_table = table2array(a(2:end,2));
101 meteo_table1 = table2array(a(2:end,3:8));
102 meteo_table2 = table2array(a(2:end,10:11));
103 meteo_table3 = table2array(a(2:end,13:17));
104 meteo_table = horzcat(meteo_table1,meteo_table2,meteo_table3); % Se une en una variable los parámetros
105 meteorológicos relevantes, excepto la velocidad y dirección del viento
106 dir_table = table2array(a(2:end,9));
107 %-----
108 %
109 % Se inicializa las matrices de las variables finales
110 %-----
111 fecha_hora = cell(length(fecha_table),1);
112 fecha_fi = zeros(length(fecha_table),1);
113 [f,c] = size(meteo_table); % La 'f' representa la cantidad de filas y 'c' la
114 % cantidad de columnas presentes en la matriz
115 dades_meteo_fi = zeros(f,c); % Se inicializa a 0 las posiciones de la tabla
116 final
117 dades_meteo_fi_interpolate = zeros(f,c);
118 espacio = ' ';
119 %-----
120 %
121 % Correlación de direcciones en formato cadena con su valor en grados
122 %-----
123 str_dir = {'N','NNE','NE','ENE','E','ESE','SE','SSE','S','SSW','SW','WSW','W','WNW','NW','NNW'};
124 num_dir = [0,22.5,45, 67.5,90, 112.5, 135, 157.5, 180, 202.5, 225, 247.5, 270, 292.5, 315, 337.5];
125 %-----
126 %
127 % Obtención de la fecha+horas, y de su equivalente numérico, y relleno con
128 % NaN de los vacíos.
129 %-----
130 %
131 % Se normaliza el formato de las horas añadiendo
132 % un 0 delante en aquellos casos que no tenga
133 % un 0 delante en aquellos casos que no tenga
134 % un 0 delante en aquellos casos que no tenga
135 % un 0 delante en aquellos casos que no tenga
136 % un 0 delante en aquellos casos que no tenga
137 % un 0 delante en aquellos casos que no tenga
138 % un 0 delante en aquellos casos que no tenga
139 % un 0 delante en aquellos casos que no tenga
140 % un 0 delante en aquellos casos que no tenga
141 % un 0 delante en aquellos casos que no tenga
142 % un 0 delante en aquellos casos que no tenga
143 % un 0 delante en aquellos casos que no tenga
144 % un 0 delante en aquellos casos que no tenga
145 % un 0 delante en aquellos casos que no tenga
146 % un 0 delante en aquellos casos que no tenga
147 % un 0 delante en aquellos casos que no tenga
148 % un 0 delante en aquellos casos que no tenga
149 % un 0 delante en aquellos casos que no tenga
150 % un 0 delante en aquellos casos que no tenga
151 % un 0 delante en aquellos casos que no tenga
152 % un 0 delante en aquellos casos que no tenga
153 % un 0 delante en aquellos casos que no tenga
154 % un 0 delante en aquellos casos que no tenga
155 % un 0 delante en aquellos casos que no tenga
156 % un 0 delante en aquellos casos que no tenga
157 % un 0 delante en aquellos casos que no tenga
158 % un 0 delante en aquellos casos que no tenga
159 % un 0 delante en aquellos casos que no tenga
160 % un 0 delante en aquellos casos que no tenga
161 % un 0 delante en aquellos casos que no tenga
162 % un 0 delante en aquellos casos que no tenga

```



```
163     end
164 end
165 %-----
166 % Se realiza el relleno numérico de los espacios que contienen los
167 % caracteres NaN de direcciones.
168 %-----
169 dir_vent_graus(dir_vent_graus==1) = NaN;
170 dir_vent_interpolate = fillmissing(dir_vent_graus,'spline');
171 %-----
172 % Se realiza el relleno numérico de los espacios que contienen los
173 % caracteres NaN, de datos meteorológicos.
174 %-----
175 for lll=1:1:c
176     dades_meteo_fi_interpolate(:,lll) = fillmissing(dades_meteo_fi(:,lll),'spline');
177 end
178 dir_vent_interpolate(dir_vent_interpolate<0) = 0; % En caso de disponer alguna dirección negativa se
179 cambia por un 0
180 %-----
181 % Se obtiene una tabla que contiene los datos meteorológicos, y las
182 % direcciones de viento rellenados.
183 %-----
184 Tabla_interpolate=horzcat(dades_meteo_fi_interpolate,dir_vent_interpolate);
185
186 %-----
187 % Se realiza un filtrado de las fechas con el objetivo de detectar las
188 % fechas que contienen horas en punto.
189 %-----
190 [filas,columnas] = size(fecha_fi);
191 fecha_fi_media = zeros(filas/2,columnas);
192 it2=1;
193 for it=1:1:filas
194     [Ano,Mes,Dia,Hora,Minuto,Segundo]=datevec( fecha_fi(it,1));
195     if(Ano>2015 ||Ano<2010)
196         continue
197     end
198     if(Mes>12 ||Mes<1)
199         continue
200     end
201     if(Dia>31 ||Dia<1)
202         continue
203     end
204     if(Segundo>0)
205         continue
206     end
207     if(Minuto==0)
208         fecha_fi_media(it2,1)=datenum(Ano,Mes,Dia,Hora,Minuto,Segundo);
209         it2=it2+1;
210     end
211 end
212 Fecha_fi_revisado=(cellstr(datestr( fecha_fi_media(:,1), 'dd/mm/yyyy HH:MM:SS')));
213 %-----
214 % Se obtiene la media de los valores de pares de filas adyacentes para
215 % todas las variables.
216 %-----
217 [filas3,columnas3] = size(Tabla_interpolate);
218 Tabla_interpolate_media=zeros(filas3/2,columnas3); % Se reduce el número de filas a la mitad.
219 for itcolumnas=1:1:columnas3
220     it3=1;
221     for itfilas=1:2:filas3
222         Tabla_interpolate_media(it3,itcolumnas)= (Tabla_interpolate(itfilas,itcolumnas)+
223 Tabla_interpolate(itfilas+1,itcolumnas))/2;
224         it3=it3+1;
225     end
226 end
227 %-----
228 % Se obtiene una tabla que dispone de datos meteorológicos para fechas con
229 % horas en punto.
230 %-----
231 Tabla_fechas_meteo_fi=[Fecha_fi_revisado,num2cell(Tabla_interpolate_media)];
232
233 %+++++
234 %
235 %
236 % DATOS DE GENERACIÓN ELÉCTRICA
237 %
238 %+++++
239
240 %-----
241 % Directorio donde se hallan los datos
242 %-----
243 str_hist_dir_Gen = strcat(str_dir_input_data,'\DatosGeneración');
244 str_out_dir_Gen = strcat(str_dir_input_data,'\out');
245 %-----
246 % Se define el tipo de archivo al que se quiere acceder
247 %-----
248 list_dir_hist_Gen = dir(fullfile(str_hist_dir_Gen, file_type));
249 names_hist_Gen = {list_dir_hist_Gen.name};
250 names_hist_size_Gen = {list_dir_hist_Gen.bytes};
251 num_excels_Gen = length(names_hist_Gen);
252 cd(str_hist_dir_Gen);
253 %-----
254 % Se guarda el contenido del archivo seleccionado en una variable
255 %-----
256 b = readtable(char(names_hist_Gen(:,5)), 'Delimiter',';');
257 %-----
258 % Se guarda el contenido de interés de la tabla en un conjunto de variables
259 %-----
260 fecha_table_Gen = table2array(b(1:end,1)); % Se obtiene las fechas y horas, y la tabla de
261 generación ( AE,AS,AS-AE)
262 hora_table_Gen = table2array(b(1:end,2));
263
```

```

264 Potencia_table = table2array(b(1:end,3:5));
265
266 % Se inicializa las matrices de las variables finales
267
268 fecha_hora_Gen = cell(length(fecha_table_Gen),1);
269 fecha_fi_Gen = zeros(length(fecha_table_Gen),1);
270 [f_Gen,c_Gen] = size(Potencia_table);
271 Potencia_table_fi_Gen = zeros(f_Gen,c_Gen);
272 Potencia_table_fi_Gen_interpolate = zeros(f_Gen,c_Gen);
273
274 % Obtención de la fechas horas, y de su equivalente numérico, y relleno con
275 % NaN de los vacios.
276
277 for i=1:length(fecha_table_Gen)
278     if (length(char(hora_table_Gen(i,1)))<8)
279         dat_Gen = ['0',char(hora_table_Gen(i,1))];
280     else
281         dat_Gen = char(hora_table_Gen(i,1));
282     end
283     fecha_hora_Gen (i,1) = strcat(char(fecha_table_Gen(i,1)), {' '}, dat_Gen);
284
285     for j=1:c_Gen
286         if not(isempty(Potencia_table(i,j)))
287             Potencia_table_fi_Gen(i,j) = (Potencia_table(i,j));
288         else
289             Potencia_table_fi_Gen(i,j) = NaN;
290         end
291     end
292
293     if not(isempty(fecha_table_Gen(i,1)))
294         fecha_fi_Gen(i,1) = datenum(char(fecha_hora_Gen (i,1)), 'dd/mm/yyyy HH:MM:SS');
295     end
296 end
297
298 % Se realiza el relleno numérico de los espacios que contienen los
299 % caracteres NaN mediante una interpolación cúbica.
300
301 for ppp=1:c_Gen
302     Potencia_table_fi_Gen_interpolate(:,ppp) = fillmissing(Potencia_table_fi_Gen(:,ppp),'spline');
303 end
304
305 % Se realiza un filtrado de las fechas con el objetivo de detectar las
306 % fechas que contienen horas en punto.
307
308 [filas_Gen,columnas_Gen] = size(fecha_fi_Gen);
309 fecha_fi_Gen_revisado = zeros(filas_Gen,columnas_Gen);
310 it4=1;
311 for itt=1:1:filas_Gen
312     [Ano_Gen,Mes_Gen,Dia_Gen,Hora_Gen,Minuto_Gen,Segundo_Gen]=datevec(fecha_fi_Gen(itt,1));
313     if(Ano_Gen>2015 ||Ano_Gen<2010)
314         continue
315     end
316     if(Mes_Gen>12 ||Mes_Gen<1)
317         continue
318     end
319     if(Dia_Gen>31 ||Dia_Gen<1)
320         continue
321     end
322     if(Segundo_Gen>0)
323         continue
324     end
325     if(Minuto_Gen>0)
326         continue
327     end
328     fecha_fi_Gen_revisado(it4,1)=datenum(Ano_Gen,Mes_Gen,Dia_Gen,Hora_Gen,Minuto_Gen,Segundo_Gen);
329     it4=it4+1;
330 end
331
332 Fecha_fi_revisado_Gen=(cellstr(datestr(fecha_fi_Gen,'dd/mm/yyyy HH:MM:SS')));
333
334 % Se obtiene la tabla de generación tratada y unida con las fechas
335 % filtradas
336
337 Tabla_fechas_generacion_fi=[Fecha_fi_revisado_Gen,num2cell(Potencia_table_fi_Gen_interpolate)];
338
339
340
341
342 %-----
343 % INTEGRACIÓN EN UN ÚNICO ARCHIVO DE LOS DATOS METEOROLÓGICOS Y DE
344 % GENERACIÓN ELÉCTRICA.
345 %-----
346
347
348 % Se realiza una comparación de las fechas. En caso de que se disponga de
349 % la fechas en la tabla de datos meteorológicos y en la de generación, se
350 % carga la fila correspondiente.
351
352 [filas_union1,columnas_union1]=size(Tabla_fechas_meteo_fi); % Se obtiene el número de filas y columnas de las
353 % tablas de cada tipo de datos.
354 [filas_union2,columnas_union2]=size(Tabla_fechas_generacion_fi);
355 Tabla_Union=zeros(filas_union2,columnas_union1+columnas_union2-1);
356 Tabla_Union_Cell=num2cell(Tabla_Union);
357 it5=1;
358 for pas=1:length(Tabla_fechas_generacion_fi)
359     [row2,column2]=find(datenum(Tabla_fechas_meteo_fi(:,1))==datenum(Tabla_fechas_generacion_fi(pas,1)));
360     if not(isempty(row2))
361         Tabla_Union_Cell(it5,:)= [Tabla_fechas_meteo_fi(row2,:) Tabla_fechas_generacion_fi(pas,2:end)];
362         it5=it5+1;
363     end
364     if (isempty(row2))

```



```
365         continue
366     end
367 end
368
369 %-----
370 % Se escribe la tabla resultante a una hoja de cálculo. Se obtiene de una
371 % serie temporal anual de datos meteorológicos y datos de generación
372 %-----
373 xlswrite('out_2014_Meteo_Gen_Ex.xls',Tabla_Union_Cell(1:6369,:));
374 csvwrite('out_2010_Meteo_Gen_Tx.csv',Tabla_Union_Cell(1:6992,:));
375
376
377 %-----
378 % Se repite el proceso expuesto para los archivos de cada año.
379 %-----
```

B.2: PREPROCESAMIENTO_FASE_2.M

Este script se ha utilizado en la segunda fase de preprocesamiento. Se realiza la transformación y el cálculo de nuevas variables de interés, se filtran los datos, se obtiene las rosas de los vientos, se obtiene la superficie de generación del parque de Es Milá, se normalizan las variables de interés, y se separan los datos según la estación a la que pertenecen para luego cortarlos en diferentes archivos tolerando huecos temporales de hasta 12 horas consecutivas.

