



GRADO EN INGENIERÍA ELECTROMECAÁNICA CON ESPECIALIDAD EN ELECTRÓNICA

TRABAJO FIN DE GRADO

MODELADO Y DETECCIÓN DE ANOMALÍAS DE UN AEROGENERADOR A TRAVÉS DE SUS CURVAS DE ESTADO

Autor: Ramón Conejero Guillén

Director: Miguel Ángel Sanz Bobi

Madrid

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título Modelado y detección de anomalías de un aerogenerador a través de sus curvas de estado

en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el curso académico 2019/20 es de mi autoría, original e inédito y no ha sido presentado con anterioridad a otros efectos. El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido tomada de otros documentos está debidamente referenciada.



Fdo.: **Ramón Conejero Guillén**

Fecha: 19 / 06 / 2020

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO



Fdo.: **Miguel Ángel Sanz Bobi**

Fecha: 19 / 06 / 2020

AUTORIZACIÓN PARA LA DIGITALIZACIÓN, DEPÓSITO Y DIVULGACIÓN EN RED DE PROYECTOS FIN DE GRADO, FIN DE MÁSTER, TESIS O MEMORIAS DE BACHILLERATO

1º. Declaración de la autoría y acreditación de la misma.

El autor D Ramón Conejero Guillén

DECLARA ser el titular de los derechos de propiedad intelectual de la obra: MODELADO Y DETECCIÓN DE ANOMALÍAS DE UN AEROGENERADOR A TRAVÉS DE SUS CURVAS DE ESTADO, que ésta es una obra original, y que ostenta la condición de autor en el sentido que otorga la Ley de Propiedad Intelectual.

2º. Objeto y fines de la cesión.

Con el fin de dar la máxima difusión a la obra citada a través del Repositorio institucional de la Universidad, el autor **CEDE** a la Universidad Pontificia Comillas, de forma gratuita y no exclusiva, por el máximo plazo legal y con ámbito universal, los derechos de digitalización, de archivo, de reproducción, de distribución y de comunicación pública, incluido el derecho de puesta a disposición electrónica, tal y como se describen en la Ley de Propiedad Intelectual. El derecho de transformación se cede a los únicos efectos de lo dispuesto en la letra a) del apartado siguiente.

3º. Condiciones de la cesión y acceso

Sin perjuicio de la titularidad de la obra, que sigue correspondiendo a su autor, la cesión de derechos contemplada en esta licencia habilita para:

- a) Transformarla con el fin de adaptarla a cualquier tecnología que permita incorporarla a internet y hacerla accesible; incorporar metadatos para realizar el registro de la obra e incorporar “marcas de agua” o cualquier otro sistema de seguridad o de protección.
- b) Reproducir la en un soporte digital para su incorporación a una base de datos electrónica, incluyendo el derecho de reproducir y almacenar la obra en servidores, a los efectos de garantizar su seguridad, conservación y preservar el formato.
- c) Comunicarla, por defecto, a través de un archivo institucional abierto, accesible de modo libre y gratuito a través de internet.
- d) Cualquier otra forma de acceso (restringido, embargado, cerrado) deberá solicitarse expresamente y obedecer a causas justificadas.
- e) Asignar por defecto a estos trabajos una licencia Creative Commons.
- f) Asignar por defecto a estos trabajos un HANDLE (URL *persistente*).

4º. Derechos del autor.

El autor, en tanto que titular de una obra tiene derecho a:

- a) Que la Universidad identifique claramente su nombre como autor de la misma
- b) Comunicar y dar publicidad a la obra en la versión que ceda y en otras posteriores a través de cualquier medio.
- c) Solicitar la retirada de la obra del repositorio por causa justificada.
- d) Recibir notificación fehaciente de cualquier reclamación que puedan formular terceras personas en relación con la obra y, en particular, de reclamaciones relativas a los derechos de propiedad intelectual sobre ella.

5º. Deberes del autor.

El autor se compromete a:

- a) Garantizar que el compromiso que adquiere mediante el presente escrito no infringe ningún derecho de terceros, ya sean de propiedad industrial, intelectual o cualquier otro.
- b) Garantizar que el contenido de las obras no atenta contra los derechos al honor, a la intimidad y a la imagen de terceros.

- c) Asumir toda reclamación o responsabilidad, incluyendo las indemnizaciones por daños, que pudieran ejercitarse contra la Universidad por terceros que vieran infringidos sus derechos e intereses a causa de la cesión.
- d) Asumir la responsabilidad en el caso de que las instituciones fueran condenadas por infracción de derechos derivada de las obras objeto de la cesión.

6º. Fines y funcionamiento del Repositorio Institucional.

La obra se pondrá a disposición de los usuarios para que hagan de ella un uso justo y respetuoso con los derechos del autor, según lo permitido por la legislación aplicable, y con fines de estudio, investigación, o cualquier otro fin lícito. Con dicha finalidad, la Universidad asume los siguientes deberes y se reserva las siguientes facultades:

- La Universidad informará a los usuarios del archivo sobre los usos permitidos, y no garantiza ni asume responsabilidad alguna por otras formas en que los usuarios hagan un uso posterior de las obras no conforme con la legislación vigente. El uso posterior, más allá de la copia privada, requerirá que se cite la fuente y se reconozca la autoría, que no se obtenga beneficio comercial, y que no se realicen obras derivadas.
- La Universidad no revisará el contenido de las obras, que en todo caso permanecerá bajo la responsabilidad exclusiva del autor y no estará obligada a ejercitar acciones legales en nombre del autor en el supuesto de infracciones a derechos de propiedad intelectual derivados del depósito y archivo de las obras. El autor renuncia a cualquier reclamación frente a la Universidad por las formas no ajustadas a la legislación vigente en que los usuarios hagan uso de las obras.
- La Universidad adoptará las medidas necesarias para la preservación de la obra en un futuro.
- La Universidad se reserva la facultad de retirar la obra, previa notificación al autor, en supuestos suficientemente justificados, o en caso de reclamaciones de terceros.

Madrid, a 19 de Julio de 2020

ACEPTA 
Fdo... RAMÓN... CONEJERO... GUILLEN

Motivos para solicitar el acceso restringido, cerrado o embargado del trabajo en el Repositorio Institucional:



GRADO EN INGENIERÍA ELECTROMECAÁNICA CON ESPECIALIDAD EN ELECTRÓNICA

TRABAJO FIN DE GRADO

MODELADO Y DETECCIÓN DE ANOMALÍAS DE UN AEROGENERADOR A TRAVÉS DE SUS CURVAS DE ESTADO

Autor: Ramón Conejero Guillén

Director: Miguel Ángel Sanz Bobi

Madrid

Agradecimientos:

Al director del proyecto Miguel Ángel Sanz Bobi por su paciencia e inestimable ayuda.

A mi familia y amigos, por su incondicional apoyo en los momentos difíciles.

MODELADO Y DETECCIÓN DE ANOMALÍAS DE UN AEROGENERADOR A TRAVÉS DE SUS CURVAS DE ESTADO

Autor: Conejero Guillen, Ramón.

Director: Sanz Bobi, Miguel Ángel.

ICAI- Universidad Pontificia de Comillas.

RESUMEN DEL PROYECTO

Palabras clave: Aerogenerador, curvas de estado, modelos de comportamiento, redes neuronales, MLP, detección de anomalías.

I. INTRODUCCIÓN

En 2010 había menos de 200.000MW de energía eólica instalada en todo el mundo. En 2018, ese número se situó en 591.000MW. La demanda de energía a nivel global no cesa de aumentar, hecho que se acentúa en el caso de la energía eléctrica. Se trata de una fuente de energía renovable, inagotable, y no contaminante. Esto último cobra gran importancia hoy en día porque cada vez se pone más el foco en buscar fuentes de energía que reduzcan el impacto ambiental, y de todas ellas, la energía eólica es la que más se ha instalado. Por todo ello, es conveniente optimizar el uso y los recursos destinados a dicho tipo de energías.

La motivación de este proyecto es crear un sistema de detección de anomalías que ayude en la operación y mantenimiento del aerogenerador o aerogeneradores a los que se aplique, y con ello, evite o mitigue el decrecimiento del rendimiento que indudablemente se produce en los sistemas eólicos con el paso del tiempo y las horas de funcionamiento por envejecimiento o degradación. Esta mejora no sólo supone un beneficio económico para el propietario del aerogenerador, sino también para el planeta, pues se trata de una energía limpia y al alza.

Este trabajo pone su atención en la creación de modelos de comportamiento de la curva de potencia y otras curvas de estado que manifiesta un aerogenerador en su operación. Existen numerosos estudios que analizan la curva de potencia del aerogenerador para analizar su funcionamiento y detectar anomalías, sin embargo, no se presta tanta atención a otras curvas que pueden ser relevantes en la detección de anomalías. La aleatoriedad de las condiciones ambientales, o el defecto o deterioro de algún componente pueden no ser correctamente detectados y analizados únicamente con la curva de potencia. Existen otras curvas, comunes a todo tipo de aerogeneradores que pueden aportar más información y, en consecuencia, generar un mejor modelo del funcionamiento de la instalación. Teniendo en cuenta lo anterior, este trabajo analizará la vida del aerogenerador en base a nueve curvas de estado con las que se controlará el funcionamiento de componentes del sistema tales como la multiplicadora, el generador eléctrico, el controlador de pitch o el grupo hidráulico.

a) EL AEROGENERADOR MODERNO

El aerogenerador es un tipo de máquina de generación eléctrica a partir de la potencia del viento que es bien conocida desde la antigüedad. A continuación, se describe el funcionamiento de los componentes monitorizados mediante los modelos de comportamiento:

- Rotor: componente del motor que gira respecto a la parte fija para conseguir energía. Las palas del rotor, construidas a partir de materiales compuestos, se diseñan para transferir la energía cinética del viento en momento torsor en el eje del equipo. Pueden alcanzar diámetros de hasta 80 metros. Criterios acústicos delimitan la velocidad máxima de rotación. El rotor a su vez se compone de:
- Sistema de inclinación de pala: también conocido como sistema de pitch. Permite variar la inclinación de las palas respecto al mismo. Se encuentra sobre el buje, en la base de los álabes.
- Multiplicadora: también llamada caja multiplicadora, se trata del conjunto de engranajes que transforman la rotación de baja velocidad del rotor a una velocidad más elevada al eje que entra en el generador eléctrico.
- Generador eléctrico: transforma la energía mecánica en eléctrica.
- Grupo hidráulico: proporciona la presión necesaria para los componentes accionados hidráulicamente, como el freno, el sistema de control de pitch o el sistema de bloqueo del rotor.

b) MODELOS DE REGRESIÓN LINEAL

El modelo que se va a utilizar en este trabajo es un modelo de regresión lineal múltiple de media móvil (MA). En este tipo de modelos se calcula la predicción de la variable dependiente Y_i en función de un término independiente X_i y de una sucesión ponderada de errores correspondientes a los instantes precedentes. El resultado consiste en una serie de valores correspondientes a la predicción de la variable dependiente a lo largo de una serie de intervalos regulares de la variable independiente.

c) MODELOS DE REDES NEURONALES

Están formados por unidades simples de procesamiento llamadas neuronas o nodos conectadas entre sí, con varias entradas a las que se les asigna un peso específico y una única salida de datos que es función de una combinación lineal de los valores de sus entradas. La principal característica de las redes neuronales es su capacidad de aprendizaje. Este aprendizaje consiste en el reajuste continuo de los pesos de las conexiones entre neuronas de una forma iterativa para así “aprender” las relaciones existentes entre pares de valores de vectores de entradas-salidas que se le suministren, o encontrar un patrón común en datos de entrada para poderlos clasificar de acuerdo a los distintos patrones.

Como ventajas del uso de redes neuronales, destaca el hecho de que no es necesario conocer la naturaleza del conjunto de datos de estudio, pues el proceso de entrenamiento aporta las características de los mismos. Asimismo, el comportamiento no lineal de las funciones de activación de las redes neuronales permite que se pueda trabajar con funciones no lineales.

d) MLP

El perceptrón multicapa o red neuronal MLP (Multilayer Perceptron) es un tipo de red neuronal unidireccional con una organización de neuronas dispuesta en capas, de tal manera que las neuronas de una capa tomen como entradas las salidas de la capa de neuronas anterior y sus salidas sean las entradas de la capa de neuronas posterior. Lógicamente esta estructura no se cumple para la primera capa (capa de entradas), ni la última (capa de salidas), sino para las intermedias (capas ocultas). El número de capas ocultas puede variar, aunque con una sola ya se podría aproximar con cierto grado de error una función continua acotada.

e) REVISIÓN DE PROYECTOS ANTERIORES

Se ha observado que existe un gran número de estudios que utilizan diferentes tipos de modelos no lineales para detectar anomalías partiendo de la curva de potencia del aerogenerador, y estudiando que sucede en el resto de componentes en los instantes de fallo en el funcionamiento

En [1] se crea un modelo no lineal de comportamiento para la curva de potencia. El tipo de redes neuronales utilizadas son MLP. A partir del modelo creado, se obtiene el error entre estimación y valor real y de ahí se extraen las anomalías. Después se utiliza un análisis de PCA con las variables que se han utilizado para entrenar el modelo y se estudia que variables son responsables de la variabilidad de los datos, para analizar después a qué se debe el fallo. La diferencia principal es que se van a estudiar más de un modelo de comportamiento, para poder estudiar directamente el rendimiento de componentes como la caja multiplicadora o el generador, lo que supone una ventaja a la hora de aplicarlo en una instalación real, ya que facilita el proceso de operación y mantenimiento, conocer directamente la fuente concreta de fallos

En [2] el método utilizado para detectar anomalías es un mapa auto-organizado (SOM). Este proyecto es parecido a [1] en el sentido en que ambos utilizan redes neuronales para crear un modelo de comportamiento sobre la curva de potencia para detectar anomalías. En este caso, la base de muestras es de 2 meses, 35 veces menor a la base de muestras de la que se dispone para el proyecto a realizar.

[3] y [4] responden al mismo problema utilizando Support Vector Machines (SVM) y Proportional hazard models (PHM) respectivamente a diferentes variables, como el ángulo del controlador de pitch o temperaturas internas. Es interesante que se busque directamente la detección de anomalías por estudio directo de estas variables, sin tener que crear antes un modelo comportamiento de la curva de potencia.

En [5] se modelan la mayoría de las variables disponibles mediante redes neuronales recurrentes (RNN). La componente de retro-programación permite tener un resultado excelente, sin embargo, una vez extraído el error no se trabaja en la detección de anomalías.

En [6] se modela el controlador del pitch creando una máquina de Boltzmann restringida (RBM), para hacer después de ella un Deep Small-World Neural Network (DSWNN) añadiendo conexiones entre las capas ocultas de la máquina. Finalmente se valida su correcto funcionamiento detectando anomalías.

En general se puede ver que hay varios métodos que se pueden utilizar para modelar el funcionamiento de un aerogenerador, sin embargo, no se ha encontrado ninguno que permita

detectar la causa de las anomalías y que modele no solo la curva de potencia del aerogenerador, sino también los componentes internos.

II. OBJETIVOS

El objetivo principal de este proyecto es crear un sistema que detecte anomalías en el funcionamiento de un aerogenerador a través de datos de su operación disponibles. A continuación, se recogen los subobjetivos que se han determinado para evaluar el desarrollo del proyecto:

- Importar los datos medidos mediante el SCADA del aerogenerador y seleccionar cuales son las variables disponibles que van a permitir una mejor monitorización del aerogenerador y sus componentes principales.
- Crear modelos de comportamiento para las curvas de estado (tanto de la curva de potencia como de otras relaciones entre la velocidad el viento y variables intermedias que muestren el comportamiento de componentes vitales para el funcionamiento del aerogenerador) creadas a partir de las variables seleccionadas anteriormente. Se realizarán modelos tanto lineales como no lineales, siempre y cuando su rendimiento justifique su implementación.
- A partir de los modelos de comportamiento se buscarán anomalías a lo largo del periodo que se simule y se comprobará la valía del estudio comparándolo con el registro de reparaciones del que se dispone.

III. METODOLOGÍA

La programación íntegra de este proyecto se realiza mediante Matlab y sus aplicaciones.

En primer lugar, se importan las mediciones de fecha y variables del aerogenerador, así como sus registros de reparaciones y alarmas (Todos provenientes del sistema SCADA del aerogenerador), a Matlab. Tras analizar los datos de las variables y acotar los registros de reparaciones y alarmas al periodo temporal del que se dispone, se seleccionan las variables con las que se van a crear las curvas de estado y los modelos de comportamiento. Para elegir las variables oportunas se ha tenido en cuenta tanto el componente del aerogenerador al que hacen referencia, tanto como la cantidad de valores de los que de ella se dispone, pues debido a errores de diversa índole durante el periodo de estudio, no se dispone de todas las medidas de todas las variables.

En segundo lugar, se aplica un pretratamiento a las variables que se van a utilizar. Tiene 3 etapas. En la primera etapa, se eliminan los valores de la variable en cuestión para los cuales el aerogenerador se encuentra fuera de servicio. En la segunda etapa, se sustituyen los valores NaN (Not a Number, valor que da Matlab a los huecos en los vectores de medida) por la media aritmética entre el valor anterior y el posterior. En la tercera etapa, se corrigen errores en las medidas de variables de entrada. Se ha corregido un offset que tiene la variable velocidad del viento durante el tercer año del periodo de estudio.

En tercer lugar, se crean los modelos lineales (ML) de media móvil de las curvas de estado. Para ello se limpian los valores más aislados que podrían alterar el comportamiento del modelo en intervalos con pocos registros, y se determina un ancho de intervalo fijo que permita captar el número máximo de anomalías, sin verse alterado por valores ligeramente aislados.

En cuarto lugar, se crean los modelos no lineales (MNL). Se han utilizado diferentes configuraciones de MLP para cada curva de estado, y se ha seleccionado aquella que minimice en MSE. Finalmente se analiza el resultado de los modelos de comportamiento, tanto ML como MNL, para eliminar aquel que no prediga correctamente. Para todos los modelos se ha utilizado el primer año de datos como periodo de referencia para crear el modelo, y el resto del tiempo (aproximadamente 5 años) para simular las curvas de estado (periodo de simulación).

En quinto y último lugar, se extrae el error del periodo de simulación de todos los modelos. Se toman aquellos valores fuera de las bandas de confianza (2 desviaciones típicas a ambos lados de la media) y se extraen las anomalías. Se ha considerado como anomalía aquellos periodos consecutivos de 30 minutos donde al menos 20 minutos, el resultado del modelo esté fuera de las bandas de confianza. Por último, se crea una curva que refleje el porcentaje de anomalías para cada modelo y se observan las reparaciones no programadas que afectan a cada componente representado en las curvas de estado.

IV. RESULTADOS

Se importan los datos del aerogenerador conforme a los explicado en el apartado anterior. Se toman las variables temperatura ambiente (var14), velocidad del viento (var22) y ángulo del viento (var23) como entradas de MNL, y var22 como entrada de los ML. Todas las curvas de estado utilizan var22 como variable de abscisas. Respecto a las variables a predecir, se toman las siguientes:

1. Potencia generada (var30)
2. Giro del rotor (var19)
3. Ángulo de pitch (var21)
4. Temperatura de los rodamientos DE del generador (var3)
5. Temperatura de los rodamientos NDE del generador (var4)
6. Temperatura del aceite de la multiplicadora (var1)
7. Temperatura de los anillos SLD del generador (var5)
8. Presión del grupo hidráulico (var11)
9. Temperatura del grupo hidráulico (var10)

Con este mismo orden y var22 se crean las curvas de estado (C1, C2, ... C9).

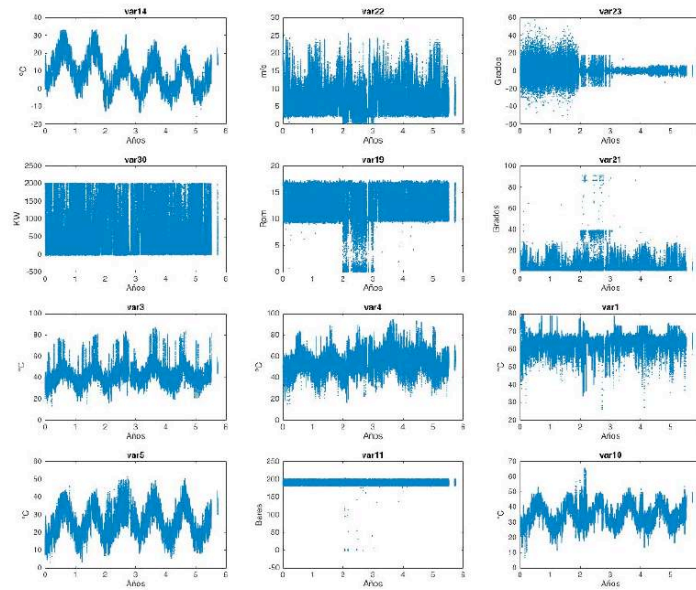


Figura 1. Variables utilizadas

Tras realizar el pretratamiento de elimina el offset negativo de la var22 y, al eliminar los valores para los que el aerogenerador se encuentra fuera de servicio, se pasa de vectores de 291312 valores a vectores de 216949 valores.

Tomando las consideraciones descritas en el apartado anterior se crean los modelos de comportamiento en base a los datos del primer año (periodo de referencia).

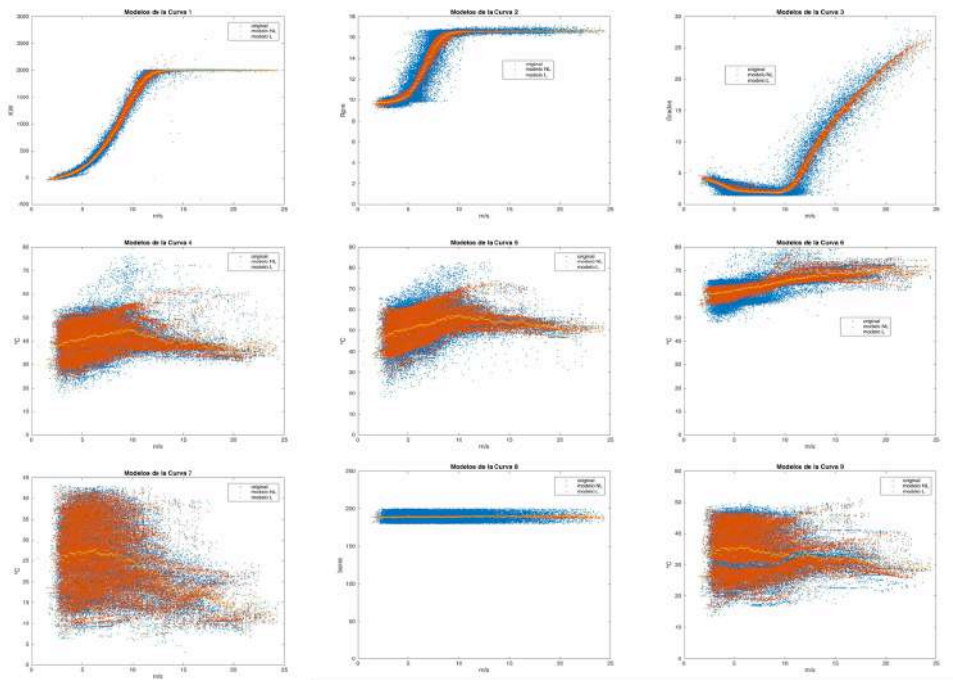


Figura 2. ML y MNL de las curvas de estado

La siguiente tabla recoge el valor de regresión lineal de los MNL para cada año. En base a los resultados obtenidos, no se tendrá en cuenta el resultado de MNL3 durante Y3, y se elimina completamente MNL8.

AÑO	C1	C2	C3	C4	C5	C6	C7	C8	C9
Y1 (Ent.)	99,7	96	96,7	88,2	84,0	72,7	97,5	10,8	93,5
Y2	99,6	95,7	95,8	84,4	80,9	63,2	96,1	2,1	89,9
Y3	93,7	77,4	16,2	78,2	74,4	56,1	86,8	2,7	73,6
Y4	99,1	95,0	92,5	82,9	78,2	74,0	95,5	1,9	87,6
Y5	99,6	95,8	95,7	80,6	77,5	73,2	95,0	1,2	84,0
Y6	99,6	95,5	93,5	76,0	66,8	72,7	97,6	1,5	84,5

Tabla1. Regresión lineal de los MNL para cada año del periodo de estudio

A partir del error durante el periodo de simulación de los modelos, se obtienen las anomalías de cada modelo y con ello se crean las curvas de anomalías de los 17 modelos utilizados.

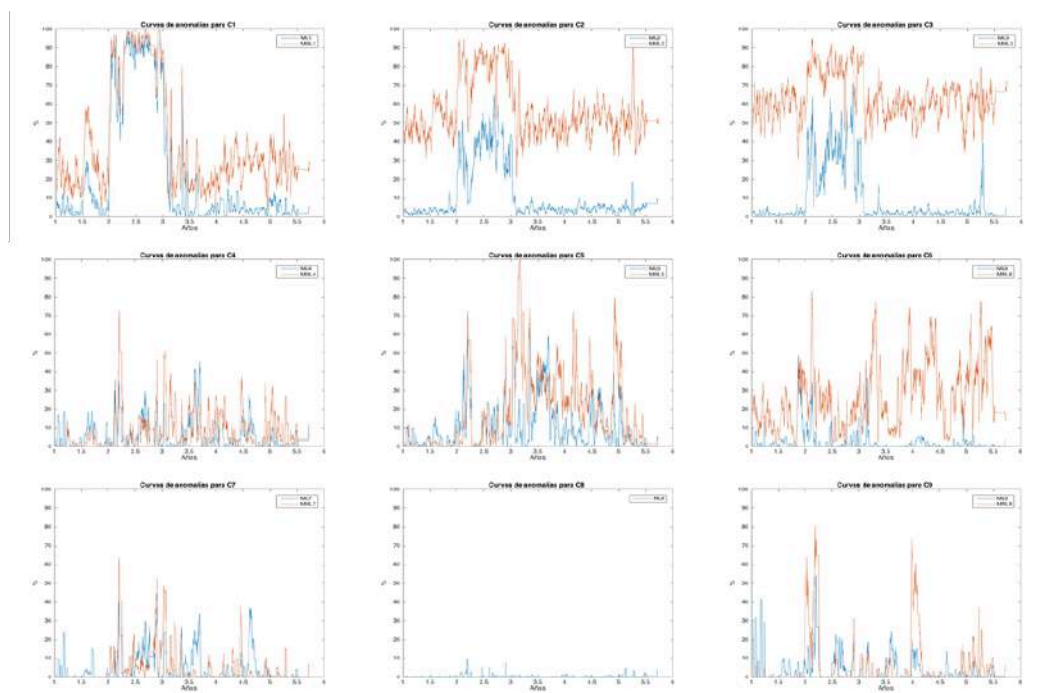


Figura3. Curvas de anomalías de las curvas de estado

Finalmente se comparan las reparaciones imprevistas durante el periodo de simulación con las curvas de anomalías de los modelos a los que estas reparaciones afectan. Aquí se muestra el ejemplo de las reparaciones que afectan directamente al grupo hidráulico (C8 y C9), donde se aprecia que el porcentaje de anomalías disminuye tras la reparación, luego se habría predicho correctamente que la reparación era necesaria.

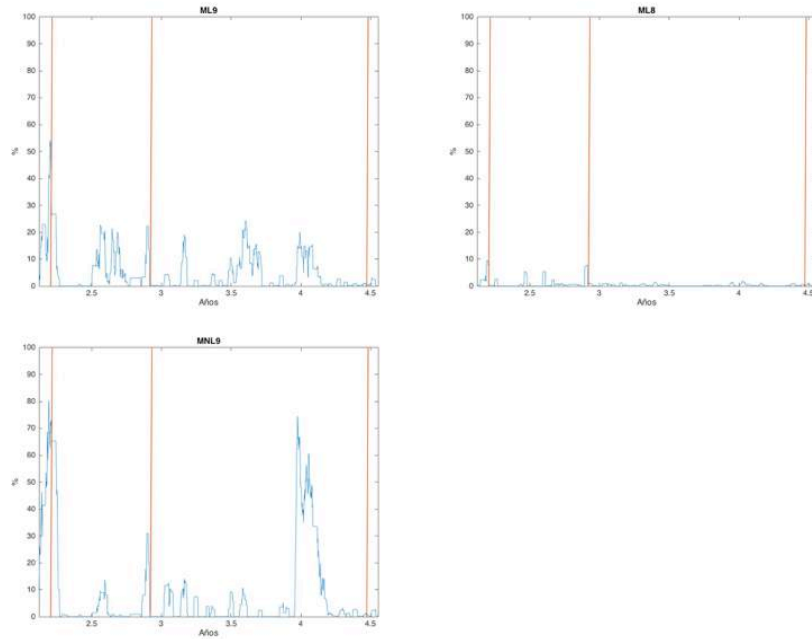


Figura4. Curvas de anomalías de C8 y C9 para reparaciones del grupo hidráulico

Tal y como se ha mostrado en la figura anterior se estudian las reparaciones que afectan directamente a la multiplicadora (C4 y C6), al generador eléctrico (C5 y C7), al sistema de medición de la velocidad del viento (todas las curvas), y las que suceden durante Y3 (periodo de peor funcionamiento general del aerogenerador).

V. CONCLUSIONES Y FUTUROS DESARROLLOS

La conclusión general es que se han abordado con éxito los 3 objetivos principales que se habían citado anteriormente, obteniendo los resultados deseados en los 2 primeros, y resultados aceptables para el tercer objetivo. Este objetivo se ha cumplido parcialmente debido a que con los métodos utilizados no se ha podido aislar completamente la variable de estudio de cada curva del resto de la instalación, por lo que al crear la curva de anomalías de cada modelo se encuentra una componente importante de ruido.

Como futuros desarrollos se plantea la utilización de diferentes tipos de modelos. Para modelos lineales se valoraría realizar modelos ANOVA, y para modelos no lineales se valoraría utilizar métodos como Deep Learning, redes CNN o una combinación de SOM y MLP.

También se tiene en cuenta que los modelos creados se podrían aplicar a otras variables en caso de que hubiese datos suficientes. Interesaría modelar el comportamiento del freno o del sistema de orientación (YAW).

VI. BIBLIOGRAFÍA

- [1] Peyman Mazidi, Student Member, IEEE, Lina Bertling Tjernberg, Senior Member, IEEE, and Miguel A. Sanz-Bobi, Senior Member, IEEE (2017). *Performance Analysis and Anomaly Detection in Wind Turbines based on Neural Networks and Principal Component Analysis*. Conference: 12th Workshop on Industrial Systems and Energy Technologies in Madrid, Spain (May 2017).
- [2] Mian Du, Lina Bertling Tjernberg, Shicong Ma, Qing He¹, Lin Cheng and Jianbo Guo (2016). *A SOM based Anomaly Detection Method for Wind Turbines Health Management through SCADA Data*. Article in the Energies Journal, published on February 20, 2017. DOI: 10.3390/en10020253.
- [3] Nassim Laouti, Nida Sheibat-Othman, Sami Othman (2011). *Support Vector Machines for Fault Detection in Wind Turbines*. Conference: 18th IFAC World Congress in Milano, Italy (August 28 – September 2, 2011).
- [4] Peyman Mazidi, Mian Du, Lina Bertling Tjernberg, and Miguel A. Sanz-Bobi (2017). *A Health condition model for wind turbine monitoring through neural networks and proportional hazard models*. Article in the SAGE Journal, published on May 4, 2011. DOI: 10.1177/1748006X17707902.
- [5] Yonglong Yan, Jian Li, and David Wenzhong Gao (2014). *Condition Parameter Modeling for Anomaly Detection in Wind Turbines*. Article in the Energies Journal, published on May 6, 2014. DOI: 10.3390/en7053104.
- [6] Meng Li ¹, Shuangxin Wang ¹, Shanxiang Fang ^{1,2} and Juchao Zhao (February 12, 2020). *Anomaly Detection of Wind Turbines Based on Deep Small-World Neural Network*. Article in the Applied Sciences Journal, published on February 12, 2020. DOI: 10.3390/app10041243. Last accessed on 04/04/2020.

MODELING AND DETECTION OF ANOMALIES OF A WIND TURBINE THROUGH ITS STATE CURVES

Author: Conejero Guillen, Ramón.

Director: Sanz Bobi, Miguel Ángel.

ICAI- Universidad Pontificia de Comillas.

PROJECT ABSTRACT

Keywords: Wind turbine, state curves, behavior models, neural networks, MLP, anomaly detection.

I. INTRODUCTION

In 2010, there were less than 200,000MW of wind power installed worldwide. In 2018, that number stood at 591,000MW. Global energy demand continues to increase, a fact that is accentuated in the case of electrical energy. It is a renewable, inexhaustible, and non-polluting energy source. The latter is of great importance today because the focus is increasingly on finding energy sources that reduce the environmental impact, and of all of them, wind energy is the one that has been installed the most. For all these reasons, it is convenient to optimize the use and resources destined for this type of energy.

The motivation of this project is to create an anomaly detection system that helps in the operation and maintenance of the wind turbine or wind turbines to which it is applied, and with it, avoids or mitigates the decrease in performance that undoubtedly occurs in wind systems with the passage of time and hours of operation due to aging or degradation. This improvement not only represents an economic benefit for the owner of the wind turbine, but also for the planet, since it is clean energy and on the rise.

This work focuses on creating behavioral models of the power curve and other state curves that a wind turbine manifests in its operation. There are numerous studies that analyze the power curve of the wind turbine to analyze its operation and detect anomalies, however, less attention is paid to other curves that may be relevant to anomaly detection. The randomness of the environmental conditions, or the defect or deterioration of some component may not be correctly detected and analyzed only with the power curve. There are other curves, common to all types of wind turbines that can provide more information and, consequently, generate a better model of the operation of the installation. Considering the above, this work will analyze the life of the wind turbine based on nine state curves with which the operation of system components such as the gearbox, the electric generator, the pitch controller or the hydraulic group will be controlled.

a) THE MODERN WIND TURBINE

The wind turbine is a type of machine for generating electricity from the power of the wind that has been well known since ancient times. The following describes the operation of the components monitored using the behavior models:

- Rotor: motor component that rotates with respect to the fixed part to obtain energy. The rotor blades, built from composite materials, are designed to transfer the kinetic energy of the wind at torque on the equipment shaft. They can reach diameters of up to 80 meters. Acoustic criteria define the maximum rotation speed. The rotor in turn consists of:
- Pitch system: it allows to vary the inclination of the blades with respect to it. It is located on the hub, at the base of the blades.
- Gearbox: also, called gearbox, it is the set of gears that transform the low-speed rotation of the rotor to a higher speed to the axis that enters the electric generator.
- Electric generator: transforms mechanical energy into electrical energy.
- Hydraulic group: it provides the necessary pressure for the hydraulically actuated components, such as the brake, the pitch control system or the rotor locking system.

b) LINEAR REGRESSION MODELS

The model to be used in this work is a moving average multiple linear regression (MA) model. In this type of model, the prediction of the dependent variable Y_i is calculated based on an independent term X_i and a weighted succession of errors corresponding to the preceding moments. The result consists of a series of values corresponding to the prediction of the dependent variable over a series of regular intervals of the independent variable.

c) NEURAL NETWORK MODELS

They are made up of simple processing units called neurons or nodes connected to each other, with several inputs assigned a specific weight and a single data output that is a function of a linear combination of the values of its inputs. The main characteristic of neural networks is their ability to learn. This learning consists of the continuous readjustment of the weights of the connections between neurons in an iterative way to "learn" the existing relationships between pairs of input-output vector values supplied to it, or find a common pattern in data from input to be able to classify them according to the different patterns.

As advantages of the use of neural networks, it highlights the fact that it is not necessary to know the nature of the study data set, since the training process provides the characteristics of the same. Likewise, the non-linear behavior of the activation functions of the neural networks allows us to work with non-linear functions.

d) MLP

The multilayer perceptron or MLP (Multilayer Perceptron) neural network is a type of unidirectional neural network with an organization of neurons arranged in layers, in such a way that the neurons of a layer take as inputs the outputs of the previous neuron layer and their outputs be the inputs to the posterior neuron layer. Logically this structure is not fulfilled for the first layer (input layer), nor the last one (output layer), but for the intermediate ones (hidden layers). The number of hidden layers can vary, although with one it could already be approximated with a certain degree of error, a function is still limited.

e) REVIEW OF PREVIOUS PROJECTS

It has been observed that there are many studies that use different types of non-linear models to detect anomalies starting from the power curve of the wind turbine, and studying what happens in the rest of the components in the instants of malfunction.

In [1] a nonlinear behavior model for the power curve is created. The type of neural networks used are MLP. From the created model, the error between estimation and real value is obtained and from there the anomalies are extracted. Then a PCA analysis is used with the variables that have been used to train the model and it is studied which variables are responsible for the variability of the data, to later analyze why the failure is due. The main difference is that more than one behavioral model will be studied, to directly study the performance of components such as the gearbox or the generator, which is an advantage when applying it in a real installation, since it facilitates the operation and maintenance process, knowing directly the specific source of faults

In [2] the method used to detect anomalies is a self-organized map (SOM). This project is like [1] in that both use neural networks to create a behavioral model on the power curve to detect abnormalities. In this case, the sample base is 2 months, 35 times less than the sample base available for the project to be carried out.

[3] and [4] answer the same problem using Support Vector Machines (SVM) and Proportional hazard models (PHM) respectively for different variables, such as pitch controller angle or internal temperatures. It is interesting that anomaly detection is directly sought by direct study of these variables, without having to first create a behavioral model of the power curve.

In [5] most of the available variables are modeled using recurrent neural networks (RNN). The back-programming component allows to have an excellent result, however, once the error has been extracted, it does not work on detecting anomalies.