```
1 %
2 %+++++
3 %+++++
4 % Trabajo de fin de grado: INGENIERÍA ELECTRÓNICA INDUSTRIAL Y AUTOMÁTICA
5 % Universidad de las Islas Baleares (UIB)
6 % Deva Murti Baer
7 % 2019/2020
8 %+++++
9 %+++++
10 %
11 % DESCRIPCIÓN DE LA FUNCIÓN DEL SCRIPT
12 %-----
13 %
14 % TRANSFORMACIÓN DE VARIABLES
15 %-----
16 % - Lectura del archivo que contiene todo el conjunto de datos meteorológicos
17 % y de generación.
18 % - Conversión de la velocidad del viento a m/s, de la temperatura exterior
19 % a Kelvin, y de la humedad en humedad relativa.
20 % - Visualización de los datos de velocidad.
21 % - Cálculo del coeficiente de rugosidad alpha del terreno de Es Milá.
22 % - Extraplación de la velocidad del viento a la altura del buje de los
23 % aerogeneradores.
24 %
25 %
26 % FASE DE FILTRADO
27 %-----
28 % -Aplicación del FILTRO 0:eliminación de las velocidades de viento
29 % negativas.
30 % Aplicación del FILTRO 1:eliminación de los datos cuando hay generación,
31 % y la velocidad del viento no llega a la mínima de operación: 3m/s
32 % Aplicación del FILTRO 2: eliminación de los datos cuando hay generación,
33 % y la velocidad del viento supera la velocidad de parada de seguridad:
34 % 25m/s.
35 % Aplicación del FILTRO 3: eliminación de datos cuando hay direcciones de
36 % viento que no estén comprendidas entre 0-360°
37 %
38 %
39 % ROSAS DE LOS VIENTOS
40 %-----
41 % - Clasificación de los datos según la estación del año a la que pertenece
42 % - Obtención de las rosas de los vientos de cada estación, y del conjunto
43 % de los datos filtrados.
44 %
45 %
46 % AJUSTE DE UNA FUNCIÓN A LA SUPERFICIE DE GENERACIÓN DEL PARQUE EÓLICO
47 %-----
48 % - Ajuste polinómico de una función a la representación en 3 dimensiones
49 % de la velocidad y dirección del viento, junto a la generación eléctrica
50 % real del parque eólico de Es Milá.
51 % - Se calcula la serie temporal de la generación del parque a partir de la
52 % función ajustada.
53 % - Se obtiene la matriz de generación generada.
```



```
54 %
55 %
56 % NORMALIZACIÓN DE LOS DATOS
57 % -----
58 % - Se obtiene los máximos y los mínimos de las variables de interés:
59 % a) Velocidad del viento.
60 % b) Dirección del viento.
61 % c) Generación eléctrica real del parque.
62 % d) Generación eléctrica obtenida a partir de la función ajustada.
63 % - Se realiza el escalado de los datos dentro del rango [0 1] o [-1 1]
64 % utilizando el método de normalización min/max.
65 %
66 %
67 % SEGMENTACIÓN DE LOS DATOS
68 % -----
69 % - Se cortan los datos que previamente se ha clasificado por estaciones
70 % cuando se detecta un hueco temporal superior a 12 horas
71 %
72 %+++++
73 %+++++
74 %
75 %-----
76 % Se inicializa el espacio de memoria.
77 %-----
78 close all;
79 clear all;
80 clc;
81 t_start = cputime;
82 date_start = now;
83
84 %-----
85 % Mensaje de bienvenida
86 %-----
87 disp('-----');
88 disp('- Universidad de las Islas Baleares');
89 disp('- Trabajo de fin de grado: Deva Murti Baer (2019-2020)');
90 disp(str_txt);
91 disp('-----');
92 disp(' ');
93
94 %-----
95 % Directorio donde se hallan los datos
96 %-----
97 str_dir_input_data = pwd; % Se lee el directorio donde está ubicado el .m
98 str_hist_dir_MG = strcat(str_dir_input_data, '\DatosMeteoDatosGeneración'); % Se obtiene la dirección de la carpeta que
99 contiene las hojas de cálculo con los datos.
100 str_out_dir_MG = strcat(str_dir_input_data, '\out'); % Se genera un directorio de salida
101
102 %-----
103 % Se define el tipo de archivo al que se quiere acceder
104 %-----
105 file_type = '*.csv';
106 list_dir_hist_MG = dir(fullfile(str_dir_hist_dir_MG, file_type)); % Se obtiene una lista de todos los excels
107 incluidos en la carpeta
108 names_hist_MG = {list_dir_hist_MG.name}; % Se obtiene una lista de todas las hojas de
109 cálculo que tienen la extensión .csv y se encuentran en la carpeta
110 names_hist_size_MG = {list_dir_hist_MG.bytes}; % Se obtiene el tamaño en bytes de cada una de las
111 hojas de cálculo
112 num_excels_MG = length(names_hist_MG); % Se obtiene la cantidad de archivos que cumplen
113 los requisitos
114 cd(str_dir_hist_dir_MG); % Se cambia el directorio de acción
115
116 %-----
117 % Se guarda el contenido del archivo seleccionado en una variable
118 %-----
119 Tabla_Ano = readtable(char(names_hist_MG(:,2)), 'Delimiter', ',');
120
121 %-----
122 % Se guarda el contenido de interés de la tabla en un conjunto de variables
123 %-----
124 fecha_MG=table2array(Tabla_Ano(1:end,1));
125 meteo_MG = table2array(Tabla_Ano(1:end,2:18));
126 [fil_meteo_MG,col_meteo_MG]=size(meteo_MG);
127 columnas_adi=zeros(fil_meteo_MG,10); % Se define una matriz con el número de filas de
128 la tabla, y columnas de 0. Servirá para poder añadir columnas con variables nuevas.
129 meteo_MG_fi=horzcat(meteo_MG,columnas_adi); % Se le han añadido columnas de 0 a la matriz que
130 contiene los datos.
131 [fil_meteo_MG_fi,col_meteo_MG_fi]=size(meteo_MG_fi);
132 datos_generacion = table2array(Tabla_Ano(1:end,16:18));
133 [fil_gen_MG,col_gen_MG]=size(datos_generacion);
134
135 %-----
136 % Se añade a las columnas de 0 añadidas las siguientes variables:
137 % - Velocidad del viento en m/s
138 % - Temperatura en Kelvin
139 % - Humedad en porcentaje
140 %-----
141 for pas3=1:1:length(meteo_MG_fi)
142 meteo_MG_fi(pas3,18)=meteo_MG_fi(pas3,6)*(1000/3600); %velocidad (m/s)
143 meteo_MG_fi(pas3,20)=meteo_MG_fi(pas3,4)/(100); %humedad (%)
144 meteo_MG_fi(pas3,21)=meteo_MG_fi(pas3,1)+(273.15); %temperatura (K)
145 end
146
147 %-----
148 % Se obtiene el gráfico de la velocidad y de la generación
149 %-----
150 vel_viento=zeros(fil_meteo_MG_fi,1);
151 Pot_generada=zeros(fil_meteo_MG_fi,1);
152 for pas2=1:1:length(meteo_MG_fi)
153 vel_viento(pas2,1)=meteo_MG_fi(pas2,18); % Matriz de velocidades de viento
154 Pot_generada(pas2,1)=meteo_MG_fi(pas2,17); % Matriz de generación
```



```
155 end
156 histogram(vel_viento) % Gráfico de la frecuencia de las velocidades
157 title('Velocidad del viento');
158 xlabel('Velocidad del viento(m/s)');
159 ylabel('Cantidad de valores(u)');
160 figure
161 scatter(vel_viento,meteo_MG(:,16));
162 title('Velocidad del viento/Potencia generada');
163 xlabel('Velocidad del viento(m/s)');
164 ylabel('Potencia generada(kWh)');
165 grid on
166 axis([0 max(vel_viento) min(meteo_MG(:,17)) max(meteo_MG(:,17))]); % No se representan las velocidades negativas
167
168 -----
169 % Se representa todos los valores, con un gradiente de color
170 -----
171 createfigure(vel_viento, Pot_generada, [], vel_viento);
172
173 -----
174 % Cálculo de la velocidad del viento a la altura del buje de un
175 % aerogenerador modelo AE-59 del parque de los Milans, Menorca.
176 % - alpha=0.0910*log10(z0)+0.016*(log10(z0))^2+0.24;
177 % - Vx=Vo*(hx/ho)^alpha
178 -----
179
180 % Cálculo del coeficiente de fricción alpha. Uso de la ecuación de
181 % Abbes&Belhadj 2014
182 -----
183
184 altura_datos_JMora=5; % [m]
185 altura_buje=55; % [m]
186 alpha=0.0910*log10(0.009)+0.016*(log10(0.009))^2+0.24; % [adimensional]
187 for pas4=1:length(meteo_MG_fi)
188     meteo_MG_fi(pas4,19)=meteo_MG_fi(pas4,18)*(altura_buje/altura_datos_JMora)^(alpha);
189 end
190
191 -----
192 % Cálculo de la densidad del aire a la altura del buje de un
193 % aerogenerador modelo AE-59 del parque de los Milans, Menorca.
194 % Suposición 1: Variación de temperatura entre altura estación JMora y
195 % altura del buje, negligible.
196 % Suposición 2: Humedad relativa constante.
197 -----
198 R=8.3144621; % Constante de los gases ideales [J/K.mol]
199 g=9.807; % Gravedad [m/s^2]
200 for pas5=1:length(meteo_MG_fi)
201     meteo_MG_fi(pas5,22)=(meteo_MG_fi(pas5,13))/(R*meteo_MG_fi(pas5,21))*exp(-
202     ((g*altura_buje*meteo_MG_fi(pas5,20))/(R*meteo_MG_fi(pas5,21))));
203 end
204
205 -----
206 % Se une la tabla con datos meteorológicos con las fechas y horas
207 -----
208 Tabla_meteo_MG=[cellstr(datestr(fecha_MG,'dd/mm/yyyy HH:MM:SS')),num2cell(meteo_MG_fi)];
209
210
211 -----
212 %
213 FASE DE FILTRADO DE LOS DATOS
214 -----
215 -----
216 [fil_tab_MG,col_tab_MG]=size(Tabla_meteo_MG);
217 aux=zeros(fil_tab_MG,col_tab_MG);
218 Tabla_meteo_MG_0=num2cell(aux); % Se define una variable en la que guardar los
219 datos que no han sido filtrados.
220 Tabla_meteo_MG_1=num2cell(aux);
221 Tabla_meteo_MG_2=num2cell(aux);
222 Tabla_meteo_MG_auxFiltro3=num2cell(aux);
223
224 -----
225 % Aplicación del Filtro 0:eliminación de las velocidades de viento
226 % negativas.
227 -----
228 it10=1;
229 fila ( it10 -1) % La nueva tabla tiene filas con números hasta la
230 for pas6=1:length(Tabla_meteo_MG)
231     if(cell2mat(Tabla_meteo_MG(pas6,20))<0) % En el caso que la velocidad es negativa, esta no
232     se guarda.
233         continue
234     else
235         Tabla_meteo_MG_0(it10,:)=Tabla_meteo_MG(pas6,:);
236         it10=it10 +1;
237     end
238 end
239 Tabla_meteo_MG_0(33060:end,:)=[]; % Se elimina las filas de 0 que sobran tras
240 aplicar el filtro.
241
242 -----
243 % Se representa los datos que se tiene una vez aplicado el filtro 0:
244 -----
245 figure
246 scatter(cell2mat(Tabla_meteo_MG_0(:,20)),cell2mat(Tabla_meteo_MG_0(:,18)), 'filled','r');
247 xlabel('Velocidad del viento (m/s)');
248 ylabel('Energía generada (kWh)');
249 grid on
250
251 -----
252 % Aplicación del Filtro 1: Se eliminan los datos cuando se dispone de
253 % generación, y la velocidad del viento no llega a la mínima de operación:
254 % 3m/s
255 -----
```