In [6] the pitch controller is modeled creating a restricted Boltzmann machine (RBM), to make after it a Deep Small-World Neural Network (DSWNN) by adding connections between the hidden layers of the machine. Finally, its correct operation is validated by detecting anomalies.

In general, there are several methods that can be used to model the operation of a wind turbine, however, none have been found that allows detecting the cause of the anomalies and that models not only the power curve of the wind turbine, but also the internal components.

II. OBJECTIVES

The main objective of this project is to create a system that detects anomalies in the operation of a wind turbine through available data on its operation. Below are the sub-objectives that have been determined to evaluate the development of the project:

- Import the data measured by the SCADA of the wind turbine and select which are the available variables that will allow better monitoring of the wind turbine and its main components.
- Create behavioral models for the state curves (both power curve and of other relationships between wind speed and intermediate variables that show the behavior of vital components for the operation of the wind turbine) created from the variables selected previously. Both linear and non-linear models will be made, if their performance justifies their implementation.
- Based on the behavioral models, anomalies will be looked for throughout the simulated period and the value of the study will be verified by comparing it with the available repair registry.

III. METHODOLOGY

The entire programming of this project is done through Matlab and its applications.

First, the date and variable measurements of the wind turbine, as well as its repair and alarm logs (all from the SCADA system of the wind turbine), are imported to Matlab. After analyzing the data of the variables and delimiting the repair and alarm records for the time available, the variables with which to create the state curves and behavior models are selected. To choose the appropriate variables, the component of the wind turbine to which they refer has been considered, as well as the amount of values available to it, because due to errors of various kinds during the study period, it is not available of all measurements of all variables.

Second, a pretreatment is applied to the variables to be used. It has 3 stages. In the first stage, the values of the variable in question for which the wind turbine is out of service are eliminated. In the second stage, NaN values (Not a Number, Matlab's value for gaps in measurement vectors) are replaced by the arithmetic mean between the previous and the subsequent value. In the third stage, errors in the input variable measurements are corrected. An offset that has the variable wind speed during the third year of the study period has been corrected.

Third, the moving average linear (ML) models of the state curves are created. For this, the most isolated values that could alter the model's behavior in intervals with few records are cleaned, and a fixed interval width is determined that allows the maximum number of anomalies to be captured, without being altered by slightly isolated values.

Fourth, nonlinear models (MNL) are created. Different MLP settings have been used for each state curve, and the one that minimizes in MSE has been selected. Finally, the result of the behavior models, both ML and MNL, is analyzed to eliminate the one that does not predict correctly. For all models, the first year of data has been used as the reference period to create

the model, and the rest of the time (approximately 5 years) to simulate the state curves (simulation period).

Fifth and last, the simulation period error is extracted from all models. Those values are taken outside the confidence bands (2 standard deviations on both sides of the mean) and the anomalies are extracted. Those consecutive periods of 30 minutes where at least 20 minutes, the model result is outside the confidence bands, have been considered as an anomaly. Finally, a curve is created that reflects the percentage of anomalies for each model and unscheduled repairs that affect each component represented in the state curves are observed.

IV. RESULTS

The wind turbine data is imported in accordance with those explained in the previous section. The variables ambient temperature (var14), wind speed (var22) and wind angle (var23) are taken as inputs for MNL, and var22 as input for ML. All state curves use var22 as the abscissa variable. Regarding the variables to be predicted, the following are taken:

1. Generated power (var30)
2. Rotation of the rotor (var19)
3. Pitch angle (var21)
4. Temperature of the generator DE bearings (var3)
5. Temperature of the generator NDE bearings (var4)
6. Gearbox oil temperature (var1)
7. Temperature of the generator SLD rings (var5)
8. Hydraulic group pressure (var11)
9. Hydraulic group temperature (var10)

With this same order and var22 the state curves (C1, C2, ... C9) are created.

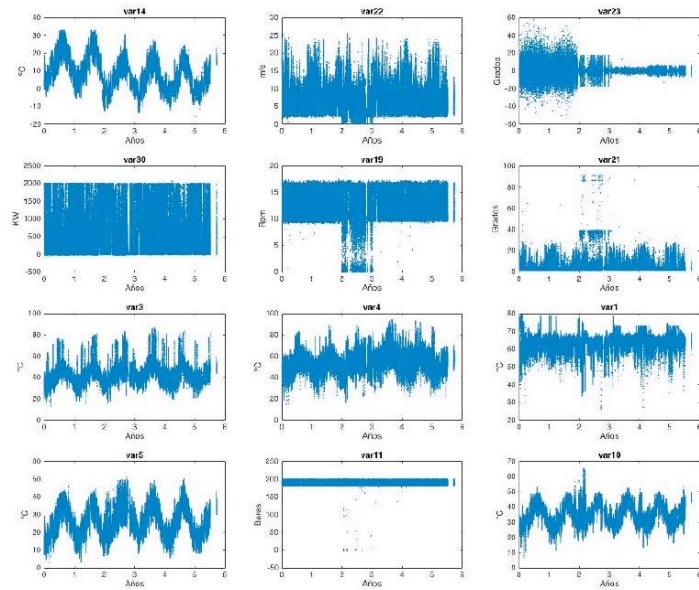


Figure 1. Variables used

After carrying out the pretreatment, the negative offset of the var22 is eliminated and, when eliminating the values for which the wind turbine is out of service, it goes from vectors of 291312 values to vectors of 216949 values.

Taking the considerations described in the previous section, behavioral models are created based on data from the first year (reference period).

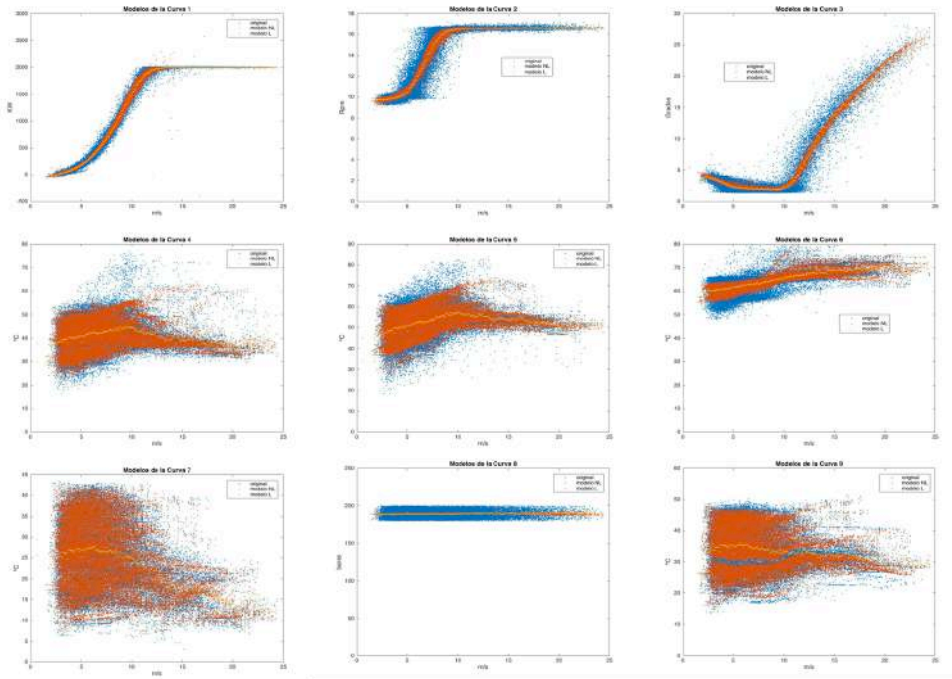


Figure 2. ML and MNL of the state curves

The following table shows the linear regression value of the MNL for each year. Based on the results obtained, the result of MNL3 during Y3 will not be considered, and MNL8 is eliminated.

AÑO	C1	C2	C3	C4	C5	C6	C7	C8	C9
Y1 (Ent.)	99,7	96	96,7	88,2	84,0	72,7	97,5	10,8	93,5
Y2	99,6	95,7	95,8	84,4	80,9	63,2	96,1	2,1	89,9
Y3	93,7	77,4	16,2	78,2	74,4	56,1	86,8	2,7	73,6
Y4	99,1	95,0	92,5	82,9	78,2	74,0	95,5	1,9	87,6
Y5	99,6	95,8	95,7	80,6	77,5	73,2	95,0	1,2	84,0
Y6	99,6	95,5	93,5	76,0	66,8	72,7	97,6	1,5	84,5

Table 1. Linear regression of the MNL for each year of the study period

From the error during the simulation period of the models, the anomalies of each model are obtained and with this the anomaly curves of the 17 models used are created.

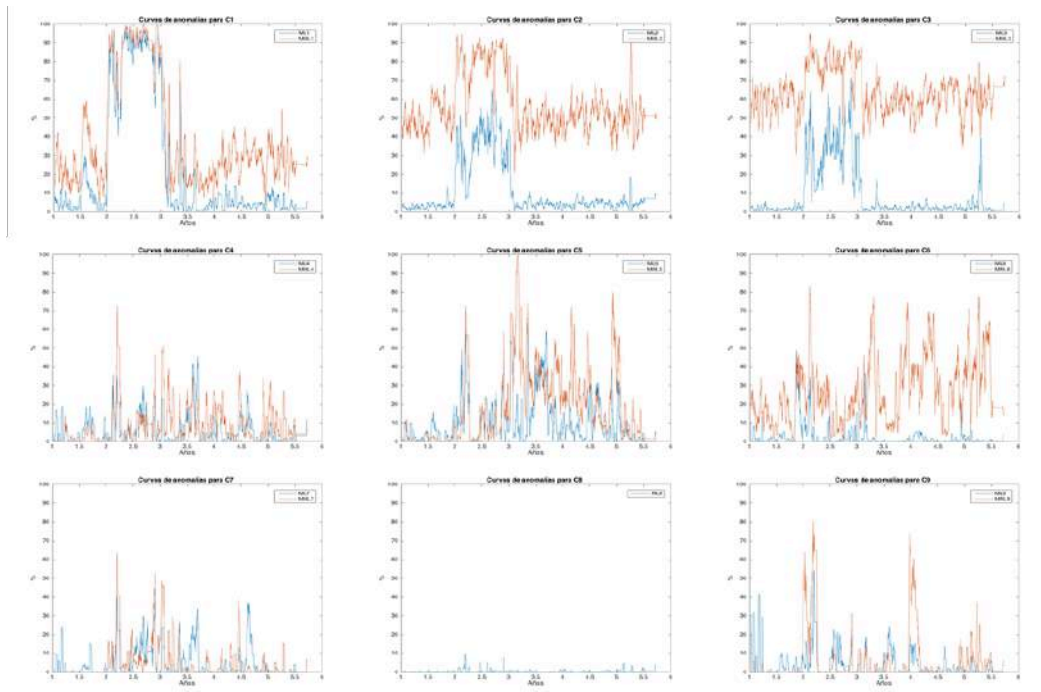


Figure 3. State curve anomaly curves

Finally, the unexpected repairs during the simulation period are compared with the anomaly curves of the models that these repairs affect. Here is the example of repairs that directly affect the hydraulic group (C8 and C9), where the percentage of anomalies decreases after the repair, then it would have been correctly predicted that the repair was necessary.

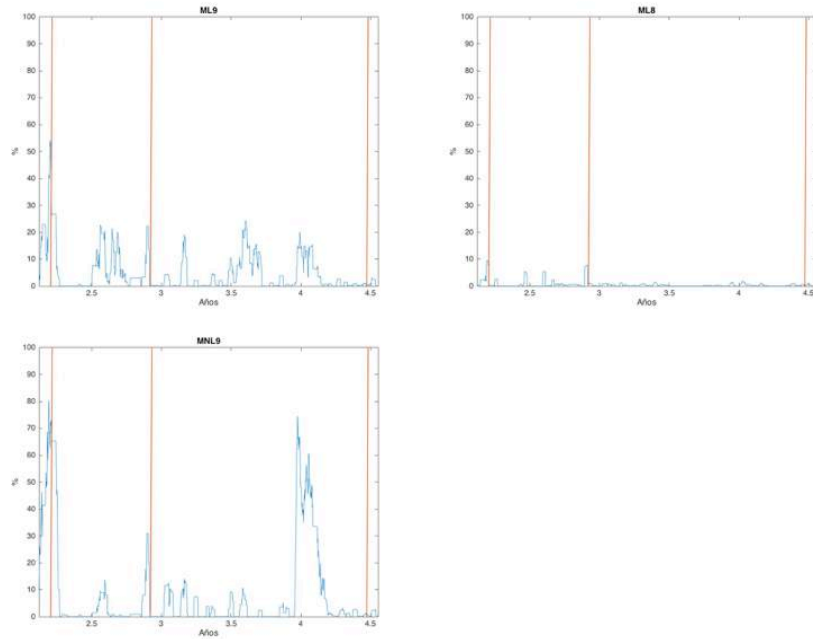


Figure 4. C8 and C9 anomaly curves for hydraulic group repairs

As shown in the previous figure, the repairs that directly affect the gearbox (C4 and C6), the electric generator (C5 and C7), the wind speed measurement system (all curves) are studied, and those that occur during Y3 (period of worst general operation of the wind turbine).

V. CONCLUSION AND FUTURE DEVELOPMENT

The general conclusion is that the 3 main objectives mentioned above have been successfully addressed, obtaining the desired results in the first 2, and acceptable results for the third objective. This objective has been partially met since with the methods used, it was not possible to completely isolate the study variable of each curve from the rest of the installation, so when creating the anomaly curve of each model, an important component of noise.

As future developments, the use of different types of models is proposed. For linear models, ANOVA models would be valued, and for non-linear models, it would be valued to use methods such as Deep Learning, CNN networks or a combination of SOM and MLP.

It is also considered that the models created could be applied to other variables in case there were sufficient data. It would be interesting to model the behavior of the brake or the orientation system (YAW).

VI. REFERENCES

- [1] Peyman Mazidi, Student Member, IEEE, Lina Bertling Tjernberg, Senior Member, IEEE, and Miguel A. Sanz-Bobi, Senior Member, IEEE (2017). *Performance Analysis and Anomaly Detection in Wind Turbines based on Neural Networks and Principal Component Analysis*. Conference: 12th Workshop on Industrial Systems and Energy Technologies in Madrid, Spain (May 2017).
- [2] Mian Du, Lina Bertling Tjernberg, Shicong Ma, Qing He1, Lin Cheng and Jianbo Guo (2016). *A SOM based Anomaly Detection Method for Wind Turbines Health Management through SCADA Data*. Article in the Energies Journal, published on February 20, 2017. DOI: 10.3390/en10020253.
- [3] Nassim Laouti, Nida Sheibat-Othman, Sami Othman (2011). *Support Vector Machines for Fault Detection in Wind Turbines*. Conference: 18th IFAC World Congress in Milano, Italy (August 28 – September 2, 2011).
- [4] Peyman Mazidi, Mian Du, Lina Bertling Tjernberg, and Miguel A. Sanz-Bobi (2017). *A Health condition model for wind turbine monitoring through neural networks and proportional hazard models*. Article in the SAGE Journal, published on May 4, 2011. DOI: 10.1177/1748006X17707902.
- [5] Yonglong Yan, Jian Li, and David Wenzhong Gao (2014). *Condition Parameter Modeling for Anomaly Detection in Wind Turbines*. Article in the Energies Journal, published on May 6, 2014. DOI: 10.3390/en7053104.
- [6] Meng Li 1, Shuangxin Wang 1, Shanxiang Fang 1,2 and Juchao Zhao (February 12, 2020). *Anomaly Detection of Wind Turbines Based on Deep Small-World Neural Network*. Article in the Applied Sciences Journal, published on February 12, 2020. DOI: 10.3390/app10041243. Last accessed on 04/04/2020.

Índice del documento

Parte I	MEMORIA	10
Capítulo 1.	Introducción	11
1.1	Motivación del proyecto	12
1.2	Objetivos del proyecto	12
1.3	Metodología del trabajo	13
1.4	Estado de la cuestión	15
1.4.1	El aerogenerador moderno	15
1.4.2	Modelos de regresión lineal	18
1.4.3	Modelos de redes neuronales artificiales	19
1.4.4	Revisión de estudios anteriores	26
1.5	Recursos a emplear	28
Capítulo 2.	Elección de variables y justificación de modelos	29
2.1	Importación de datos	29
2.1.1	Variables	32
2.1.2	Fecha	32
2.1.3	Listados de alarmas y reparaciones	35
2.2	Estudio general de datos	36
2.3	Selección de variables	43
Capítulo 3.	Pretratamiento	49
Capítulo 4.	Creación de los modelos de comportamiento	57
4.1	Modelos lineales (ML)	57
4.1.1	ML1 (Potencia total respecto a la velocidad del viento)	60
4.1.2	ML2 (Giro del rotor respecto a la velocidad del viento)	61
4.1.3	ML3 (Ángulo del controlador de Pitch respecto a la velocidad del viento)	62
4.1.4	ML4 (Temperatura de los rodamientos DE del generador respecto a la velocidad del viento)	63
4.1.5	ML5 (Temperatura de los rodamientos NDE respecto a la velocidad del viento)	64
4.1.6	ML6 (Temperatura del aceite de la multiplicadora respecto a la velocidad del viento)	65

4.1.7 ML7 (Temperatura de los anillos SLD del generador respecto a la velocidad del viento)	66
4.1.8 ML8 (Presión del grupo hidráulico respecto a la velocidad del viento)	67
4.1.9 ML9 (Temperatura del grupo hidráulico respecto a la velocidad del viento)	68
4.2 Modelos no lineales (MNL)	69
4.2.1 MNL1 (Potencia generada respecto a la velocidad del viento)	71
4.2.2 MNL2 (Giro del rotor respecto a la velocidad del viento)	75
4.2.3 MNL3 (Ángulo del controlador de Pitch respecto a la velocidad del viento)	79
4.2.4 MNL4 (Temperatura de los rodamientos DE del generador respecto a la velocidad del viento)	83
4.2.5 MNL5 (Temperatura de los rodamientos NDE respecto a la velocidad del viento)	87
4.2.6 MNL6 (Temperatura del aceite de la multiplicadora respecto a la velocidad del viento)	91
4.2.7 MNL7 (Temperatura de los anillos SLD del generador respecto a la velocidad del viento)	95
4.2.8 MNL8 (Presión del grupo hidráulico respecto a la velocidad del viento)	99
4.2.9 MNL9 (Temperatura del grupo hidráulico respecto a la velocidad del viento)	103
4.3 Observaciones generales	107
4.4 Conclusiones	108
Capítulo 5. Detección de anomalías	112
5.1 Introducción	112
5.2 Extracción de anomalías	112
5.3 Curvas de anomalías	119
5.4 Reparaciones	123
Capítulo 6. Resultados y conclusiones	130
Capítulo 7. Futuros desarrollos	132
Capítulo 8. Bibliografía	133
Parte II Estudio económico	134
Capítulo 1. Presupuesto	135
1.1 Mediciones	135
1.1.1 Equipo	135

1.1.2	Software y licencias.....	135
1.1.3	Mano de obra.....	136
1.2	Precios unitarios.....	136
1.2.1	Equipo.....	136
1.2.2	Software y licencias.....	136
1.2.3	Mano de obra.....	137
1.3	Sumas parciales.....	137
1.3.1	Equipo.....	137
1.3.2	Software y licencias.....	137
1.3.3	Mano de obra.....	138
1.4	Presupuesto General.....	138
Parte III	Apéndice I.....	139
Capítulo 1.	lista de variables.....	140
Capítulo 2.	Nomenclatura.....	142
Parte IV	Código fuente.....	143
Capítulo 1.	Descripción de la programación.....	144
1.1	Etapa 1.....	145
1.2	Etapa 2.....	145
1.3	Etapa 3.....	146
1.4	Etapa 4.....	147
1.5	Etapa 5.....	148
Capítulo 2.	Listado de archivos.....	150
Capítulo 3.	Enlace a archivos.....	152

Índice de figuras

Figura 1. Partes del aerogenerador.....	16
Figura 2. Esquema de una red neuronal tipo.....	19
Figura 3. Esquema del MLP	22
Figura 4. Contenido de ANG_analog_short_2.mat	30
Figura 5. Contenido de _ANG_alarms.mat	31
Figura 6. Contenido de repara_ANG.mat.....	31
Figura 7. Esquema de la división temporal del proyecto.....	35
Figura 8. Rosa de los vientos y distribución anual de la velocidad del viento	36
Figura 9. Potencia generada anual en función de la velocidad del viento	37
Figura 10. Coeficiente global de potencia frente a la velocidad del viento	39
Figura 11. Coeficiente global de potencia frente a la velocidad específica.....	40
Figura 12. Fecha de inicio y de fin de reparaciones y alarmas	41
Figura 13. Duración desde el inicio hasta el fin de alarmas y reparaciones	42
Figura 14. Variables utilizadas para la construcción de los modelos	48
Figura 15. Corrección del offset de la velocidad del viento desde 09/06/10 a 28/3/11	51
Figura 16. Nube de puntos de C1 en cada año del periodo de estudio	52
Figura 17. Nube de puntos de C2 en cada año del periodo de estudio	53
Figura 18. Nube de puntos de C3 en cada año del periodo de estudio	53
Figura 19. Nube de puntos de C4 en cada año del periodo de estudio	54
Figura 20. Nube de puntos de C5 en cada año del periodo de estudio	54
Figura 21. Nube de puntos de C6 en cada año del periodo de estudio	55
Figura 22. Nube de puntos de C7 en cada año del periodo de estudio	55
Figura 23. Nube de puntos de C8 en cada año del periodo de estudio	56
Figura 24. Nube de puntos de C9 en cada año del periodo de estudio	56
Figura 25. Limpieza de valores aislados en C4	58
Figura 26. Bandas de confianza de ML1 para diferentes anchos de intervalo	59
Figura 27. ML1 en periodo de referencia	60
Figura 28. ML2 en periodo de referencia	61

Figura 29. ML3 en periodo de referencia	62
Figura 30. ML4 en periodo de referencia	63
Figura 31. ML5 en periodo de referencia	64
Figura 32. ML6 en periodo de referencia	65
Figura 33. ML7 en periodo de referencia	66
Figura 34. ML8 en periodo de referencia	67
Figura 35. ML9 en periodo de referencia	68
Figura 36. Cuadro de diálogo de Neural Net Fitting	69
Figura 37. Cuadro de diálogo de la red neuronal de MNL1	72
Figura 38. MNL1 en periodo de referencia	73
Figura 39. Histograma del error de MNL1 en el periodo de referencia	73
Figura 40. Evolución del MSE de MNL1 en el periodo de referencia	74
Figura 41. Regresión E/V/T y total de MNL1 en el periodo de referencia	74
Figura 42. Cuadro de diálogo de la red neuronal de MNL2	76
Figura 43. MNL2 en periodo de referencia	77
Figura 44. Histograma del error de MNL2 en el periodo de referencia	77
Figura 45. Evolución del MSE de MNL2 en el periodo de referencia	78
Figura 46. Regresión E/V/T y total de MNL2 en el periodo de referencia	78
Figura 47. Cuadro de diálogo de la red neuronal de MNL3	80
Figura 48. MNL3 en periodo de referencia	81
Figura 49. Histograma del error de MNL3 en el periodo de referencia	81
Figura 50. Evolución del MSE de MNL3 en el periodo de referencia	82
Figura 51. Regresión E/V/T y total de MNL3 en el periodo de referencia	82
Figura 52. Cuadro de diálogo de la red neuronal de MNL4	84
Figura 53. MNL4 en periodo de referencia	85
Figura 54. Histograma del error de MNL4 en el periodo de referencia	85
Figura 55. Evolución del MSE de MNL4 en el periodo de referencia	86
Figura 56. Regresión E/V/T y total de MNL4 en el periodo de referencia	86
Figura 57. Cuadro de diálogo de la red neuronal de MNL5	88
Figura 58. MNL5 en periodo de referencia	89

Figura 59 Histograma del error de MNL5 en el periodo de referencia	89
Figura 60. Evolución del MSE de MNL5 en el periodo de referencia	90
Figura 61. Regresión E/V/T y total de MNL5 en el periodo de referencia	90
Figura 62. Cuadro de diálogo de la red neuronal de MNL6	92
Figura 63. MNL6 en periodo de referencia	93
Figura 64 Histograma del error de MNL6 en el periodo de referencia	93
Figura 65. Evolución del MSE de MNL6 en el periodo de referencia	94
Figura 66. Regresión E/V/T y total de MNL6 en el periodo de referencia	94
Figura 67. Cuadro de diálogo de la red neuronal de MNL7	96
Figura 68. MNL7 en periodo de referencia	97
Figura 69. Histograma del error de MNL7 en el periodo de referencia	97
Figura 70. Evolución del MSE de MNL7 en el periodo de referencia	98
Figura 71. Regresión E/V/T y total de MNL7 en el periodo de referencia	98
Figura 72. Cuadro de diálogo de la red neuronal de MNL8	100
Figura 73. MNL8 en periodo de referencia	101
Figura 74. Histograma del error de MNL8 en el periodo de referencia	101
Figura 75. Evolución del MSE de MNL8 en el periodo de referencia	102
Figura 76. Regresión E/V/T y total de MNL8 en el periodo de referencia	102
Figura 77. Cuadro de diálogo de la red neuronal de MNL9	104
Figura 78. MNL9 en periodo de referencia	105
Figura 79. Histograma del error de MNL9 en el periodo de referencia	105
Figura 80. Evolución del MSE de MNL9 en el periodo de referencia	106
Figura 81. Regresión E/V/T y total de MNL9 en el periodo de referencia	106
Figura 82. Comparación de ML y MNL en las curvas de estado	108
Figura 83. Variables de entrada divididas por años y estaciones	118
Figura 84. Mapa de puntos de anomalías en los ML	119
Figura 85. Mapa de puntos de anomalías en los MNL	120
Figura 86. Curvas de anomalías de las curvas C1, C2, C3 y C4	121
Figura 87. Curvas de anomalías de las curvas C5, C6, C7, C8 y C9	122
Figura 88. Curva de anomalías de C8 y C9 para reparaciones del grupo hidráulico	125

Figura 89. Curva de anomalías de C5 y C7 para reparaciones del generador	126
Figura 90. Curva de anomalías de C4 y C6 para reparaciones de la caja multiplicadora...	127
Figura 91. Curva de anomalías de C1, C4, C5 y C7 para reparaciones en el sistema de medida de la velocidad del viento.....	128
Figura 92. Curva de anomalías de C1, C2 y C3 para reparaciones durante Y3.....	129
Figura 93. Esquema de programación	144

Índice de tablas

Tabla 1. Comparación del proyecto con estudios anteriores	27
Tabla 2. Correlación entre años del estudio y fechas.....	34
Tabla 3. Factor de carga del aerogenerador durante cada año	38
Tabla 4. Rendimiento de un aerogenerador en función del factor de carga	38
Tabla 5. Principales grupos de alarmas y reparaciones	43
Tabla 6. Variables del aerogenerador y porcentaje de datos disponibles	45
Tabla 7. Fechas antes y después de eliminar valores de aerogenerador fuera de servicio ..	50
Tabla 8. Datos de las curvas de estado en el periodo de referencia.....	52
Tabla 9. Resultados de MNL1 para diferentes configuraciones de MLP	72
Tabla 10. Resultados de MNL2 para diferentes configuraciones de MLP	76
Tabla 11. Resultados de MNL3 para diferentes configuraciones de MLP	80
Tabla 12. Resultados de MNL4 para diferentes configuraciones de MLP	84
Tabla 13. Resultados de MNL5 para diferentes configuraciones de MLP	88
Tabla 14. Resultados de MNL6 para diferentes configuraciones de MLP	92
Tabla 15. Resultados de MNL7 para diferentes configuraciones de MLP	96
Tabla 16. Resultados de MNL8 para diferentes configuraciones de MLP	100
Tabla 17. Resultados de MNL9 para diferentes configuraciones de MLP	104
Tabla 18. Resultados de los modelos durante el periodo de referencia	109
Tabla 19. Regresión lineal de los MNL por años	109
Tabla 20. Error medio de los todos los modelos durante el periodo de entrenamiento	110
Tabla 21. Desviación típica de los todos los modelos durante el periodo de entrenamiento	111
Tabla 22. Valores fuera de las bandas de confianza por años y modelos	113
Tabla 23. Distribución temporal del registro de anomalías	115
Tabla 24. Distribución temporal de anomalías en ML	116
Tabla 25. Distribución temporal de anomalías en MNL.....	117

Tabla 26. Distribución temporal de todos los modelos por estaciones.....	117
Tabla 27. Listado de reparaciones tipo “correc” en el periodo de entrenamiento	124
Tabla 28. Mediciones de equipo	135
Tabla 29. Mediciones de softwares y licencias.....	135
Tabla 30. Mediciones de mano de obra directa	136
Tabla 31. Precios unitarios de equipos	136
Tabla 32. Precios unitarios de softwares y licencias.....	136
Tabla 33. Precios unitarios de mano de obra directa	137
Tabla 34. Suma parcial de equipos	137
Tabla 35. Suma parcial de softwares y licencias	137
Tabla 36. Suma parcial de mano de obra directa.....	138
Tabla 37. Presupuesto general	138
Tabla 38. Variables del aerogenerador	141
Tabla 39. Variables temporales.....	141
Tabla 40. Variables de alarmas y reparaciones.....	141
Tabla 41. Acrónimos.....	142
Tabla 42. Símbolos	142

Parte I MEMORIA

Capítulo 1. INTRODUCCIÓN

En 2010 había menos de 200.000MW de energía eólica instalada en todo el mundo. En 2018, ese número se situó en 591.000MW. La demanda de energía a nivel global no cesa de aumentar, hecho que se acentúa en el caso de la energía eléctrica. Se trata de una fuente de energía renovable, inagotable, y no contaminante. Esto último cobra gran importancia hoy en día porque cada vez se pone más el foco en buscar fuentes de energía que reduzcan el impacto ambiental, y de todas ellas, la energía eólica es la que más se ha instalado. Por todo ello, es conveniente optimizar el uso y los recursos destinados a dicho tipo de energías. Este proyecto se centra en la energía eólica, más concretamente la elaboración de un sistema que permita detectar las anomalías en el funcionamiento de un aerogenerador y a través del estudio de sus curvas de estado.

Este trabajo pone su atención en la creación de modelos de comportamiento de la curva de potencia y otras curvas de estado que manifiesta un aerogenerador en su operación. Existen numerosos estudios que analizan la curva de potencia del aerogenerador para analizar su funcionamiento y detectar anomalías, sin embargo, no se presta tanta atención a otras curvas que pueden ser relevantes en la detección de anomalías. La aleatoriedad de las condiciones ambientales, o el defecto o deterioro de algún componente pueden no ser correctamente detectados y analizados únicamente con la curva de potencia. Existen otras curvas, comunes a todo tipo de aerogeneradores que pueden aportar más información y, en consecuencia, generar un mejor modelo del funcionamiento de la instalación. Teniendo en cuenta lo anterior, este trabajo analizará la vida del aerogenerador en base a nueve curvas de estado con las que se controlará el funcionamiento de componentes del sistema tales como la multiplicadora, el generador eléctrico, el controlador de pitch o el grupo hidráulico.

1.1 MOTIVACIÓN DEL PROYECTO

La motivación de este proyecto es crear un sistema de detección de anomalías que ayude en la operación y mantenimiento del aerogenerador o aerogeneradores a los que se aplique, y con ello, evite o mitigue el decrecimiento del rendimiento que indudablemente se produce en los sistemas eólicos con el paso del tiempo y las horas de funcionamiento por envejecimiento o degradación. Esta mejora no sólo supone un beneficio económico para el propietario del aerogenerador, sino también para el planeta, pues se trata de una energía limpia y al alza.

1.2 OBJETIVOS DEL PROYECTO

El objetivo principal de este proyecto es crear un sistema que detecte anomalías en el funcionamiento de un aerogenerador a través de datos de su operación disponibles. A continuación, se recogen los subobjetivos que se han determinado para evaluar el desarrollo del proyecto:

- Importar los datos medidos mediante el SCADA del aerogenerador y seleccionar cuales son las variables disponibles que van a permitir una mejor monitorización del aerogenerador y sus componentes principales.
- Crear modelos de comportamiento para las curvas de estado (tanto de la curva de potencia como de otras relaciones entre la velocidad el viento y variables intermedias que muestren el comportamiento de componentes vitales para el funcionamiento del aerogenerador) creadas a partir de las variables seleccionadas anteriormente. Se realizarán modelos tanto lineales como no lineales, siempre y cuando su rendimiento justifique su implementación.
- A partir de los modelos de comportamiento se buscarán anomalías a lo largo del periodo que se simule y se comprobará la valía del estudio comparándolo con el registro de reparaciones del que se dispone.

1.3 METODOLOGÍA DEL TRABAJO

En primer lugar, se procede a importar los datos disponibles del sistema SCADA que monitoriza el funcionamiento del aerogenerador. Estos datos consisten en 32 variables del aerogenerador, la fecha de las mediciones de las variables y 2 registros, uno de alarmas y otro de reparaciones realizadas durante el periodo de estudio.

Seguidamente se procede a estudiar las variables disponibles. En primer lugar, hay que tener en cuenta los valores disponibles de cada una, pues debido a anomalías o errores en el funcionamiento de los componentes del sistema, hay periodos en los que algunas variables miden de manera errónea o no se ha registrado ninguna medida, esto último se guarda en Matlab como registro “Not a Number” (NaN). Tras desechar las variables de las que no se dispone de un número razonable de mediciones, se han de seleccionar las variables más representativas del sistema y de los componentes más importantes del régimen de funcionamiento. Estos componentes pueden ser:

- Rotor
- Controlador del pitch
- Multiplicadora
- Generador eléctrico
- Freno
- Sistema de orientación (YAW)
- Sistemas de refrigeración
- Transformador

Una vez seleccionadas las variables de estudio, se generan las curvas de estado que servirán como base para crear modelos de comportamiento. Para ello se adecúan los datos con anterioridad para poder trabajarlos. Esta adecuación consiste en aislar los valores NaN para poder analizar el resto de datos, desechar los datos medidos cuando el aerogenerador se encuentra fuera de servicio y corregir aquellos con mediciones erróneas. A continuación, se generan los modelos de comportamiento. Estos modelos se dividen en:

- Modelos de comportamiento lineal (ML): Se crea para cada curva un modelo de regresión lineal estadístico. Para ello se toman los valores más representativos de cada curva (en torno al 95-99%) y se obtiene una media móvil y una desviación típica móvil de un número determinado de puntos, y a través de estas dos variables se crean unas bandas de confianza de valor igual a la media más/menos 2 desviaciones típicas, similar a las bandas de Bollinger que se utilizan en el análisis bursátil. El número de valores de la media móvil y la desviación típica móvil depende del ancho de intervalo que se tome para construir el modelo. Este ancho debe ser lo suficientemente amplio como para que no se desvíe notoriamente ante valores aislados fuera de lo común, y lo suficientemente estrecho como para que entre sus bandas de confianza no se encuentren valores anómalos.
- Modelos de comportamiento no lineal (MNL): Mediante el uso de redes neuronales se procede a crear modelos de predicción que compitan en rendimiento con los modelos lineales citados anteriormente. Estos modelos deberían mejorar la precisión de los homólogos lineales. Se utilizan perceptrones multicapa (MLP) que tomarán como entrada las variables de entrada al sistema (velocidad del viento, temperatura y ángulo del viento) y darán como salida la predicción de la variable de estado de la curva que estén representando.

El periodo de tiempo disponible se ha de dividir en 2 partes. En cada caso se pasa por una primera etapa en la que, a partir de las medidas de las variables de entrada y la variable a predecir implicadas en ese modelo, se construye el modelo de comportamiento. Después se llega a una segunda etapa donde las medidas de las variables de entrada se pasan por el modelo para obtener una simulación de la variable a predecir.

Se crean curvas de anomalías, obtenidas de evaluar el error entre el valor medido y el valor simulado mediante los modelos de comportamiento. Se superponen las curvas de anomalías al registro de reparaciones y se comprueba que los modelos creados predicen la necesidad de una reparación en el componente que determine el modelo.

1.4 ESTADO DE LA CUESTIÓN

La energía eólica, y su aprovechamiento mediante aerogeneradores está cada vez más presente en nuestra sociedad. Es un tipo de energía renovable, limpia, con una amplia aceptación social y que, dadas las políticas energéticas de los gobiernos mundiales, aumentará su presencia e importancia en los años venideros. En esta parte del documento, se procede a describir los componentes del aerogenerador y su funcionamiento. Después se realiza una introducción de los modelos de regresión lineal y de las redes neuronales, detallando sus beneficios y aplicaciones, y finalmente se mostrarán estudios acerca de las curvas de estado del aerogenerador para predecir su funcionamiento y detectar posibles anomalías.

1.4.1 EL AEROGENERADOR MODERNO

El aerogenerador es un tipo de máquina de generación eléctrica a partir de la potencia del viento que es bien conocida desde la antigüedad. Brevemente se describe el papel que tienen sus componentes, pues sobre ellos versa gran parte del contenido del proyecto.