```
256 it11=1; % La nueva tabla tiene filas con números hasta la
257 fila ( it10 -1)
258 for pas7=1:length(Tabla_meteo_MG_0)
259     if(cell2mat(Tabla_meteo_MG_0(pas7,20))<3 && cell2mat(Tabla_meteo_MG_0(pas7,17))>0)
260         continue
261     else
262         Tabla_meteo_MG_1(it11,:)=Tabla_meteo_MG_0(pas7,:);
263         it11=it11 +1;
264     end
265 end
266 Tabla_meteo_MG_1(25255:end,:)=[]; % Se elimina las filas de 0 que sobran tras
267 aplicar el filtro.
268
269 %-----
270 % Se representa el resultado una vez aplicado el filtro 1:
271 %-----
272 figure
273 scatter(cell2mat(Tabla_meteo_MG_1(:,20)),cell2mat(Tabla_meteo_MG_1(:,18)), 'filled','r');
274 xlabel('Velocidad del viento (m/s)');
275 ylabel('Energía generada (kWh)');
276 grid on
277
278 %-----
279 % Aplicación del Filtro 2: Se eliminan los datos cuando hay generación,
280 % y la velocidad del viento supera la velocidad de parada de seguridad:
281 % 25m/s.
282 %-----
283 it12=1;
284 for pas8=1:length(Tabla_meteo_MG_1)
285     if(cell2mat(Tabla_meteo_MG_1(pas8,20))>25 && cell2mat(Tabla_meteo_MG_1(pas8,17))>0) % Cuando la velocidad es mayor a 25
286     m/s no se guarda la fila de la serie temporal.
287         continue
288     else
289         Tabla_meteo_MG_auxFiltro3(it12,:)=Tabla_meteo_MG_1(pas8,:);
290         it12=it12 +1;
291     end
292 end
293 Tabla_meteo_MG_auxFiltro3(25255:end,:)=[]; % Se elimina las filas de 0 que sobran tras
294 aplicar el filtro.
295
296 %-----
297 % Se representa el resultado una vez aplicado el filtro 2:
298 %-----
299 figure
300 scatter(cell2mat(Tabla_meteo_MG_auxFiltro3(:,20)),cell2mat(Tabla_meteo_MG_auxFiltro3(:,18)), 'filled','g');
301 xlabel('Velocidad del viento (m/s)');
302 ylabel('Energía generada (kWh)');
303 grid on
304
305 %-----
306 % Aplicación del Filtro 3: Se elimina los datos cuando se dispone de una
307 % dirección de viento que no se incluya en el rango [0-360]°
308 %-----
309 it13=1;
310 for pas9=1:length(Tabla_meteo_MG_auxFiltro3)
311     if(cell2mat(Tabla_meteo_MG_auxFiltro3(pas9,15))>360)
312         continue
313     else
314         Tabla_meteo_MG_2(it13,:)=Tabla_meteo_MG_auxFiltro3(pas9,:);
315         it13=it13 +1;
316     end
317 end
318 Tabla_meteo_MG_2(25254:end,:)=[];
319
320
321
322 %-----
323 %
324 % SUPERFICIE DE GENERACIÓN DEL PARQUE EÓLICO
325 %-----
326 %-----
327 [fil_tab MG_2b,col_tab MG_2b]=size(Tabla_meteo_MG_2);
328 Tabla_Vel_Viento=zeros(fil_tab MG_2b,1);
329 Tabla_Direc_Viento=zeros(fil_tab MG_2b,1);
330 Tabla_Gen=zeros(fil_tab MG_2b,1);
331 Tabla_Gen_fit=zeros(fil_tab MG_2b,1);
332 for vari=1:length(Tabla_meteo_MG_2) % Se obtiene las matrices de la velocidad,
333     dirección y generación a partir de la tabla filtrada.
334         Tabla_Vel_Viento(vari,1)=cell2mat(Tabla_meteo_MG_2(vari,20));
335         Tabla_Direc_Viento(vari,1)=cell2mat(Tabla_meteo_MG_2(vari,15));
336         Tabla_Gen(vari,1)=cell2mat(Tabla_meteo_MG_2(vari,18));
337     end
338
339 %-----
340 % Se define una serie de parámetros del ajuste
341 %-----
342 ft = fittype('poly55'); % Se realiza un ajuste polinómico.
343 opts = fitoptions('Method','LinearLeastSquares'); % Este método ajusta los datos minimizando la suma
344 de los residuos d ellos puntos de la curva.
345 opts.Robust = 'LAR';
346
347 %-----
348 % Se realiza el ajuste del modelo a los datos
349 %-----
350 [fitobject, gof] = fit([Tabla_Vel_Viento,Tabla_Direc_Viento], Tabla_Gen, ft, opts); % fitobject es una función que recibe
351 como entradas la velocidad y dirección del viento, y proporciona como salida la generación.
352 figure(2)
353 plot(fitobject,[Tabla_Vel_Viento,Tabla_Direc_Viento],Tabla_Gen);
354 xlim([0 25])
355 ylim([0 360])
356 zlim([-100 4000])
```



```
357 % Titulos
358 xlabel Velocidad Viento(m/s)
359 ylabel Dirección_Viento(°)
360 zlabel Generación(kWh)
361 grid on
362 view( -88.7, 65.2 );
363
364 -----
365 % Se añade una nueva columna a la tabla de datos, con los datos de
366 % generación obtenidos a partir de la función fitobject.
367 -----
368 for fila_gen_fit=1:1:length(Tabla_meteo_MG_2)
369
370 Tabla_meteo_MG_2(fila_gen_fit,24)=num2cell(fitobject(cell2mat(Tabla_meteo_MG_2(fila_gen_fit,20)),cell2mat(Tabla_meteo_MG_2(fila_gen_fit,22))));
371
372 end
373 for vari2=1:1:length(Tabla_meteo_MG_2) % Se obtiene un vector que contiene los datos de
374 % generación que proporciona la función generada.
375 Tabla_Gen_fit(vari2,1)=cell2mat(Tabla_meteo_MG_2(vari2,24));
376 end
377
378 -----
379 %+++++
380 %
381 % NORMALIZACIÓN DE LOS VALORES DE LAS VARIABLES DE ENTRADA
382 % A LA RED NEURONAL ARTIFICIAL
383 %
384 %+++++
385 -----
386 %
387 % Se obtiene los máximos y los mínimos de las variables de interés
388 -----
389 % Generación eléctrica real del parque eólico
390 Max_Gen=max(Tabla_Gen);
391 Min_Gen=min(Tabla_Gen);
392 % Generación eléctrica obtenida a partir de la función ajustada
393 Max_Gen_Fit=max(Tabla_Gen_fit);
394 Min_Gen_Fit=min(Tabla_Gen_fit);
395 % Velocidad del viento
396 Max_Vel=max(Tabla_Vel_Viento);
397 Min_Vel=min(Tabla_Vel_Viento);
398 % Dirección del viento
399 Max_Dir=max(Tabla_Direc_Viento);
400 Min_Dir=min(Tabla_Direc_Viento);
401
402 -----
403 % Se procede a escalar los valores de las variables mediante el método del
404 % min-max.
405 %  $v_{jnormalizado} = a + (b-a) \cdot (v_j - \min(v_k)_{(k=1\dots N)}) / (\max(v_k)_{(k=1\dots N)} - \min(v_k)_{(k=1\dots N)})$ 
406 -----
407 for pnorm=1:1:length(Tabla_meteo_MG_2)
408 % Generación real y generación de la función fit [-1,1]
409 % Fórmula : Dato= -1 + ((1-(-1))*(dato-Min_Gen))/(Max_Gen-Min_Gen);
410 Tabla_meteo_MG_2(pnorm,26)=num2cell(-1+((1-(-1))*(cell2mat(Tabla_meteo_MG_2(pnorm,24))-Min_Gen_Fit))/(Max_Gen_Fit-Min_Gen_Fit));
411
412 Tabla_meteo_MG_2(pnorm,25)=num2cell(-1+((1-(-1))*(cell2mat(Tabla_meteo_MG_2(pnorm,18))-Min_Gen))/(Max_Gen-Min_Gen));
413 % Velocidad del viento [0 1]
414 % Fórmula : Dato= 0 + ((1-0)*(dato-Min_Vel))/(Max_Vel-Min_Vel);
415 Tabla_meteo_MG_2(pnorm,27)=num2cell(0+((1-0)*(cell2mat(Tabla_meteo_MG_2(pnorm,20))-Min_Vel))/(Max_Vel-Min_Vel));
416 % Dirección del viento [0 1]
417 % Fórmula : Dato= 0 + ((1-0)*(dato-Min_Dir))/(Max_Dir-Min_Dir);
418 Tabla_meteo_MG_2(pnorm,28)=num2cell(0+((1-0)*(cell2mat(Tabla_meteo_MG_2(pnorm,15))-Min_Dir))/(Max_Dir-Min_Dir));
419 end
420
421 %+++++
422 %
423 % ROSAS DEL VIENTO DEL PARQUE EÓLICO ES MILÁ
424 %
425 %+++++
426 -----
427 %
428 % Ordenación de los datos según la estación a la que pertenecen.
429 -----
430 [fil_tab_MG_2,col_tab_MG_2]=size(Tabla_meteo_MG_2); % Se guarda el número de filas y columnas de la
431 % tabla de datos filtrados.
432 aux2=zeros(fil_tab_MG_2,col_tab_MG_2);
433 Tabla_Primavera=num2cell(aux2); % Se define una tabla para cada estación del año
434 Tabla_Verano=num2cell(aux2);
435 Tabla_Otono=num2cell(aux2);
436 Tabla_Invierno=num2cell(aux2);
437 it_Primavera=1; % Se inician los contadores que permiten recorrer
438 % las filas de las tablas de las estaciones
439 it_Verano=1;
440 it_Otono=1;
441 it_Invierno=1; % Se sigue la siguiente clasificación de las
442 % estaciones del año.
443 %
444 % • Primavera: 20/03 - 20/06
445 % • Verano: 21/06 - 22/09
446 % • Otoño: 23/09 - 21/12
447 % • Invierno: 22/12 -19/03
448
449 for paso_Estaciones=1:1:length(Tabla_meteo_MG_2)
450 [Ano_Es,Mes_Es,Dia_Es,Hora_Es,Minuto_Es,Segundo_Es]=datevec(Tabla_meteo_MG_2(paso_Estaciones,1),'dd/mm/yyyy HH:MM:SS');
451
452 % Primavera
453 if (Mes_Es==3 && Dia_Es>=20) || Mes_Es==4 || Mes_Es==5 || (Mes_Es==6 && Dia_Es<=20))
454 Tabla_Primavera(it_Primavera,:)=Tabla_meteo_MG_2(paso_Estaciones,:);
455 it_Primavera=it_Primavera +1;
456 end
457 % Verano
458 if (Mes_Es==6 && Dia_Es>=21) || Mes_Es==7 || Mes_Es==8 || (Mes_Es==9 && Dia_Es<=22))
```



```
458 Tabla_Verano(it_Verano,:)=Tabla_meteo_MG_2(paso_Estaciones,:);
459 it_Verano=it_Verano +1;
460 end
461 % Otoño
462 if((Mes_Es==9 && Dia_Es>=23) || Mes_Es==10 || Mes_Es==11 || (Mes_Es==12 && Dia_Es<=21))
463 Tabla_Otono(it_Otono,:)=Tabla_meteo_MG_2(paso_Estaciones,:);
464 it_Otono=it_Otono +1;
465 end
466 % Invierno
467 if((Mes_Es==12 && Dia_Es>=22) || Mes_Es==1 || Mes_Es==2 || (Mes_Es==3 && Dia_Es<=19))
468 Tabla_Invierno(it_Invierno,:)=Tabla_meteo_MG_2(paso_Estaciones,:);
469 it_Invierno=it_Invierno +1;
470 end
471 end
472
473 %-----
474 % Se eliminan las filas sobrantes, de 0, de las tablas de cada estación.
475 %-----
476 Tabla_Invierno(6240:end,:)=[];
477 Tabla_Primavera(6233:end,:)=[];
478 Tabla_Verano(6705:end,:)=[];
479 Tabla_Otono(6079:end,:)=[];
480
481 %-----
482 % Se guarda la tabla generada para cada estación en una hoja de cálculo de
483 % Excel en formato xls y csv.
484 %-----
485 xlswrite('Tabla_Datos_Primavera_Filtrados',Tabla_Primavera);
486 writetable(cell2table(Tabla_Primavera),'Tabla_Datos_Primavera_Filtrados.csv','Delimiter','');
487 xlswrite('Tabla_Datos_Verano_Filtrados',Tabla_Verano);
488 writetable(cell2table(Tabla_Verano),'Tabla_Datos_Verano_Filtrados.csv','Delimiter','');
489 xlswrite('Tabla_Datos_Otono_Filtrados',Tabla_Otono);
490 writetable(cell2table(Tabla_Otono),'Tabla_Datos_Otono_Filtrados.csv','Delimiter','');
491 xlswrite('Tabla_Datos_Invierno_Filtrados',Tabla_Invierno);
492 writetable(cell2table(Tabla_Invierno),'Tabla_Datos_Invierno_Filtrados.csv','Delimiter','');
493
494 %-----
495 % Se procede a graficar la rosa de los vientos de cada estación y la global
496 %-----
497 % 1. Todos los datos
498 figure(1)
499 wind_rose(cell2mat(Tabla_meteo_MG_2(:,15)),cell2mat(Tabla_meteo_MG_2(:,20)));
500
501 % 2. Invierno
502 figure(2)
503 wind_rose(cell2mat(Tabla_Invierno(:,15)),cell2mat(Tabla_Invierno(:,20)));
504
505 % 3. Primavera
506 figure(3)
507 wind_rose(cell2mat(Tabla_Primavera(:,15)),cell2mat(Tabla_Primavera(:,20)));
508
509 % 4. Verano
510 figure(4)
511 wind_rose(cell2mat(Tabla_Verano(:,15)),cell2mat(Tabla_Verano(:,20)));
512
513 % 5. Otoño
514 figure(5)
515 wind_rose(cell2mat(Tabla_Otono(:,15)),cell2mat(Tabla_Otono(:,20)));
516
517 %-----
518 % SEGMENTACIÓN DE LOS DATOS: se corta los datos de las diferentes
519 % estaciones cuando se detecta huecos temporales mayores a 12 horas.
520 %-----
521 %-----
522 %-----
523 Inicio_Dataset=1;
524 Fin_Dataset=0;
525 Nombre_Excel=0;
526 Tabla_Fechas_Global = readtable(char(names_hist_MG(:,1)),'Delimiter','');
527 Fechas_Global=table2array(Tabla_Fechas_Global(1:end,1));
528 it_Fechas_Global=1;
529 my_directory='C:\Users\murti\Desktop\TFG\Análisis de dades\DatosMeteoDatosGeneración\CorteDatos';
530 cd(my_directory);
531
532 for paso_filtro6=1:1:length(Tabla_Otono) % Cambiar el nombre según la tabla de que estación
533 se quiere fragmentar.
534 cont_hueco=0;
535
536 [Ano_1,Mes_1,Dia_1,Hora_1,Minuto_1,Segundo_1]=datevec(Tabla_Otono(paso_filtro6),1,'dd/mm/yyyy HH:MM:SS');
537 [Ano_2,Mes_2,Dia_2,Hora_2,Minuto_2,Segundo_2]=datevec(Tabla_Otono(paso_filtro6+1),1,'dd/mm/yyyy HH:MM:SS');
538 Hora_1=Hora_1+1;
539 if(datetime(datestr([Ano_1,Mes_1,Dia_1,Hora_1,Minuto_1,Segundo_1]))==datetime(datestr([Ano_2,Mes_2,Dia_2,Hora_2,Minuto_2,Segund
540 o_2])))
541 continue
542 end
543 %Como mínimo hay un hueco de una hora.
544 Hora_1=Hora_1+1;
545
546 if(datetime(datestr([Ano_1,Mes_1,Dia_1,Hora_1,Minuto_1,Segundo_1]))==datetime(datestr([Ano_2,Mes_2,Dia_2,Hora_2,Minuto_2,Segund
547 o_2])))
548 cont_hueco=1;
549 end
550 Hora_1=Hora_1+1;
551
552 if(datetime(datestr([Ano_1,Mes_1,Dia_1,Hora_1,Minuto_1,Segundo_1]))==datetime(datestr([Ano_2,Mes_2,Dia_2,Hora_2,Minuto_2,Segund
553 o_2])))
554 cont_hueco=2;
555 end
556 Hora_1=Hora_1+1;
```



```
557
558 if (datetime(datestr([Ano_1,Mes_1,Dia_1,Hora_1,Minuto_1,Segundo_1]))==datetime(datestr([Ano_2,Mes_2,Dia_2,Hora_2,Minuto_2,Segund
559 o_2])))
560     cont_hueco=3;
561     end
562     Hora_1=Hora_1+1;
563
564 if (datetime(datestr([Ano_1,Mes_1,Dia_1,Hora_1,Minuto_1,Segundo_1]))==datetime(datestr([Ano_2,Mes_2,Dia_2,Hora_2,Minuto_2,Segund
565 o_2])))
566     cont_hueco=4;
567     end
568     Hora_1=Hora_1+1;
569
570 if (datetime(datestr([Ano_1,Mes_1,Dia_1,Hora_1,Minuto_1,Segundo_1]))==datetime(datestr([Ano_2,Mes_2,Dia_2,Hora_2,Minuto_2,Segund
571 o_2])))
572     cont_hueco=5;
573     end
574     Hora_1=Hora_1+1;
575
576 if (datetime(datestr([Ano_1,Mes_1,Dia_1,Hora_1,Minuto_1,Segundo_1]))==datetime(datestr([Ano_2,Mes_2,Dia_2,Hora_2,Minuto_2,Segund
577 o_2])))
578     cont_hueco=6;
579     end
580     Hora_1=Hora_1+1;
581
582 if (datetime(datestr([Ano_1,Mes_1,Dia_1,Hora_1,Minuto_1,Segundo_1]))==datetime(datestr([Ano_2,Mes_2,Dia_2,Hora_2,Minuto_2,Segund
583 o_2])))
584     cont_hueco=7;
585     end
586     Hora_1=Hora_1+1;
587
588 if (datetime(datestr([Ano_1,Mes_1,Dia_1,Hora_1,Minuto_1,Segundo_1]))==datetime(datestr([Ano_2,Mes_2,Dia_2,Hora_2,Minuto_2,Segund
589 o_2])))
590     cont_hueco=8;
591     end
592     Hora_1=Hora_1+1;
593
594 if (datetime(datestr([Ano_1,Mes_1,Dia_1,Hora_1,Minuto_1,Segundo_1]))==datetime(datestr([Ano_2,Mes_2,Dia_2,Hora_2,Minuto_2,Segund
595 o_2])))
596     cont_hueco=9;
597     end
598     Hora_1=Hora_1+1;
599
600 if (datetime(datestr([Ano_1,Mes_1,Dia_1,Hora_1,Minuto_1,Segundo_1]))==datetime(datestr([Ano_2,Mes_2,Dia_2,Hora_2,Minuto_2,Segund
601 o_2])))
602     cont_hueco=10;
603     end
604     Hora_1=Hora_1+1;
605
606 if (datetime(datestr([Ano_1,Mes_1,Dia_1,Hora_1,Minuto_1,Segundo_1]))==datetime(datestr([Ano_2,Mes_2,Dia_2,Hora_2,Minuto_2,Segund
607 o_2])))
608     cont_hueco=11;
609     end
610     Hora_1=Hora_1+1;
611
612 if (datetime(datestr([Ano_1,Mes_1,Dia_1,Hora_1,Minuto_1,Segundo_1]))==datetime(datestr([Ano_2,Mes_2,Dia_2,Hora_2,Minuto_2,Segund
613 o_2])))
614     cont_hueco=12;
615     end
616     Hora_1=Hora_1+1;
617
618 if (datetime(datestr([Ano_1,Mes_1,Dia_1,Hora_1,Minuto_1,Segundo_1]))==datetime(datestr([Ano_2,Mes_2,Dia_2,Hora_2,Minuto_2,Segund
619 o_2])))
620     cont_hueco=13;
621     end
622     if (cont_hueco==0)
623         cont_hueco=100;
624     end
625
626 %-----
627 % Descomentar uno de los if en función del tamaño de los huecos temporales
628 % que se quiere tolerar.
629 %-----
630 % Cortamos cuando tenemos un hueco temporal de más de una hora:
631 % if (cont_hueco>1)
632 %     Fin_Dataset=paso_filtro6;
633 %     Nombre_Excel=Nombre_Excel+1;
634 %     fileName1=[strcat('Data_Set_Corte_Datos_Excel',num2str(Nombre_Excel)),'.xls'];
635 %     fileName2=[strcat('Data_Set_Corte_Datos_Texto',num2str(Nombre_Excel)),'.csv'];
636 %     writetable(cell2table(Tabla_meteo_MG_2(Inicio_Dataset:Fin_Dataset,:)),fileName2,'Delimiter','');
637 %     xlswrite(fileName1,Tabla_meteo_MG_2(Inicio_Dataset:Fin_Dataset,:));
638 %     Inicio_Dataset=paso_filtro6+1;
639 % end
640 % Cortamos cuando tenemos un hueco temporal de más de dos horas. Toleramos huecos de dos horas:
641 % if (cont_hueco>2)
642 %     Fin_Dataset=paso_filtro6;
643 %     Nombre_Excel=Nombre_Excel+1;
644 %     fileName1=[strcat('Data_Set_Corte_Datos_Excel',num2str(Nombre_Excel)),'.xls'];
645 %     fileName2=[strcat('Data_Set_Corte_Datos_Texto',num2str(Nombre_Excel)),'.csv'];
646 %     writetable(cell2table(Tabla_meteo_MG_2(Inicio_Dataset:Fin_Dataset,:)),fileName2,'Delimiter','');
647 %     xlswrite(fileName1,Tabla_meteo_MG_2(Inicio_Dataset:Fin_Dataset,:));
648 %     Inicio_Dataset=paso_filtro6+1;
649 % end
650 % Cortamos cuando tenemos un hueco temporal de tres horas o más:
651 % if (cont_hueco>3)
652 %     Fin_Dataset=paso_filtro6;
653 %     Nombre_Excel=Nombre_Excel+1;
654 %     fileName1=[strcat('Data_Set_Corte_Datos_Excel',num2str(Nombre_Excel)),'.xls'];
655 %     fileName2=[strcat('Data_Set_Corte_Datos_Texto',num2str(Nombre_Excel)),'.csv'];
656 %     writetable(cell2table(Tabla_meteo_MG_2(Inicio_Dataset:Fin_Dataset,:)),fileName2,'Delimiter','');
657 %     xlswrite(fileName1,Tabla_meteo_MG_2(Inicio_Dataset:Fin_Dataset,:));
658 %     Inicio_Dataset=paso_filtro6+1;
```



```
658 % end
659 %Cortamos cuando tenemos un hueco temporal de cuatro horas o más:
660 % if(cont_hueco>4)
661 %     Fin_Dataset=paso_filtro6;
662 %     Nombre_Excel=Nombre_Excel+1;
663 %     fileName1=[strcat('Data_Set_Corte_Datos_Excel',num2str(Nombre_Excel)),'.xls'];
664 %     fileName2=[strcat('Data_Set_Corte_Datos_Texto',num2str(Nombre_Excel)),'.csv'];
665 %     writetable(cell2table(Tabla_meteo_MG_2(Inicio_Dataset:Fin_Dataset,:)),fileName2,'Delimiter','');
666 %     xlswrite(fileName1,Tabla_meteo_MG_2(Inicio_Dataset:Fin_Dataset,:));
667 %     Inicio_Dataset=paso_filtro6+1;
668 % end
669 %Cortamos cuando tenemos un hueco temporal de cuatro horas o más:
670 % if(cont_hueco>5)
671 %     Fin_Dataset=paso_filtro6;
672 %     Nombre_Excel=Nombre_Excel+1;
673 %     fileName1=[strcat('Data_Set_Corte_Datos_Excel',num2str(Nombre_Excel)),'.xls'];
674 %     fileName2=[strcat('Data_Set_Corte_Datos_Texto',num2str(Nombre_Excel)),'.csv'];
675 %     writetable(cell2table(Tabla_meteo_MG_2(Inicio_Dataset:Fin_Dataset,:)),fileName2,'Delimiter','');
676 %     xlswrite(fileName1,Tabla_meteo_MG_2(Inicio_Dataset:Fin_Dataset,:));
677 %     Inicio_Dataset=paso_filtro6+1;
678 % end
679 % if(cont_hueco>6)
680 %     Fin_Dataset=paso_filtro6;
681 %     Nombre_Excel=Nombre_Excel+1;
682 %     fileName1=[strcat('Data_Set_Corte_Datos_Excel',num2str(Nombre_Excel)),'.xls'];
683 %     fileName2=[strcat('Data_Set_Corte_Datos_Texto',num2str(Nombre_Excel)),'.csv'];
684 %     writetable(cell2table(Tabla_meteo_MG_2(Inicio_Dataset:Fin_Dataset,:)),fileName2,'Delimiter','');
685 %     xlswrite(fileName1,Tabla_meteo_MG_2(Inicio_Dataset:Fin_Dataset,:));
686 %     Inicio_Dataset=paso_filtro6+1;
687 % end
688 % if(cont_hueco>12)
689 %     Fin_Dataset=paso_filtro6;
690 %     Nombre_Excel=Nombre_Excel+1;
691 %     fileName1=[strcat('Data_Set_Corte_Datos_Excel',num2str(Nombre_Excel)),'.xls'];
692 %     fileName2=[strcat('Data_Set_Corte_Datos_Texto',num2str(Nombre_Excel)),'.csv'];
693 %     writetable(cell2table(Tabla_Otono(Inicio_Dataset:Fin_Dataset,:)),fileName2,'Delimiter','');
694 %     xlswrite(fileName1,Tabla_Otono(Inicio_Dataset:Fin_Dataset,:));
695 %     Inicio_Dataset=paso_filtro6+1;
696 % end
697 end
698
699 function createfigure(vel_viento1, Pot_generada1, S1, vel_viento2)
700 figure1 = figure;
701 axes1 = axes('Parent',figure1);
702 hold(axes1,'on');
703 scatter(vel_viento1,Pot_generada1,S1,vel_viento2,'MarkerFaceColor','auto');
704 ylabel('Energía generada(kWh)');
705 xlabel('Velocidad del viento(m/s)');
706 title('Velocidad del viento/Potencia generada');
707 ylim(axes1,[0 3500]);
708 end
709
710 function wind_rose(wind_direction,wind_speed)
711 figure
712 pax = polaraxes;
713 polarhistogram(deg2rad(wind_direction(wind_speed<150)),deg2rad(0:10:360),'FaceColor','black','displayname','50 - 150 m/s')
714 hold on
715 polarhistogram(deg2rad(wind_direction(wind_speed<50)),deg2rad(0:10:360),'FaceColor','red','displayname','25 - 50 m/s')
716 polarhistogram(deg2rad(wind_direction(wind_speed<25)),deg2rad(0:10:360),'FaceColor','yellow','displayname','11 - 25 m/s')
717 polarhistogram(deg2rad(wind_direction(wind_speed<11)),deg2rad(0:10:360),'FaceColor','green','displayname','3 - 11 m/s')
718 polarhistogram(deg2rad(wind_direction(wind_speed<3)),deg2rad(0:10:360),'FaceColor','blue','displayname','0 - 3 m/s')
719 pax.ThetaDir = 'clockwise';
720 pax.ThetaZeroLocation = 'top';
721 legend('Show')
722 title('Wind Rose')
723 end
```

B.3: REGENERADOR_AUTOMATICO_SERIES_TEMPORALES.M

Una vez finalizada la tarea de segmentación se ha obtenido múltiples archivos para cada estación. Estos archivos contienen huecos temporales, muy perjudiciales para aplicaciones de predicción. Este script se encarga de reestablecer la continuidad temporal de las variables.