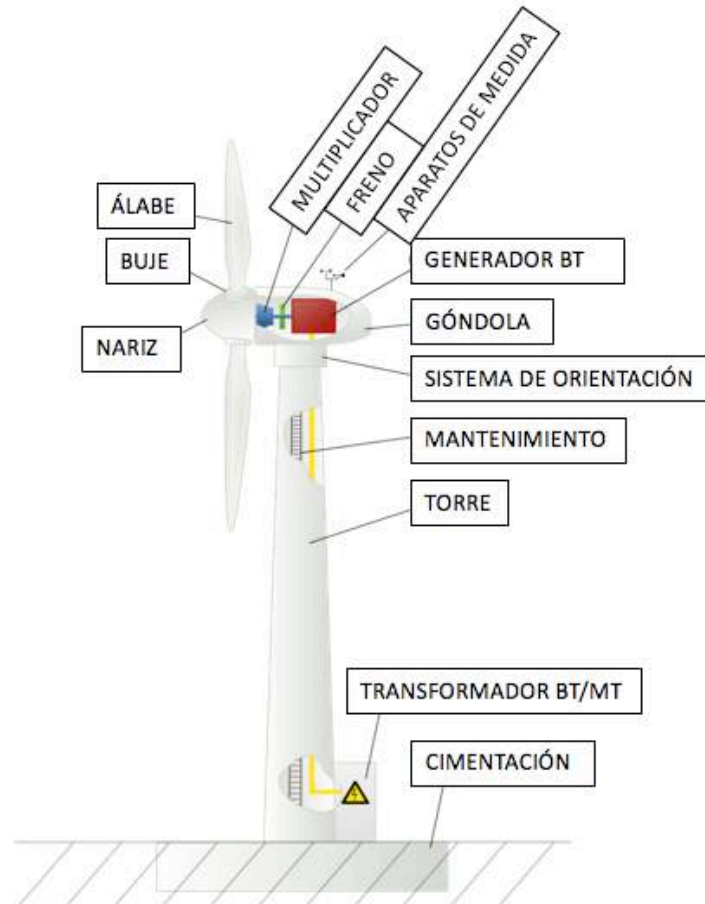


Figura 1. Partes del aerogenerador

- 1) Rotor: componente del motor que gira respecto a la parte fija para conseguir energía. Las palas del rotor, construidas a partir de materiales compuestos, se diseñan para transferir la energía cinética del viento en momento torsor en el eje del equipo. Pueden alcanzar diámetros de hasta 80 metros. Criterios acústicos delimitan la velocidad máxima de rotación. El rotor a su vez se compone de:
 - a) Buje: elemento de unión entre las palas que se encuentra acoplado al eje de baja velocidad del aerogenerador.
 - b) Álabes: nombre técnico que reciben las palas del aerogenerador.

- c) Sistema de inclinación de pala: también conocido como sistema de pitch. Permite variar la inclinación de las palas respecto al mismo. Se encuentra sobre el buje, en la base de los álabes.
- 2) Góndola: conjunto de bastidor y caja que sirve de alojamiento para los elementos mecánicos y eléctricos del aerogenerador. La carcasa suele ser de fibra de vidrio y poliéster, reforzada con perfiles de acero inoxidable.
- a) Eje de baja velocidad: conecta el buje con la multiplicadora y transmite la energía proveniente de las palas.
- b) Freno: se instala un freno en el eje de baja velocidad para evitar que el rotor alcance velocidades no deseadas.
- c) Multiplicadora: también llamada caja multiplicadora, se trata del conjunto de engranajes que transforman la rotación de baja velocidad del rotor a una velocidad más elevada al eje que entra en el generador eléctrico.
- d) Eje de alta velocidad: conecta la multiplicadora con el generador eléctrico.
- e) Generador eléctrico: transforma la energía mecánica en eléctrica.
- f) Grupo hidráulico: proporciona la presión necesaria para los componentes accionados hidráulicamente, como el freno, el sistema de control de pitch o el sistema de bloqueo del rotor.
- g) Sistema de orientación (YAW): situado en la base superior de la torre, permite orientar las aspas hacia la posición más eficaz para el aprovechamiento del viento.
- h) Sistema de control: se hace cargo del funcionamiento seguro y eficiente del equipo.
- i) Anemómetro y veleta: para medir la velocidad y dirección del viento respectivamente.
- 3) La torre: sitúa el generador a una mayor altura, donde los vientos son de mayor intensidad, así como para permitir el giro de las palas, y transmite las cargas del equipo al suelo mediante el cableado interno. En ella hay una escalera de acceso hacia la góndola.

Cimentación: Plataforma de alta resistencia sobre la que se instala la torre.

1.4.2 MODELOS DE REGRESIÓN LINEAL

Son un tipo de modelos matemáticos que se utilizan para aproximar una relación de dependencia entre una variable dependiente Y , un número determinado de variables independientes X_i y un término aleatorio ε .

$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

Ecuación 1. Función general de la regresión lineal.

Cuando sólo se maneja una variable independiente se denomina modelo de regresión simple, mientras que cuando el número de variables aumenta, ya sea porque tomado un mayor número de variables independientes o porque se trabaja con una única variable independiente dividida en intervalos, se le denomina modelo de regresión lineal múltiple.

El modelo que se va a utilizar en este trabajo es un modelo de regresión lineal múltiple de media móvil (MA). En este tipo de modelos se calcula la predicción de la variable dependiente en función de un término independiente y de una sucesión ponderada de errores correspondientes a los instantes precedentes.

Estos modelos se suelen combinar con un componente auto-regresivo (AR), y un componente integrado (I) para formar un modelo ARIMA. ARIMA ha sido el método más usado para identificar y construir modelos de predicción para series temporales univariantes y estacionarias hasta los últimos años, cuando los modelos de predicción mediante redes neuronales han demostrado tener una mayor precisión y fiabilidad al abordar el problema de la no linealidad.

1.4.3 MODELOS DE REDES NEURONALES ARTIFICIALES

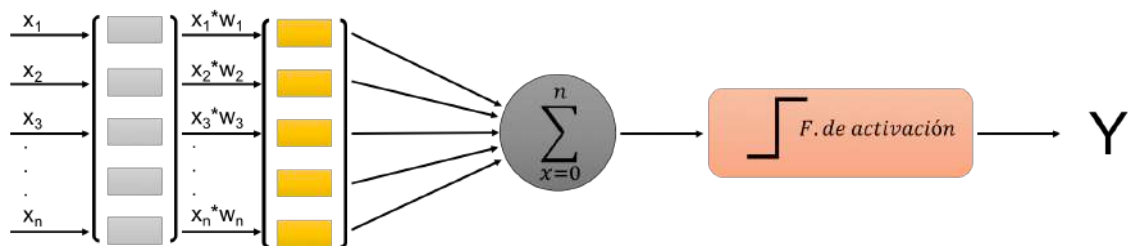


Figura 2. Esquema de una red neuronal tipo

Se trata de sistemas inspirados en el cerebro humano para el modelado de todo tipo de funciones, ya sean lineales o no lineales (principal inconveniente de los modelos ARIMA). Están formados por unidades simples de procesamiento llamadas neuronas o nodos conectadas entre sí, con varias entradas a las que se les asigna un peso específico y una única salida de datos que es función de una combinación lineal de los valores de sus entradas.

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

Ecuación 2. Función de una red neuronal tipo

Las redes neuronales pueden dividirse en redes unidireccionales (feedforward neural networks) y redes recurrentes (backprogramation neural networks), dependiendo de cómo esté realizada la interconexión de neuronas: en las redes unidireccionales las neuronas

están conectadas siguiendo una única dirección entrada - salida; mientras que las redes recurrentes puede haber cualquier dirección de conexión entre neuronas.

La estructura de una red neuronal viene determinada por la disposición de las neuronas y de sus funciones de activación. Las neuronas se agrupan en capas, considerándose como “ocultas” aquellas que no corresponden a la capa de entradas (las neuronas que actúan como entrada de señales o valores externos) y a la capa de salidas (las neuronas que proporcionan su salida al exterior). El conjunto de valores que acceden a las neuronas de la capa de entradas (un valor por neurona), correspondientes a las entradas de la propia red neuronal, recibe la denominación de valor del vector de entradas.

La principal característica de las redes neuronales es su capacidad de aprendizaje. Este aprendizaje consiste en el reajuste continuo de los pesos de las conexiones entre neuronas de una forma iterativa para así “aprender” las relaciones existentes entre pares de valores de vectores de entradas-salidas que se le suministren, o encontrar un patrón común en datos de entrada para poderlos clasificar de acuerdo a los distintos patrones. Se pueden distinguir tres tipos de aprendizajes en redes neuronales comentados seguidamente.

- **Aprendizaje supervisado:** Se suministran a la red neuronal tanto el valor del vector de entradas, como el valor del vector de salidas correspondientes a dichas entradas. Durante el proceso de aprendizaje, la red neuronal ajusta sucesivamente el peso de las conexiones de forma que se minimiza el error entre los valores del vector de salidas de la red neuronal y los valores del vector de salidas esperado. Su uso principal es la aproximación de funciones no lineales con redes neuronales.
- **Aprendizaje no supervisado:** Se suministran a la red neuronal únicamente los valores del vector de entradas, sin las salidas esperada (de ahí lo de aprendizaje no supervisado), para que busque en los valores del vector de entradas alguna propiedad o correlación que los permita clasificar de acuerdo a la misma.
- **Aprendizaje reforzado:** Mezcla de los 2 anteriores. Se trata de un caso especial de aprendizaje supervisado en el que el valor del vector de salidas esperado no es

suministrado: únicamente se suministra a la red neuronal la información precisa de si el valor del vector de salidas es igual o no al esperado.

Como ventajas del uso de redes neuronales, destaca el hecho de que no es necesario conocer la naturaleza del conjunto de datos de estudio, pues el proceso de entrenamiento aporta las características de los mismos. Asimismo, el comportamiento no lineal de las funciones de activación de las redes neuronales permite que se pueda trabajar con funciones no lineales. Finalmente, son más sencillas de programar que otros modelos como el ARIMA.

Pero todo no son ventajas, ya que necesitan de una gran cantidad de datos de entrada y salida para optimizar su funcionamiento, y es posible que el entrenamiento conduzca a un mínimo local. Para este caso, se dispone de vectores de 291312 valores por variable, luego la falta de datos no resulta un problema, y para evitar el problema de los mínimos locales no supervisados.

1.4.3.1 Perceptrón multicapa (MLP)

El perceptrón multicapa o red neuronal MLP (Multilayer Perceptron) es un tipo de red neuronal unidireccional con una organización de neuronas dispuesta en capas, de tal manera que las neuronas de una capa tomen como entradas las salidas de la capa de neuronas anterior y sus salidas sean las entradas de la capa de neuronas posterior. Lógicamente esta estructura no se cumple para la primera capa (capa de entradas), ni la última (capa de salidas), sino para las intermedias (capas ocultas). El número de capas ocultas puede variar, aunque con una sola ya se podría aproximar con cierto grado de error una función continua acotada.

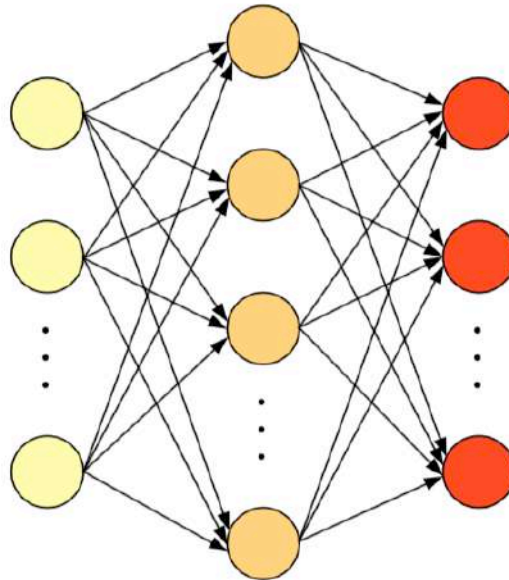


Figura 3. Esquema del MLP

El MLP se aplica tanto para casos de clasificación como para aproximación de funciones. Durante su diseño hay que seleccionar el número de capas ocultas y de neuronas por capa. Por un lado, al aumentar dichos números, aumenta la capacidad del MLP para predecir el resultado deseado, pero aumenta el tiempo necesario de entrenamiento de la red, y lo que es más peligroso, se corre el riesgo de sobre-entrenar la red. Con el sobre-entrenamiento (over-fitting) una red puede obtener un valor muy cercano a la salida deseada para los datos con los que se le ha entrenado, pero al haberse adaptado tanto a ellos, no podrá adaptarse a nuevos valores de entrada, dando malos resultados.

Existen diversas técnicas para el aprendizaje o entrenamiento de las redes neuronales MLP. El más extendido es el de retropropagación del error (backpropagation) que ha sido adaptado al entrenamiento de más tipos de redes neuronales, aparte de las MLP. La técnica de backpropagation consiste, básicamente, en actualizar los pesos de las conexiones entre neuronas proporcionalmente a su contribución al error. Su aplicación puede realizarse actualizando el peso de las conexiones neuronales tras la presentación de cada valor del vector de entradas “instance-based”, o realizando la actualización del peso de las conexiones tras haber presentado todos los valores del vector de entradas “batch”. En este

último caso, a cada actualización se le denomina “epoch”.

Existen dos versiones en la aplicación de esta técnica, la conocida como versión “instance-based” u “on-line” y la versión “batch” u “off-line”. En la versión “on-line” de entrenamiento, la actualización de los pesos de la red neuronal se realiza tras la presentación de cada valor del vector de entradas. En la versión “off-line”, la actualización de los pesos de la red neuronal se realiza tras la presentación de todo el conjunto de valores del vector de entradas de los datos de entrenamiento (lo que constituye un “epoch”). En los trabajos de investigación recogidos en este documento se ha utilizado siempre este segundo método.

El gradiente de error local para una neurona de la capa de salidas es el producto de la derivada de la función de error escogida (normalmente el error cuadrático) y de la función de activación de la neurona. El gradiente de error local para una neurona de una capa oculta es igual a la suma ponderada de los gradientes de error local de las neuronas a las que conectan las conexiones que salen de dicha neurona (el peso de cada conexión actúa de factor de ponderación).

$$\Delta w_{ij}(n) = \eta \delta_j o_i(n) + \alpha \Delta w_{ij}(n-1)$$

Ecuación 3. Gradiente de error local de una neurona

Los coeficientes, η y α juegan un papel relevante en el entrenamiento de la red neuronal. Ambos son dos constantes numéricas de importancia en el entrenamiento de la red neuronal.

Un alto valor del factor de aprendizaje, η , hace que la red neuronal reaccione rápidamente a cambios en las entradas, pero también que sea inestable si es demasiado elevado. Por el contrario, un factor de aprendizaje bajo hace que el tiempo necesario para el entrenamiento de la red neuronal se incremente. El coeficiente α o momento describe la porción de

cambio de peso que se añade en la siguiente actualización de pesos. Un bajo valor del momento hace que el entrenamiento se vuelva inestable, ocasionando oscilaciones en los valores de los pesos de las conexiones e impidiendo el aprendizaje de la red neuronal. El máximo valor para el momento es la unidad, y los valores usuales los comprendidos entre 0.5 y 0.9.

Para evitar el sobre-ajuste o sobre-entrenamiento (over-fitting) de la red neuronal, existen varias técnicas como la limitación en el número de epochs en el proceso de entrenamiento (“early-stopping”), la adición de ruido blanco sobre los valores de las variables de entrada, o la utilización de la validación cruzada, siendo esta última, la técnica más extendida. Con ella se dividen los datos disponibles en dos grupos: entrenamiento y validación cruzada (validation). Los datos del grupo de entrenamiento se usan para el ajuste de los parámetros de la red neuronal utilizando el método de entrenamiento elegido. Los datos del grupo de validación cruzada no se emplea en la determinación del error que se pretende minimizar con el entrenamiento de la red neuronal, sino que se presentan a la red neuronal tras cada epoch de entrenamiento. El proceso de entrenamiento se interrumpe cuando haya transcurrido un número predeterminado de epochs sin reducción del error con los datos de validación cruzada.

1.4.3.2 Otros tipos de redes neuronales

Mapa auto-organizado (SOM)

Se trata de redes neuronales en cuyo entrenamiento se utiliza aprendizaje no supervisado. Se realiza una transformación de un espacio de entrada continuo y con dimensión arbitraria, a un espacio discreto de una o dos dimensiones al que se llama mapa, manteniendo las propiedades topológicas de la entrada. Esto se realiza por reconocimiento, en los valores del vector de entradas, de unas características o patrones que son representadas después en el plano como un mapa de dichas características o patrones.

Los SOM se basan en las llamadas redes neuronales competitivas, donde en función del valor del vector de entradas se activa únicamente una neurona de la capa de salidas (la

unidad ganadora). En una red neuronal competitiva las neuronas de la capa de salidas que implementa el funcionamiento competitivo, suelen ser de tipo binario, en el sentido de que sólo tienen dos posibles valores de salida: 0 cuando están desactivadas y 1 cuando están activadas. En un SOM además se modifican también los valores de los pesos de las conexiones (vectores sinápticos) de las neuronas vecinas, aunque en menor medida, según su proximidad a la neurona ganadora, acercándose al patrón de entradas. Ello garantiza que las neuronas vecinas tengan sus vectores sinápticos parecidos, es decir, se preserve la topología del espacio de entradas. Para el entrenamiento de una red neuronal SOM se utiliza el método de aprendizaje no supervisado de Kohonen

El objetivo de una red neuronal SOM es clasificar los valores del vector de entradas en grupos de características similares, de tal forma que los valores del vector de entradas del mismo grupo activarán siempre las mismas neuronas de la capa de salidas del mapa auto-organizado. De ahí su principal aplicación como clasificador, o sistema para agrupamiento o clusterización de datos.

Red neuronal recurrente (RNN)

Son aquellas redes neuronales en las que las neuronas de una o varias capas se realimentan a sí mismas por medio de unos pesos ajustables. Se puede distinguir entre redes neuronales totalmente recurrentes y redes parcialmente recurrentes. Son redes neuronales totalmente recurrentes aquellas en las que cada una de las neuronas de una capa se realimenta a sí misma. Una red neuronal es parcialmente recurrente cuando sólo algunas neuronas de la capa tienen realimentación.

Respecto a las redes MLP, aumentan la profundidad de la memoria, pero pueden ser inestables durante el proceso de entrenamiento, y éste es más lento que el del resto de redes neuronales, en parte debido a la retro propagación del error a través del tiempo.

1.4.4 REVISIÓN DE ESTUDIOS ANTERIORES

Se ha observado que existe un gran número de estudios que utilizan diferentes tipos de modelos no lineales para detectar anomalías partiendo de la curva de potencia del aerogenerador, y estudiando que sucede en el resto de componentes en los instantes de fallo en el funcionamiento

En [1] se crea un modelo no lineal de comportamiento para la curva de potencia. El tipo de redes neuronales utilizadas son MLP. A partir del modelo creado, se obtiene el error entre estimación y valor real y de ahí se extraen las anomalías. Después se utiliza un análisis de PCA con las variables que se han utilizado para entrenar el modelo y se estudia que variables son responsables de la variabilidad de los datos, para analizar después a qué se debe el fallo. Este proyecto resulta muy útil, pues buena parte de la programación sigue el mismo esquema que se pretende implementar en el proyecto a realizar. La diferencia principal es que se van a estudiar más de un modelo de comportamiento, para poder estudiar directamente el rendimiento de componentes como la caja multiplicadora o el generador, lo que supone una ventaja a la hora de aplicarlo en una instalación real, ya que facilita el proceso de operación y mantenimiento, conocer directamente la fuente concreta de fallos

En [2] el método utilizado para detectar anomalías es un mapa auto-organizado (SOM). Este proyecto es parecido a [1] en el sentido en que ambos utilizan redes neuronales para crear un modelo de comportamiento sobre la curva de potencia para detectar anomalías. En este caso, la base de muestras es de 2 meses, 35 veces menor a la base de muestras de la que se dispone para el proyecto a realizar.

[3] y [4] responden al mismo problema utilizando Support Vector Machines (SVM) y Proportional hazard models (PHM) respectivamente a diferentes variables, como el ángulo del controlador de pitch o temperaturas internas. Es interesante que se busque directamente

la detección de anomalías por estudio directo de estas variables, sin tener que crear antes un modelo comportamiento de la curva de potencia.

En [5] se modelan la mayoría de las variables disponibles mediante redes neuronales recurrentes (RNN). La componente de retro-programación permite tener un resultado excelente, sin embargo, una vez extraído el error no se trabaja en la detección de anomalías.

En [6] se modela el controlador del pitch creando una máquina de Boltzmann restringida (RBM), para hacer después de ella un Deep Small-World Neural Network (DSWNN) añadiendo conexiones entre las capas ocultas de la máquina. Finalmente se valida su correcto funcionamiento detectando anomalías.

En general se puede ver que hay varios métodos que se pueden utilizar para modelar el funcionamiento de un aerogenerador, sin embargo, no se ha encontrado ninguno que permita detectar la causa de las anomalías y que modele no solo la curva de potencia del aerogenerador, sino también los componentes internos.

	Modelos realizados	Modelo/s lineal/es	Modelo/s no lineal/es	Detección de anomalías en curva de potencia	Detección de la causa de anomalía	Método
[1]	Curva de potencia		✓	✓	* (Sólo para variables de entrada del modelo)	MLP + PCA
[2]	Curva de potencia		✓	✓	✓	SOM
[3]	Varios	✓				SVM
[4]	Varios	✓				PHM
[5]	Varios		✓			RNN
[6]	Controlador de pitch		✓	✓		DSWNN
TFG	Varios	✓	✓	✓	✓	Lineal (MA) + MLP

Tabla 1. Comparación del proyecto con estudios anteriores

1.5 RECURSOS A EMPLEAR

Los datos utilizados para el proyecto son los aportados por el director del mismo, y para la programación del mismo se ha realizado íntegramente en lenguaje M mediante Matlab. La versión utilizada es Matlab_R2015b, y se han utilizado las siguientes cajas de herramientas.

- Statistics and Machine Learning Toolbox
- Curve Fitting Toolbox

Cabe destacar que Matlab no era la única opción para programar, pues con los datos y objetivos de los que se dispone, este proyecto podría perfectamente haberse realizado en R o en Python. Finalmente se decidió utilizar Matlab debido a que era el lenguaje con el que más familiarizado me encuentro y a que Matlab dispone de una gran cantidad de recursos online que sirven de gran ayuda a la hora afrontar las dudas y problemas que surgen durante el periodo de programación.

Capítulo 2. ELECCIÓN DE VARIABLES Y JUSTIFICACIÓN DE MODELOS

En este capítulo se exponen los primeros pasos de este proyecto. En primer lugar, se importan los datos provenientes del sistema SCADA que monitoriza el aerogenerador y se transforman en vectores y matrices con los que se pueda operar. Después se estudian dichos datos, para obtener una idea general del rendimiento del aerogenerador en el periodo de datos disponible y observar si hay algún fallo en la medida de alguna variable importante para la consecución del proyecto. Por último, se estudia con qué variables se pueden crear modelos de comportamiento y cuáles de ellas son las más oportunas para satisfacer los objetivos del proyecto.

2.1 IMPORTACIÓN DE DATOS

En este apartado se expone como se han importado las medidas del sistema SCADA que reporta el funcionamiento del aerogenerador que se estudia.

Los datos con los que se trabaja se obtienen de 3 campos distintos:

- Variables y fecha: El fichero ANG_analog_short_2.mat contiene una matriz llamada ANG. Se trata de una estructura formada por 2 campos: Un vector double de 291312x1 correspondiente a la fecha, y otra subestructura de nombre aero. En su interior se encuentran 32 subestructuras, correspondientes a las 32 variables de las que se dispone. Cada subestructura contiene el vector single de 291312x1 registros correspondiente a las medidas de la variable que representa. Cabe destacar que en el caso de las variables vel_viento y pot_total se dispone de tres tipos de medidas: val, que mide el valor para cada instante de medida,

avg, que mide la media de valores entre 2 instantes consecutivos de medida, y avgPI, que a efectos prácticos aporta la misma medida que avg.

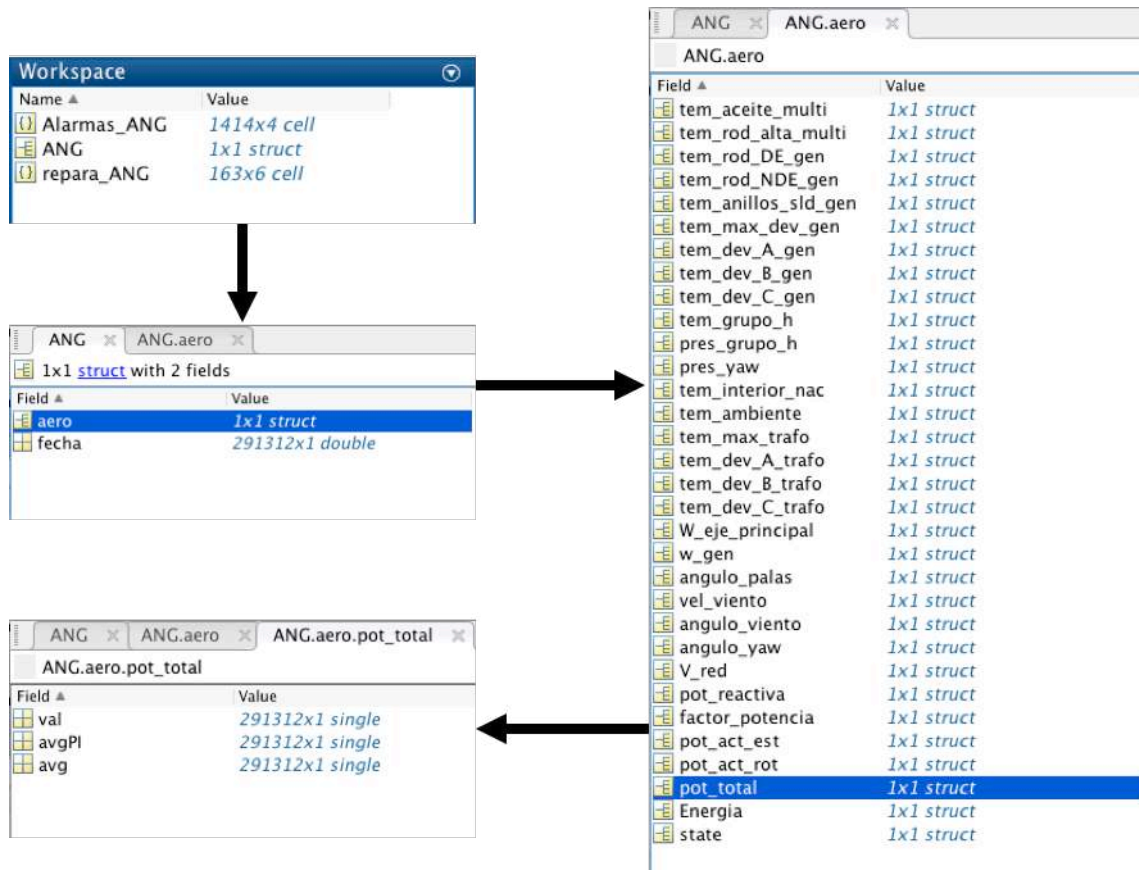
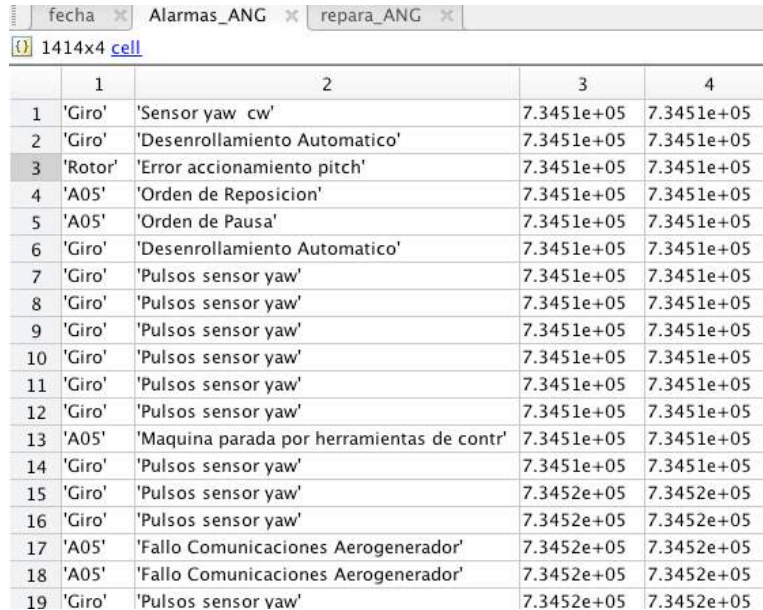


Figura 4. Contenido de `ANG_analog_short_2.mat`

- Listado de alarmas: El fichero `_ANG_alarms.mat` contiene `Alarmas_ANG`. Una matriz de formato cell de 1414x4. Las columnas 3 y 4 corresponden a las fechas de inicio y fin de las alarmas respectivamente y las columnas 1 y 2 contienen información correspondiente a dichas alarmas.



	1	2	3	4
1	'Giro'	'Sensor yaw cw'	7.3451e+05	7.3451e+05
2	'Giro'	'Desenrollamiento Automatico'	7.3451e+05	7.3451e+05
3	'Rotor'	'Error accionamiento pitch'	7.3451e+05	7.3451e+05
4	'A05'	'Orden de Reposicion'	7.3451e+05	7.3451e+05
5	'A05'	'Orden de Pausa'	7.3451e+05	7.3451e+05
6	'Giro'	'Desenrollamiento Automatico'	7.3451e+05	7.3451e+05
7	'Giro'	'Pulsos sensor yaw'	7.3451e+05	7.3451e+05
8	'Giro'	'Pulsos sensor yaw'	7.3451e+05	7.3451e+05
9	'Giro'	'Pulsos sensor yaw'	7.3451e+05	7.3451e+05
10	'Giro'	'Pulsos sensor yaw'	7.3451e+05	7.3451e+05
11	'Giro'	'Pulsos sensor yaw'	7.3451e+05	7.3451e+05
12	'Giro'	'Pulsos sensor yaw'	7.3451e+05	7.3451e+05
13	'A05'	'Maquina parada por herramientas de contr'	7.3451e+05	7.3451e+05
14	'Giro'	'Pulsos sensor yaw'	7.3451e+05	7.3451e+05
15	'Giro'	'Pulsos sensor yaw'	7.3452e+05	7.3452e+05
16	'Giro'	'Pulsos sensor yaw'	7.3452e+05	7.3452e+05
17	'A05'	'Fallo Comunicaciones Aerogenerador'	7.3452e+05	7.3452e+05
18	'A05'	'Fallo Comunicaciones Aerogenerador'	7.3452e+05	7.3452e+05
19	'Giro'	'Pulsos sensor yaw'	7.3452e+05	7.3452e+05

Figura 5. Contenido de `_ANG_alarms.mat`

- Listado de reparaciones: El fichero `repara_ANG.mat` contiene `repara_ANG`. Una matriz de formato `cell` de `163x6`. Las columnas 1 y 2 corresponden a las fechas de inicio y fin de las reparaciones respectivamente y las columnas 3, 4, 5 y 6 contienen información correspondiente a dichas reparaciones.



	1	2	3	4	5	6
1	7.3363e+05	7.3363e+05	'MDY'	'MDY03'	'KR220-A-SERIE DE EMERG. HUB'	'Rearme_loc'
2	7.3363e+05	7.3363e+05	'UMD'	'UMD(SUB)'	'TORRE SOPORTE AEROGENERADOR'	'Correc'
3	7.3364e+05	7.3364e+05	'MDY'	'MDY01'	'MODULO 3.8/3.7-ILC200_UNI PLC AUXILIAR CAB.3'	'Rearme_loc'
4	7.3364e+05	7.3364e+05	'MDY'	'MDY03'	'KR920-SERIE DE EMERG.'	'Rearme_loc'
5	7.3365e+05	7.3365e+05	'MDY'	'MDY03'	'FM230-ALIM. ELECT. VALV. EMERG. PALAS'	'Rearme_loc'
6	7.3365e+05	7.3365e+05	'G87(SIST)'	'G87(SIST)(SUB)'	'G87(SIST)(SUB)(EC)'	'Rearme_loc'
7	7.3365e+05	7.3365e+05	'G87(SIST)'	'G87(SIST)(SUB)'	'G87(SIST)(SUB)(EC)'	'Rearme_loc'
8	7.3365e+05	7.3365e+05	'MDY'	'MDY02'	'A433-DDCS TARJETA DE COMUNICACIONES'	'Correc'
9	7.3365e+05	7.3365e+05	'MDY'	'MDY(SUB)'	'CABLE DE TIERRA'	'Correc'
10	7.3365e+05	7.3365e+05	'MDY'	'MDY03'	'KR920-SERIE DE EMERG.'	'Correc'
11	7.3365e+05	7.3365e+05	'MDY'	'MDY03'	'KR920-SERIE DE EMERG.'	'Rearme_loc'
12	7.3365e+05	7.3365e+05	'MDY'	'MDY03'	'KR330-H-DETECCION 0°C CON OFF DELAY'	'Correc'
13	7.3367e+05	7.3367e+05	'CHI'	'CHI01'	'SENSOR INDUCTIVO-SOBRE VELOCIDAD (VOG)'	'Rearme_loc'
14	7.3367e+05	7.3367e+05	'CHI'	'CHI01'	'SENSOR INDUCTIVO-SOBRE VELOCIDAD (VOG)'	'Rearme_loc'
15	7.3367e+05	7.3367e+05	'MDY'	'MDY06'	'MODULO 0.0-CPU'	'Correc'
16	7.3368e+05	7.3368e+05	'MDY'	'MDY03'	'KR330-H-DETECCION 0°C CON OFF DELAY'	'Rearme_loc'
17	7.3368e+05	7.3368e+05	'MDY'	'MDY03'	'KR330-H-DETECCION 0°C CON OFF DELAY'	'Rearme_loc'
18	7.3369e+05	7.3369e+05	'MDY'	'MDY03'	'KR920-SERIE DE EMERG.'	'Rearme_loc'
19	7.3369e+05	7.3369e+05	'MDY'	'MDY06'	'KR330-G-DESCONEX. POR 0°C'	'Correc'
20	7.3370e+05	7.3370e+05	'MDY'	'MDY06'	'UPS001-CARGADOR DE BATERIAS'	'Correc'
21	7.3370e+05	7.3370e+05	'MDY'	'MDY03'	'KR920-SERIE DE EMERG.'	'Rearme_loc'
22	7.3369e+05	7.3369e+05	'MDY'	'MDY01'	'MODULO 3.8/3.7-ILC200_UNI PLC AUXILIAR CAB.3'	'Correc'
23	7.3370e+05	7.3370e+05	'MDY'	'MDY03'	'KR920-SERIE DE EMERG.'	'Rearme_loc'

Figura 6. Contenido de `repara_ANG.mat`

El fichero de variables con el que principalmente se trabaja en este proyecto. En las conclusiones del proyecto se utiliza también el listado de reparaciones para comprobar que se detectan correctamente anomalías, pues un aumento en el número de anomalías debería desencadenar una reparación del componente responsable de dicha anomalía.

Se crea un script de Matlab que cargue y pase los datos de estas matrices y estructuras a vectores con los que se pueda trabajar. Seguidamente se explica cómo se han modificado los datos para su uso.

2.1.1 VARIABLES

Las 32 variables de ANG.aero se copian como variables individuales numeradas del 1 al 32 en formato double, de la misma manera, se copian los valores del vector ANG.fecha en la variable fecha, de 291312x1 valores double. En el Apéndice I se encuentra la Tabla 38 que recoge la información relativa a cada variable.

2.1.2 FECHA

Los datos del aerogenerador de los que se dispone se registran mediante un sistema SCADA. Este tipo de sistemas SCADA mide los valores de cada parámetro a frecuencias inferiores de 10 minutos, y se realiza una media de estos valores para guardar una medida final cada periodo de 10 minutos. Se dispone de 291312 medidas para cada variable, teniendo en cuenta que el día tiene 24 horas y que cada hora tiene 6 periodos de 10 minutos, en un día hay 144 periodos de 10 minutos, y por lo tanto se dispone de 2023 días de registros.

Las fechas originales se encuentran en formato “datenum”, que convierte una fecha a un valor numérico donde cada unidad es un día. Utilizando el comando “ConvertFrom” se puede pasar este tipo de fechas al formato “datetime”, que devuelve el valor temporal de la fecha. Por ejemplo, el valor numérico 1 en “datenum” es equivalente al valor temporal “01/01/00 00:00”.

El hecho de que la fecha se encuentre en un valor numérico facilita la programación con la misma, sin embargo, la escala "datenum" es del todo intuitiva, y al trabajar con decimales pequeños se pueden producir errores. Por ello se procede a utilizar una escala numérica más sencilla, en la que la unidad pase a ser cada periodo de 10 minutos, y la primera medición disponible valga 1. Con la nueva escala la primera fecha de la que se dispone pasa de ser la fecha 733757 a la fecha 1, y cada fecha posterior suma una unidad, de tal manera que la fecha que originalmente valía 733578 y se encontraba en la posición 145 del vector en formato "datenum", en la nueva escala vale 145 y se encuentra en la posición 145 del vector.

Conviene destacar que en los datos originales hay un salto temporal. Entre el valor 290304 y 290305 (735772.9931 y 735843 "datenum" respectivamente) transcurren 70 días y 10 minutos, 70 días más de lo habitual en el resto de valores. Para facilitar la programación se omite este salto temporal, pero se tendrá en cuenta en el desarrollo del proyecto que ahí sucede un salto temporal.

Utilizando la función "ConvertFrom" para obtener el formato "datetime" del primer y el último valor del vector fecha, se obtiene que la fecha inicial es el 15 de diciembre de 2008 a las 00:00 y la fecha final es el 7 de septiembre de 2014 a las 23:50. Para asegurarse de que esto es correcto, basta con observar la Figura 14, concretamente la variable temperatura ambiental (var14) donde se ve que las temperaturas inicial y final corresponden con las fechas del año señaladas, y también se observa que suceden 6 veranos entre ambos puntos.

Se subdivide el periodo de datos disponibles en 6 años, teniendo en cuenta las siguientes observaciones:

- Se le da a cada año (Y) una duración estándar de 52560 valores, lo que corresponde exactamente a 365 días y 0 horas. El inicio de los 4 primeros años es el 15 de diciembre, en los años 5 y 6 es el 14 de diciembre debido a que el 2012 fue un año bisiesto. No se tiene en cuenta esto debido a que tiene un efecto ínfimo en los datos.