```
1 %
2 %+++++
3 %+++++
4 % Trabajo de fin de grado: INGENIERÍA ELECTRÓNICA INDUSTRIAL Y AUTOMÁTICA
5 % Universidad de las Islas Baleares (UIB)
6 % Deva Murti Baer
```

```

7          % 2019/2020
8          %+++++
9          %+++++
10         %
11         % ----- DESCRIPCIÓN DE LA FUNCIÓN DEL SCRIPT -----
12         %
13         %
14         % - Se lee de forma sucesiva los archivos presentes en el directorio
15         % establecido.
16         % - Se determina la cantidad de huecos temporales presentes.
17         % - Se genera una variable que contiene todas las fechas (incluidas las
18         % faltantes).
19         % - Se rellena los huecos de los valores de la serie temporal con NaN.
20         % - Se substituye los NaN por valores obtenidos usando la media móvil.
21         % - Se representa la serie temporal regenerada y con huecos, de una
22         % variable (en este caso la temperatura) con el objetivo de visualizar la
23         % la senyal regenerada.
24         % - Este proceso se repite para todas las series temporales presentes en la
25         % carpeta.
26         %
27         %+++++
28         %+++++
29         %
30         % -----
31         % Se inicializa el espacio de memoria.
32         % -----
33         close all;
34         clear all;
35         clc;
36         t_start = cputime;
37         date_start = now;
38         %
39         % -----
40         % Mensaje de bienvenida
41         % -----
42         disp('-----');
43         disp('- Universidad de las Islas Baleares');
44         disp('- Trabajo de fin de grado: Deva Murti Baer (2019-2020)');
45         disp(str_txt);
46         disp('-----');
47         disp(' ');
48         %
49         % -----
50         % Directorio donde se hallan los datos
51         % -----
52         str_dir_input = pwd; % Se lee el directorio donde está ubicado el .m
53         str_hist_dir = strcat(str_dir_input, '\Corte_Datos_Primavera_Tolerancia_12h'); % Se obtiene la dirección de la carpeta que
54         contiene las hojas de cálculo con los datos.
55         % -----
56         % Se define el tipo de archivo al que se quiere acceder
57         % -----
58         file_type = '*.csv';
59         list_direc = dir(fullfile(str_hist_dir, file_type)); % Se obtiene una lista de todos los excels
60         incluidos en la carpeta
61         names_hist = {list_direc.name}; % Se obtiene una lista de todas las hojas de
62         cálculo que tienen la extensión .csv y se encuentran en la carpeta
63         names_hist_size = {list_direc.bytes}; % Se obtiene el tamaño en bytes de cada una de las
64         hojas de cálculo
65         num_archivos = length(names_hist); % Se obtiene la cantidad de archivos que cumplen
66         los requisitos
67         cd(str_hist_dir); % Se cambia el directorio de acción
68         % -----
69         % Se regenera todas las series temporales de todos los archivos presentes
70         % en la carpeta seleccionada mediante un bucle.
71         % -----
72         for Numero_Archivo=1:num_archivos
73             Nombre_Archivo=[strcat('Serie_Temp_Regenerada_Texto', num2str(Numero_Archivo)), '.csv'];
74             Nombre_Archivo2=[strcat('Serie_Temp_Regenerada_Excel', num2str(Numero_Archivo)), '.xls'];
75             % -----
76             % Se accede a los datos del archivo en bruto correspondiente.
77             % -----
78             Tabla_Huecos = readtable(char(names_hist(:, Numero_Archivo)));
79             [F_in, C_in] = size(Tabla_Huecos);
80             data.temp = zeros(F_in, C_in);
81             data.temp(:,1) = datenum(Char(table2array(Tabla_Huecos(:,1))), 'dd/mm/yyyy HH:MM:SS'); % Se convierte el tipo de datos Table a
82             double (en bruto), y la columna de las fechas en formato numérico.
83             for i=2:1:C_in
84                 data.temp(:,i) = table2array(Tabla_Huecos(:,i)); % Se carga los datos de la tabla en formato
85             double.
86             end
87             date_start = min(data.temp(:,1)); % Se obtiene el valor numérico de la primera fecha
88             de la serie temporal presente en el archivo
89             date_end = max(data.temp(:,1)); % Se obtiene el valor numérico de la última fecha
90             de la serie temporal presente en el archivo
91             % -----
92             % Se crea una nueva variable que contiene todas las fechas de hora en hora
93             % -----
94             date_list = [date_start:1/24:date_end]';
95             [F_total, C_total] = size(date_list);
96             data.in = zeros(F_total, C_in);
97             data.in(:,1) = date_list;
98             lost_data = F_total - F_in;
99             str_txt_1 = ['- Missing data: ', num2str(lost_data), ' [1-h steps]']; % Se muestra por pantalla la cantidad de huecos
100            temporales presentes.
101            disp(str_txt_1);
102            disp('-----');
103            [data.in(:, :)] = Create_full_time_series_v01(data.temp(:, :), date_start, date_end); % Se genera una nueva serie temporal con
104            todas las fechas y se rellena los huecos con NaN
105            window_size = 16; % Se define la cantidad de datos que se consideran
106            para hacer la media 'moviente' a la hora de rellenar los datos.
107

```



```
108 data.filled = fillmissing(data.in,'movmedian',window_size, 1); % data.filled contiene la totalidad d elas series
109 temporales rellenas con valores coherentes.
110
111 %-----
112 % Se representa la temperatura del archivo relleno respecto al archivo en
113 % bruto con el objetivo de comparar el aspecto de la regenerada y la que
114 % contiene huecos.
115 %-----
116 x=[1:1:F_total];
117 figure(Numero_Archivo);
118 subplot(2,1,1);
119 plot(data.in(:,1), data.in(:,2), 'b-');
120 grid on;
121 grid minor;
122 datetick('x','dd/mm/yyyy HH:MM');
123 xlabel('Date');
124 ylabel('Temperature [°C]');
125 title([strcat('DataSetCorteDatosTexto',num2str(Numero_Archivo)),'.csv']);
126 subplot(2,1,2);
127 plot(data.filled(:,1), data.filled(:,2), 'ro-', 'MarkerSize',2)
128 grid on;
129 grid minor;
130 datetick('x','dd/mm/yyyy HH:MM');
131 xlabel('Date');
132 ylabel('Temperature [°C]');
133 title([strcat('Serie Temporal Regenerada usando la Media Movable ',num2str(Numero_Archivo)),'.csv']);
134
135 %-----
136 % Se prepara la serie temporal regenerada para ser exportada
137 %-----
138 Tabla_Fechas=zeros(F_total,1);
139 Tabla_Fechas=(cellstr(datestr(data.filled(:,1),'dd/mm/yyyy HH:MM:SS')));
140 Tabla_Datos_Meteo=zeros(F_total,C_in-1);
141 Tabla_Datos_Meteo=data.filled(:,2:end);
142 Ser_Temp_Regen=[Tabla_Fechas,num2cell(Tabla_Datos_Meteo)];
143 writetable(cell2table(Ser_Temp_Regen),Nombre_Archivo,'Delimiter',' ');
144 xlswrite(Nombre_Archivo2,Ser_Temp_Regen);
145 end
146
147 %-----
148 % Rutina encargada para meter NaN en las horas donde faltan datos
149 %-----
150 function [data_out] = Create_full_time_series_v01(data_raw, date_start, date_end)
151 [F_in, C_in] = size(data_raw);
152 date_list = [date_start:1/24:date_end]';
153 [F_total, C_total] = size(date_list);
154 data_out = zeros(F_total,C_in);
155 data_out(:,1) = date_list;
156
157 index_inicial = 1;
158 index_final = 1;
159 num_lost_detect=0; % numero de datos que faltan
160
161 data_out(1,2:end)=data_raw(1,2:end);
162 for i=2:1:F_in
163     if (round(data_raw(i,1),5)>round(data_raw(i-1,1)+(1/24),5))
164         index_final = find(round(data_out(:,1),5)==round(data_raw(i,1),5));
165         index_inicial = find(round(data_out(:,1),5)==round(data_raw(i-1,1),5));
166         num_lost_detect = num_lost_detect + (index_final-index_inicial)-1; % Acumula el numero de datos faltantes
167         data_out(index_inicial,2:end) = data_raw(i-1,2:end);
168         for j=index_inicial+1:1:index_final-1
169             data_out(j,2:end)=NaN;
170         end
171         data_out(index_final,2:end)=data_raw(i,2:end);
172     else
173         index_data = find(round(data_out(:,1),5)==round(data_raw(i,1),5));
174         data_out(index_data,2:end) = data_raw(i,2:end);
175     end
176 end
177 end
```

B.4: ENTRENAMIENTO_VALIDACION_NARX.M

Este script se utiliza para obtener la configuración óptima de las cuatro redes neuronales que conforman el modelo de predicción desarrollado. Se entrena la red neuronal, y posteriormente se realiza una predicción de prueba que sirve para validar y optimizar los pesos sinápticos que se han obtenido durante la fase de entrenamiento.



```
1 %
2 %+++++
3 %+++++
4 % Trabajo de fin de grado: INGENIERÍA ELECTRÓNICA INDUSTRIAL Y AUTOMÁTICA
5 % Universidad de las Islas Baleares (UIB)
6 % Deva Murti Baer
7 % 2019/2020
8 %+++++
9 %+++++
10 %
11 % DESCRIPCIÓN DE LA FUNCIÓN DEL SCRIPT
12 % -----
13 %
14 % SELECCIÓN DE LA CONFIGURACIÓN DE LA NARX
15 % -----
16 % - Se lee el archivo de prueba, y se obtiene las variables de entrada y
17 % salida objetivo, para la red neuronal.
18 % - Se determina a que estación pertenece el archivo leído, y se carga el
19 % la configuración de la red obtenida entrenando con datos de esa estación.
20 %
21 % RED NEURONAL RECURRENTE NARX
22 % -----
23 % - Se define los parámetros de la red (tamaño de la ventana de retardos,
24 % el tamaño de la capa oculta, la función de aprendizaje, las entradas, las
25 % conexiones entre capas, etc.)
26 % - Se preparan los datos usando la función preparets
27 % - Se entrena la red.
28 %
29 % OPTIMIZACIÓN DE LA CONFIGURACIÓN DE LA RED
30 % -----
31 % - Se lee los datos del archivo de validación, se cierra el lazo de la red
32 % y se realiza una predicción de prueba.
33 % - Se define el periodo de predicción, el horizonte de predicción, etc.
34 % - Se evalúa el rendimiento de la predicción calculando diferentes
35 % métricas de error (RMSE, RMSPE,MAE,MAPE,NMSE...)
36 % - Se compara el valor del RMSE obtenido, con el que se ha obtenido
37 % anteriormente. Si es menor se guarda, y se guarda la configuración de la
38 % red.
39 % Este proceso se repite múltiples veces con el objetivo de optimizar la
40 % fase de entrenamiento. Se guarda la mejor configuración de la red en un
41 % .mat, y los resultados estadísticas obtenidos.
42 %
43 % Una vez guardado la configuración óptima de la red neuronal para una
44 % estación, se repite el proceso con datos de diferentes estaciones. Se
45 % cambia los archivos que de datos que se usa, y se guarda un nuevo .mat.
46 %+++++
47 %+++++
48 %
49 % -----
50 % Se inicializa el espacio de memoria.
51 % -----
52 %
53 % close all;
54 % clear all;
55 % clc;
56 % t_start = cputime;
57 % date_start = now;
58 %
59 % -----
60 % Mensaje de bienvenida
61 % -----
62 % disp('-----');
63 % disp('- Universidad de las Islas Baleares');
64 % disp('- Trabajo de fin de grado: Deva Murti Baer (2019-2020)');
65 % disp(str_txt);
66 % disp('-----');
67 %
68 % -----
69 % Directorio donde se hallan los datos
70 % -----
71 % str_dir_input = pwd; % Se lee el directorio donde está ubicado el .m
72 % str_hist_dir = strcat(str_dir_input, '\DATA ENTRENAMIENTO TEST PRIMAVERA'); % Se obtiene la dirección de la carpeta que
73 % contiene las hojas de cálculo con los datos.
74 % -----
75 % Se define el tipo de archivo al que se quiere acceder
76 % -----
77 % file_type = '*.csv';
78 % list_direc = dir(fullfile(str_hist_dir, file_type)); % Se obtiene una lista de todos los excels
79 % incluidos en la carpeta
80 % names_hist = {list_direc.name}; % Se obtiene una lista de todas las hojas de
81 % cálculo que tienen la extensión .csv y se encuentran en la carpeta
82 % names_hist_size = {list_direc.bytes}; % Se obtiene el tamaño en bytes de cada una de las
83 % hojas de cálculo
84 % num_archivos = length(names_hist); % Se obtiene la cantidad de archivos que cumplen
85 % los requisitos
86 % cd(str_hist_dir); % Se cambia el directorio de acción
87 %
88 % Se define las variables que se utilizaran para guardar las variables de
89 % entrada y los targets, de los diferentes archivos que se van a utilizar
90 % para la fase de entrenamiento.
91 % -----
92 %
93 % Se definen 4 matrices de entradas
94 % -----
95 % x1 = {};
96 % x2 = {};
97 % x3 = {};
98 % x4 = {};
99 % x5 = {};
100 % -----
101 % Se definen 4 matrices de salidas objetivo (targets)
```