- Los años de estudios no se corresponden a años naturales (2009, 2010, 2011...), pero aproximadamente corresponden al inicio de cada estación (entre 6 y 7 días de diferencia).
- Durante el 6º año se produce el salto temporal de 70 días. Dado que en esa época no se producen reparaciones que sean objeto de este proyecto no resulta un problema.

AÑO	Inicio		Fin	
Y1	1	15/12/08 0:00	52560	14/12/09 23:50
Y2	52561	15/12/09 0:00	105120	14/12/10 23:50
Y3	105121	15/12/10 0:00	157680	14/12/11 23:50
Y4	157681	15/12/11 0:00	210240	13/12/12 23:50
Y5	201241	14/12/12 0:00	262800	13/12/13 23:50
Y6	262801	14/12/13 0:00	*291312 (301391)	07/09/14 23:50

Tabla 2. Correlación entre años del estudio y fechas

La Tabla 2 recoge la correspondencia entre la escala de tiempos utilizada y la fecha real para el comienzo y el final de cada año, y en la Figura 7 se puede ver eso mismo, así como la división por estaciones y periodo en el que no se tiene registro de datos. Como se ha dicho en el apartado 1.3, el rango temporal disponible se ha de dividir en 2 etapas, una primera para crear modelos de comportamiento, y una segunda para aplicar esos modelos. A lo largo de esta memoria estas etapas van a recibir los siguientes nombres:

- Periodo de estudio: corresponde a todo el rango de datos, desde el principio hasta el final, incluyendo el periodo en el que no se dispone de información durante el verano de 2014. En la escala utilizada transcurre desde 1 hasta 291312.
- Periodo de referencia: Corresponde al primer año (Y1) de mediciones. Con los valores registrados durante este periodo se crean los modelos de comportamiento. Transcurre desde 1 hasta 52560.
- Periodo de simulación: la diferencia entre el periodo de estudio y el periodo de referencia (Y2, Y3, Y4, Y5 y Y6). Contiene el periodo sin datos del verano de 2014

y a lo largo de él se simulan las variables deseadas con los modelos de comportamiento entrenados en el periodo de referencia. Transcurre desde 52561 hasta 291312.

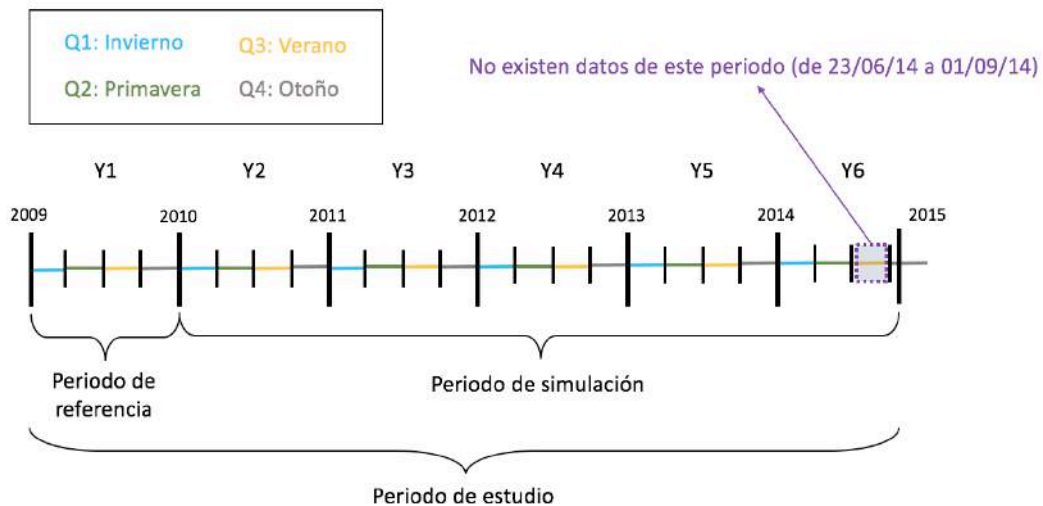


Figura 7. Esquema de la división temporal del proyecto

2.1.3 LISTADOS DE ALARMAS Y REPARACIONES

En ambos casos se dispone de una matriz en formato cell con las fechas de inicio y de fin de cada acción, y otras columnas con información al respecto de cada acción (2 columnas de información para alarmas, y 4 columnas de información para reparaciones).

De forma similar a como se ha hecho con la fecha de las variables del aerogenerador, se convierte la escala “datenum” de los registros temporales de ambas matrices a la escala utilizada del proyecto. Se observa que en ambos casos hay acciones que suceden fuera del periodo de estudio, habiendo reparaciones que suceden antes de la fecha 1, y alarmas que suceden después de la fecha 291312. Se eliminan los registros de las acciones fuera del periodo de referencia. Son eliminadas 7 alarmas y 10 reparaciones, quedando así la matriz de alarmas como 1407x4 y la matriz de reparaciones como 153x6.

2.2 ESTUDIO GENERAL DE DATOS

En este apartado se realiza un primer análisis de los datos provenientes del SCADA, con los que se hace un estudio general del funcionamiento del aerogenerador a nivel macroscópico durante el periodo de estudio de 2023 días.

En primer lugar, se analizará la “materia prima” que requiere el aerogenerador para generar energía, que no es otra que el viento. A continuación, se muestra una rosa de los vientos y un histograma de velocidades de viento para cada año de estudio. Para realizar la rosa de los vientos se utiliza una función creada por otro autor en Matlab, se observa que la dirección del viento es correcta, pero los pesos porcentuales de la misma están divididos entre 4.

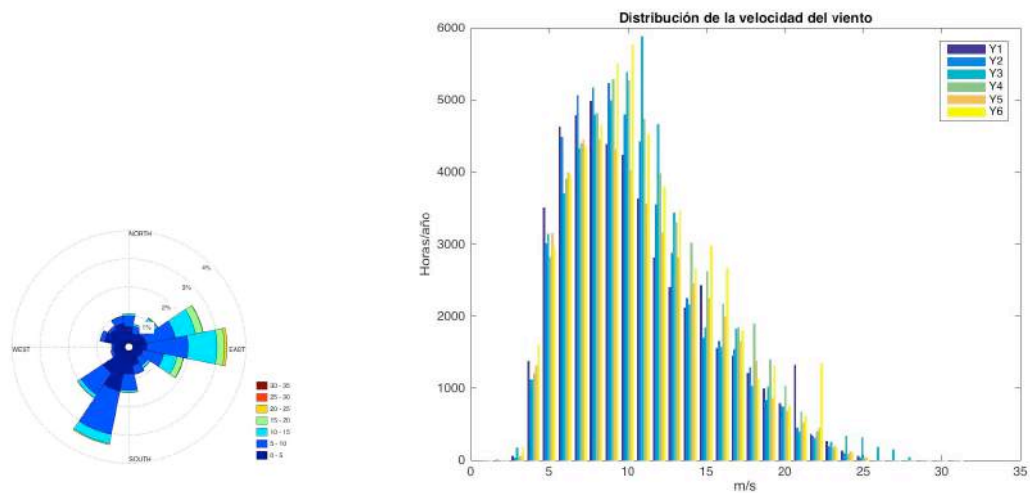


Figura 8. Rosa de los vientos y distribución anual de la velocidad del viento

Dividiendo el rango de velocidades de viento según los criterios de la curva de potencia, se observa que la zona de arranque, de entre 3 y 12 m/s corresponde al 80% de los datos, la zona de trabajo nominal de una velocidad superior a 12 m/s es de un 13,8% y la zona de

parado, menor de 3m/s, de un 6,2%. Esta predominancia de la zona de arranque se ve ligeramente mitigada por los datos del tercer año, en el que se está un 21% del tiempo en la zona de arranque. Para hacer estas medidas en el sexto año se ha escalado para obtener su valor anual, aunque realmente sólo se dispone de datos de los primeros 198 días

Conocida la dirección y la velocidad del viento en cada instante, se puede relacionar la energía eléctrica generada. En primer lugar, se observa que la potencia nominal del aerogenerador es de 2000 kW. Ya que se conocen los datos de potencia en función del viento, se muestra a continuación la energía producida por año para cada m/s de velocidad del viento.

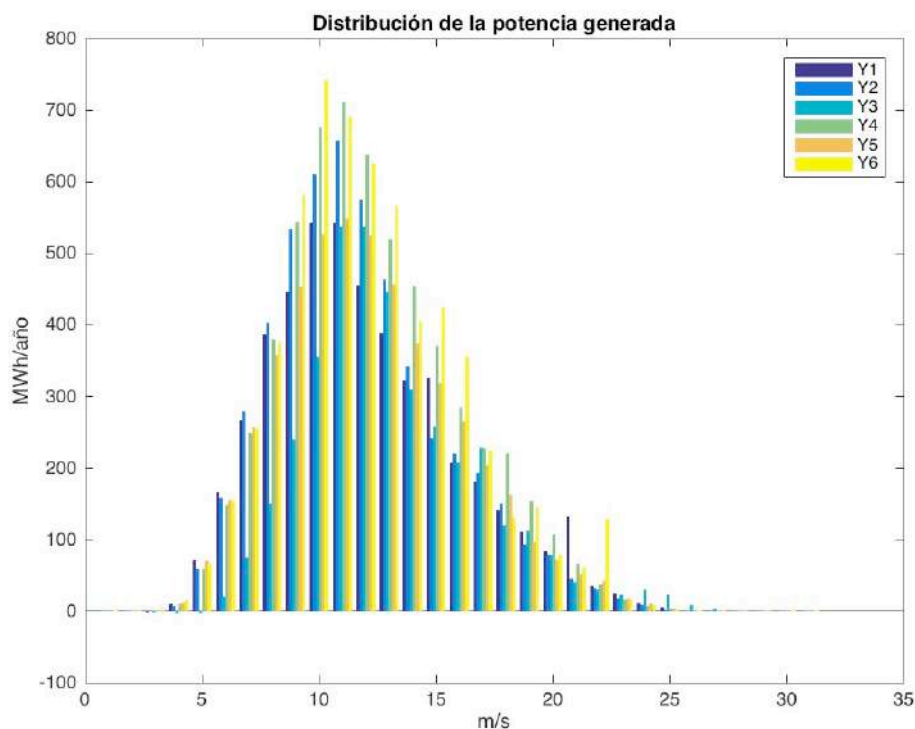


Figura 9. Potencia generada anual en función de la velocidad del viento

En la figura anterior se aprecia claramente que en los años 4 y 6 la generación de energía es máxima, mientras que el tercer año es el de menor rendimiento eléctrico.

Siendo la potencia nominal del aerogenerador 2000kW, en el caso hipotético e irreal de que se pudiese mantener ese valor durante todo un año se generarían 17,52 GW. Comparando con este valor se obtiene los siguientes FC y HE.

AÑO	Factor de Carga (FC)	HE (horas equivalentes a plena carga/año)
Y 1	0,276	2425,5
Y 2	0,294	2582,0
Y 3	0,218	1912,0
Y 4	0,335	2937,5
Y 5	0,285	2486,0
Y 6	0,343	3007,2

Tabla 3. Factor de carga del aerogenerador durante cada año

El rendimiento de la instalación, pese a tener una media más que aceptable de entorno al 28,8 % se ve fuertemente afectado en el Y3. Está claro que entre Y3 y Y4 se produjeron reparaciones que incrementaron el rendimiento general de la máquina, por lo que el estudio de las reparaciones en esta franja temporal cobra gran importancia (5.4)

Factor de carga (FC)	Valoración	Factor de carga (FC)	Valoración
> 0,50	Extraordinario	0,25 – 0,30	Bueno
0,4 – 0,5	Excelente	0,20 – 0,25	Aceptable
0,3 – 0,4	Muy bueno	<0,20	Insuficiente

Tabla 4. Rendimiento de un aerogenerador en función del factor de carga

A continuación, se calcula el coeficiente de potencia (C_p) para cada instante siguiendo la ecuación que a continuación se muestra. Se representa el coeficiente de potencia sobre la velocidad del viento y sobre la velocidad específica (λ). Para realizar estos cálculos se ha tomado la densidad del aire como 1.225 kg/m³ y que el diámetro de las palas es de 80 metros. Este diámetro se ha seleccionado basándose en la potencia nominal, la velocidad de giro del rotor nominal y el hecho de que tiene un controlador de pitch. Comparando los

modelos en funcionamiento en España durante el periodo de estudio, se ha tomado el aerogenerador Gamesa G80/2000 como modelo, y el diámetro de palas de este es de 80 metros. Utilizando las formulas se construyen las curvas de coeficiente de potencia frente a velocidad de viento y frente a velocidad específica para cada año.

$$\text{Coef. Potencia} = C_e = \frac{P}{\frac{1}{2}\rho A v^3} \quad A = \pi R^2 \quad V. \text{ Específica} = \lambda = \frac{\Omega R}{v}$$

Ecuación 4. Coeficiente de potencia y velocidad específica

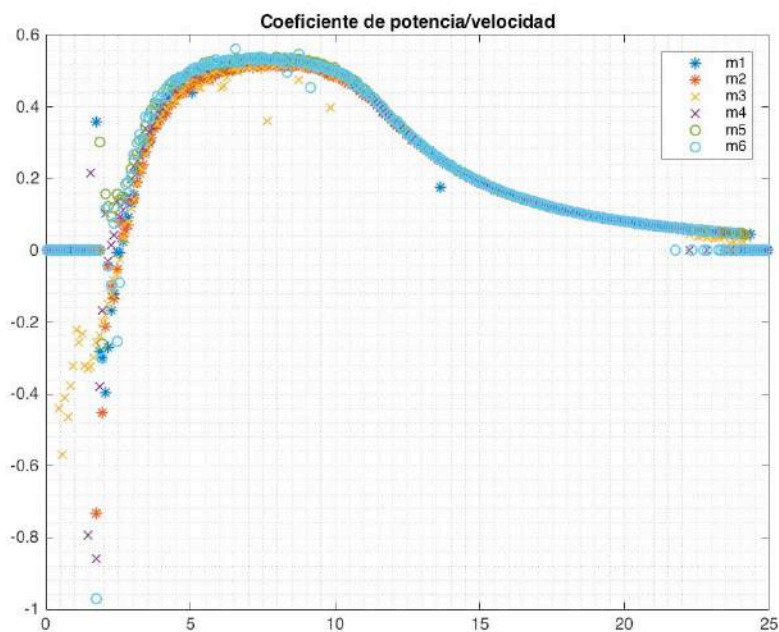


Figura 10. Coeficiente global de potencia frente a la velocidad del viento

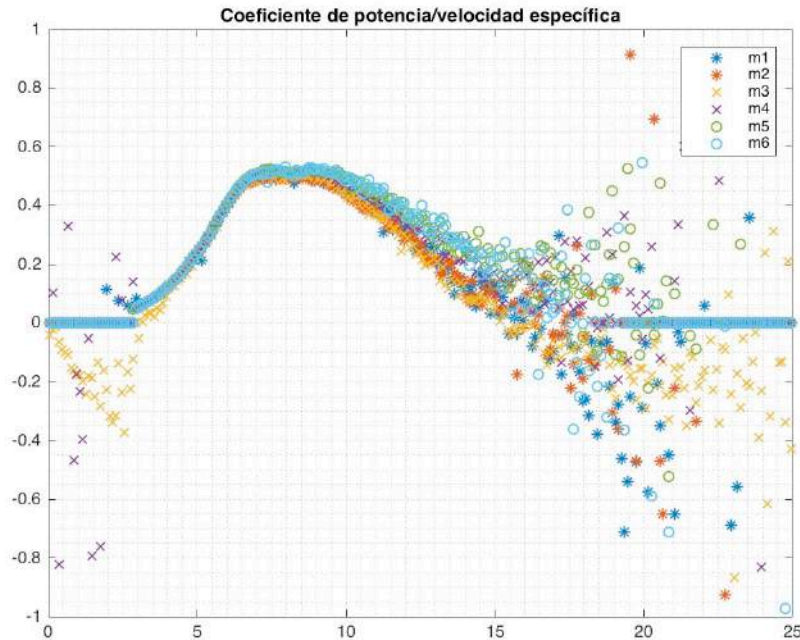


Figura 11. Coeficiente global de potencia frente a la velocidad específica

Se observa que ciertos valores superan el límite de Betz ($16/27=0.5926$). Esto es teóricamente imposible, luego en esos instantes algún componente asociado del SCADA no ha realizado correctamente la medición. Dado que se trata de contados casos aislados, no requiere un mayor estudio.

A continuación, se atiende a los datos de alarmas y reparaciones. Se observa que durante Y1 no sucede ninguna alarma, de la misma manera no se produce ninguna reparación durante Y6. Es oportuno señalar que en un principio se disponía de fechas de reparación anteriores al primer instante de datos del que se dispone, de la misma manera había valores de fecha de alarmas posteriores al último instante medido. Para ambos casos se desechan los casos que empezaron y terminaron antes o después del periodo de estudio. Finalmente quedan 153 reparaciones y 1407 alarmas.

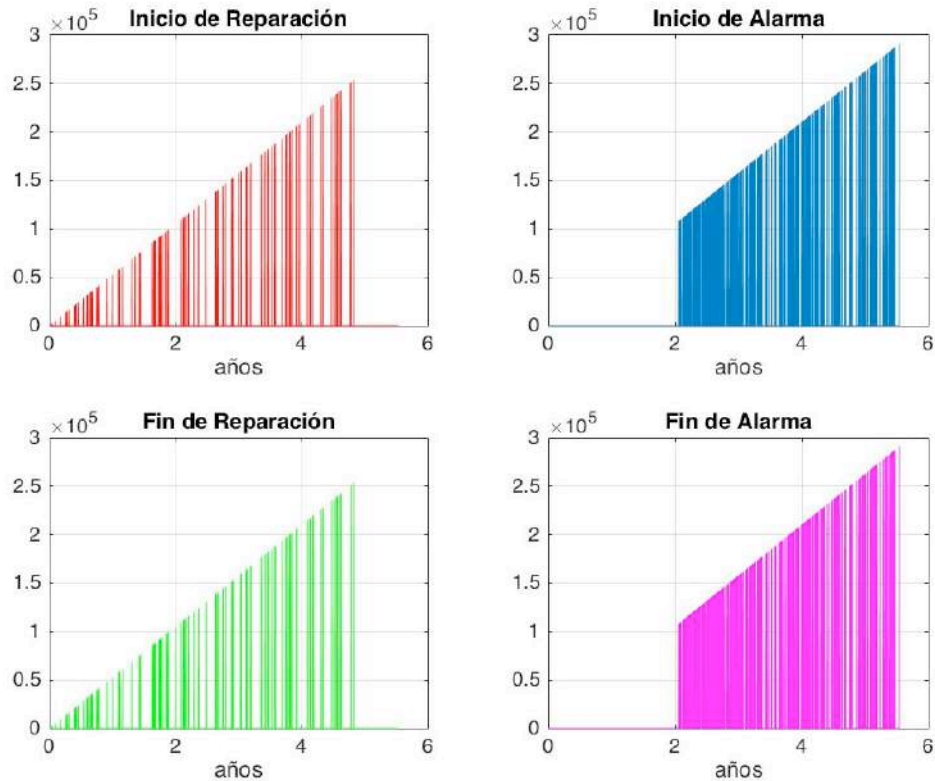


Figura 12. Fecha de inicio y de fin de reparaciones y alarmas

La Figura 13 muestra la diferencia de duración entre alarmas y reparaciones, donde las primeras tienden a durar notablemente menos que las segundas. Los valores del eje y de la Figura 13 representan:

1. Duración entre 0 minutos y 10 minutos.
2. Duración entre 10 minutos y 1 hora.
3. Duración entre 1 hora y 2 horas.
4. Duración entre 2 horas y 1 día.
5. Duración superior a 1 día.

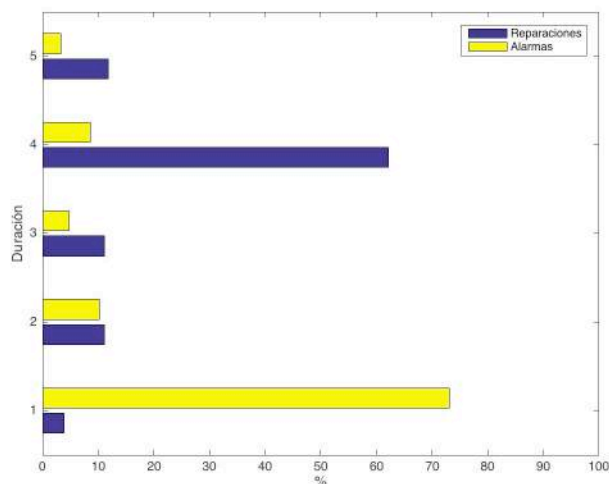


Figura 13. Duración desde el inicio hasta el fin de alarmas y reparaciones

Cada fecha de evento viene acompañada de etiquetas descriptivas del mismo. En el caso de las alarmas hay 2 etiquetas, la primera etiqueta nombra el grupo o zona en que sucede dicho evento y la segunda señala el evento en cuestión. Mientras tanto, se dispone de 4 etiquetas por evento. La primera señala por qué se realiza la reparación, la segunda señala el tipo de reparación que se va a realizar, y la tercera y la cuarta aportan información del componente a reparar. La Tabla 5 recoge información de las primeras etiquetas en ambos casos y el número total de casos de cada categoría.

REPARACIONES			ALARMAS		
Correc	Sustitución o corrección de componente	57	A05	Control general aerogenerador	866
Insp	Inspección	13	Amb	Condiciones ambientales	11
MDF		15	Gen	Generador	8
MDP	Maintenace Data Procesing	6	Giro	Sistema yaw	151
MPE		2	Gon	Góndola	1
Prev	Mantenimiento preventivo	30	Gro	Sensores inferiores	96
Rearme_loc	Reset sistema	30	Hidr	Sistema hidráulico	42
			Mul	Caja Multiplicadora	9
			Rotor	Rotor y control de pitch	53

		Top	Sensores externos góndola	167
		Varios	Resetear sistema	4

Tabla 5. Principales grupos de alarmas y reparaciones

En el caso de las alarmas destaca que más del 60% de ellas corresponden a A05, y dentro de esta categoría, “Fallo Comunicaciones Aerogenerador”, ”Orden de marcha” y “Orden de Pausa” son las más comunes.

Las reparaciones cobran una mayor importancia en este proyecto, pues un buen sistema de detección de anomalías debería predecir con éxito cuando suceden las reparaciones correspondientes a un funcionamiento defectuoso. Por ello el apartado 5.4 se va a centrar en las reparaciones de tipo “Correc”.

2.3 SELECCIÓN DE VARIABLES

En este apartado se seleccionan las variables que se han elegido para modelar el comportamiento global del aerogenerador, y particular de ciertos componentes de vital importancia para el sistema.

En primer lugar, es conveniente hacer una clasificación del tipo de variables que se dispone.

- **Variables de entrada del aerogenerador:** Condiciones de entrada del aerogenerador tales como las medidas de viento, temperatura, ... Son factores que influyen en la velocidad de las palas, y por tanto en el comportamiento de todos los componentes internos del molino.
- **Variables internas del aerogenerador:** Medidas del comportamiento de componentes del aerogenerador en las condiciones de entrada correspondientes. Aquí es donde se va poner más atención en el desarrollo del estudio de predicción detección, no tanto en relaciones directas como puede ser la velocidad de giro del roto o el ángulo del pitch, sino en las relaciones inesperadas que se espera obtener

de los otros componentes de la góndola

- Variables de salida: son aquellas que indican diferentes aspectos potencia o la energía generadas.

En segundo lugar, hay que tener en cuenta de que no se dispone de todos los valores de todas las variables. Conviene que haya un alto porcentaje de datos en el periodo de estudio para maximizar el rendimiento de los modelos de predicción.

Tomando estas aproximaciones pasamos a disponer de 3 variables de entrada, 21 variables internas y 8 variables de salida. En el apéndice I se encuentra la Tabla 38 donde se expone la información de cada variable como su descripción o su rango de valores.

Previo a seleccionar las variables para los modelos de comportamiento, hay que hacer hincapié en que no se dispone de todas las medidas de todas las variables para el periodo de estudio. Cuando un aparato no ha podido realizar una medida sea cual sea el motivo, se ha registrado como NaN, siglas de Not a Number. Estos datos no aportan ninguna información de la variable a la que pertenecen, por lo tanto, si una variable susceptible de ser elegida para crear un modelo de comportamiento tiene un alto porcentaje de valores NaN entre sus mediciones, no es conveniente seccionarla. Teniendo esto en mente se ha hecho la Tabla 6, que relaciona todas las variables del aerogenerador con el porcentaje de valores que no sean NaN que tienen, con el tipo de variables al que pertenecen, y en el caso de las variables intermedias, se expone a que componente del aerogenerador pertenecen. Conviene recordar que cada variable parte con 291312 mediciones.

VARIABLES IMPORTADAS AL WORKSPACE				
Nombre	Tipo de variable	Zona	Valores disponibles	
var1	Interna	Multiplicadora	282135	96,85%
var2	Interna	Multiplicadora	282076	96,83%
var3	Interna	Multiplicadora	282043	96,82%
var4	Interna	Generador	282287	96,90%
var5	Interna	Generador	282090	96,83%

var6	Interna	Generador-Transformador	97756	33,56%
var7	Interna	Generador-Transformador	66618	22,87%
var8	Interna	Generador-Transformador	66618	22,87%
var9	Interna	Generador-Transformador	66617	22,87%
var10	Interna	Grupo Hidráulico	282206	96,87%
var11	Interna	Grupo Hidráulico	281551	96,65%
var12	Interna	YAW	254364	87,32%
var13	Interna	Góndola	282173	96,86%
var14	Entrada		282050	96,82%
var15	Interna	Transformador	97699	33,54%
var16	Interna	Transformador	66621	22,87%
var17	Interna	Transformador	66616	22,87%
var18	Interna	Transformador	66618	22,87%
var19	Interna	Rotor	281405	96,60%
var20	Interna	Eje Generador	254628	87,41%
var21	Interna	Control de pitch	281375	96,59%
var22	Entrada		282654	97,03%
var23	Entrada		282447	96,96%
var24	Interna	Sistema de orientación (YAW)	66624	22,87%
var25	Salida		281124	96,50%
var26	Salida		281437	96,61%
var27	Salida		254708	87,43%
var28	Salida		281446	96,61%
var29	Salida		281258	96,55%
var30	Salida		282683	97,04%
var31	Salida		282111	96,84%
var32	Salida		291312	100,00%

Tabla 6. Variables del aerogenerador y porcentaje de datos disponibles

Hay 3 variables de entrada: temperatura ambiente, velocidad del viento y ángulo del viento (var14, var22 y var23 respectivamente), todas ellas con un porcentaje de NaN en torno al 3%. Estas variables se utilizarán en los modelos no lineales (MNL) como predictores, y la velocidad el viento se utilizará como variable de abscisas para todas las curvas de estado

porque es la relación con la que mejor se diferencian los periodos del aerogenerador: parado, arranque y nominal.

Tal y como se ha expuesto en el apartado 1.3. Se va a crear un modelo de comportamiento de la potencia (var30). Se trata de la relación más estudiada del aerogenerador y para diagnosticar el comportamiento general del aerogenerador, pues cada fallo de un componente repercute en la producción total. Además, se dispone de más del 97% de los datos durante el periodo de referencia.

Respecto a las variables intermedias, estableciendo el 90% como la cantidad mínima de datos necesaria para modelar, quedan 10 variables intermedias para modelar a razón de 3 para la caja multiplicadora, 2 para el generador, 2 para el grupo hidráulico, 1 para la góndola, 1 para el rotor, y 1 para el controlador de pitch.

1. Temperatura del aceite de la multiplicadora (var1): esta variable muestra el comportamiento de la caja de manera aislada.
2. Temperatura de los rodamientos de la multiplicadora en el lado de generación (var2): esta variable, siendo parte del componente de la multiplicadora, permite relacionar este componente con el generador eléctrico.
3. Temperatura de los rodamientos DE en el lado de generación (var3): los rodamientos DE se encuentran en el eje de generación entre el generador eléctrico de la góndola y la multiplicadora. Junto con la var4, permite observar la diferencia de comportamiento entre la multiplicadora y el generador eléctrico.
4. Temperatura de los rodamientos NDE en el lado de generación (var4): estos rodamientos se encuentran en el lado opuesto a los rodamientos DE del generador eléctrico y muestran de manera aislada el comportamiento del mismo.
5. Temperatura de los anillos SLD en el lado de generación (var5): análogamente a la var1, permite observar de manera aislada el comportamiento del generador eléctrico.
6. Temperatura máxima del grupo electrógeno (var10): esta variable muestra de manera aislada el funcionamiento del grupo eléctrico.

7. Presión máxima del grupo electrógeno (var11): Similar a var10, pero midiendo la presión del grupo electrógeno. El reto que supone esta variable es que como se muestra en la Figura 14, la inmensa mayoría de sus valores son casi constantes durante todos los años, por los que será difícil para un modelo no lineal (MNL) predecir con gran precisión los valores de esta variable en el periodo de simulación.
8. Temperatura interior de la góndola (var13): esta variable está relacionada con todos los componentes que se encuentran dentro de la góndola, luego la información que pueda aportar para la detección de anomalías será amplia pero no muy concreta.
9. Giro del rotor (var19): tiene una relación directa con la potencia generada, e informa del estado del buje.
10. Controlador de pitch (var21): Existen muchos trabajos previos que analizan el comportamiento del pitch. Nuevamente, este componente tiene una relación directa con la potencia que se genera, y suele tratarse del componente que más fallos da en la mayoría de instalaciones.

En caso de que hubiese datos suficientes de variables relacionadas con otros componentes como el transformador o el YAW hubiese sido interesante estudiarlas.

Se decide no abordar el modelo de comportamiento de la var13 porque es consecuencia directa de las temperaturas de los componentes de góndola y la temperatura exterior. Aporta una información general que no resulta de tanta utilidad como la que pueden proporcionar el resto de variables.

Se decide no abordar el modelo de comportamiento de la var2 porque se trata de una medición intermedia entre la var1 y la var3. Con 2 modelos de comportamiento se cree que es suficiente para modelar el comportamiento de la multiplicadora.

En resumen, se han seleccionado 9 variables para crear 9 curvas de estado, y 3 variables de entrada para entrenar dichos modelos. En la Figura 14 se puede ver la nube de puntos de todas ellas a lo largo del periodo de estudio.

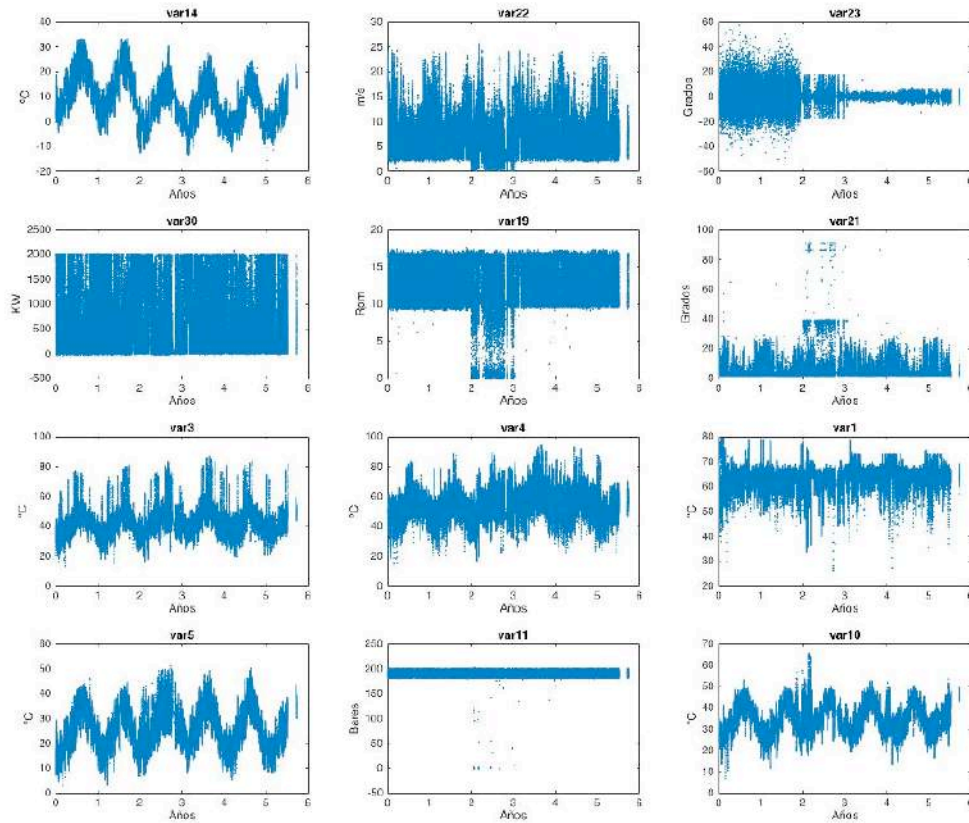


Figura 14. Variables utilizadas para la construcción de los modelos

Capítulo 3. PRETRATAMIENTO

En este capítulo se describen los pasos realizados para poder trabajar con los datos originales.

En primer lugar, se observa que hay una gran parte de tiempo en el que el aerogenerador se encuentra fuera de servicio (cuando $var32 = 0$). Las medidas registradas en esos momentos no son de interés del estudio, y enturbian la visión del resto, por consiguiente, se decide eliminarlos. Se crea una función que, para los instantes temporales en los que el aerogenerador esté fuera de servicio ($var32=0$), sustituya el registro original de todas las variables por 0, de esta manera al graficar las curvas de estudio, todos los puntos de fuera de servicio se superpondrán en el origen de coordenadas y no dificultarán el análisis.

Una vez seleccionadas las variables que se van a utilizar para crear las curvas de estado del aerogenerador, se procede a seleccionar el rango de datos que se van a utilizar. No todos los datos son representativos del funcionamiento del sistema. Mediante la $var32$ se puede discernir cuando el aerogenerador está en funcionamiento y cuando no lo está. Generalmente un aerogenerador se encuentra en no funcionamiento cuando la velocidad del viento es demasiado baja como para que los álabes comiencen a girar o demasiado alta como para que tenga que saltar el freno interno entre la multiplicadora y el generador de baja tensión de la góndola. Cuando no está en funcionamiento el sistema SCADA no para de realizar mediciones, luego hay una gran cantidad de valores que no son útiles para el desarrollo del proyecto.

Se ha observado que el porcentaje de tiempo en el que la máquina esta fuera de servicio es de media de un 25%, siendo en Y6 cuando porcentualmente más tiempo está en funcionamiento (79%) y en Y5 cuando menos (70%). Se eliminan las mediciones para aerogenerador fuera de servicio pasando de esta manera de vectores de 291312 datos a vectores de 216949 datos. Se crea un vector con las fechas correspondientes a cada uno de estos valores.

FECHAS DE LOS DATOS			
AÑO	Antes de pretratamiento	Después de pretratamiento	Fecha real
Y1 Inicio	1	1	15/12/08 0:00
Y1 Fin	52560	38641	15/12/09 0:00
Y2 Inicio	52561	38642	15/12/09 0:10
Y2 Fin	105120	77799	15/12/10 0:00
Y3 Inicio	105121	77800	15/12/10 0:10
Y3 Fin	157680	117815	15/12/11 0:00
Y4 Inicio	157681	117816	15/12/11 0:10
Y4 Fin	210240	157823	14/12/12 0:00
Y5 Inicio	201241	157824	14/12/12 0:10
Y5 Fin	262800	194442	14/12/13 0:00
Y6 Inicio	262801	194443	14/12/13 0:10
Y6 Fin	*291312 (301391)	216949	7/9/14 23:50

Tabla 7. Fechas antes y después de eliminar valores de aerogenerador fuera de servicio

De esta manera se crea una nueva variable llamada `fecha_on`, que recoge las fechas antes del pretratamiento en la longitud de vector de después del pretratamiento. En otras palabras, se pasa de un vector fecha de 291312 valores (1,2,3...291312) a otro llamado `fecha_on` de 216949 con las fechas del anterior para los puntos que tienen en común, que son aquellos en los que el aerogenerador está en funcionamiento.

En segundo lugar, se han de tratar los valores NaN (not a number). Este es el valor que Matlab le da a los elementos vacíos de un vector, en este caso se trata de los instantes en los que no se realizó ninguna medición. En el apartado 2.3 se ha mostrado como en base al porcentaje de valores NaN, entre otros factores, se han elegido las variables a estudiar. Para las variables de estudio el porcentaje de valores NaN no supera el 4%, luego se ha decidido sustituir el valor NaN por la media aritmética de los valores reales inmediatamente anterior y posterior. En el caso del primer valor de cada variable, todos los valores NaN, se sustituye por 0.

En tercer lugar, se procede a corregir *var22*. La primera corrección consiste en acotar el rango de valores entre 0 y 25 m/s. Esto se debe a que hay 11 valores (0,0051%) comprendidos entre 25 y 30 m/s, sin embargo, claramente pertenecen a fallos de operación, pues a esa velocidad se debería activar el freno del eje de generación y parar el giro del rotor y en consecuencia la potencia generada sería igual a 0. Para mantener constante el tamaño de los vectores de estudio, se sustituyen los valores mayores de 25m/s por 25m/s. Para segunda corrección se ha observado que entre los valores 78000 y 120000 (del 09/06/10 16:00 al 28/03/11 7:40) hay un error constante. Para solucionar este inconveniente, se añade en este periodo un valor constante de 1.7 m/s. Este valor se ha calculado como la media de diferencias entre mínimos y máximos locales de este periodo con el resto de valores.