```
102 %-----
103 t1 = {};
104 t2 = {};
105 t3 = {};
106 t4 = {};
107 t5 = {};
108
109 num_entramientos = 10; % Se define las rondas de entrenamiento que se
110 quiere realizar.
111 rmse_best = 10000000000000000; % Se inicializa el valor del RMSE que se pretende
112 mejorar en las rondas de entrenamiento.
113
114 for kkk=1:1:num_entramientos
115 %-----
116 % Se accede a los datos del archivo de interés. Se accederá a los datos de
117 % cuatro archivos de la estación de invierno.
118 %-----
119 for Numero_Archivo=1:5 %Prueba
120 % for Numero_Archivo=1:num_archivos-1
121 Tabla_Archivo = readtable(char(names_hist(:,Numero_Archivo))); % Se accede a la tabla de los datos.
122 [F_in, C_in] = size(Tabla_Archivo); % Se determina el número de filas y columnas de la
123 tabla.
124 data.fecha = zeros(F_in, C_in); % Se crea una variable de fechas.
125 data.fecha(:,1) = datenum(char(table2array(Tabla_Archivo(:,1)), 'dd/mm/yyyy HH:MM:SS')); % Se convierte el tipo de datos
126 Table a double (en bruto), y la columna de las fechas en formato numérico.
127 for i=2:1:C_in
128 data.fecha(:,i) = table2array(Tabla_Archivo(:,i)); % Se carga los datos de la tabla en formato
129 double.
130 end
131 date_start = min(data.fecha(:,1)); % Se obtiene el valor numérico de la primera
132 fecha de la serie temporal presente en el archivo
133 date_end = max(data.fecha(:,1)); % Se obtiene el valor numérico de la última
134 fecha de la serie temporal presente en el archivo
135 %-----
136 % Se obtiene los datos de la tabla.
137 %-----
138 Tabla_Fechas=zeros(F_in,1);
139 Tabla_Fechas=(cellstr(datestr(data.fecha(:,1), 'dd/mm/yyyy HH:MM:SS'))); % Se obtiene la columna de fechas de la tabla
140 Tabla_Datos_Meteo=zeros(F_in,C_in-1);
141 Tabla_Datos_Meteo=table2array(Tabla_Archivo(:,2:end)); % Se obtiene la tabla de datos numéricos de la
142 tabla.
143 Ser_Temp_Regen=[Tabla_Fechas,num2cell(Tabla_Datos_Meteo)]; % Se concatenan las dos tablas anteriores, en
144 formato de celda ya que es el único que puede leer la red neuronal.
145
146 %-----
147 % Se define las variables de entrada y las salidas objetivo, para la fase de
148 % entrenamiento. Los valores cambiarán en función del archivo al que se
149 % está accediendo.
150 %-----
151 X_Input_train = Ser_Temp_Regen(:,26:28)';
152 T_target_train = Ser_Temp_Regen(:,25)';
153 %-----
154 % SWITCH-CASE: el objetivo es obtener las variables de entrada y salida
155 % objetivo de todos los archivos a los que se quiere acceder.
156 %-----
157 Archivo_Abierto = Numero_Archivo;
158 switch Archivo_Abierto
159 %-----
160 % Se guardan las variables de entrada y salida de los diferentes
161 % archivos a los que se va accediendo.
162 %-----
163 case 1
164 x1 = X_Input_train;
165 t1 = T_target_train;
166 case 2
167 x2 = X_Input_train;
168 t2 = T_target_train;
169 case 3
170 x3 = X_Input_train;
171 t3 = T_target_train;
172 case 4
173 x4 = X_Input_train;
174 t4 = T_target_train;
175 case 5
176 x5 = X_Input_train;
177 t5 = T_target_train;
178 end
179 end
180 %-----
181 % Se utiliza la función de MATLAB catesamples para concatenar las variables
182 % de entrada y salidas de los diferentes archivos, con el objetivo de poder
183 % entrenar la red neuronal con más datos.
184 %-----
185 X_Input_train = catsamples(x1,x2,x3,x4,x5, 'pad');
186 T_target_train = catsamples(t1,t2,t3,t4,t5, 'pad');
187 %-----
188 % Se definen los parámetros de la red neuronal
189 %-----
190 inputDelays = 1:24; % Se almacena los 36 últimos valores de las
191 entradas en el shiftregister
192 feedbackDelays = 1:24; % Se realimenta los valores de las 36 últimas
193 predicciones
194 hiddenLayerSize = 15; % Numero de neuronas Tansig en la capa hidden, a
195 la salida siempre hay una neurona lineal
196 trainFcn = 'trainlm'; % Algoritmo de aprendizaje utilizado: Levenberg-
197 Marquardt
198 net = narnet(inputDelays,feedbackDelays,hiddenLayerSize, 'open',trainFcn); % Se define la topología de la red neuronal
199 view(net)
200 net.performFcn = 'mse'; % Mean squared error (MSE)
201 net.numInputs = 4; % Se define el número de entradas y el nombre de
202 cada una de ellas
```

```

203 net.inputs{1}.name='x1';
204 net.inputs{2}.name='y';
205 net.inputs{3}.name='x2';
206 net.inputs{4}.name='x3';
207 net.inputConnect = logical([1 1 1 1 ; 0 0 0 0]); % Se define las conexiones de cada entrada hacia
208 las diferentes capas de la red neuronal
209 %net.inputConnect=logical([1 1 1 1 ; 0 0 0 0]); % Se define las conexiones de cada entrada hacia
210 las diferentes capas de la red neuronal
211 net.inputWeights{1,1}.delays=inputDelays; % Se define los retardos de las variables
212 net.inputWeights{1,2}.delays=feedbackDelays;
213 net.inputWeights{1,3}.delays=inputDelays;
214 net.inputWeights{1,4}.delays=inputDelays;
215 view(net); % Se visualiza la red neuronal definida hasta este
216 punto.
217 [Xs,Xi,Ai,Ts] = preparets(net, X_Input_train, {}, T_target_train);
218 %Xs: contiene los datos de entrada, menos los que
219 se utilizan para rellenar la ventana temporal
220 %Xi: contiene los valores de entrada de la ventana
221 temporal. En este caso se trata de un vector de 4 filas de 36 valores
222 %Ai: en el caso expuesto se encuentra vacío. Se
223 utiliza cuando se define una NARX sin realimentación
224 %Ts: contiene los valores de salida objetivo
225 desplazados
226
227 %rng('default') % Añadido por reproducibilidad
228 net.divideFcn='';
229 net.trainParam.min_grad = 0.0003; % Se define el valor del gradiente mínimo
230 objetivo.
231 %net.divideFcn='divideblock';
232 % net.trainParam.show = 50;
233 % net.trainParam.lr = 0.05;
234 net.trainParam.epochs = 1000; % Se define el número de iteraciones que se
235 realizarán en cada etapa de entrenamiento.
236 % net.trainParam.goal = 1e-3;
237 [neto, tro, Yo, Eo, Xof, Aof] = train(net, Xs, Ts, Xi, Ai); % Entrenamiento de la red neuronal. neto contiene
238 la estructura de la red entrenada.
239 view(neto); % Se presenta la arquitectura de la red neuronal
240 generada.
241 % Red_Entrenada_Pasol=neto;
242 % save ('Red_Entrenada_Pasol.mat', 'Red_Entrenada_Pasol');
243 w1_Pasol = neto.IW{1}; % Pesos de la entrada hacia las neuronas de la
244 capa oculta.
245 w2_Pasol = neto.LW{2}; % Pesos de la capa oculta hacia la salida.
246 nnet.guis.closeAllViews();
247
248
249
250 %-----
251 % Testeo de la red neuronal: el objetivo es realizar múltiples
252 % entrenamientos que permitan optimizar la configuración de la red neuronal
253 % recurrente.
254 %-----
255 cd(str_hist_dir) % Se inicializa la dirección de memoria.
256 Test_file_number = 6; % Se selecciona el archivo que se utilizará en la
257 prueba.
258 Tabla_Archivo = readtable(char(names_hist(:, Test_file_number)));
259 [F_in, C_in] = size(Tabla_Archivo);
260 data.fecha = zeros(F_in, C_in);
261 data.fecha(:,1) = datenum(Char(table2array(Tabla_Archivo(:,1))), 'dd/mm/yyyy HH:MM:SS'); % Se convierte el tipo de datos
262 Table a double (en bruto), y la columna de las fechas en formato numérico.
263 for i=2:1:C_in
264     data.fecha(:,i) = table2array(Tabla_Archivo(:,i)); % Se carga los datos de la tabla en formato
265     double.
266 end
267 date_start = min(data.fecha(:,1)); % Se obtiene el valor numérico de la primera fecha
268 de la serie temporal presente en el archivo
269 date_end = max(data.fecha(:,1)); % Se obtiene el valor numérico de la última fecha
270 de la serie temporal presente en el archivo
271 %-----
272 % Se obtiene los datos de la tabla.
273 %-----
274 Tabla_Fechas = zeros(F_in,1);
275 Tabla_Fechas = (cellstr(datestr(data.fecha(:,1), 'dd/mm/yyyy HH:MM:SS')));
276 Tabla_Datos_Meteo = zeros(F_in, C_in-1);
277 Tabla_Datos_Meteo = table2array(Tabla_Archivo(:,2:end));
278 Ser_Temp_Regen = [Tabla_Fechas, num2cell(Tabla_Datos_Meteo)]; % Se obtiene la tabla de datos del archivo en
279 formato celda.
280
281 %-----
282 % Se define las variables de entrada y las salidas objetivo, para la fase
283 % de testeo de la red neuronal recurrente.
284 %-----
285 X_Input_test = Ser_Temp_Regen(:,26:28); % Se seleccionan las columnas de las variables de
286 entrada (26=generación de la función ajustada, % 27 y 28 velocidad y dirección del viento
287 normalizados.)
288 T_target_test = Ser_Temp_Regen(:,25); % Se selecciona los valores normalizados de la
289 generación eléctrica real del parque eólico, como variables objetivo
290 %-----
291 % Se cierra el lazo de la red neuronal entrenada anteriormente y se realiza
292 % la predicción para los datos del archivo de prueba.
293 %-----
294 netc = closeloop(neto); % neto contiene la estructura de la NARX entrenada
295 anteriormente, pero con lazo cerrado.
296 netc.name = [net.name ' - Closed Loop']; % Se define el nombre de la red netc.
297 view(netc); % Se visualiza la arquitectura de la NARX en lazo
298 cerrado.
299 %-----
300 % Se define el horizonte de predicción, el tamaño del archivo de testeo y
301 % el tamaño de la ventana temporal de delays.
302 %-----
303 Num_hour_forecast = 24; % Se define el horizonte de predicción

```



```
304 step_test = 24;
305 Num_lags = 24; % Se define el tamaño de la ventana temporal de la
306 red neuronal.
307 %-----
308 % Se define el rango temporal para el cual se quiere realizar las
309 % predicciones.
310 %-----
311 start_hour = 48; % Se define la hora de inicio de predicción
312 end_hour = 336; % Se define la hora de finalización de la
313 predicción
314 %-----
315 % Se define el valor mínimo y el máximo de la generación eléctrica real,
316 % para poder reconvertir los valores escalados durante la fase de
317 % normalización
318 %-----
319 power_min = -101; % Mínimo de la generación eléctrica del parque[kW]
320 power_max = 3138; % Máximo de la generación eléctrica del parque[kW]
321
322 %-----
323 % Se define el tamaño de una variable que contendrá los valores de las
324 % métricas de evaluación de las predicciones, y una variable que permitirá
325 % saber el paso de predicción en que se encuentra el programa
326 %-----
327 resultats_estadistics = zeros(length(start_hour:24:end_hour),7);
328 medias_result_estad= zeros (1,7);
329 index_result = 1;
330
331 %-----
332 % Se define un bucle de predicción, con el objetivo de realizar las
333 % predicciones para el rango de datos definido anteriormente, de forma
334 % automática.
335 %-----
336 for kk = start_hour:step_test:end_hour
337 [xc,xic,aic,tc] = preparets(netc, X_Imput_test(:,kk-24-1:kk+Num_hour_forecast-2), {},T_target_test(:,kk-24-
338 1:kk+Num_hour_forecast-2));
339 %-----
340 % Se utiliza la función de preparets para preparar los datos de cada
341 % predicción que se realiza.
342 %-----
343 % [xc,xic,aic,tc] = preparets(netc, X_Imput_train(:,kk-24-1:kk+Num_hour_forecast-1), {},T_target_train(:,kk-24-
344 1:kk+Num_hour_forecast-1));
345 yc = netc(xc, xic, aic); % Se obtiene los valores de predicción
346 perfc = perform(netc,tc,yc); % Se obtiene el MSE de la predicción
347
348 %hores_axis = [kk-24:1:kk+Num_hour_forecast-1] % Se definen ejes de representación para poder
349 graficar las predicciones
350 % %hores_axis_2 = [kk:1:kk+Num_hour_forecast-1];
351 % hores_axis_2 = [kk:1:kk+Num_hour_forecast];
352
353 predicco = cell2mat(yc); % Se obtiene los valores de predicción.
354 target = cell2mat(tc); % Se obtiene los valores de la salida objetivo.
355 predicco(predicco<-1.0) = -1; % Se limita el rango de posibles valores de
356 predicción, ya que en algunos caso diverge.
357
358 errors_fore = (1619.5*target+1518.5)-(1619.5*predicco+1518.5); % Se calcular el error como la diferencia entre la
359 predicción y el objetivo [kW]
360 tc_2 = (1619.5*cell2mat(tc)'+1518.5); % [kW] de error
361 MSE00o2 = mean(var((1619.5*target+1518.5)',1)); % Referencia de normalización.
362 MSE_Eo2 = mse(errors_fore); % MSE para la salida de entrenamiento
363 NMSEo2 = mse(errors_fore)/MSE00o2; % NMSE (normalized mean square error) -> link:
364 https://math.stackexchange.com/questions/488964/the-definition-of-nmse-normalized-mean-square-error
365 R2o2 = 1 - NMSEo2; % Rsquared
366
367 %-----
368 % Se calcula las siguientes métricas de error:
369 % MAE (Mean Absolute Error)
370 % MAPE (Mean Absolute Percentage Error)
371 % RMSE (Root Mean Square Error)
372 % MSPE
373 % RMSPE (Root Mean Square Percentage Error)
374 %-----
375
376 mape_res1 = abs((errors_fore)./tc_2);
377 mape_fore = ((1/length(errors_fore))*sum(mape_res1))*100; % Sus unidades son [%]
378 mae_fore = (1/length(errors_fore))*sum(abs(errors_fore)); % Sus unidades son [W]
379 rmse_fore = (sum((errors_fore).^2)/length(errors_fore))^0.5; % Sus unidades son [W]
380 mspe_res1=((errors_fore)./tc_2).^2;
381 mspe_fore=((1/length(errors_fore))*sum(mspe_res1))*100;
382 rmspe_fore=sqrt(mspe_fore);
383 %-----
384 % Se muestra por pantalla las métricas de error calculadas para la previsión.
385 %-----
386 str_txt = ['Start hour: ', num2str(kk), ' [h] | End hour: ', num2str(kk+Num_hour_forecast-1), ' [h] |', 'Iteración: ',
387 num2str(index_result)];
388 disp(str_txt);
389 disp('-----')
390 str_rmse_fore=['RMSE (Root Mean Square Error) de la previsión: ', num2str(rmse_fore),' [kWh]'];
391 disp(str_rmse_fore);
392 str_mse_fore=['MSE (Mean Squared Error) de la previsión: ', num2str(MSE_Eo2)];
393 disp(str_mse_fore);
394 str_nmse_fore=['NMSE (Normaliced Mean Squared Error) de la previsión: ', num2str(NMSEo2)];
395 disp(str_nmse_fore);
396 str_mae_fore=['MAE (Mean Absolute Error) de la previsión: ', num2str(mae_fore),' [kWh]'];
397 disp(str_mae_fore);
398 str_mape_fore=['MAPE (Mean Absolute Percentage Error) de la previsión: ', num2str(mape_fore),' [%]'];
399 disp(str_mape_fore);
400 str_rmspe_fore=['RMSPE (Root Mean Square Percentage Error)de la previsión: ', num2str(rmspe_fore),' [%]'];
401 disp(str_rmspe_fore);
402 str_rsquared_fore=['R^2 (R squared)de la previsión: ', num2str(R2o2)];
403 disp(str_rsquared_fore);
404 disp('-----')
```

```

405 %-----
406 % Se guarda las diferentes métricas de error en la variable creada.
407 %-----
408 resultats_estadistics(index_result,1) = MSE_Eo2;
409 resultats_estadistics(index_result,2) = NMSEo2;
410 resultats_estadistics(index_result,3) = R2o2;
411 resultats_estadistics(index_result,4) = mape_fore;
412 resultats_estadistics(index_result,5) = mae_fore;
413 resultats_estadistics(index_result,6) = rmse_fore;
414 resultats_estadistics(index_result,7) = rmspe_fore;
415
416
417 index_result = index_result + 1; % Se incrementa el indice.
418 end
419
420 %-----
421 % Se obtiene la media final de los errores para el archivo
422 % (para previsiones de 24h)
423 %-----
424 disp('-----');
425 disp('- Valores medios de los resultados');
426 disp('-----');
427
428 %-----
429 % Media de los MSE
430 %-----
431 str_mse_fore=['MSE: ', num2str(mean(resultats_estadistics(:,1))), ' '];
432 disp(str_mse_fore);
433 %-----
434 % Media de los NMSE
435 %-----
436 str_nmse_fore=['NMSE: ', num2str(mean(resultats_estadistics(:,2))), ' '];
437 disp(str_nmse_fore);
438 %-----
439 % Media de los R^2
440 %-----
441 str_rsquared_fore=['RSQUARED: ', num2str(mean(resultats_estadistics(:,3))), ' '];
442 disp(str_rsquared_fore);
443 %-----
444 % Media de los MAPE
445 %-----
446 str_mape_fore=['MAPE (mean absolute percentage error) de la prevision: ', num2str(mean(resultats_estadistics(:,4))), ' [%]'];
447 disp(str_mape_fore);
448 %-----
449 % Media de los MAE
450 %-----
451 str_mae_fore=['MAE (mean absolute error) de la prevision: ', num2str(mean(resultats_estadistics(:,5))), ' [kWh]'];
452 disp(str_mae_fore);
453 %-----
454 % Media de los RMSE
455 %-----
456 str_rmse_fore=['RMSE (root-mean-square error) de la prevision: ', num2str(mean(resultats_estadistics(:,6))), ' [kWh]'];
457 disp(str_rmse_fore);
458 disp('-----');
459 %-----
460 % Se guarda los valores medios de las métricas de error para aquella
461 % predicción.
462 %-----
463 medias_result_estad(1,1) = mean(resultats_estadistics(:,1)); % MSE
464 medias_result_estad(1,2) = mean(resultats_estadistics(:,2)); % NMSE
465 medias_result_estad(1,3) = mean(resultats_estadistics(:,3)); % R^2
466 medias_result_estad(1,4) = mean(resultats_estadistics(:,4)); % MAPE
467 medias_result_estad(1,5) = mean(resultats_estadistics(:,5)); % MAE
468 medias_result_estad(1,6) = mean(resultats_estadistics(:,6)); % RMSE
469 medias_result_estad(1,7) = mean(resultats_estadistics(:,7)); % RMSPE
470 %-----
471 % Se compara la media de los RMSE (MSE_Eo2^0.5) con el mejor valor de RMSE
472 % disponible. Si la media es mejor, se guarda la configuración de la red
473 % como nueva mejor estructura. En caso contrario, no se guarda.
474 %-----
475 if (mean(resultats_estadistics(:,6)) < rmse_best)
476     rmse_best = mean(resultats_estadistics(:,6));
477     Mejor_Arquitectura_RNN_Primavera=neto;
478     save ('Mejor_Arquitectura_RNN_Primavera.mat','Mejor_Arquitectura_RNN_Primavera');
479     xlswrite('Tabla_Medias_Mejores_Resultados_Primavera',medias_result_estad);
480 %-----
481 % Se guarda las métricas de error para la fase de testeo para el
482 % mejor entrenamiento.
483 %-----
484 xlswrite('Resultados_Estadisticos_Mejor_Entrenamiento_Primavera',resultats_estadistics);
485 end
486 end

```

B.5: TEST_DE_NARX_ENTRENADAS.M

Este script se utiliza para realizar las predicciones de generación eléctrica del parque eólico. Se carga una de las redes neuronales entrenadas, en función de la estación a la que pertenece el archivo de prueba seleccionado, y se utiliza para predecir. Se obtiene las curvas de predicción en función del horizonte seleccionado, y las métricas de error para poder valorar los resultados.

```

1 %
2 %+++++
3 %+++++
4 % Trabajo de fin de grado: INGENIERÍA ELECTRÓNICA INDUSTRIAL Y AUTOMÁTICA
5 % Universidad de las Islas Baleares (UIB)
6 % Deva Murti Baer
7 % 2019/2020
8 %+++++
9 %+++++
10 %
11 % DESCRIPCIÓN DE LA FUNCIÓN DEL SCRIPT
12 % -----
13 %
14 % SELECCIÓN DE LA CONFIGURACIÓN DE LA NARX
15 % -----
16 % - Se lee el archivo de prueba, y se obtiene las variables de entrada y
17 % salida objetivo, para la red neuronal.
18 % - Se determina a que estación pertenece el archivo leído, y se carga el
19 % la configuración de la red obtenida entrenando con datos de esa estación.
20 %
21 % PARÁMETROS DE PREDICCIÓN
22 % -----
23 % - Se establece parámetros de predicción como el horizonte de predicción,
24 % el número de lags (depende de la ventana temporal de retardos
25 % establecida), la hora de inicio y la hora final de predicción, el valor
26 % mínimo y máximo de la generación eléctrica real, etc.
27 %
28 % PREDICCIÓN
29 % -----
30 % - Se realiza la predicción, se obtiene las métricas de error
31 % (RMSE,RMSPE,MAE,MAPE, MSE..), y se representa la curva de generación
32 % real y la predicha para cada una de las predicciones.
33 %
34 %+++++
35 %+++++
36 %
37 %-----
38 % Se inicializa el espacio de memoria.
39 %-----
40 close all;
41 clear all;
42 clc;
43 t_start = cputime;
44 date_start = now;
45 %-----
46 % Mensaje de bienvenida
47 %-----
48 disp('-----');
49 disp('- Universidad de las Islas Baleares');
50 disp('- Trabajo de fin de grado: Deva Murti Baer (2019-2020)');
51 disp(str_txt);
52 disp('-----');
53 disp(' ');
54 %-----
55 % Directorio donde se hallan los datos
56 %-----
57 str_dir_input = pwd; % Se lee el directorio donde está ubicado el .m
58 str_hist_dir = strcat(str_dir_input, '\DATA_ENTRENAMIENTO_TEST_INVIERNO'); % Se obtiene la dirección de la carpeta que
59 contiene las hojas de cálculo con los datos.
60 %-----
61 % Se define el tipo de archivo al que se quiere acceder
62 %-----
63 file_type = '*.csv';
64 list_direc = dir(fullfile(str_hist_dir, file_type)); % Se obtiene una lista de todos los excels
65 incluidos en la carpeta
66 names_hist = {list_direc.name}; % Se obtiene una lista de todas las hojas de
67 cálculo que tienen la extensión .csv y se encuentran en la carpeta
68 names_hist_size = {list_direc.bytes}; % Se obtiene el tamaño en bytes de cada una de las
69 hojas de cálculo
70 num_archivos = length(names_hist); % Se obtiene la cantidad de archivos que cumplen
71 los requisitos
72 cd(str_hist_dir); % Se cambia el directorio de acción
73 cd(str_hist_dir)
74 %-----
75 % Se accede a los datos del archivo que se va a usar para testear la red
76 % neuronal NARX, que ha sido entrenada previamente.
77 %-----
78 Test_file_number = 6;
79 Tabla_Archivo = readtable(char(names_hist(:,Test_file_number)));

```



```
80 [F_in, C_in] = size(Tabla_Archivo);
81 data.fecha = zeros(F_in, C_in);
82 data.fecha(:,1) = datenum(char(table2array(Tabla_Archivo(:,1))), 'dd/mm/yyyy HH:MM:SS'); % Se convierte el tipo de datos
83 Table a double (en bruto), y la columna de las fechas en formato numérico.
84 for i=2:1:C_in
85     data.fecha(:,i) = table2array(Tabla_Archivo(:,i)); % Se carga los datos de la tabla en formato
86     double.
87 end
88 date_start = min(data.fecha(:,1)); % Se obtiene el valor numérico de la primera fecha
89 de la serie temporal presente en el archivo
90 date_end = max(data.fecha(:,1)); % Se obtiene el valor numérico de la última fecha
91 de la serie temporal presente en el archivo
92 %-----
93 % Se obtiene los datos de la tabla.
94 %-----
95 Tabla_Fechas = zeros(F_in,1);
96 Tabla_Fechas = (cellstr(datestr(data.fecha(:,1), 'dd/mm/yyyy HH:MM:SS')));
97 Tabla_Datos_Meteo = zeros(F_in,C_in-1);
98 Tabla_Datos_Meteo = table2array(Tabla_Archivo(:,2:end));
99 Ser_Temp_Regen = [Tabla_Fechas,num2cell(Tabla_Datos_Meteo)];
100 %-----
101 % Se determina la estación a la que pertenece el archivo para saber que
102 % estructura de red neuronal entrenada hay que cargar para realizar as
103 % predicciones
104 %-----
105 [Ano_Es,Mes_Es,Dia_Es,Hora_Es,Minuto_Es,Segundo_Es]=datevec(Tabla_Fechas(1,1), 'dd/mm/yyyy HH:MM:SS');
106
107 Variable_Estacion=0;
108
109 % Primavera
110 if((Mes_Es==3 && Dia_Es>=20) || Mes_Es==4 || Mes_Es==5 || (Mes_Es==6 && Dia_Es<=20))
111     Variable_Estacion=1;
112 end
113 % Verano
114 if((Mes_Es==6 && Dia_Es>=21) || Mes_Es==7 || Mes_Es==8 || (Mes_Es==9 && Dia_Es<=22))
115     Variable_Estacion=2;
116 end
117 % Otoño
118 if((Mes_Es==9 && Dia_Es>=23) || Mes_Es==10 || Mes_Es==11 || (Mes_Es==12 && Dia_Es<=21))
119     Variable_Estacion=3;
120 end
121 % Invierno
122 if((Mes_Es==12 && Dia_Es>=22) || Mes_Es==1 || Mes_Es==2 || (Mes_Es==3 && Dia_Es<=19))
123     Variable_Estacion=4;
124 end
125
126 %-----
127 % Se carga la red neuronal de la estación que correspondiente.
128 %-----
129 switch Variable_Estacion
130 case 1
131     cd('C:\Users\murti\Desktop\TFG\Análisis de
132     datos\DatosMeteoDatosGeneración\Archivos_Entrenamiento_Test\DATA_ENTRENAMIENTO_TEST_PRIMAVERA')
133     load ('Mejor_Arquitectura_RNN_Primavera.mat')
134     netc = closeloop(Mejor_Arquitectura_RNN_Primavera);
135     view(netc);
136 case 2
137     cd('C:\Users\murti\Desktop\TFG\Análisis de
138     datos\DatosMeteoDatosGeneración\Archivos_Entrenamiento_Test\DATA_ENTRENAMIENTO_TEST_VERANO')
139     load ('Mejor_Arquitectura_RNN_Verano.mat')
140     netc = closeloop(Mejor_Arquitectura_RNN_Verano);
141     view(netc);
142 case 3
143     cd('C:\Users\murti\Desktop\TFG\Análisis de
144     datos\DatosMeteoDatosGeneración\Archivos_Entrenamiento_Test\DATA_ENTRENAMIENTO_TEST_OTONO')
145     load ('Mejor_Arquitectura_RNN_Otono.mat')
146     netc = closeloop(Mejor_Arquitectura_RNN_Otono);
147     view(netc);
148 case 4
149     cd('C:\Users\murti\Desktop\TFG\Análisis de
150     datos\DatosMeteoDatosGeneración\Archivos_Entrenamiento_Test\DATA_ENTRENAMIENTO_TEST_INVIERNO')
151     load ('Mejor_Arquitectura_RNN_Invierno.mat')
152     netc = closeloop(Mejor_Arquitectura_RNN_Invierno);
153     view(netc);
154 end
155 %-----
156 % Se define las variables de entrada y las salidas objetivo, para la fase
157 % de test.
158 %-----
159 X_Input_Test = Ser_Temp_Regen(:,26:28)';
160 T_target_Test = Ser_Temp_Regen(:,25)';
161 %
162 %-----
163 % Se define el horizonte de predicción, el tamaño del archivo de testeo y
164 % el tamaño de la ventana temporal de delays.
165 %-----
166 Num_hour_forecast = 72;
167 step_test = 24;
168 Num_lags = 24;
169 %-----
170 % Se define el rango temporal para el cual se quiere realizar las
171 % predicciones.
172 %-----
173 start_hour = 48;
174 end_hour = 120;
175 %-----
176 % Se define el valor mínimo y el máximo de la generación eléctrica real,
177 % para poder reconvertir los valores escalados durante la fase de
178 % normalización
179 %-----
180 power_min = -101; % [kW]
```