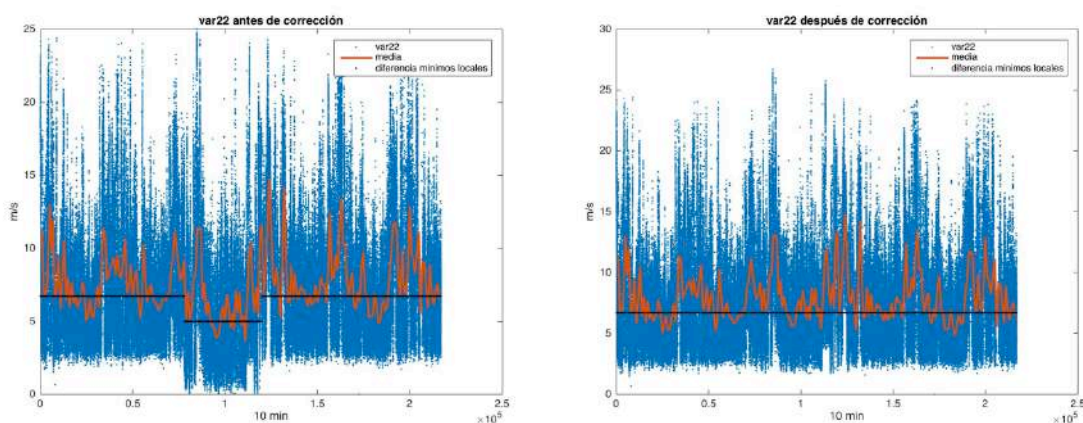


Figura 15. Corrección del offset de la velocidad del viento desde 09/06/10 a 28/3/11

El orden utilizado para numerar las curvas se ha tomado de manera aleatoria. La tabla recoge los datos correspondientes a cada curva de estado durante el periodo de referencia.

Curvas de estado					
	Variable x	Variable y	Udm	Rango (1% – 99%)	Zona
C1	var22	var30	kW	-9,6 – 2001,9	General
C2	var22	var19	Rpm	0,5 – 16,8	Rotor

C3	var22	var21	Grados	1,5 – 39,1	Pitch
C4	var22	var30	°C	26,9 – 65,6	Multiplicadora
C5	var22	var4	°C	33,8 – 77,5	Genereador
C6	var22	var1	°C	55,2 – 73,0	Multiplicadora
C7	var22	var5	°C	10,1 – 42,9	Genereador
C8	var22	var11	Bares	181,0 – 199,3	G. Hidráulico
C9	var22	var10	°C	20,2 – 47,9	G. Hidráulico

Tabla 8. Datos de las curvas de estado en el periodo de referencia

Finalmente se muestra como queda cada curva en cada año tras en pretratamiento. Se observa como durante el Y3 todas las curvas presentan un comportamiento notablemente diferente, especialmente en las curvas C1, C2 y C3.

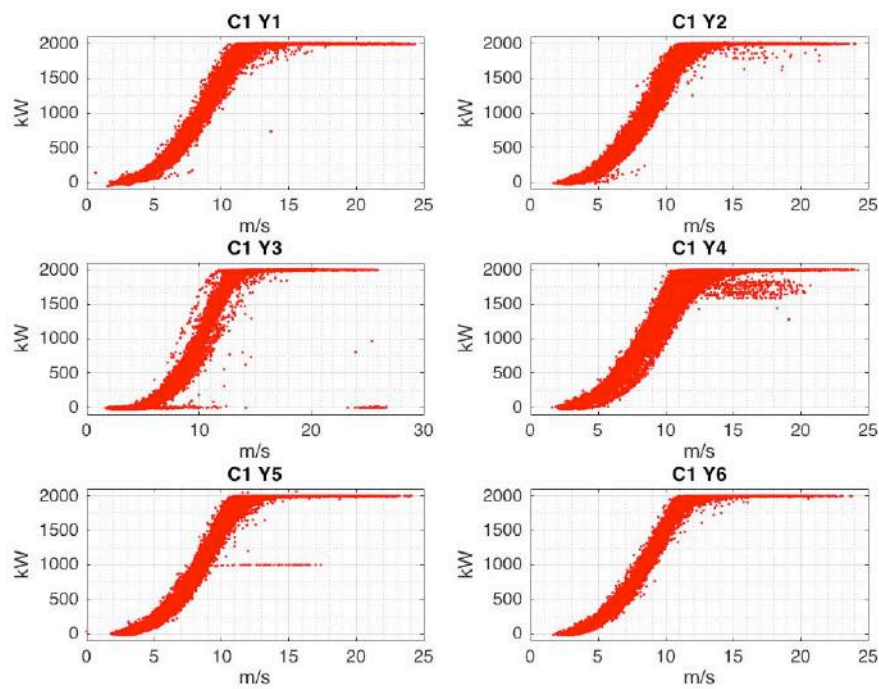


Figura 16. Nube de puntos de C1 en cada año del periodo de estudio

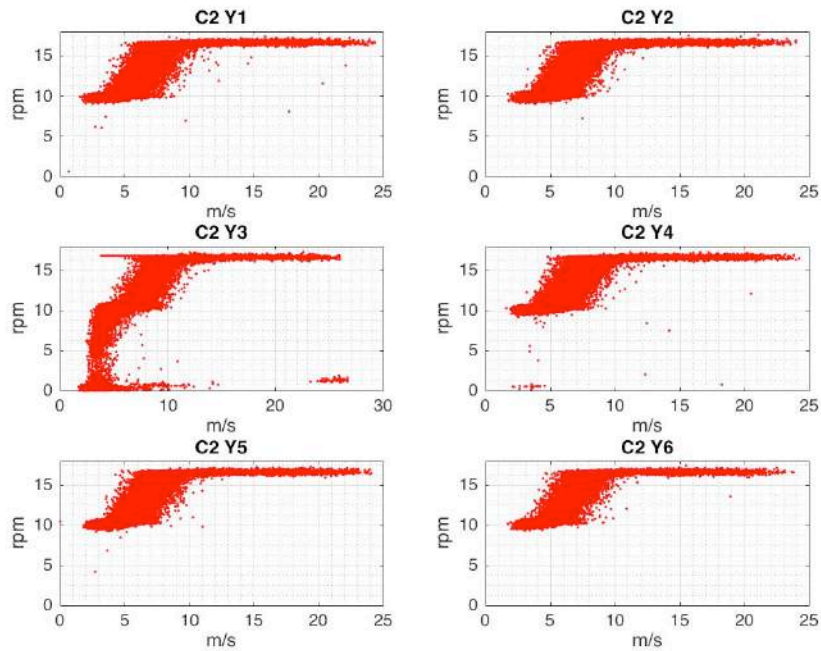


Figura 17. Nube de puntos de C2 en cada año del periodo de estudio

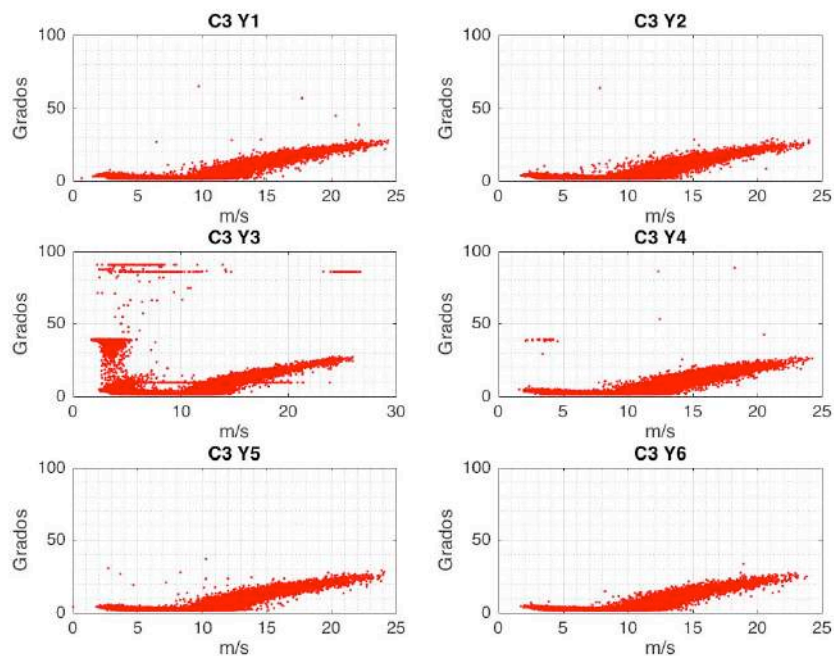


Figura 18. Nube de puntos de C3 en cada año del periodo de estudio

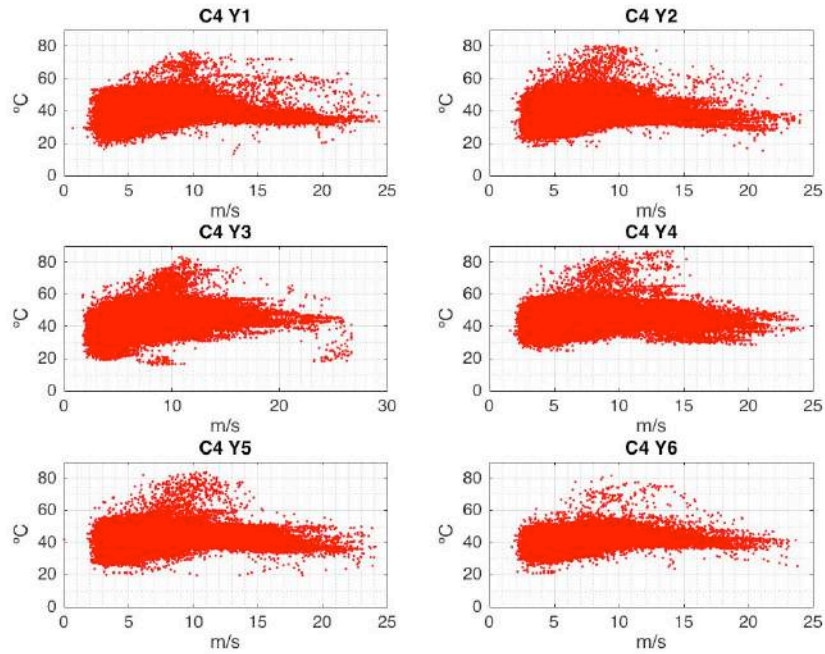


Figura 19. Nube de puntos de C4 en cada año del periodo de estudio

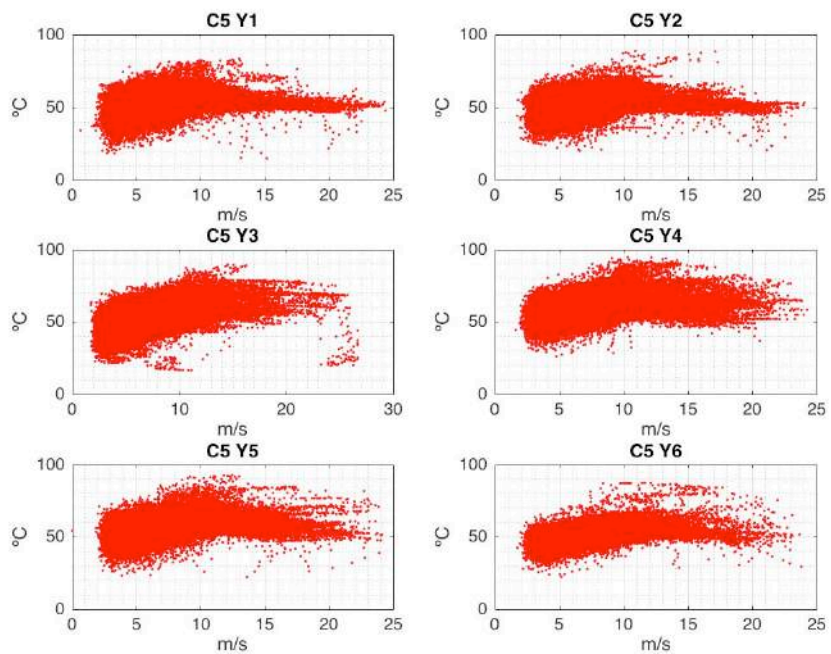


Figura 20. Nube de puntos de C5 en cada año del periodo de estudio

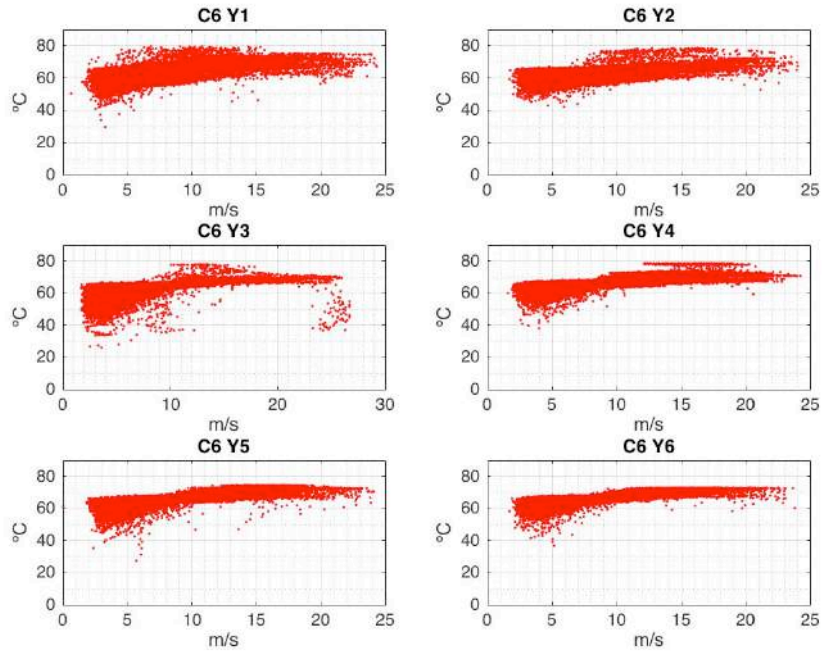


Figura 21. Nube de puntos de C6 en cada año del periodo de estudio

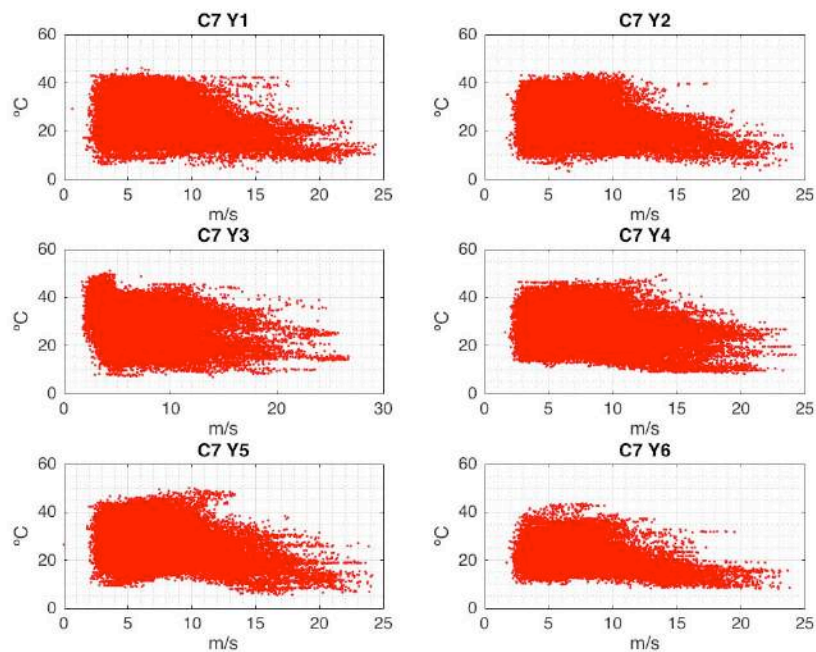


Figura 22. Nube de puntos de C7 en cada año del periodo de estudio

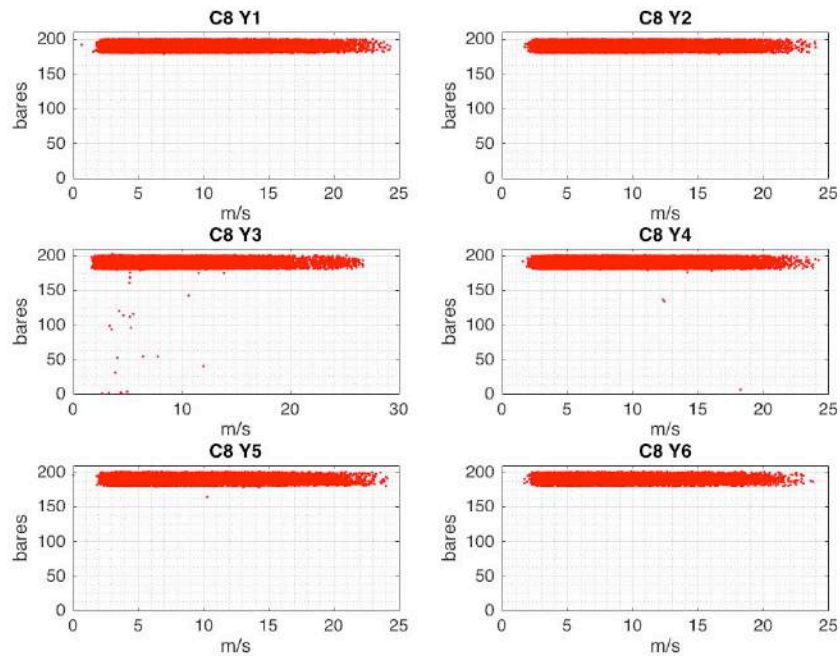


Figura 23. Nube de puntos de C8 en cada año del periodo de estudio

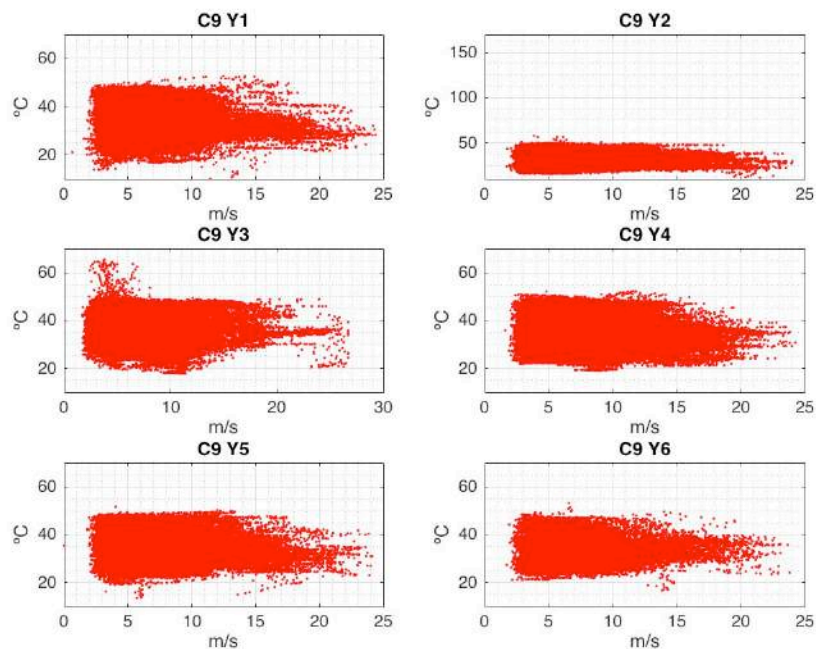


Figura 24. Nube de puntos de C9 en cada año del periodo de estudio

Capítulo 4. CREACIÓN DE LOS MODELOS DE COMPORTAMIENTO

En este apartado se explican los pasos que se han seguido para construir los modelos de comportamiento. En los apartados 4.1 y 4.2 se exponen las consideraciones que se han tomado para crear cada tipo de modelos, sin embargo, hay algunos parámetros que son similares para ambos. A saber:

- Se ha utilizado el $Y1$ como periodo de referencia para crear los modelos lineales y entrenar los modelos no lineales, y $Y2$, $Y3$, $Y4$, $Y5$ y $Y6$ como periodo de simulación.
- En todos los casos todas las variables utilizadas constan de 216949 valores.

4.1 MODELOS LINEALES (ML)

Para cada modelo lineal se ha utilizado únicamente 2 variables, la variable correspondiente al eje de ordenadas de la curva a modelar y la variable del eje de abscisas, que es *var22* en todos los casos.

Los modelos lineales que aquí se construyen consisten en crear para cada curva una media y desviación típica móviles que permita crear unas bandas de confianza que contengan en torno a un 90-95% de los datos. Con ello se pretende dejar los datos más inesperados como errores para poder buscar anomalías. Para obtener estos valores estadísticos se tienen que dividir los valores de la variable de ordenadas intervalos en regulares de la variable de abscisas. De esta manera acotando el eje x , automáticamente se obtienen valores estadísticos del eje y .

En este tipo de modelos el hecho de que haya valores NaN no supone un problema para la programación, por lo que se han utilizado cuando ha sido necesario obviar ciertos datos. En la siguiente lista se explica el proceso realizado:

- Se toma el primer año de datos como periodo de referencia para crear los modelos. De esta manera, a partir de la información que tenga el Y1, se modela el comportamiento general del modelo y se puede predecir su funcionamiento en el periodo de simulación y obtener el error de funcionamiento y posteriormente las anomalías detectadas por ese modelo.
- Para mejorar la aproximación que crea el modelo lineal de comportamiento es conveniente no tener en cuenta los valores aislados que puedan alterar los valores estadísticos (media móvil y desviación móvil) con los que se crean las bandas de confianza. Se sustituyen los valores más alejados del comportamiento natural de la curva en cuestión por valores *NaN*. Este proceso se aplica manualmente y en ningún caso el número de datos alterados supera el 5% el total. A este proceso se e llamará limpieza preventiva.

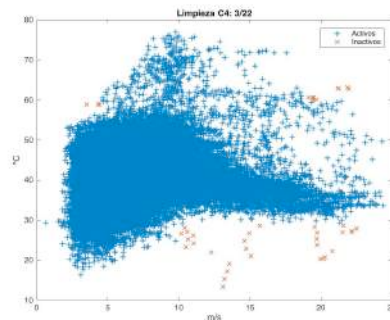


Figura 25. Limpieza de valores aislados en C4

- Se aplica una función que obtiene la media móvil y desviación móvil de la variable de estudio en el espectro de valores de velocidad del viento. Para las medidas móviles se establece un intervalo constante y común para los 9 modelos. Se ha probado a utilizar intervalos de amplitud desde 5m/s (5 valores), hasta 0.01 m/s (2500 valores), decantándose finalmente por intervalos de 0.1 m/s (250 valores). Este intervalo es lo suficientemente pequeño como para ajustar correctamente

durante la zona intermedia de las curvas de estado, y lo suficientemente amplio como para no ser susceptible a valores aislados en zonas con poca cantidad de datos. En la Figura 26 se muestran las diferentes bandas de confianza que se crean con diferentes anchos de intervalo.

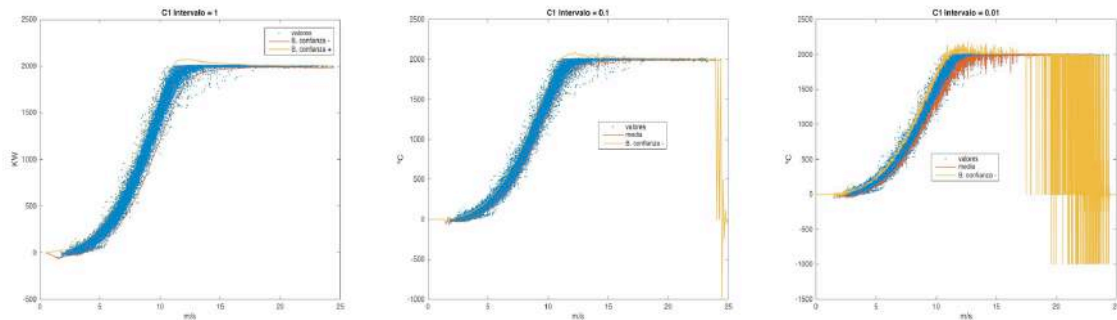


Figura 26. Bandas de confianza de ML1 para diferentes anchos de intervalo

- Se crean unas bandas de confianza de 2 desviaciones típicas de ancho a cada lado de la media móvil. Estas bandas de confianza en el análisis bursátil toman el nombre de bandas de Bollinger y son uno de los datos más estudiados en ese campo. Hay que tener en cuenta que los valores modificados en la limpieza preventiva se encontrarían fuera de las bandas de confianza, por lo tanto, el número total de valores fuera de las bandas de confianza es el que resulta tras aplicar las bandas de confianza más el número de valores que se han modificado en la limpieza preventiva.
- Se crea un programa que transforma los 250 valores de la media móvil y las bandas de confianza en 216949 valores de manera que corresponda el valor en cada instante al valor del intervalo correspondiente. De esta manera, para cada instante del periodo de simulación (Y2, Y3, Y4, Y5 y Y6) se puede obtener tanto el error entre la predicción y el valor real, como si se encuentra o no en el rango de valores de confianza.
- Para tener en cuenta el error del modelo, se extrae el error como la diferencia entre la predicción y el valor real. Se obtiene la media aritmética del error y se suma a ambos lados de las bandas de confianza. Este valor se le suma a las bandas de confianza y se vuelve a calcular el número de valores fuera de las bandas,

4.1.1 ML1 (POTENCIA TOTAL RESPECTO A LA VELOCIDAD DEL VIENTO)

- Variable de abscisas: var22
- Variable de ordenadas: var30
- Periodo de referencia: 15/12/08 00:00 → 15/12/09 (38641 valores)
- Número de valores modificados durante la limpieza preventiva en el periodo de referencia: 104 (0,27%)
- Número valores fuera de las bandas de confianza en el periodo de referencia: 1056 (2,73%)
- Error medio del modelo en el periodo de referencia: 2,269 kW (0,276%)

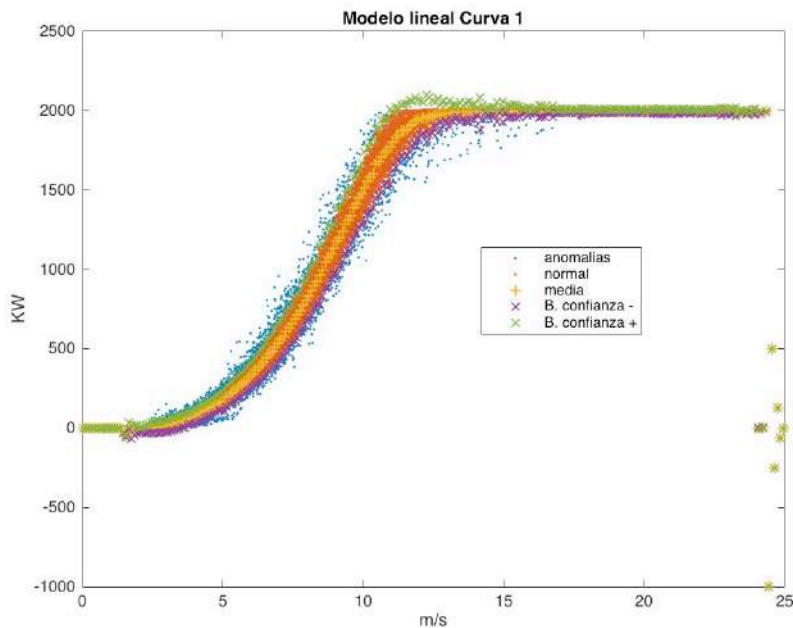


Figura 27. ML1 en periodo de referencia

4.1.2 ML2 (GIRO DEL ROTOR RESPECTO A LA VELOCIDAD DEL VIENTO)

- Variable de abscisas: var22
- Variable de ordenadas: var19
- Periodo de referencia: 15/12/08 00:00 → 15/12/09 (38641 valores)
- Número de valores modificados durante la limpieza preventiva en el periodo de referencia: 21 (0,05 %)
- Número valores fuera de las bandas de confianza en el periodo de referencia: 881 (2,28%)
- Error medio del modelo en el periodo de referencia: -0,009 rpm (0,066%)

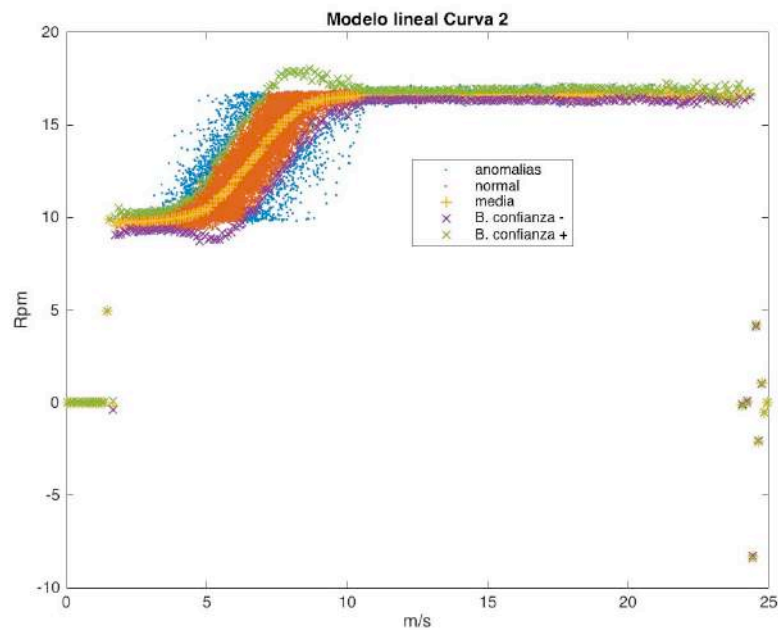


Figura 28. ML2 en periodo de referencia

4.1.3 ML3 (ÁNGULO DEL CONTROLADOR DE PITCH RESPECTO A LA VELOCIDAD DEL VIENTO)

- Variable de abscisas: var22
- Variable de ordenadas: var21
- Periodo de referencia: 15/12/08 00:00 → 15/12/09 (38641 valores)
- Número de valores modificados durante la limpieza preventiva en el periodo de referencia: 13 (0,04%)
- Número valores fuera de las bandas de confianza en el periodo de referencia: 459 (1,19%)
- Error medio del modelo en el periodo de referencia: 0,010 ° (0,245%)

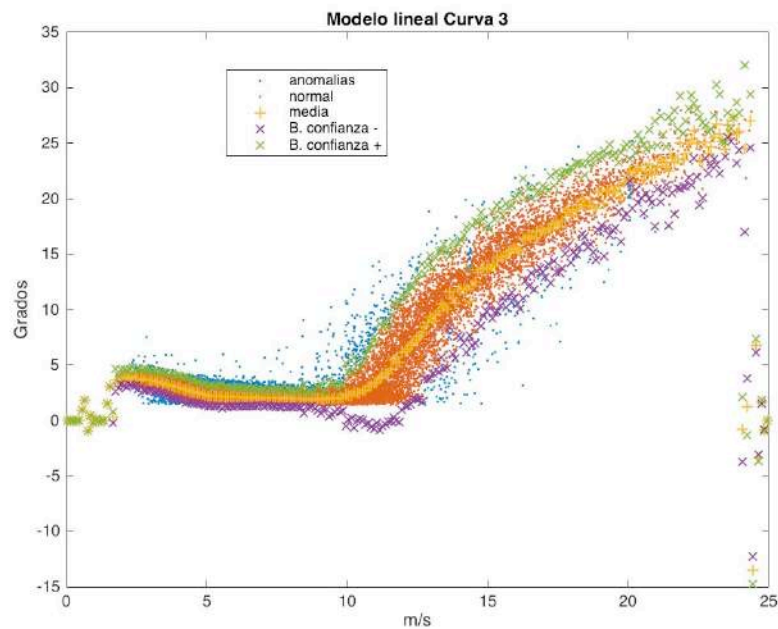


Figura 29. ML3 en periodo de referencia

4.1.4 ML4 (TEMPERATURA DE LOS RODAMIENTOS DE DEL GENERADOR RESPECTO A LA VELOCIDAD DEL VIENTO)

- Variable de abscisas: var22
- Variable de ordenadas: var3
- Periodo de referencia: 15/12/08 00:00 → 15/12/09 (38641 valores)
- Número de valores modificados durante la limpieza preventiva en el periodo de referencia: 42 (0,11%)
- Número valores fuera de las bandas de confianza en el periodo de referencia: 580 (1,50%)
- Error medio del modelo en el periodo de referencia: $-0,025$ °C (0,060%)

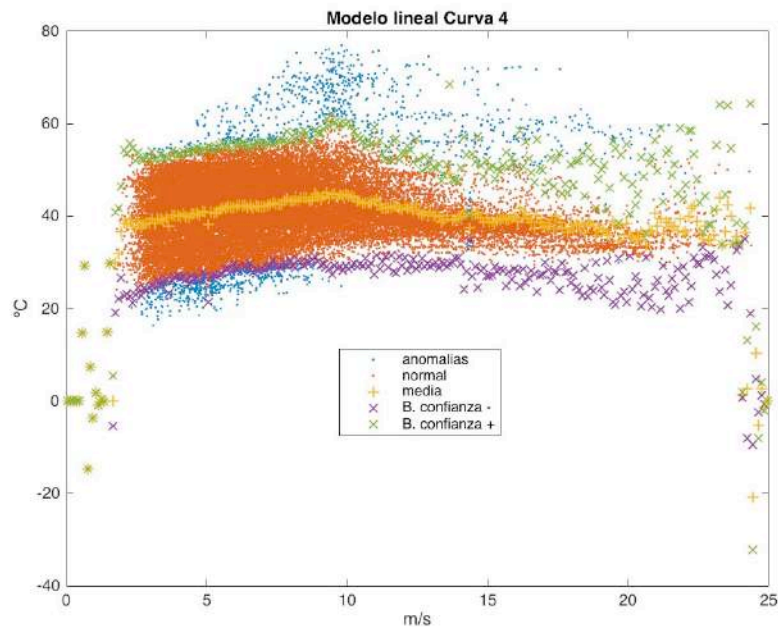


Figura 30. ML4 en periodo de referencia

4.1.5 ML5 (TEMPERATURA DE LOS RODAMIENTOS NDE RESPECTO A LA VELOCIDAD DEL VIENTO)

- Variable de abscisas: var22
- Variable de ordenadas: var4
- Periodo de referencia: 15/12/08 00:00 → 15/12/09 (38641 valores)
- Número de valores modificados durante la limpieza preventiva en el periodo de referencia: 20 (0,05%)
- Número valores fuera de las bandas de confianza en el periodo de referencia: 856 (2,22%)
- Error medio del modelo en el periodo de referencia: -0,035 °C (0,043%)

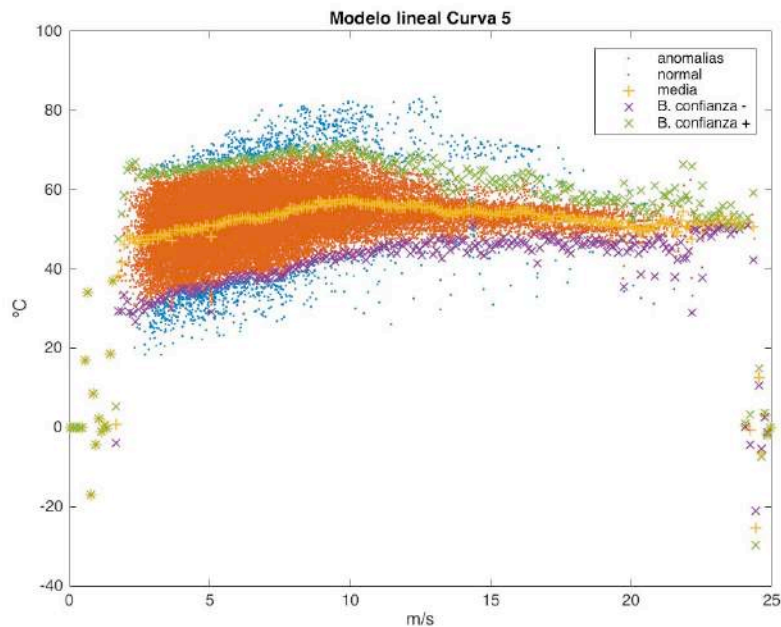


Figura 31. ML5 en periodo de referencia

4.1.6 ML6 (TEMPERATURA DEL ACEITE DE LA MULTIPLICADORA RESPECTO A LA VELOCIDAD DEL VIENTO)

- Variable de abscisas: var22
- Variable de ordenadas: var1
- Periodo de referencia: 15/12/08 00:00 → 15/12/09 (38641 valores)
- Número de valores modificados durante la limpieza preventiva en el periodo de referencia: 89 (0,23 %)
- Número valores fuera de las bandas de confianza en el periodo de referencia: 1418 (3,67 %)
- Error medio del modelo en el periodo de referencia: -0,015 °C (0,024 %)

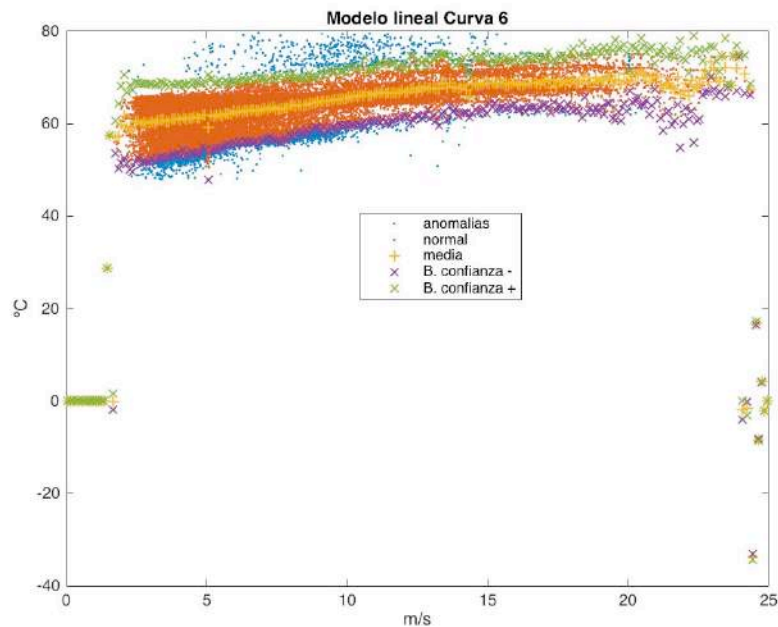


Figura 32. ML6 en periodo de referencia

4.1.7 ML7 (TEMPERATURA DE LOS ANILLOS SLD DEL GENERADOR RESPECTO A LA VELOCIDAD DEL VIENTO)

- Variable de abscisas: var22
- Variable de ordenadas: var5
- Periodo de referencia: 15/12/08 00:00 → 15/12/09 (38641 valores)
- Número de valores modificados durante la limpieza preventiva en el periodo de referencia: 74 (0,19 %)
- Número valores fuera de las bandas de confianza en el periodo de referencia: 486 (1,26 %)
- Error medio del modelo en el periodo de referencia: -0,056 °C (0,224 %)