```
181 power_max = 3138; % [kW]
182 %-----
183 % Se define el tamaño de una variable que contendrá los valores de las
184 % métricas de evaluación de las predicciones, y una variable que permitirá
185 % saber el paso de predicción en que se encuentra el programa.
186 %-----
187 resultats_estadistics = zeros(length(start_hour:24:end_hour),7);
188 medias_result_estad= zeros(1,7);
189 index_result = 1;
190 %-----
191 % Se define un bucle de predicción, con el objetivo de realizar las
192 % predicciones para el rango de datos definido anteriormente, de forma
193 % automática.
194 %-----
195 for kk = start_hour:step_test:end_hour
196 %-----
197 % Se utiliza la función de preparets para preparar los datos de cada
198 % predicción que se realiza.
199 %-----
200 [xc,xic,aic,tc] = preparets(netc, X_Imput_Test(:,kk-24-1:kk+Num_hour_forecast-2), {}, T_target_Test(:,kk-24-
201 1:kk+Num_hour_forecast-2));
202 [xc,xic,aic,tc] = preparets(netc, X_Imput_Test(:,kk-24-1:kk+Num_hour_forecast-1), {}, T_target_Test(:,kk-24-
203 1:kk+Num_hour_forecast-1));
204 yc = netc(xc, xic, aic); % Se obtiene los valores de predicción
205 perfc = perform(netc,tc,yc); % Se obtiene el MSE de la predicción
206 %-----
207 % Se definen ejes de representación acorde al horizonte de predicción
208 %-----
209 horas_axis = [kk-Num_lags:1:kk+Num_hour_forecast-1];
210 %horas_axis_2 = [kk:1:kk+Num_hour_forecast-1];
211 horas_axis_2 = [kk:1:kk+Num_hour_forecast];
212 %-----
213 % Se calculan los resultados de la predicción.
214 %-----
215 predicc = cell2mat(yc)'; % Se obtiene los valores de predicción.
216 target = cell2mat(tc)'; % Se obtiene los valores de la salida objetivo.
217 predicc(predicc<-1.0) = -1; % Se limita el rango de posibles valores de
218 predicción, ya que en algunos caso diverge.
219 %-----
220 errors_fore = (1619.5*target+1518.5)-(1619.5*predicc+1518.5); % Se calcular el error como la diferencia entre la
221 predicción y el objetivo [kW]
222 tc_2 = (1619.5*cell2mat(tc)'+1518.5); % [kW] de error
223 MSE00o2 = mean(var((1619.5*target+1518.5)',1)); % Referencia de normalización.
224 MSE_Eo2 = mse(errors_fore); % MSE para la salida de entrenamiento
225 NMSEo2 = mse(errors_fore)/MSE00o2; % NMSE (normalized mean square error) -> link:
226 https://math.stackexchange.com/questions/488964/the-definition-of-nmse-normalized-mean-square-error
227 R2o2 = 1 - NMSEo2; % Rsquared
228 %-----
229 % Se calcula las siguientes métricas de error:
230 % MAE (Mean Absolute Error)
231 % MAPE (Mean Absolute Percentage Error)
232 % RMSE (Root Mean Square Error)
233 % MSPE
234 % RMSPE (Root Mean Square Percentage Error)
235 %-----
236 %-----
237 mape_res1 = abs((errors_fore)./tc_2);
238 mape_fore = (1/length(errors_fore))*sum(mape_res1)*100; % Sus unidades son [%]
239 mae_fore = (1/length(errors_fore))*sum(abs(errors_fore)); % Sus unidades son [W]
240 rmse_fore = (sum((errors_fore).^2)/length(errors_fore))^0.5; % Sus unidades son [W]
241 mspe_res1=(errors_fore)./tc_2.^2;
242 mspe_fore=(1/length(errors_fore))*sum(mspe_res1)*100;
243 rmspe_fore=sqrt(mspe_fore);
244 %-----
245 % Se muestra por pantalla las métricas de error calculadas para la previsión.
246 %-----
247 str_txt = ['Start hour: ', num2str(kk), ' [h] | End hour: ', num2str(kk+Num_hour_forecast-1), ' [h] |', 'Iteración: ',
248 num2str(index_result)];
249 disp(str_txt);
250 %-----
251 str_rmse_fore=['RMSE (Root Mean Square Error) de la previsión: ', num2str(rmse_fore),' [kWh]'];
252 disp(str_rmse_fore);
253 str_mse_fore=['MSE (Mean Squared Error) de la previsión: ', num2str(MSE_Eo2)];
254 disp(str_mse_fore);
255 str_nmse_fore=['NMSE (Normaliced Mean Squared Error) de la previsión: ', num2str(NMSEo2)];
256 disp(str_nmse_fore);
257 str_mae_fore=['MAE (Mean Absolute Error) de la previsión: ', num2str(mae_fore),' [kWh]'];
258 disp(str_mae_fore);
259 str_mape_fore=['MAPE (Mean Absolute Percentage Error) de la previsión: ', num2str(mape_fore),' [%]'];
260 disp(str_mape_fore);
261 str_rmspe_fore=['RMSPE (Root Mean Square Percentage Error)de la previsión: ', num2str(rmspe_fore),' [%]'];
262 disp(str_rmspe_fore);
263 str_rsquared_fore=['R^2 (R squared)de la previsión: ', num2str(R2o2)];
264 disp(str_rsquared_fore);
265 disp('-----')
266 %-----
267 % Se guarda las diferentes métricas de error en la variable de resultados
268 % estadísticos.
269 %-----
270 resultats_estadistics(index_result,1) = MSE_Eo2;
271 resultats_estadistics(index_result,2) = NMSEo2;
272 resultats_estadistics(index_result,3) = R2o2;
273 resultats_estadistics(index_result,4) = mape_fore;
274 resultats_estadistics(index_result,5) = mae_fore;
275 resultats_estadistics(index_result,6) = rmse_fore;
276 resultats_estadistics(index_result,7) = rmspe_fore;
277 %-----
278 % Se representa la generación eléctrica que se debe predecir junto a los
279 % resultados de predicción
280 %-----
281
```



```
282 clear figure(kk);
283 Grafico = figure(kk);
284 hold on
285 a11 = plot(hores_axis,1619.5*cell2mat(T_target_Test(:,kk-24-1:kk+Num_hour_forecast-2))+1518.5,'color','r');
286 M11 = 'Generación Eléctrica(Dato Histórico)';
287 a22 = plot(hores_axis_2,1619.5*predico+1518.5,'color','b','LineStyle',':');
288 M22 = 'Predicción de la Generación Eléctrica';
289 legend([a11; a22], M11,M22);
290 %datetick('x','dd/mm/yyyy HH:MM', 'keepticks')
291 xlabel('Fechas de Predicción');
292 ylabel('Generación Eléctrica del Parque Eólico de Es Milá [kWh]');
293 ylim([-500 3500])
294 fileName_Graficos=[strcat('Resultado_Grafico_Predicción_',num2str(kk),'_h_', 'Horizonte
295 ',num2str(Num_hour_forecast)),'.jpg'];
296 saveas(Grafico,fileName_Graficos)
297 index_result = index_result + 1;
298 end
299 %-----
300 % Se obtiene la media final de los errores para el archivo
301 %-----
302 disp('-----');
303 disp('- Valores medios de los resultados');
304 disp('-----');
305
306 %-----
307 % Media de los MSE
308 %-----
309 str_mse_fore=['MSE: ', num2str(mean(resultats_estadistics(:,1))), ' '];
310 disp(str_mse_fore);
311 %-----
312 % Media de los NMSE
313 %-----
314 str_nmse_fore=['NMSE: ', num2str(mean(resultats_estadistics(:,2))), ' '];
315 disp(str_nmse_fore);
316 %-----
317 % Media de los R^2
318 %-----
319 str_rsquared_fore=['RSQUARED: ', num2str(mean(resultats_estadistics(:,3))), ' '];
320 disp(str_rsquared_fore);
321 %-----
322 % Media de los MAPE
323 %-----
324 str_mape_fore=['MAPE (Mean Absolute Percentage Error) de la prevision: ', num2str(mean(resultats_estadistics(:,4))), ' [%]'];
325 disp(str_mape_fore);
326 %-----
327 % Media de los MAE
328 %-----
329 str_mae_fore=['MAE (Mean Absolute Error) de la prevision: ', num2str(mean(resultats_estadistics(:,5))), ' [kWh]'];
330 disp(str_mae_fore);
331 %-----
332 % Media de los RMSE
333 %-----
334 str_rmse_fore=['RMSE (Root Mean Square Error) de la prevision: ', num2str(mean(resultats_estadistics(:,6))), ' [kWh]'];
335 disp(str_rmse_fore);
336 disp('-----');
337 %-----
338 % Media de los RMSPE
339 %-----
340 str_rmse_fore=['RMSPE (Root Mean Square Percentage Error) de la prevision: ', num2str(mean(resultats_estadistics(:,7)))]];
341 disp(str_rmse_fore);
342 disp('-----');
343 %-----
344 % Se guarda los valores medios de las métricas de error para aquella
345 % predicción.
346 %-----
347 medias_result_estad(1,1)= mean(resultats_estadistics(:,1)); % MSE
348 medias_result_estad(1,2)= mean(resultats_estadistics(:,2)); % NMSE
349 medias_result_estad(1,3)= mean(resultats_estadistics(:,3)); % R^2
350 medias_result_estad(1,4)= mean(resultats_estadistics(:,4)); % MAPE
351 medias_result_estad(1,5)= mean(resultats_estadistics(:,5)); % MAE
352 medias_result_estad(1,6)= mean(resultats_estadistics(:,6)); % RMSE
353 medias_result_estad(1,7)= mean(resultats_estadistics(:,7)); % RMSPE
354 %-----
355 % Se procede a guardar los resultados de la predicción
356 %-----
357 fileName1=[strcat('Resultados Estadísticos Predicción_',num2str(Num_hour_forecast)), ' h', '.xls'];
358 fileName2=[strcat('Resultados Estadísticos Medias Predicción_',num2str(Num_hour_forecast)), ' h', '.xls'];
359 xlswrite(fileName1,resultats_estadistics);
360 xlswrite(fileName2,medias_result_estad);
361
```

B.6: COEFICIENTE DE FRICCIÓN.M

La función de este script es calcular el coeficiente de fricción del emplazamiento de un parque eólico, a partir de datos de velocidad del viento a dos alturas.

```

1  %
2  %+++++
3  %+++++
4  % Trabajo de fin de grado: INGENIERÍA ELECTRÓNICA INDUSTRIAL Y AUTOMÁTICA
5  % Universidad de las Islas Baleares (UIB)
6  % Deva Murti Baer
7  % 2019/2020
8  %+++++
9  %+++++
10 %
11 % DESCRIPCIÓN DE LA FUNCIÓN DEL SCRIPT
12 % -----
13 %
14 % CÁLCULO DEL COEFICIENTE DE FRICCIÓN DEL TERRENO DEL PARQUE EÓLICO ES MILÁ
15 % -----
16 % - Aplicación para obtener el coeficiente de fricción de forma
17 % experimental del emplazamiento de el parque eólico, a partir de datos de
18 % velocidad del viento a dos alturas diferentes realizando un ajuste no
19 % lineal de los datos.
20 %
21 %+++++
22 %
23 %
24 % -----
25 % Se inicializa el espacio de memoria.
26 % -----
27 close all;
28 clear all;
29 clc;
30 % -----
31 % Mensaje de bienvenida
32 % -----
33 disp('-----');
34 disp('- Universidad de las Islas Baleares');
35 disp('- Trabajo de fin de grado: Deva Murti Baer (2019-2020)');
36 disp(str_txt);
37 disp('-----');
38 disp(' ');
39 % -----
40 % Altura de los datos de velocidad del parque
41 % -----
42 hx = 45; % [m] altura de la velocidad usada para ajustar
43 "alpha"
44 h0 = 35; % [m] altura de referencia a ajustar
45 % -----
46 %
47 % Se carga los datos de un fichero csv previamente tratado, se copia los
48 % datos de interes a un nuevo libro de excel y se carga en este script
49 % -----
50 T = readtable('dades_2007.csv'); % Parece que las columnas estan giradas, seria
51 entonces col1: 45m col2: 30m
52 T1 = table2array(T);
53 vel_hx = T1(:,1); % [m/s] a la altura "hx"
54 vel_h0 = T1(:,2); % [m/s] a la altura "h0"
55 % vel_hx = T1(:,2); % [m/s] a la altura "hx"
56 % vel_h0 = T1(:,1); % [m/s] a la altura "h0"
57 figure(1);
58 plot(vel_h0,vel_hx, '.');
59 xlabel('Vel. a 30m [m/s]');
60 ylabel('Vel. a 45m [m/s]');
61 grid on;
62 grid minor;
63 title('Wind Vel. at 45m Vs Wind Vel. at 30m (2007)');
64 hold off;
65 % -----
66 % Se ajusta una función matemática al conjunto de datos
67 % -----
68 startingVals = [0.00001];
69 % Definición de la función matemática a ajustar
70 modelFun = @(p,x) x.*(hx/h0)^p(1);
71 % Se determina los parametros del ajuste, mediante un ajuste no lineal
72 [alpha,R,J,CovB,MSE,ErrorModelInfo] = nlinfit(vel_h0, vel_hx, modelFun, startingVals);
73 vel_funcion_fit = modelFun(alpha, vel_h0); % Se calcula los valores de la función ajustada
74 % -----
75 % Se dibuja la función ajustada sobre los datos
76 % -----
77 figure(2);
78 plot(vel_h0,vel_hx, '.', 'Color', 'b');
79 line(vel_h0, vel_funcion_fit, 'Color', 'r');
80 xlabel('Vel. a 30m [m/s]');
81 ylabel('Vel. a 45m [m/s]');
82 grid on;
83 grid minor;
84 title('Wind Vel. at 45m Vs Wind Vel. at 30m + function fitted (Year 2007)');
85 hold off;
86
87

```