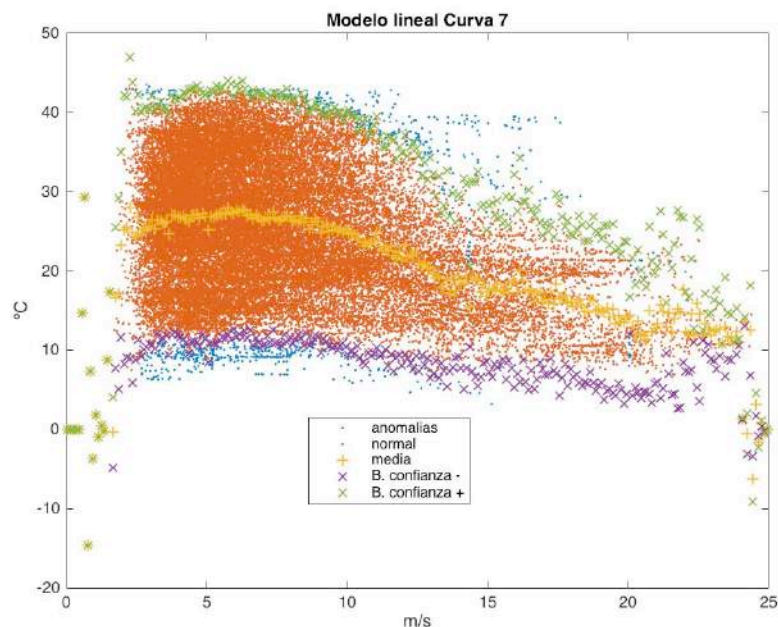


Figura 33. ML7 en periodo de referencia

4.1.8 ML8 (PRESIÓN DEL GRUPO HIDRÁULICO RESPECTO A LA VELOCIDAD DEL VIENTO)

- Variable de abscisas: var22
- Variable de ordenadas: var11
- Periodo de referencia: 15/12/08 00:00 → 15/12/09 (38641 valores)
- Número de valores modificados durante la limpieza preventiva en el periodo de referencia: 48 (0,12%)
- Número valores fuera de las bandas de confianza en el periodo de referencia: 206 (0,53 %)
- Error medio del modelo en el periodo de referencia: -0,147 bares (0,077%)

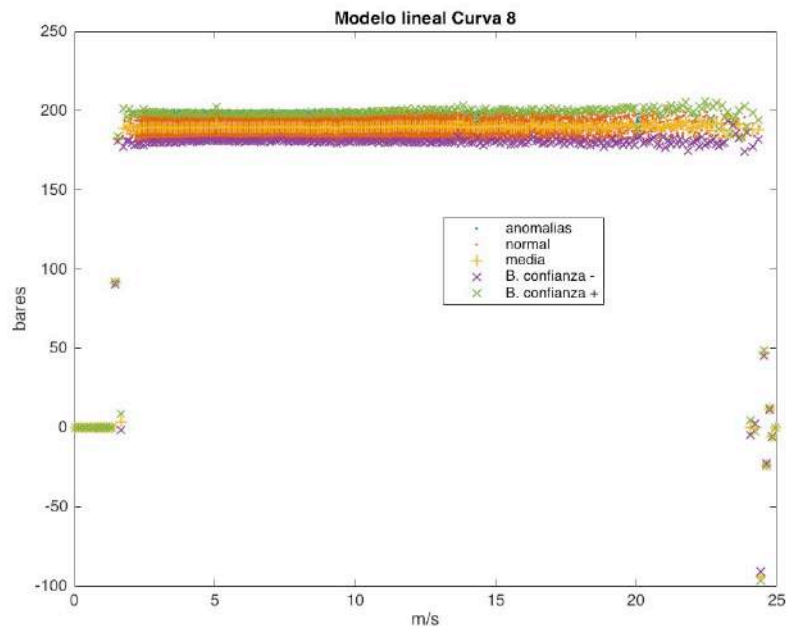


Figura 34. ML8 en periodo de referencia

4.1.9 ML9 (TEMPERATURA DEL GRUPO HIDRÁULICO RESPECTO A LA VELOCIDAD DEL VIENTO)

- Variable de abscisas: var22
- Variable de ordenadas: var10
- Periodo de referencia: 15/12/08 00:00 → 15/12/09 (38641 valores)
- Número de valores modificados durante la limpieza preventiva en el periodo de referencia: 50 (0,13 %)
- Número valores fuera de las bandas de confianza en el periodo de referencia: 615 (1,59 %)
- Error medio del modelo en el periodo de referencia: -0,030 °C (0,088%)

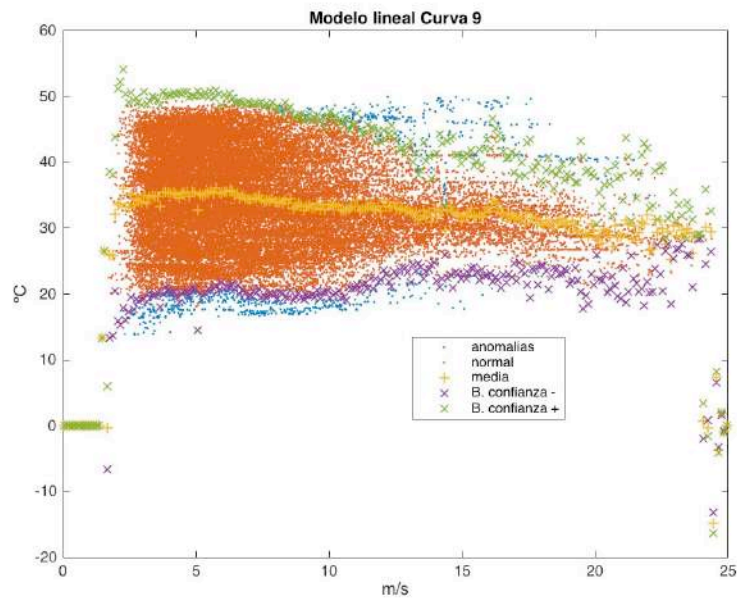


Figura 35. ML9 en periodo de referencia

4.2 MODELOS NO LINEALES (MNL)

Este tipo de modelo es más complejo que el primero, pero se espera que sus resultados sean mejores que los que aportan los modelos lineales anteriores. Se ha decidido utilizar redes neuronales para estos modelos, concretamente perceptrones multicapa (MLP). Este tipo de redes neuronales son relativamente sencillas de programar en Matlab, y como se ha explicado en el apartado 1.5 aportan resultados satisfactorios.

Para poder programar redes neuronales MLP es necesario tener el paquete de Matlab de matemáticas, estadística y optimización. Una forma fácil y rápida de crear estas redes es mediante la aplicación “Neural Net Fitting”. Dentro de la aplicación explica cómo crear diversos tipos de redes neuronales en función de la utilidad que se les quiera dar (Clustering, búsqueda de patrones ...). Para el caso que atañe se seleccionaría “nftool” y se seguirían las instrucciones de la aplicación, sin embargo, se ha optado por programar íntegramente cada MLP. Se ha tomado como base el código fuente de “nftool”.

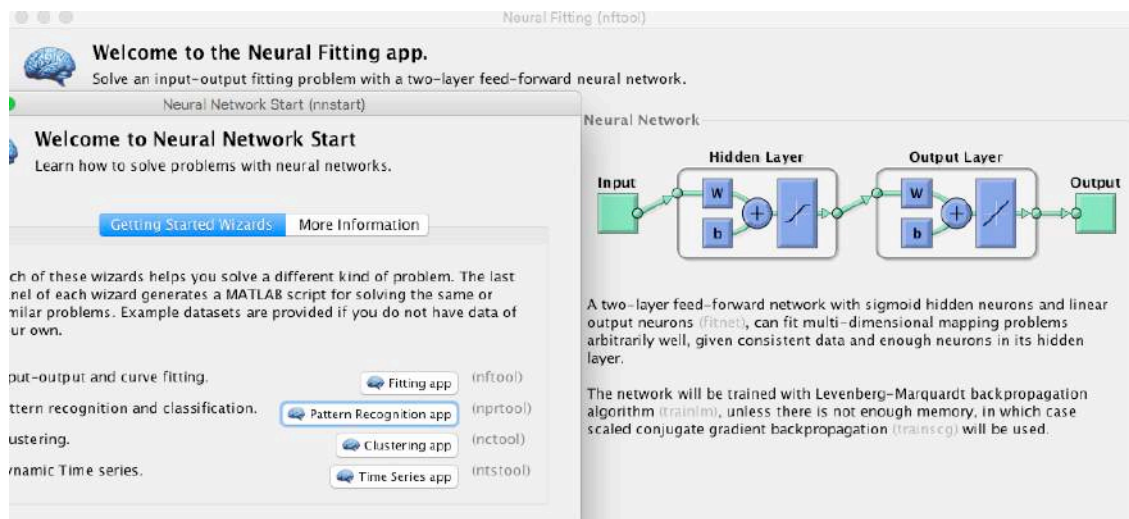


Figura 36. Cuadro de diálogo de Neural Net Fitting

Se utilizan como entrada de los modelos las variables de entrada var22, var14 y var23 (velocidad y ángulo del viento, y temperatura ambiente respectivamente) y var30 (potencia

generada). En el caso de C1, cuya variable objeto var30, las entradas son únicamente var22, var14 y var23, mientras que el resto de modelos utilizan estas 3 mismas y la predicción de la var30 que hace el MNL1 (Pred1).

A continuación, se lista el procedimiento seguido:

- Análogamente a como sucedía con los modelos lineales, se divide el rango de datos entre periodo de referencia y periodo de simulación. A través de la información de las variables utilizadas en cada modelo en el periodo de referencia, se crea el modelo que se simulará en el periodo de simulación y de él se extraerá el error y posteriormente las anomalías que detecte.
- Se crean 8 redes neuronales para cada modelo con diferentes arquitecturas. Esto se debe a que hay múltiples configuraciones posibles de MLP. Mediante prueba y error se obtuvo una configuración básica, y se varía el número de neuronas y el algoritmo de entrenamiento para dar 8 configuraciones diferentes. Se selecciona aquella que resulte tener un error menor cuadrático medio (MSE) y que consistentemente tarde menos de 10 segundos en modelarse, pues se ha observado que los modelos de larga duración dan resultados más aleatorios.
- Se obtienen los valores de la regresión lineal de cada modelo para cada año, tanto de referencia (Y1) como de simulación (Y2, Y3, Y4, Y5, y Y6) para comprobar su rendimiento.
- Se aplica el programa utilizado en los modelos lineales para obtener las bandas de confianza. Para ello se introducen la predicción del modelo (Pred) y la var22, y se obtienen la media y desviación móviles para 250 intervalos. Nuevamente se da 2 desviaciones de ancho a las bandas de confianza y se analiza el porcentaje de datos que quedan fuera de ellas
- Para tener en cuenta el error del modelo, se extrae el error como la diferencia entre la predicción y el valor real. Se obtiene la media aritmética del error y se suma a ambos lados de las bandas de confianza.

4.2.1 MNL1 (POTENCIA GENERADA RESPECTO A LA VELOCIDAD DEL VIENTO)

- Variables de entrada: var14, var22, var23
- Variable estimada: var30
- Estimación: Pred1
- Tipo de red: MLP
- Porcentajes de E/V/T (%): 70/15/15 (aleatorio)
- Número de neuronas en capa oculta: 12
- Algoritmo de entrenamiento: Regularización Bayesiana (BR)
- Tasa de aprendizaje: 0.002
- Error medio del modelo en el periodo de referencia: -0.278 kW (0.033%)

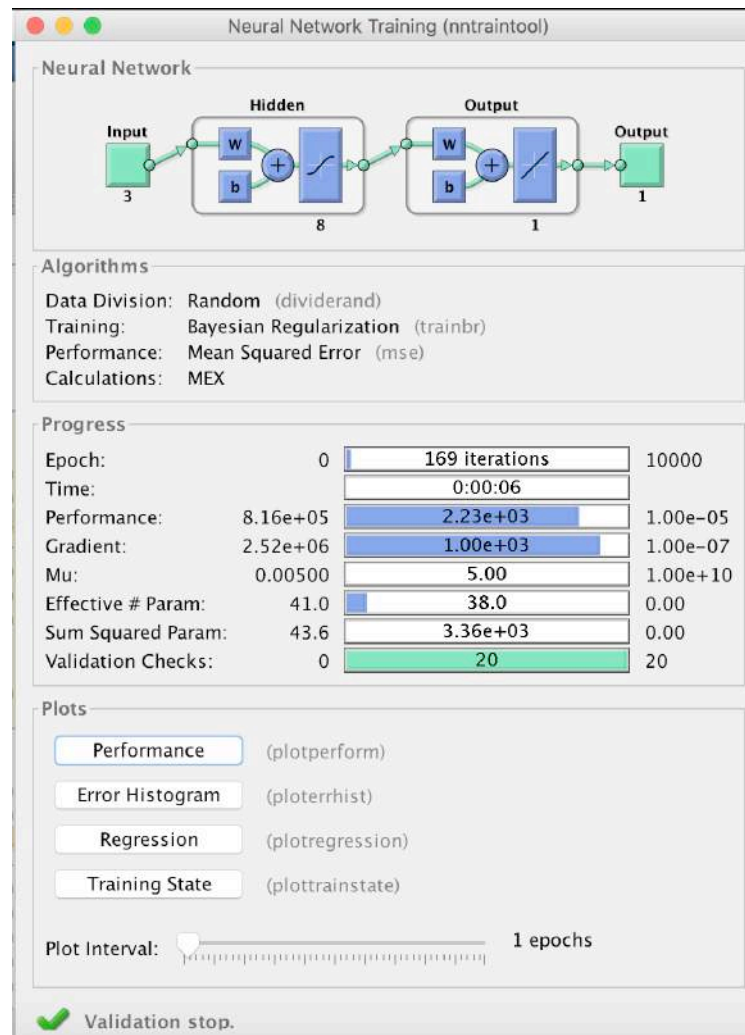


Figura 37. Cuadro de diálogo de la red neuronal de MNL1

Modelos no lineales de C1										
Algoritmo	LM	LM	LM	LM	LM	BR	BR	BR	BR	BR
N Neuronas	2	4	8	12	16	2	4	8	12	16
Regresión	0,994	0,994	0,997	0,997	0,996	0,993	0,995	0,997	0,997	0,998
MSE	6072	5822	3074	2850	3540	6700	4578	2532	3462	2091
N ciclos	51	375	236	212	121	77	3498	169	106	1217
Duración (seg)	1,3	8,6	8,5	9,9	7,3	1,8	84	6,6	5,4	75,8

Tabla 9. Resultados de MNL1 para diferentes configuraciones de MLP

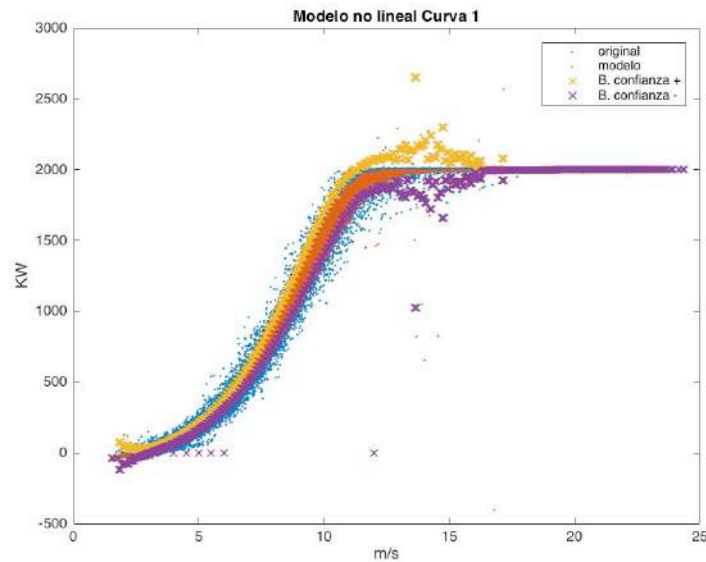


Figura 38. MNL1 en periodo de referencia

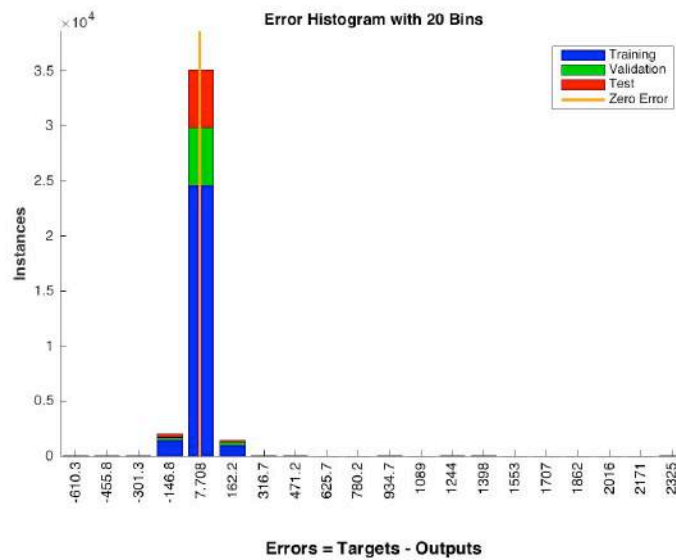


Figura 39. Histograma del error de MNL1 en el periodo de referencia

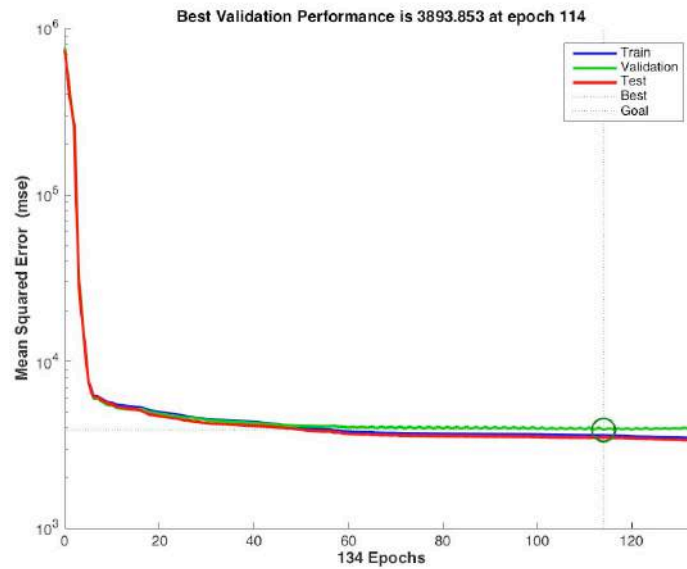


Figura 40. Evolución del MSE de MNL1 en el periodo de referencia

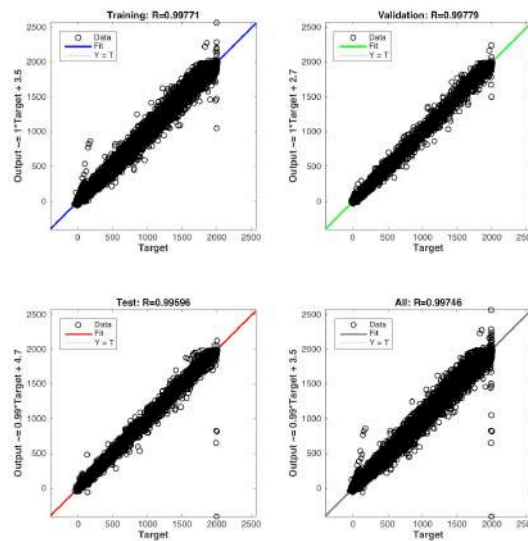


Figura 41. Regresión E/V/T y total de MNL1 en el periodo de referencia

4.2.2 MNL2 (GIRO DEL ROTOR RESPECTO A LA VELOCIDAD DEL VIENTO)

- Variables de entrada: var14, var22, var23, Pred1
- Variable estimada: var19
- Estimación: Pred2
- Tipo de red: MLP
- Porcentajes de E/V/T (%): 70/15/15 (aleatorio)
- Número de neuronas en capa oculta: 12
- Algoritmo de entrenamiento: Levenberg-Marquardt (LM)
- Tasa de aprendizaje: 0.002
- Error medio del modelo en el periodo de referencia: -0.005 rpm (0.036%)

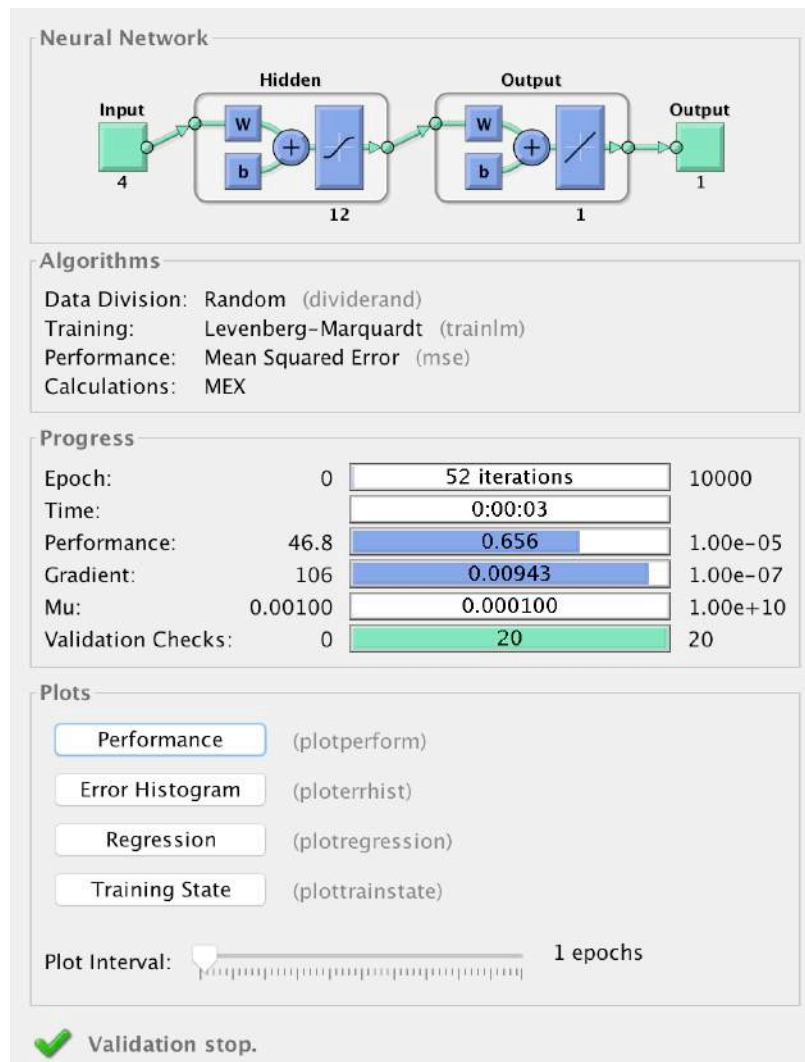


Figura 42. Cuadro de diálogo de la red neuronal de MNL2

Modelos no lineales de C2										
Algoritmo	LM	LM	LM	LM	LM	BR	BR	BR	BR	BR
N Neuronas	2	4	8	12	16	2	4	8	12	16
Regresión	0,959	0,959	0,96	0,96	0,96	0,959	0,959	0,96	0,96	0,96
MSE	0,669	0,666	0,657	0,655	0,655	0,669	0,661	0,659	0,657	0,657
N ciclos	44	53	66	52	98	89	41	60	58	44
Duración (seg)	1	1,3	2,5	3,5	6,5	2	1,1	2,3	3,3	3

Tabla 10. Resultados de MNL2 para diferentes configuraciones de MLP

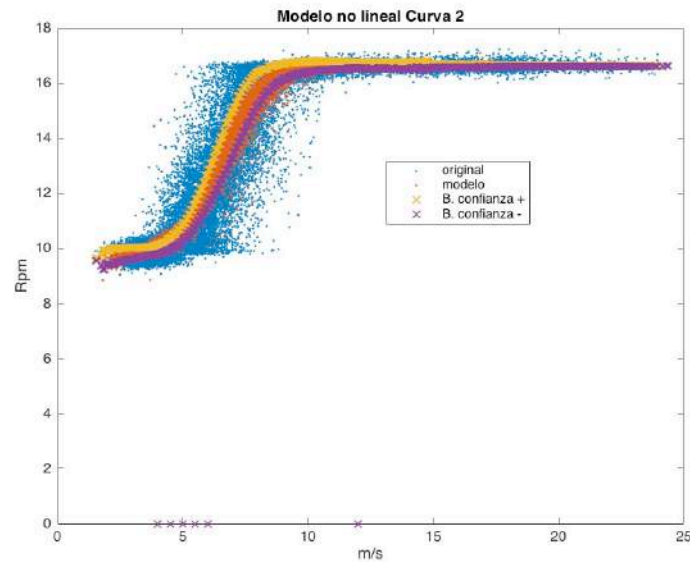


Figura 43. MNL2 en periodo de referencia

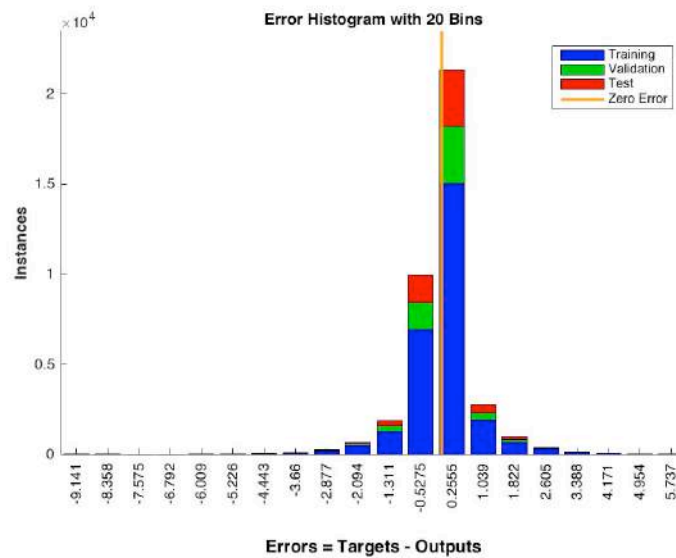


Figura 44. Histograma del error de MNL2 en el periodo de referencia

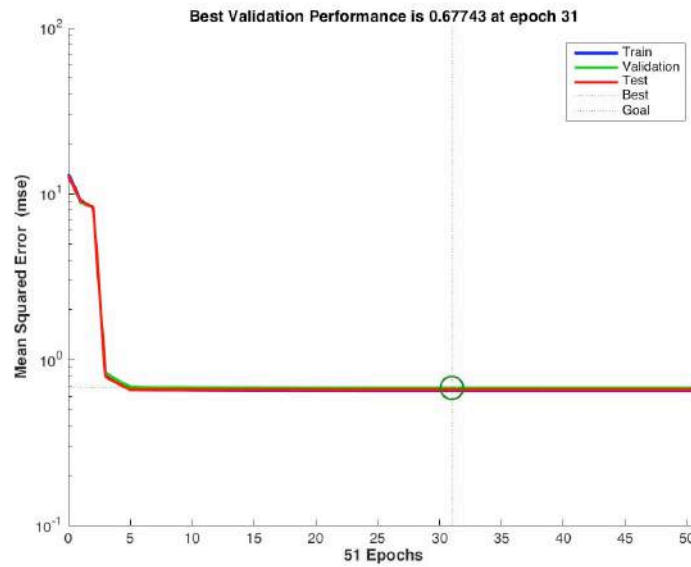


Figura 45. Evolución del MSE de MNL2 en el periodo de referencia

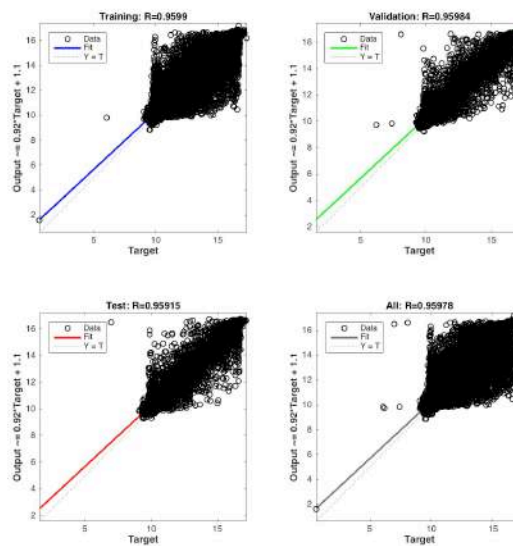


Figura 46. Regresión E/V/T y total de MNL2 en el periodo de referencia

4.2.3 MNL3 (ÁNGULO DEL CONTROLADOR DE PITCH RESPECTO A LA VELOCIDAD DEL VIENTO)

- Variables de entrada: var14, var22, var23, Pred1
- Variable estimada: var21
- Estimación: Pred3
- Tipo de red: MLP
- Porcentajes de E/V/T (%): 70/15/15 (aleatorio)
- Número de neuronas en capa oculta: 12
- Algoritmo de entrenamiento: Levenberg-Marquardt (LM)
- Tasa de aprendizaje: 0.002
- Error medio del modelo en el periodo de referencia: -0.007° (0.166%)

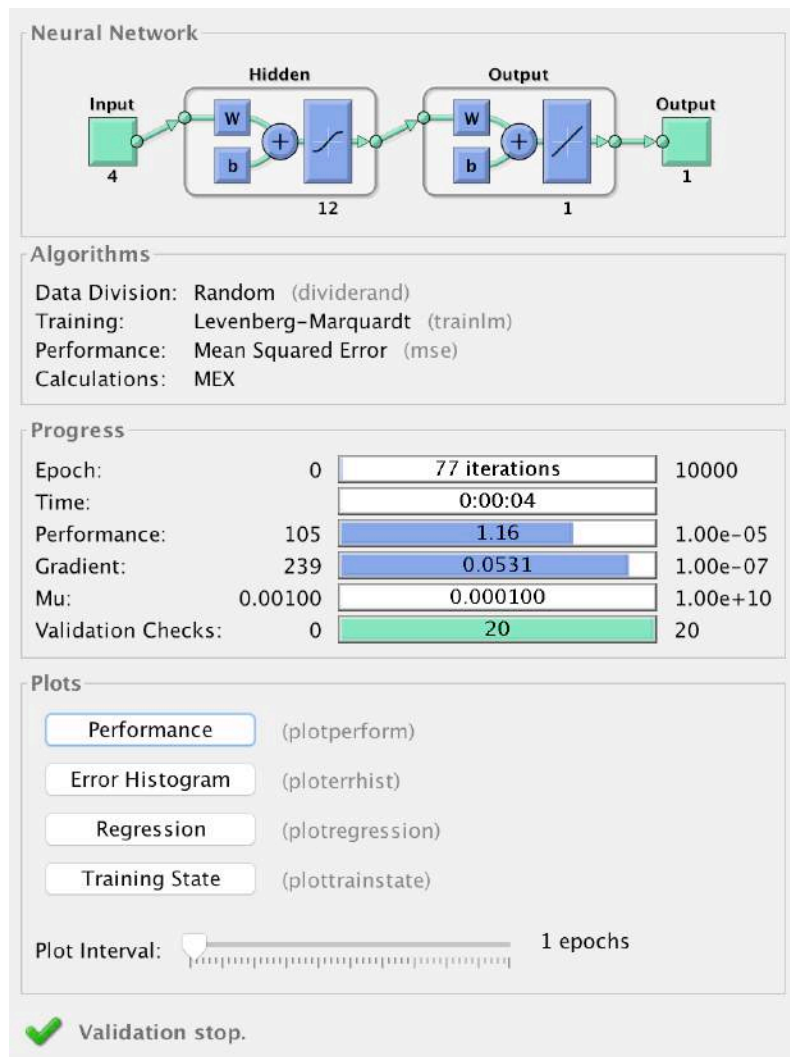


Figura 47. Cuadro de diálogo de la red neuronal de MNL3

Modelos no lineales de C3										
Algoritmo	LM	LM	LM	LM	LM	BR	BR	BR	BR	BR
N Neuronas	2	4	8	12	16	2	4	8	12	16
Regresión	0,966	0,966	0,967	0,967	0,967	0,965	0,966	0,967	0,967	0,967
MSE	1,389	1,352	1,345	1,337	1,34	1,39	1,358	1,346	1,35	1,34
N ciclos	3722	286	792	77	33	532	113	32	44	29
Duración (seg)	65	6,5	30,1	5	2,2	10	2,7	1,3	2,4	2,4

Tabla 11. Resultados de MNL3 para diferentes configuraciones de MLP

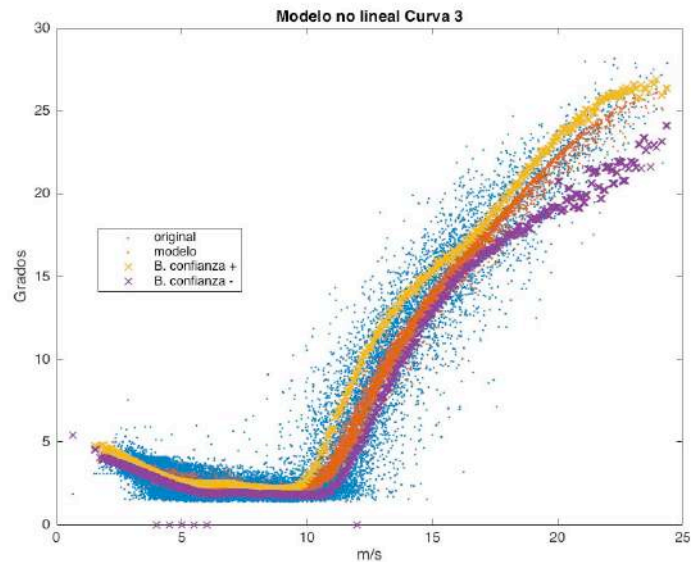


Figura 48. MNL3 en periodo de referencia

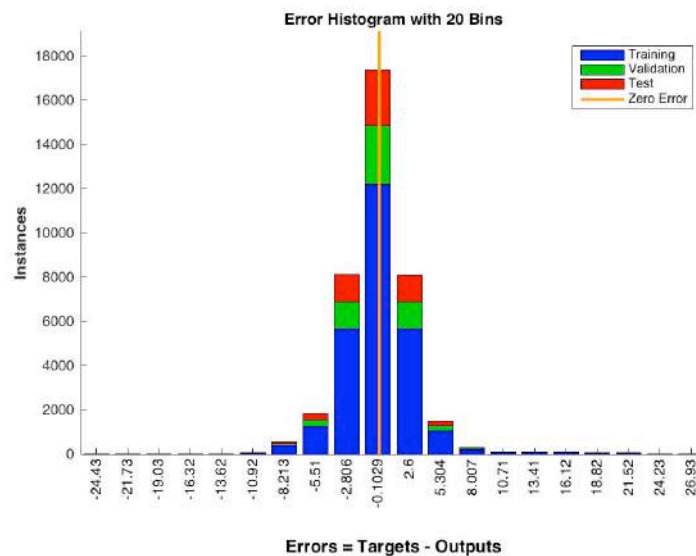


Figura 49. Histograma del error de MNL3 en el periodo de referencia

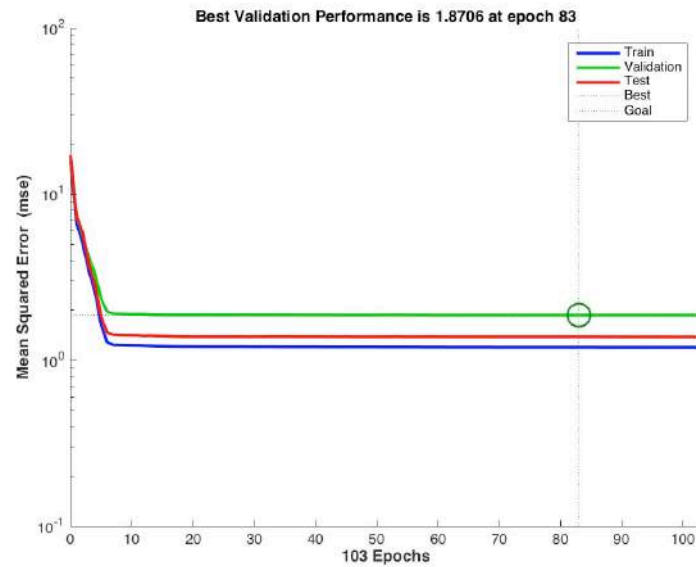


Figura 50. Evolución del MSE de MNL3 en el periodo de referencia

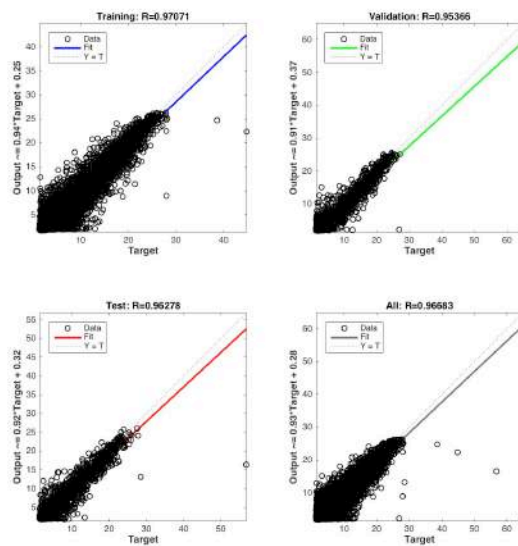


Figura 51. Regresión E/V/T y total de MNL3 en el periodo de referencia

4.2.4 MNL4 (TEMPERATURA DE LOS RODAMIENTOS DE DEL GENERADOR RESPECTO A LA VELOCIDAD DEL VIENTO)

- Variables de entrada: var14, var22, var23, Pred1
- Variable estimada: var3
- Estimación: Pred4
- Tipo de red: MLP
- Porcentajes de E/V/T (%): 70/15/15 (aleatorio)
- Número de neuronas en capa oculta: 16
- Algoritmo de entrenamiento: Levenberg-Marquardt (LM)
- Tasa de aprendizaje: 0.002
- Error medio del modelo en el periodo de referencia: 0.003 °C (0.007%)

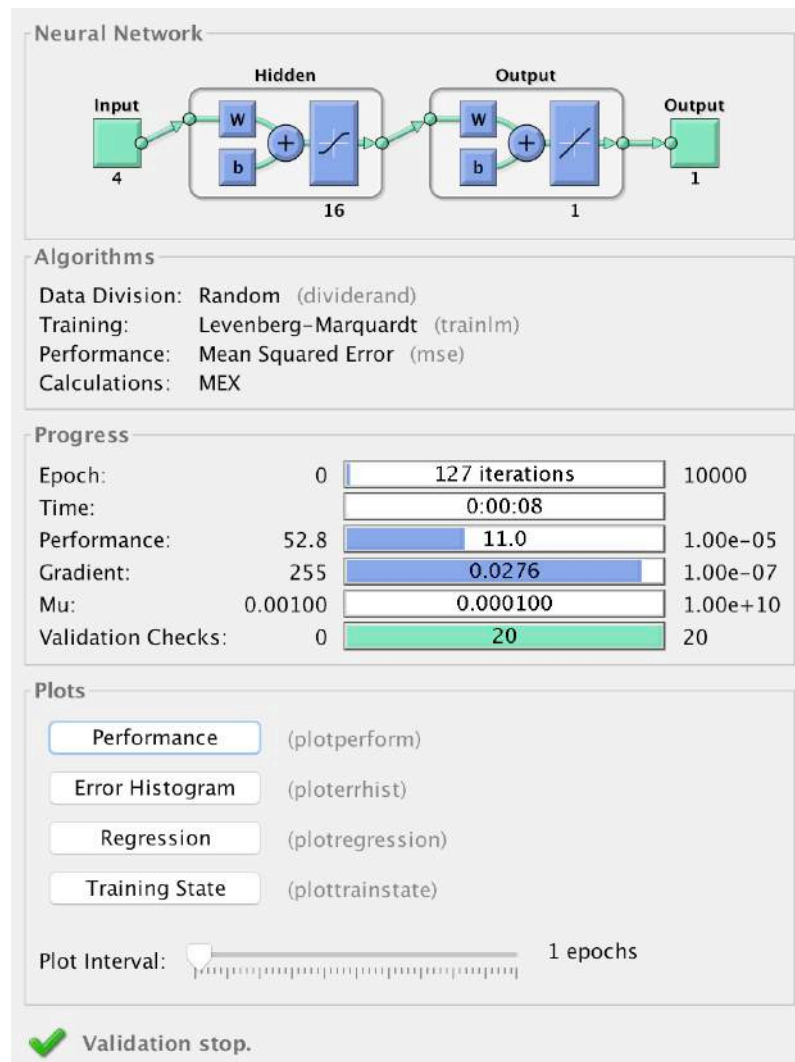


Figura 52. Cuadro de diálogo de la red neuronal de MNL4

Modelos no lineales de C4										
Algoritmo	LM	LM	LM	LM	LM	BR	BR	BR	BR	BR
N Neuronas	2	4	8	12	16	2	4	8	12	16
Regresión	0,865	0,871	0,877	0,879	0,882	0,866	0,873	0,875	0,878	0,88
MSE	12,69	12,16	11,62	11,45	11,18	12,54	11,99	11,76	11,48	11,31
N ciclos	174	136	169	211	127	78	305	89	132	204
Duración (seg)	3,2	3,3	6,6	10,8	8,6	1,6	7,5	3,4	7	13,3

Tabla 12. Resultados de MNL4 para diferentes configuraciones de MLP

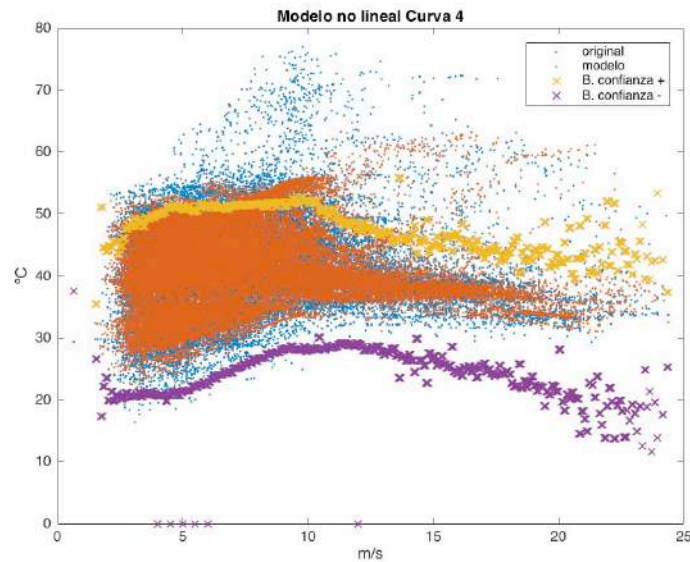


Figura 53. MNL4 en periodo de referencia

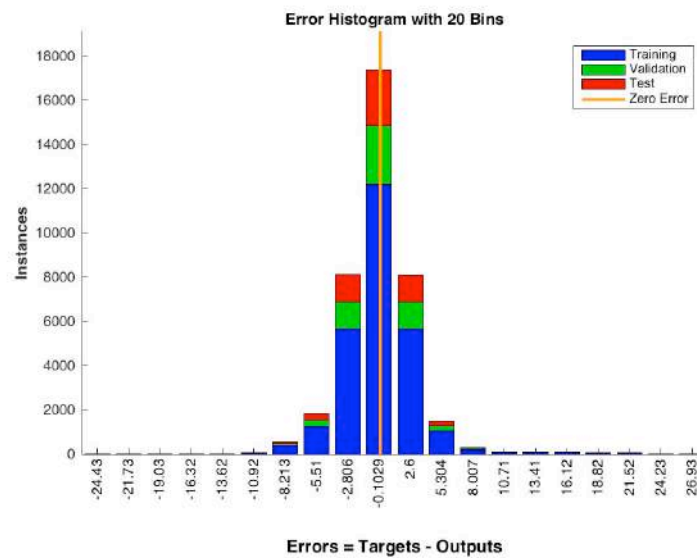


Figura 54. Histograma del error de MNL4 en el periodo de referencia

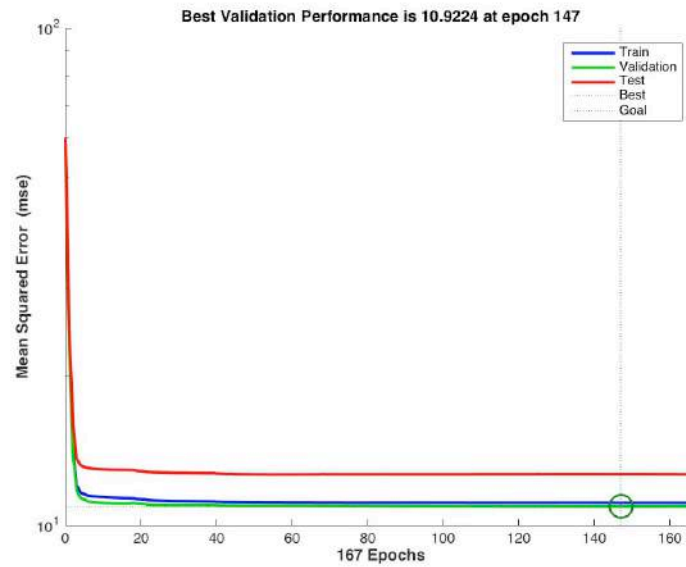


Figura 55. Evolución del MSE de MNL4 en el periodo de referencia

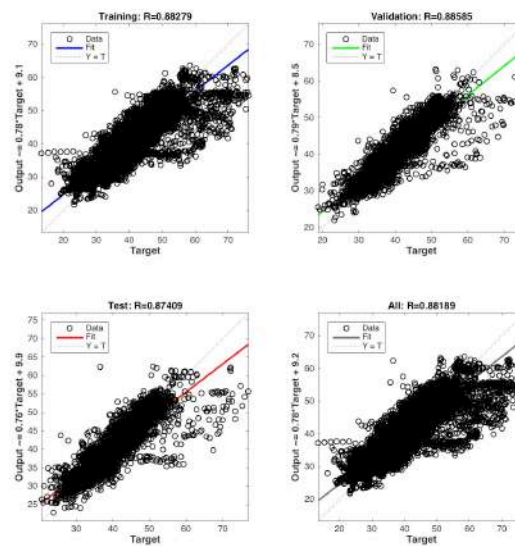


Figura 56. Regresión E/V/T y total de MNL4 en el periodo de referencia

4.2.5 MNL5 (TEMPERATURA DE LOS RODAMIENTOS NDE RESPECTO A LA VELOCIDAD DEL VIENTO)

- Variables de entrada: var14, var22, var23, Pred1
- Variable estimada: var4
- Estimación: Pred5
- Tipo de red: MLP
- Porcentajes de E/V/T (%): 70/15/15 (aleatorio)
- Número de neuronas en capa oculta: 16
- Algoritmo de entrenamiento: Levenberg-Marquardt (LM)
- Tasa de aprendizaje: 0.002
- Error medio del modelo en el periodo de referencia: 0.021 °C (0.030%)

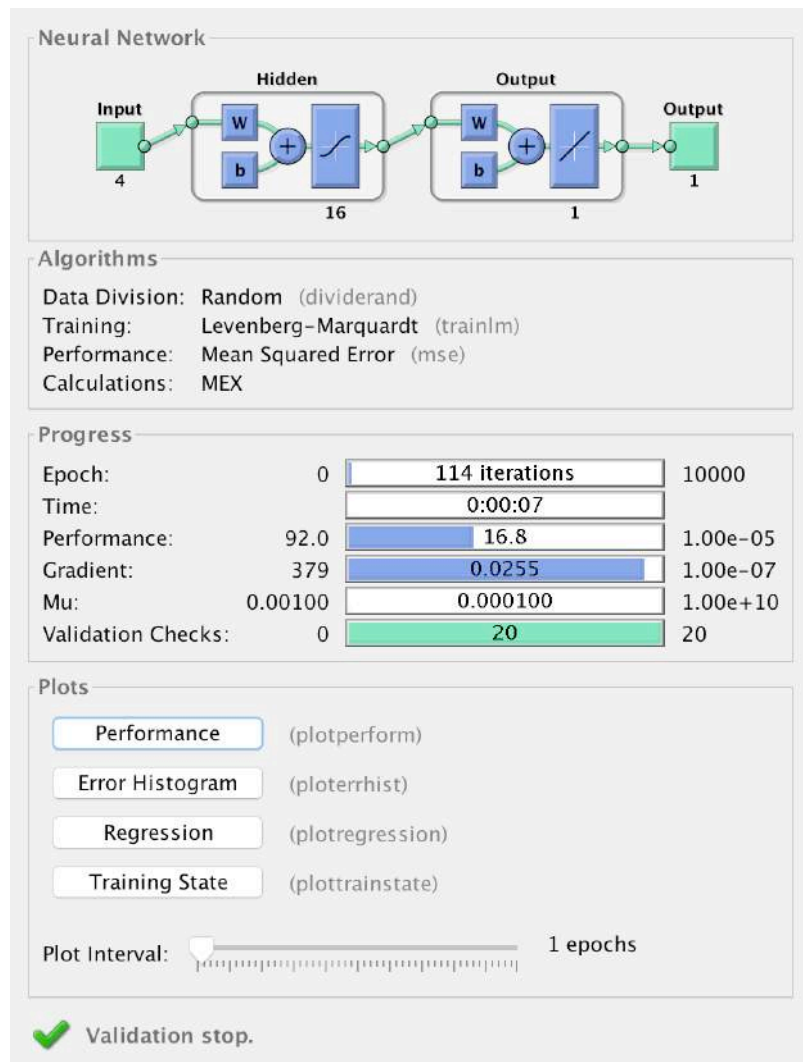


Figura 57. Cuadro de diálogo de la red neuronal de MNL5

Modelos no lineales de C5										
Algoritmo	LM	LM	LM	LM	LM	BR	BR	BR	BR	BR
N Neuronas	2	4	8	12	16	2	4	8	12	16
Regresión	0,832	0,834	0,836	0,838	0,84	0,832	0,836	0,836	0,839	0,838
MSE	17,84	17,56	17,37	17,23	16,97	17,8	17,39	17,38	17,13	17,23
N ciclos	80	41	46	102	114	37	169	29	115	30
Duración (seg)	1,6	1	1,7	5,4	7,9	0,9	4,1	1,2	7,4	2

Tabla 13. Resultados de MNL5 para diferentes configuraciones de MLP

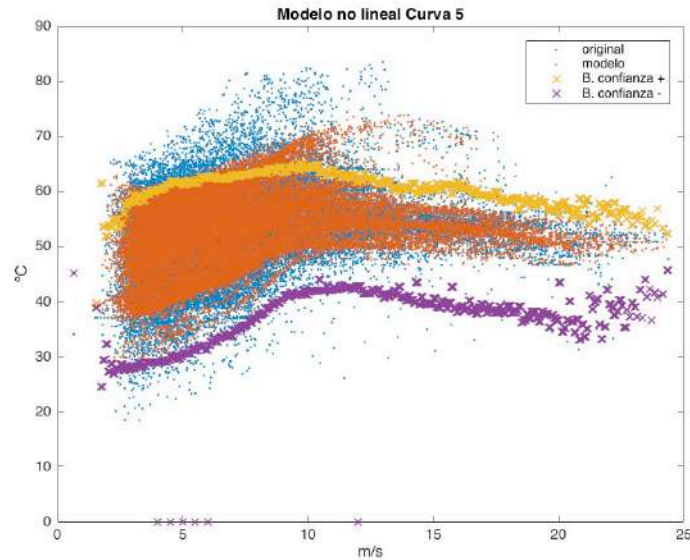


Figura 58. MNL5 en periodo de referencia

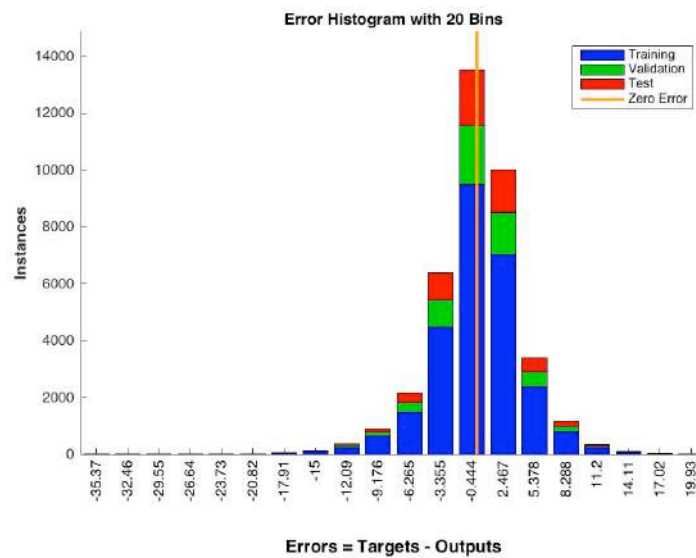


Figura 59 Histograma del error de MNL5 en el periodo de referencia

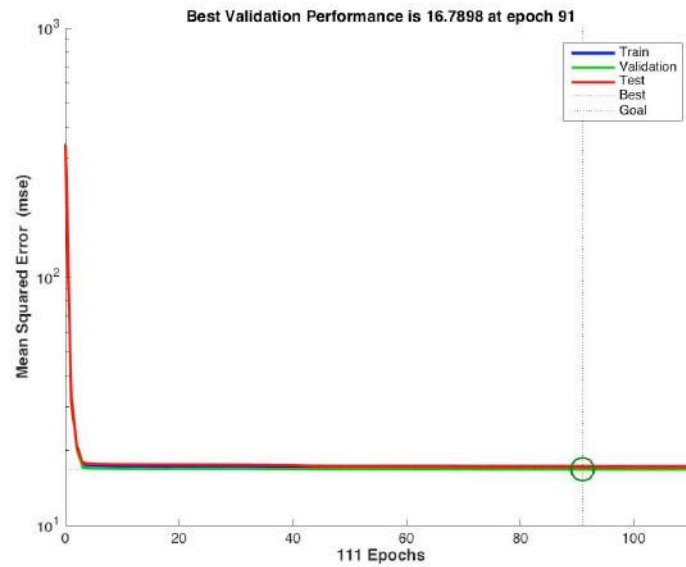


Figura 60. Evolución del MSE de MNL5 en el periodo de referencia

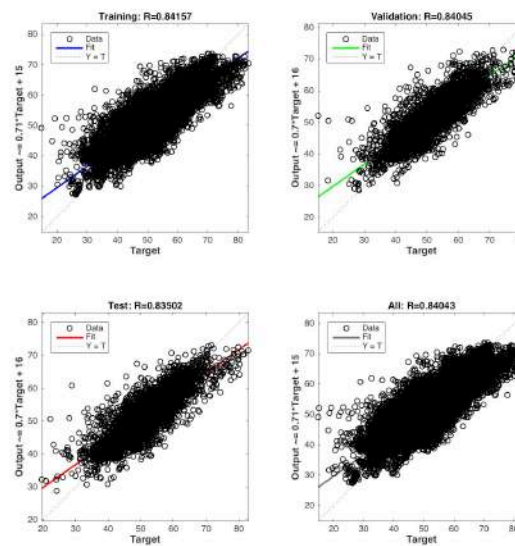


Figura 61. Regresión E/V/T y total de MNL5 en el periodo de referencia

4.2.6 MNL6 (TEMPERATURA DEL ACEITE DE LA MULTIPLICADORA RESPECTO A LA VELOCIDAD DEL VIENTO)

- Variables de entrada: var14, var22, var23, Pred1
- Variable estimada: var1
- Estimación: Pred6
- Tipo de red: MLP
- Porcentajes de E/V/T (%): 70/15/15 (aleatorio)
- Número de neuronas en capa oculta: 16
- Algoritmo de entrenamiento: Regularización Bayesiana (BR)
- Tasa de aprendizaje: 0.002
- Error medio del modelo en el periodo de referencia: 0.022 °C (0.034%)

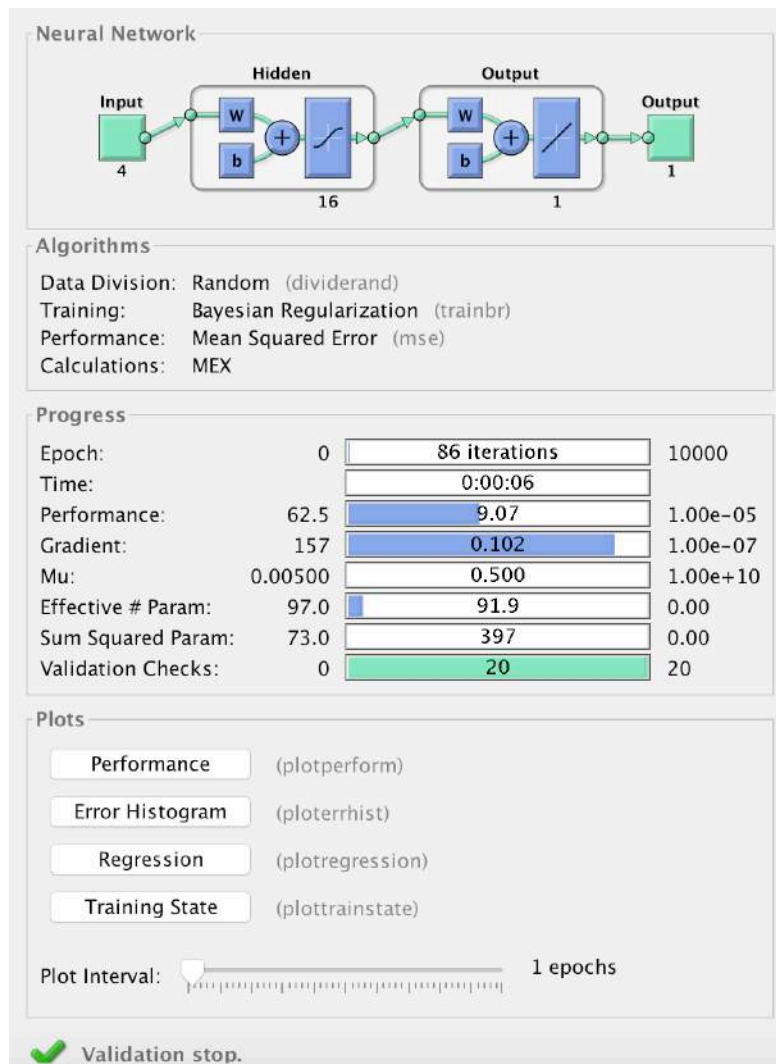


Figura 62. Cuadro de diálogo de la red neuronal de MNL6

Modelos no lineales de C6										
Algoritmo	LM	LM	LM	LM	LM	BR	BR	BR	BR	BR
N Neuronas	2	4	8	12	16	2	4	8	12	16
Regresión	0,671	0,691	0,72	0,725	0,722	0,672	0,71	0,712	0,721	0,727
MSE	10,67	10,12	9,32	9,22	9,27	10,62	9,62	9,54	9,31	9,14
N ciclos	27	53	65	72	51	39	39	60	67	86
Duración (seg)	0,6	1,2	2,4	4,3	3,8	0,8	0,9	2,3	4,1	6

Tabla 14. Resultados de MNL6 para diferentes configuraciones de MLP

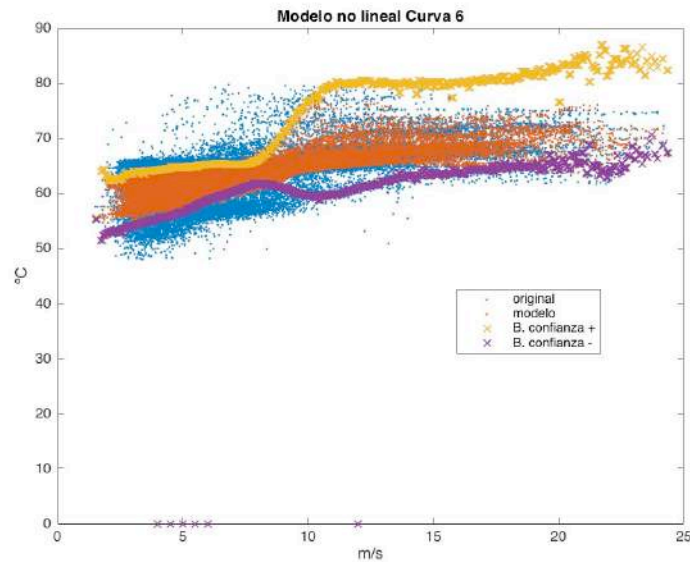


Figura 63. MNL6 en periodo de referencia

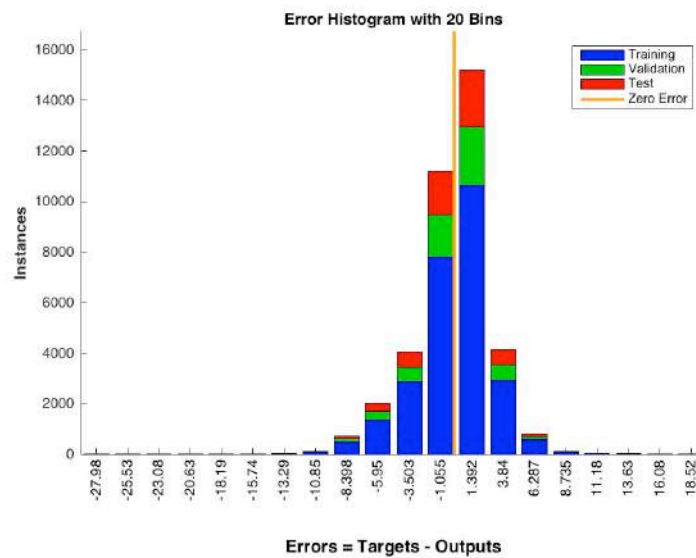


Figura 64 Histograma del error de MNL6 en el periodo de referencia

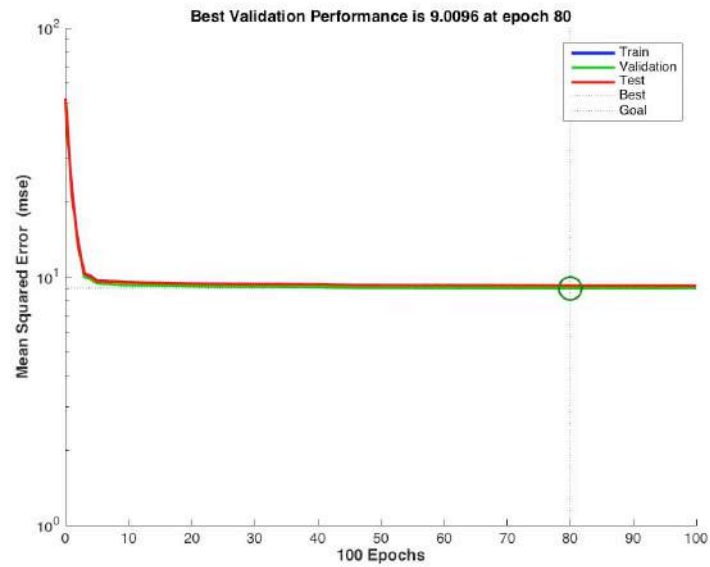


Figura 65. Evolución del MSE de MNL6 en el periodo de referencia

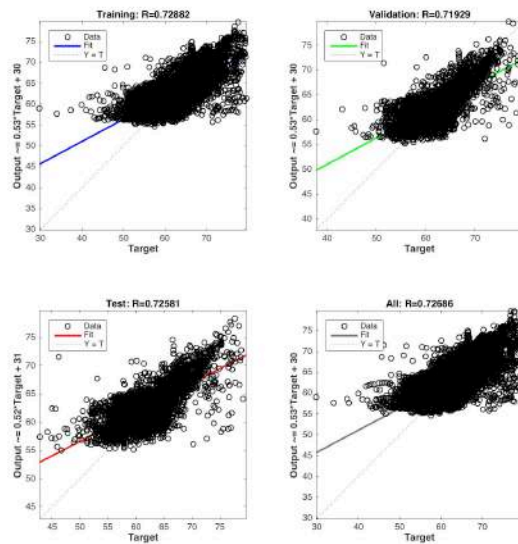


Figura 66. Regresión E/V/T y total de MNL6 en el periodo de referencia

4.2.7 MNL7 (TEMPERATURA DE LOS ANILLOS SLD DEL GENERADOR RESPECTO A LA VELOCIDAD DEL VIENTO)

- Variables de entrada: var14, var22, var23, Pred1
- Variable estimada: var5
- Estimación: Pred7
- Tipo de red: MLP
- Porcentajes de E/V/T (%): 70/15/15 (aleatorio)
- Número de neuronas en capa oculta: 16
- Algoritmo de entrenamiento: Levenberg-Marquardt (LM)
- Tasa de aprendizaje: 0.002
- Error medio del modelo en el periodo de referencia: 0.002 °C (0.010%)

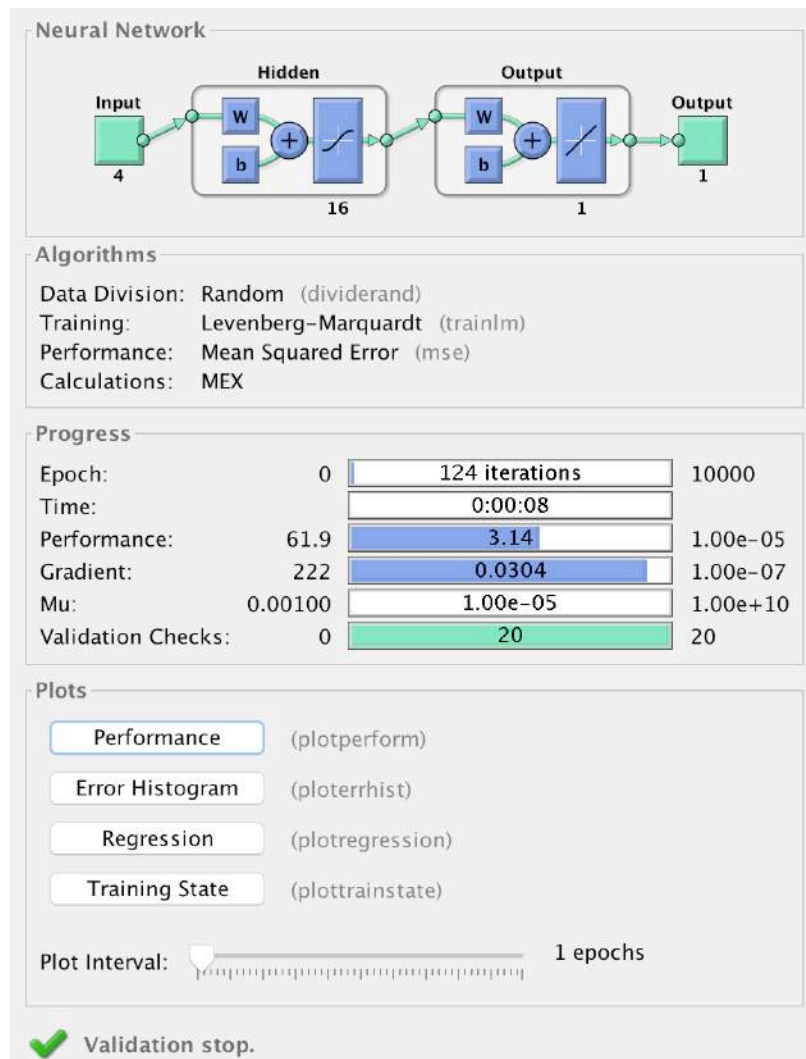


Figura 67. Cuadro de diálogo de la red neuronal de MNL7

Modelos no lineales de C7										
Algoritmo	LM	LM	LM	LM	LM	BR	BR	BR	BR	BR
N Neuronas	2	4	8	12	16	2	4	8	12	16
Regresión	0,974	0,974	0,975	0,975	0,975	0,974	0,974	0,975	0,976	0,975
MSE	3,381	3,318	3,209	3,231	3,195	3,363	3,304	3,257	3,115	3,239
N ciclos	29	87	130	133	124	33	84	215	277	140
Duración (seg)	0,7	1,8	4,9	7	8,1	0,6	2,1	8,2	14,7	9,1

Tabla 15. Resultados de MNL7 para diferentes configuraciones de MLP

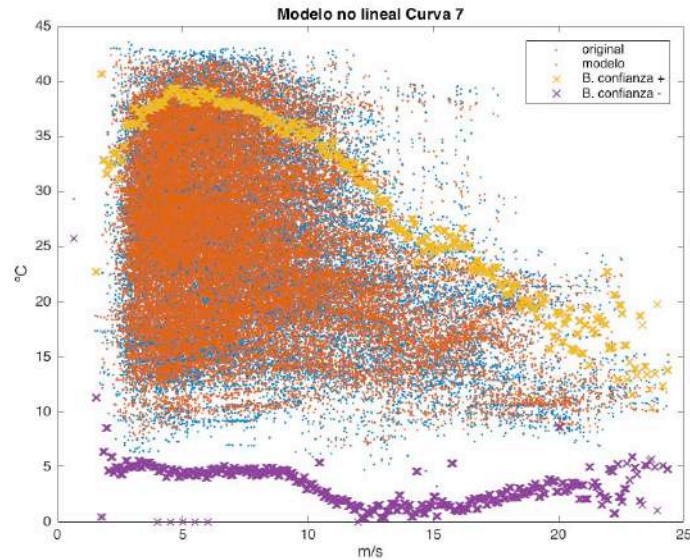


Figura 68. MNL7 en periodo de referencia

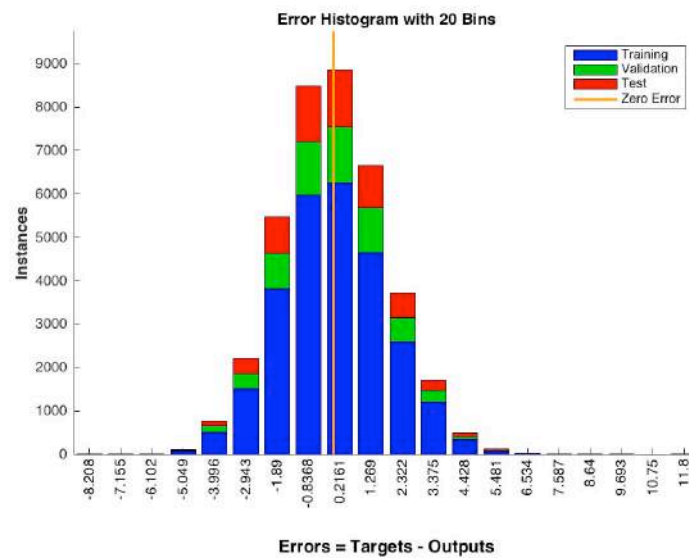


Figura 69. Histograma del error de MNL7 en el periodo de referencia

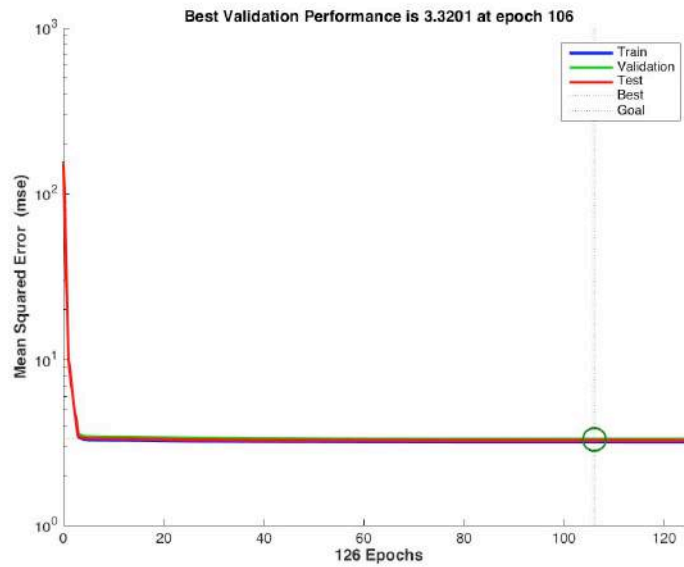


Figura 70. Evolución del MSE de MNL7 en el periodo de referencia

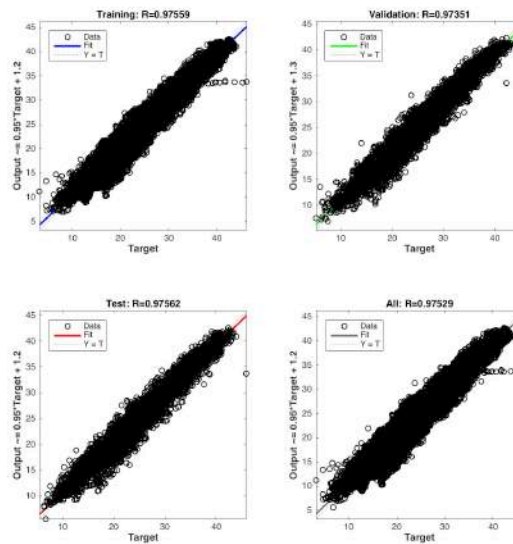


Figura 71. Regresión E/V/T y total de MNL7 en el periodo de referencia

4.2.8 MNL8 (PRESIÓN DEL GRUPO HIDRÁULICO RESPECTO A LA VELOCIDAD DEL VIENTO)

- Variables de entrada: var14, var22, var23, Pred1
- Variable estimada: var11
- Estimación: Pred8
- Tipo de red: MLP
- Porcentajes de E/V/T (%): 70/15/15 (aleatorio)
- Número de neuronas en capa oculta: 12
- Algoritmo de entrenamiento: Regularización Bayesiana (BR)
- Tasa de aprendizaje: 0.002
- Error medio del modelo en el periodo de referencia: -0.008 bares (0.004%)

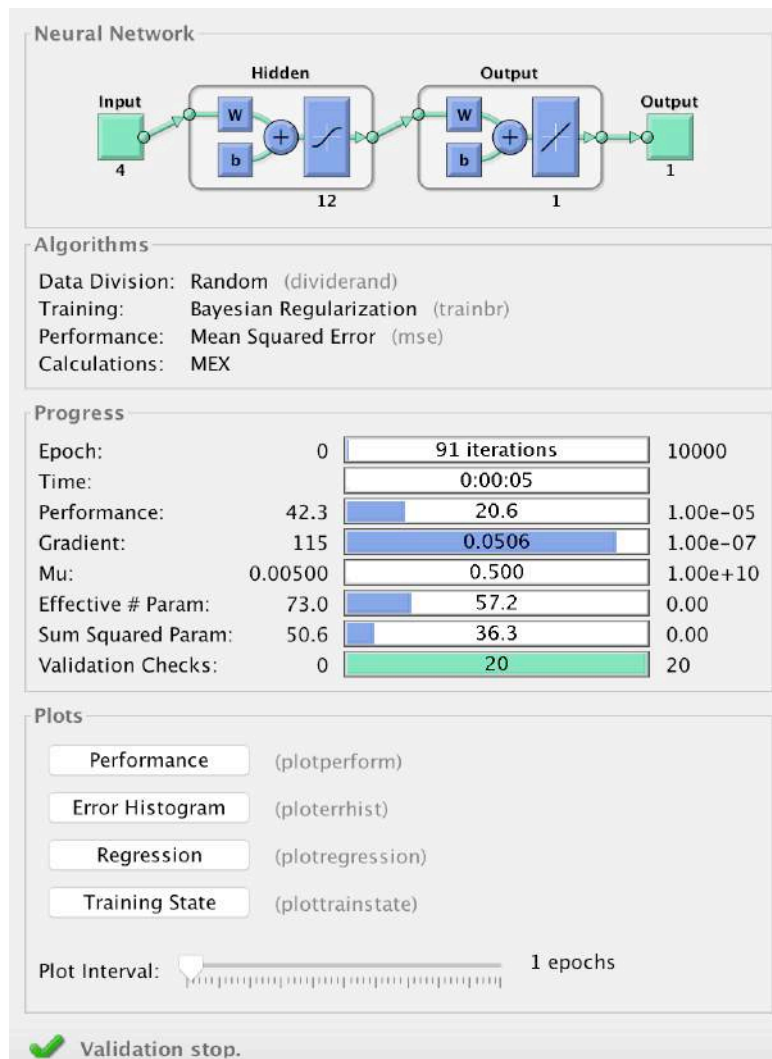


Figura 72. Cuadro de diálogo de la red neuronal de MNL8

Modelos no lineales de C8										
Algoritmo	LM	LM	LM	LM	LM	BR	BR	BR	BR	BR
N Neuronas	2	4	8	12	16	2	4	8	12	16
Regresión	0,051	0,067	0,097	0,141	0,097	0,095	0,084	0,095	0,108	0,103
MSE	20,9	20,86	20,76	20,54	20,75	20,76	20,8	20,76	20,71	20,73
N ciclos	33	28	49	229	27	99	63	44	91	57
Duración (seg)	0,7	0,7	1,8	12,3	2,4	2	1,6	1,7	5,8	3,9

Tabla 16. Resultados de MNL8 para diferentes configuraciones de MLP

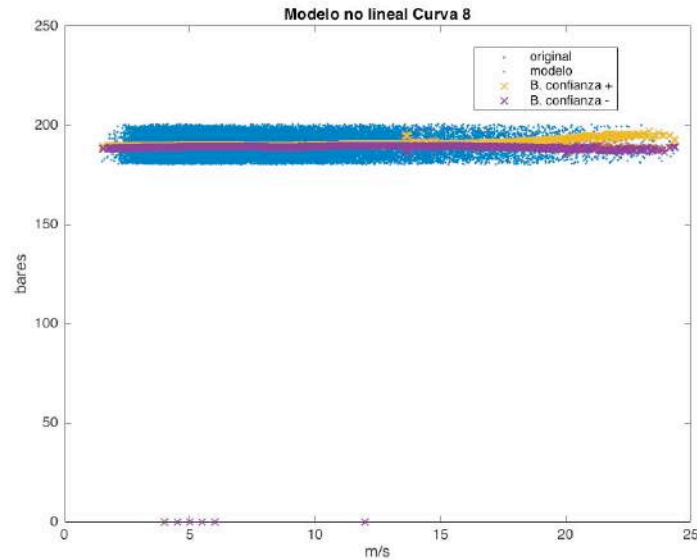


Figura 73. MNL8 en periodo de referencia

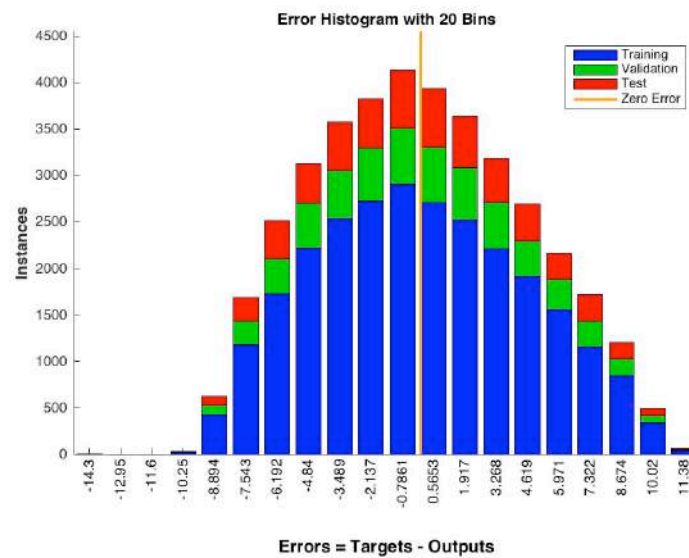


Figura 74. Histograma del error de MNL8 en el periodo de referencia

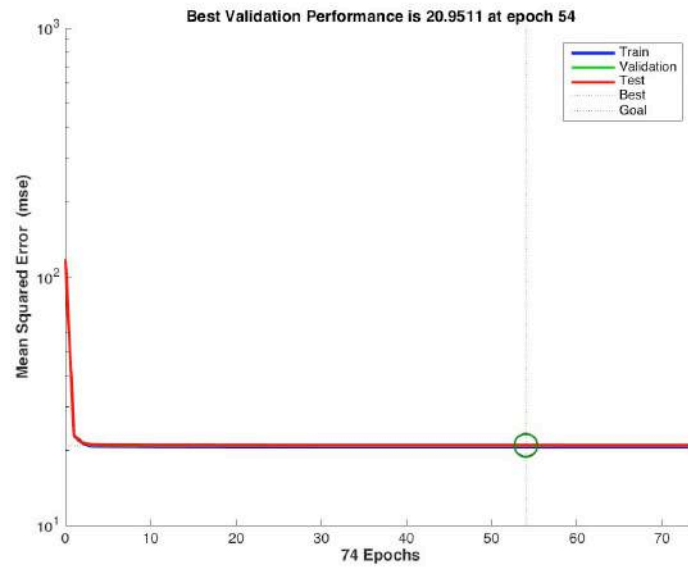


Figura 75. Evolución del MSE de MNL8 en el periodo de referencia

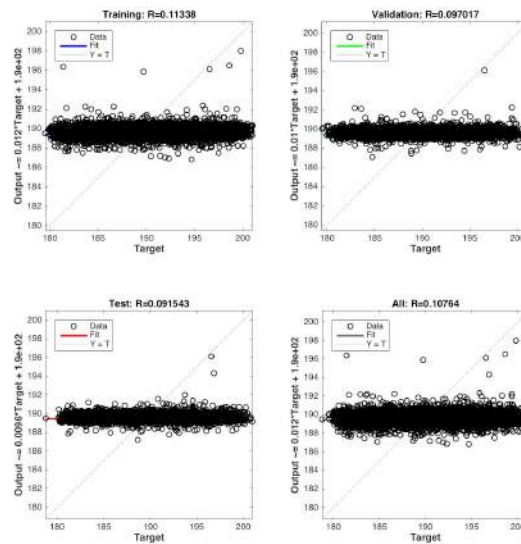


Figura 76. Regresión E/V/T y total de MNL8 en el periodo de referencia

4.2.9 MNL9 (TEMPERATURA DEL GRUPO HIDRÁULICO RESPECTO A LA VELOCIDAD DEL VIENTO)

- Variables de entrada: var14, var22, var23, Pred1
- Variable estimada: var10
- Estimación: Pred9
- Tipo de red: MLP
- Porcentajes de E/V/T (%): 70/15/15 (aleatorio)
- Número de neuronas en capa oculta: 16
- Algoritmo de entrenamiento: Levenberg-Marquardt (LM)
- Tasa de aprendizaje: 0.002
- Error medio del modelo en el periodo de referencia: 0.002 °C (0.007%)

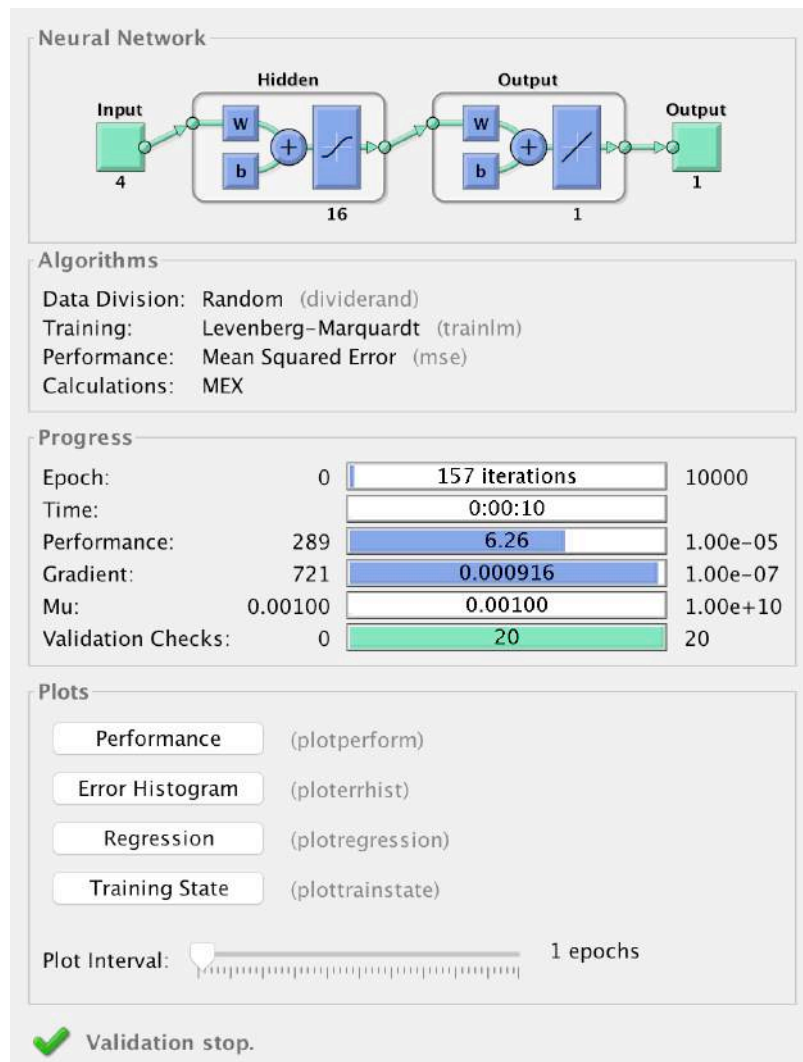


Figura 77. Cuadro de diálogo de la red neuronal de MNL9

Modelos no lineales de C9										
Algoritmo	LM	LM	LM	LM	LM	BR	BR	BR	BR	BR
N Neuronas	2	4	8	12	16	2	4	8	12	16
Regresión	0,927	0,932	0,934	0,935	0,935	0,927	0,932	0,934	0,935	0,935
MSE	7,042	6,588	6,398	6,355	6,279	7,044	6,637	6,385	6,336	6,315
N ciclos	198	44	54	83	157	68	106	122	319	120
Duración (seg)	3,1	1	1,9	4,1	10,5	1,4	2,6	4,7	16,7	8

Tabla 17. Resultados de MNL9 para diferentes configuraciones de MLP

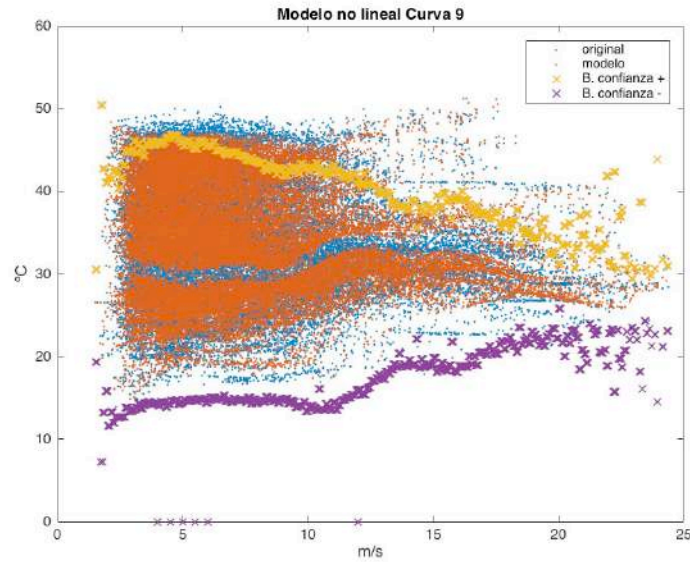


Figura 78. MNL9 en periodo de referencia

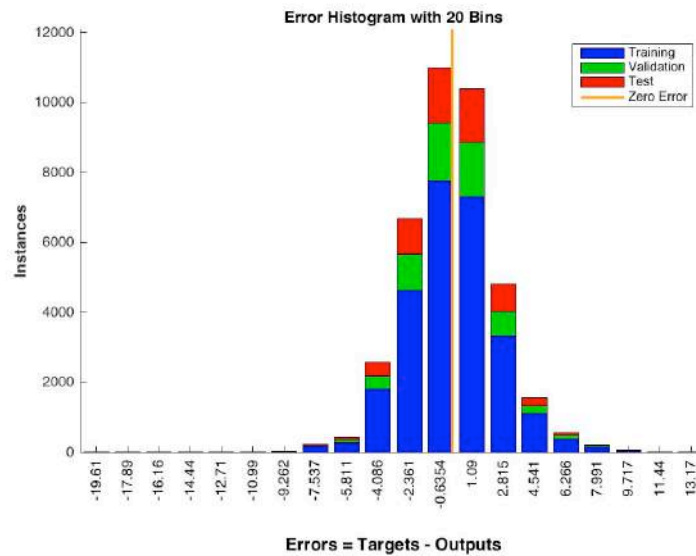


Figura 79. Histograma del error de MNL9 en el periodo de referencia

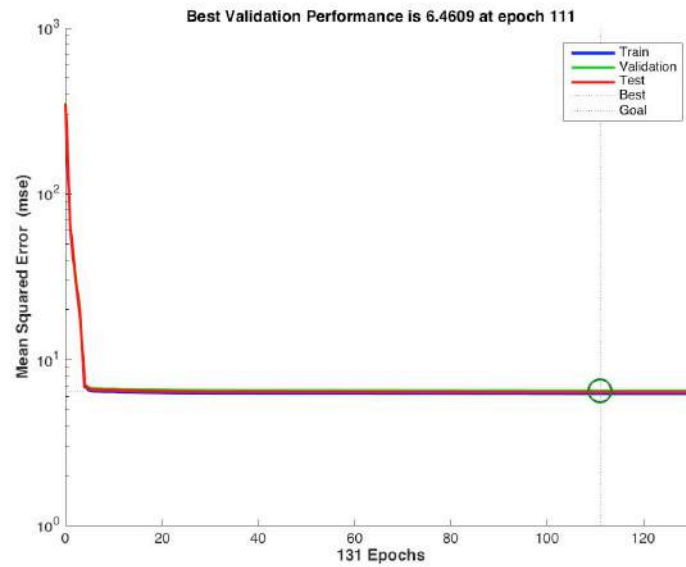


Figura 80. Evolución del MSE de MNL9 en el periodo de referencia

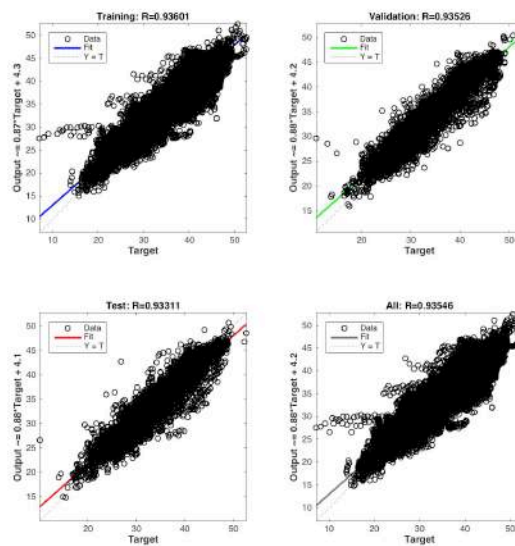


Figura 81. Regresión E/V/T y total de MNL9 en el periodo de referencia

4.3 OBSERVACIONES GENERALES

En general el error de los modelos se aproxima a una distribución normal, a excepción del MNL8, donde el histograma del error está mucho más plano, lo que indica que hay un error entre el valor simulado y el valor real mucho más constante, luego la precisión del modelo es peor. Esto también se puede observar poniendo atención a su valor de regresión lineal.

Tal y como era de esperar, los MNL (a excepción de la C8, de la que se hablará más adelante) aproximan mejor el comportamiento de las curvas de estudio que los ML. Tal y como se muestra en la Figura 82, mientras que los ML se aproximan a una curva, los MNL se aproximan a una nube de puntos con la forma aproximada de la curva de estudio, por lo que en caso de funcionar correctamente permite que el error en cada instante sea menor, pues se ha creado simulado cada punto del periodo de estudio, mientras que en ML se crea una predicción unitaria para cada punto dentro de un intervalo de velocidad de viento. En consecuencia, el error medio de los MNL es menor que el de los ML.

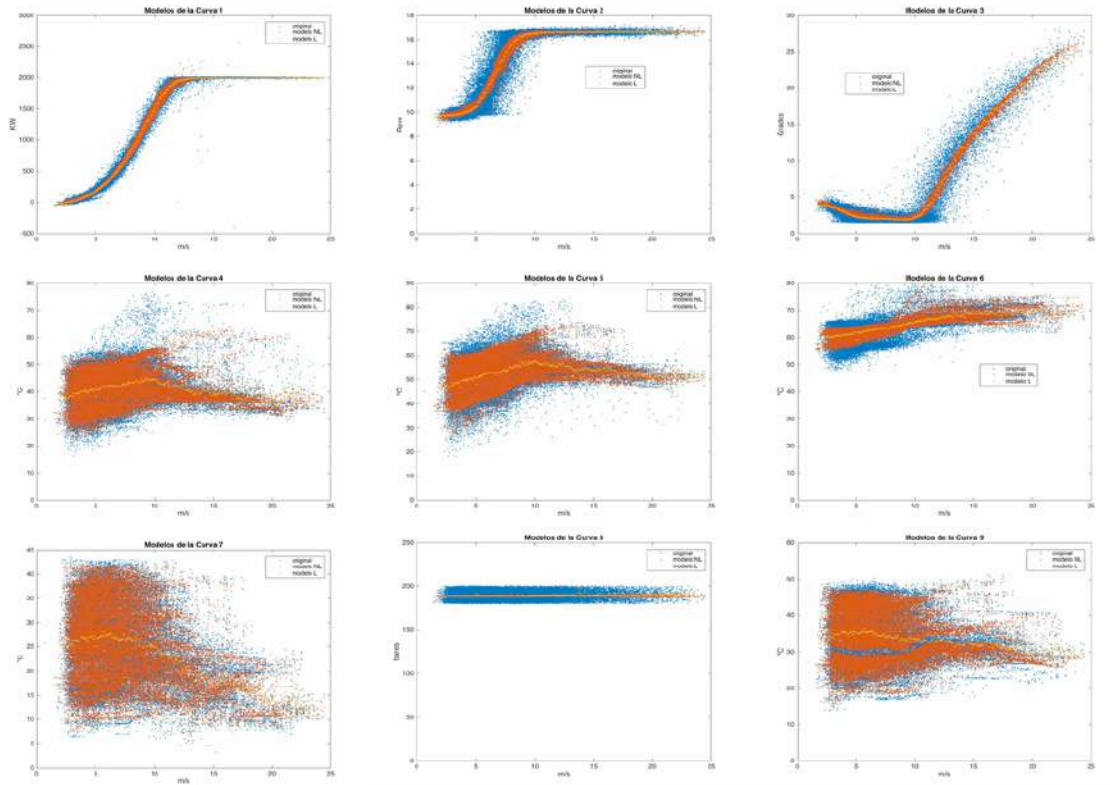


Figura 82. Comparación de ML y MNL en las curvas de estado

4.4 CONCLUSIONES

En la Tabla 18, Tabla 19, Tabla 20 y Tabla 21 se recogen los datos con el rendimiento de todos los modelos. Cabe destacar que se pueden extraer más datos del funcionamiento de los modelos no lineales que de los modelos lineales, como el valor de regresión lineal. Este dato no verifica que un modelo es mejor que otro, pero permite observar si la red funciona bien en un determinado periodo de tiempo. Por ejemplo, en MNL3 se puede ver a través del valor de regresión lineal que durante el Y3 el rendimiento de la red baja significativamente. pero permite observar cómo se comporta el modelo, tanto en el periodo de entrenamiento como en el período de simulación, para cualquier intervalo de tiempo dado. Se ha obtenido el valor de regresión lineal de cada modelo no lineal para cada año para verificar si las anomalías que encuentren estos modelos son o no fiables.

CURVA	Tipo de modelo	Valores fuera de las bandas de confianza	% valores entre las bandas de confianza	Valores de regresión lineal
C1	ML	3846	97,3%	
	MNL	1847	95,2%	99,7%
C2	ML	3535	97,7%	
	MNL	2088	94,6%	96,0%
C3	ML	3189	98,8%	
	MNL	1329	96,6%	96,7%
C4	ML	2800	98,5%	
	MNL	2740	92,9%	88,2%
C5	ML	3346	97,8%	
	MNL	1896	95,1%	84,0%
C6	ML	4027	96,3%	
	MNL	388	99,0%	72,7%
C7	ML	1816	98,7%	
	MNL	2987	92,3%	97,5%
C8	ML	1674	99,5%	
	MNL	3404	91,2%	10,8%
C9	ML	2032	98,4%	
	MNL	2307	94,0%	93,5%

Tabla 18. Resultados de los modelos durante el periodo de referencia

AÑO	C1	C2	C3	C4	C5	C6	C7	C8	C9
Y1 (Ent.)	99,7	96	96,7	88,2	84,0	72,7	97,5	10,8	93,5
Y2	99,6	95,7	95,8	84,4	80,9	63,2	96,1	2,1	89,9
Y3	93,7	77,4	16,2	78,2	74,4	56,1	86,8	2,7	73,6
Y4	99,1	95,0	92,5	82,9	78,2	74,0	95,5	1,9	87,6
Y5	99,6	95,8	95,7	80,6	77,5	73,2	95,0	1,2	84,0
Y6	99,6	95,5	93,5	76,0	66,8	72,7	97,6	1,5	84,5

Tabla 19. Regresión lineal de los MNL por años

CURVA	Tipo de modelo	Error medio en periodo de entrenamiento	
C1	ML	2,269 kW	0,276%
	MNL	0,278 kW	0,033%
C2	ML	-0,009 rpm	0,066%
	MNL	0,005 rpm	0,036%
C3	ML	-0,010 °	0,245%
	MNL	-0,007 °	0,166%
C4	ML	-0,025 °C	0,060%
	MNL	0,003 °C	0,007%
C5	ML	-0,035 °C	0,043%
	MNL	0,021 °C	0,030%
C6	ML	-0,015 °C	0,024%
	MNL	0,022 °C	0,034%
C7	ML	-0,056 °C	0,224%
	MNL	0,002 °C	0,010%
C8	ML	-0,147 bares	0,077%
	MNL	-0,008 bares	0,004%
C9	ML	-0,030 °C	0,088%
	MNL	0,002 °C	0,007%

Tabla 20. Error medio de los todos los modelos durante el periodo de entrenamiento

CURVA	Tipo de modelo	Desviación típica en periodo de entrenamiento	
C1	ML	93,495 kW	11,37%
	MNL	50,324 kW	5,99%
C2	ML	0,918 rpm	6,76%
	MNL	0,810 rpm	5,91%
C3	ML	1,164 °	28,50%
	MNL	1,156 °	27,33%
C4	ML	6,907 °C	16,66%
	MNL	3,343 °C	8,98%
C5	ML	7,248 °C	13,70%
	MNL	4,120 °C	8,49%
C6	ML	3,930 °C	6,18%
	MNL	3,023 °C	4,70%

C7	ML	7,460 °C	2985%
	MNL	1,787 °C	9,09%
C8	ML	6,883 bares	36,36%
	MNL	4,551bares	24,01%
C9	ML	7,023 °C	20,64%
	MNL	2,506 °C	8,54%

Tabla 21. Desviación típica de los todos los modelos durante el periodo de entrenamiento

A la vista de los datos arrojados en las tablas mostradas, se toman las siguientes conclusiones respecto a los modelos aquí mostrados:

- Se deshecha en MNL8. No supera el 11% de regresión lineal y da los valores de error más elevados. Generalmente los modelos no lineales tienen más problemas que los lineales para simular rectas. Mientras que en el resto de curvas los modelos no lineales simulan una nube de puntos, el resultado del MNL8 es prácticamente una recta, y la amplitud de sus bandas de confianza es mínima, por lo que los puntos que se encuentran fuera de ellas tienen un error muy elevado y con una distribución que poco se ajusta a una distribución normal (Figura 74)
- Se concluye que, debido al valor de regresión lineal durante Y3, el MNL3 es muy poco fiable para este periodo. Esto ya se señaló en el Capítulo 3. . No obstante, hay que resaltar que el resto de años la regresión lineal es muy buena (superior al 92%), y como se observó en la Figura 18 el comportamiento de esta curva durante este año es notablemente diferente al resto de años, luego este valor tan bajo de regresión lineal, más que a un fallo en el modelado, se debe a un comportamiento atípico del componente durante todo el Y3.
- Se concluye que, debido al valor de regresión lineal durante Y2 y Y3, el MNL6 es muy poco fiable para este periodo.
- Se concluye que el resto de modelos creados simulan adecuadamente las curvas de estudio.

Capítulo 5. DETECCIÓN DE ANOMALÍAS

5.1 INTRODUCCIÓN

En esta sección se mostrarán las anomalías detectadas mediante los modelos de comportamiento anteriormente construidos. Se analizará el origen de las anomalías de cada modelo, y las consecuencias que tienen en el funcionamiento del aerogenerador.

Conviene recordar que los cálculos realizados se hacen sobre los valores medidos por el sistema SCADA de la máquina para los instantes en que esta se encontraba en funcionamiento, lo que corresponde a más de 1506 días de funcionamiento repartidos entre las 0:00 del 16 de abril de 2010 hasta las 23:50 del 20 de marzo de 2016.

5.2 EXTRACCIÓN DE ANOMALÍAS

Se han obtenido los vectores error de cada modelo durante el periodo de simulación y se han agrupado en dos matrices, una correspondiente a los modelos lineales, y otra correspondiente a los no lineales. Estos vectores muestran la diferencia entre el valor predicho por cada modelo para cada instante del periodo de simulación (Y2, Y3, Y4, Y5 y Y6) y el valor real de dicho instante. Se toman los valores de aquellos instantes en los que el error en valor absoluto es mayor al ancho de las bandas de confianza (2 desviaciones típicas más el error medio de cada modelo). La Tabla 22 recoge la cantidad de puntos fuera de las bandas de confianza en cada modelo y cada año de simulación.

Modelo	Y2		Y3		Y4		Y5		Y6	
	Total	%	Total	%	Total	%	Total	%	Total	%
ML1	3344	8,54	33200	82,97	4547	11,37	2983	8,15	1769	7,86
MNL1	11079	28,29	35727	89,28	11878	29,69	11604	31,69	7096	31,53
ML2	2763	7,06	15249	38,11	3189	7,97	2952	8,06	2152	9,56
MNL2	19295	49,27	30816	77,01	19624	49,05	18516	50,56	11854	52,67
ML3	2051	5,24	14623	36,54	2672	6,68	2280	6,23	2091	9,29
MNL3	22221	56,75			23280	58,19	21062	57,52	13391	59,5
ML4	1932	4,93	3253	8,13	3034	7,58	1616	4,41	576	2,56
MNL4	796	2,03	4194	10,48	5071	12,67	3467	9,47	2091	9,29
ML5	1586	4,05	5192	12,97	7655	19,13	3906	10,67	1156	5,14
MNL5	613	1,57	5216	13,03	16402	41	10397	28,39	2500	11,11
ML6	2064	5,27	1917	4,79	778	1,94	832	2,27	361	1,6
MNL6	7427	18,97	9186	22,96	11480	28,69	11578	31,62	9059	40,25
ML7	951	2,43	3493	8,73	2504	6,26	1751	4,78	304	1,35
MNL7	13	0,03	4648	11,62	2440	6,1	1478	4,04	283	1,26
ML8	800	2,04	1234	3,08	911	2,28	853	2,33	726	3,23
ML9	2209	5,64	2600	6,5	1754	4,38	1011	2,76	367	1,63
MNL9	386	0,99	4413	11,03	2168	5,42	2464	6,73	1670	7,42
Promedio ML	1967	5,02	8973	22,42	3005	7,51	2020	5,52	1056	4,69
Promedio MNL	7729	19,74	13457	33,63	11543	28,85	10071	27,5	5993	26,63

Tabla 22. Valores fuera de las bandas de confianza por años y modelos

Se observa que durante el Y3 el número de valores fuera de las bandas de confianza es notablemente mayor al resto de años, siendo el porcentaje de ambos modelos de la curva de potencia (C1) mayor del 80%. Dado que se trata de la curva de potencia que últimamente modela el funcionamiento global de la máquina, es de suma importancia analizar más adelante este periodo.

Todos los puntos señalados en la tabla muestran comportamientos atípicos, que se podrían catalogar de anomalías, pero hay que tener en cuenta que estos datos provienen de más de 10 variables durante casi 6 años. Muchos de los valores anómalos registrados pueden

tratarse de fallos en la medida del sistema SCADA, reparaciones o simplemente acontecimientos puntuales externos al sistema que alteran la medida durante un instante pero que tienen efectos en el rendimiento del aerogenerador. Por ello se ha creído conveniente distinguir entre valor fuera de las bandas de confianza y anomalía.

Una anomalía se define como una desviación respecto a lo que es normal o predecible. Se podrían haber tomado todos los valores señalados en la Tabla 22 como anomalías, pero se ha creído conveniente acotar más rígidamente lo que se está tomando como anomalía. Para evitar tomar como anomalía un instante aislado en el que un punto de una curva de estado no es el esperado, se ha establecido como condición primera que una anomalía tiene que afectar a más de un instante.

El sistema SCADA hace mediciones diezminutales, que en el caso de las variables de velocidad del viento y potencia (var22 y var30 respectivamente) son promedios durante los 10 minutos, y que para el resto de variables toma el valor que hay en el momento de la medición. Para descartar valores aislados se estima que una anomalía debe durar más de 10 minutos. También se tiene en cuenta que, en casos limítrofes al borde de las bandas de confianza, un error en el funcionamiento de varias horas de duración puede saltar entre dentro y fuera de las bandas de confianza.

Como solución se ha concluido que una anomalía es aquel conjunto de 30 minutos (3 valores consecutivos) en el que al menos 2 de las 3 mediciones realizadas estén fuera de las bandas de confianza. Es preciso que el aerogenerador este en operación durante los 30 minutos en los que se evalúa la anomalía, por ello hay que tomar sólo valores consecutivos en los que la var32 sea 100 (aerogenerador en operación). Utilizando esta c como sesgo se obtienen 56820 instantes de 30 minutos de duración, equivalente al 78.57% de los valores de encendido, y el 58.51% de los valores totales. A estos instantes se les denomina registro de anomalías.

Cobra importancia además del año de simulación, la estación del año del mismo, pues 5 curvas de estudio modelan variables de temperatura. Los ML están indirectamente relacionados con la temperatura ambiente y los MNL de manera directa, pues se ha

utilizado esa variable (var14) para entrenarlos. Las estaciones que se han denominado como Q1, Q2, Q3 y Q4 corresponden a una aproximación de lo que serían invierno, primavera, verano y otoño, pues se ha tomado como inicio del invierno el primer instante del primer año de simulación (Y2) y cada estación dura 3 meses. De esta manera el invierno comienza en el 14/15 de diciembre, la primavera en el 14/15 de marzo, el verano en el 14/15 de junio y el otoño en el 14/15 de septiembre. El día 15 corresponde a las fechas anteriores al 2014 y el día 14 a las fechas del 2014 y posteriores. Esto se debe a que 2014 fue un año bisiesto, pero no se tuvo en cuenta para que la duración de los años Y1 a Y5 fuese la misma.

El registro de anomalías se divide en años y estaciones, la Tabla 23 refleja esta división. Como era de esperar, las subdivisiones no tienen el mismo tamaño, pues era improbable que en cada cuadrante hubiese el mismo número de jornadas de 30 minutos encendido. Sin embargo, a excepción del Q3 del Y6 (cabe recordar que es la franja temporal en la que faltan 70 días de datos, como ya se ha explicado previamente), se mueven en un rango de entre 50 y 80 días de datos.

	Q1	Q2	Q3	Q4
Y2	1-3501	3502-6645	6646-9107	9908-12335
Y3	12336-15197	15198-19058	19059-22761	22762-25603
Y4	25604-28875	28876-32063	32064-34987	34988-38257
Y5	38258-40977	40978-43955	43956-46447	46448-49679
Y6	49680-53252	53253-56372	56372-56813	

Tabla 23. Distribución temporal del registro de anomalías

Las siguientes tablas desglosan la distribución de las anomalías, tanto para modelos ML como para modelos MNL. Se observa que los modelos MNL son capaces de detectar más anomalías en las mismas curvas que los ML. En general, los modelos presentan un mayor número de anomalías en el Q1 (invierno), especialmente los modelos de la C9.

Recuento de anomalías para modelos ML										
Y	Q	C1	C2	C3	C4	C5	C6	C7	C8	C9
Y2	Q1	197	89	62	202	144	124	157	7	475
	Q2	94	87	57	65	60	21	7	6	10
	Q3	348	78	36	228	115	45	82	7	80
	Q4	117	145	79	59	120	428	18	10	102
Y3	Q1	2042	835	895	327	691	311	193	49	396
	Q2	3412	1429	1312	105	176	102	94	29	14
	Q3	3354	1555	1315	472	438	22	495	35	331
	Q4	2315	1017	1059	157	388	195	335	33	98
Y4	Q1	560	143	126	104	592	166	152	14	115
	Q2	452	145	141	165	588	16	181	13	61
	Q3	106	130	59	557	1003	0	400	1	251
	Q4	109	99	50	99	285	22	37	8	93
Y5	Q1	127	96	76	75	236	73	42	13	129
	Q2	170	132	76	109	280	25	100	10	25
	Q3	92	123	47	231	407	16	317	6	72
	Q4	127	102	60	53	288	97	68	15	80
Y6	Q1	224	184	158	79	279	51	32	22	47
	Q2	99	141	193	47	43	11	23	24	43
	Q3	24	37	14	39	19	11	23	13	21
TOTAL		13969	6567	5815	3173	6152	1736	2756	315	2443
%		24,6	11,6	10,2	5,6	10,8	3,1	4,9	0,6	4,3

Tabla 24. Distribución temporal de anomalías en ML

Recuento de anomalías para modelos MNL										
Y	Q	C1	C2	C3	C4	C5	C6	C7	C8	C9
Y2	Q1	798	1700	1986	53	53	649	1	NaN	35
	Q2	559	1421	1862	50	35	343	0	NaN	5
	Q3	958	1379	1530	103	18	422	0	NaN	0
	Q4	618	1604	1766	35	60	812	1	NaN	74
Y3	Q1	2326	2244	2351	482	583	1075	458	NaN	1215
	Q2	3611	3140	3038	213	216	907	311	NaN	33

	Q3	3483	2947	2837	347	289	373	199	NaN	82
	Q4	2485	2364	2183	334	610	597	534	NaN	134
Y4	Q1	1091	1561	1979	683	2023	880	546	NaN	176
	Q2	1008	1547	1943	286	1294	1054	154	NaN	57
	Q3	575	1477	1720	320	951	349	19	NaN	44
	Q4	545	1525	1877	346	1087	1248	72	NaN	399
Y5	Q1	675	1381	1542	363	967	919	57	NaN	541
	Q2	833	1458	1752	332	873	1339	305	NaN	29
	Q3	821	1415	1629	136	375	414	26	NaN	11
	Q4	801	1472	1779	260	1147	866	75	NaN	126
Y6	Q1	1031	1858	2210	380	562	1369	20	NaN	321
	Q2	814	1674	1877	241	197	1425	59	NaN	199
	Q3	116	222	295	23	15	61	11	NaN	12
TOTAL		23148	32389	36156	4987	11355	15102	2848	0	3493
%		40,7	57	63,6	8,8	20	26,6	5	0	6,1

Tabla 25. Distribución temporal de anomalías en MNL

Anomalías agrupadas entre Y2 y Y5 en modelos ML									
	C1	C2	C3	C4	C5	C6	C7	C8	C9
Q1	2926	1163	1159	708	1663	674	544	83	1115
Q2	4128	1793	1586	444	1104	164	382	58	110
Q3	3900	1886	1457	1488	1963	83	1294	49	734
Q4	2668	1363	1248	368	1081	742	458	66	373
Anomalías agrupadas entre Y2 y Y5 en modelos MNL									
	C1	C2	C3	C4	C5	C6	C7		C9
Q1	4890	6886	7858	1581	3626	3523	1062		1967
Q2	6011	7566	8595	881	2418	3643	770		124
Q3	5837	7218	7716	906	1633	1558	244		137
Q4	4449	6965	7605	975	2904	3523	682		733

Tabla 26. Distribución temporal de todos los modelos por estaciones

En invierno principalmente y también en verano (Q1 y Q3 respectivamente) es cuando más anomalías suceden, esto está relacionado con que generalmente son los momentos donde las variables de entrada al sistema varían más. En la Figura 83 se aprecia la división por años y estaciones de las variables de entrada. Lógicamente las temperaturas más elevadas suceden en verano y las más frías en invierno, mientras que el viento de entrada al aerogenerador es más veloz en otoño y en invierno que en verano y primavera. Respecto al ángulo del viento, la época del año no parece alterar su valor.

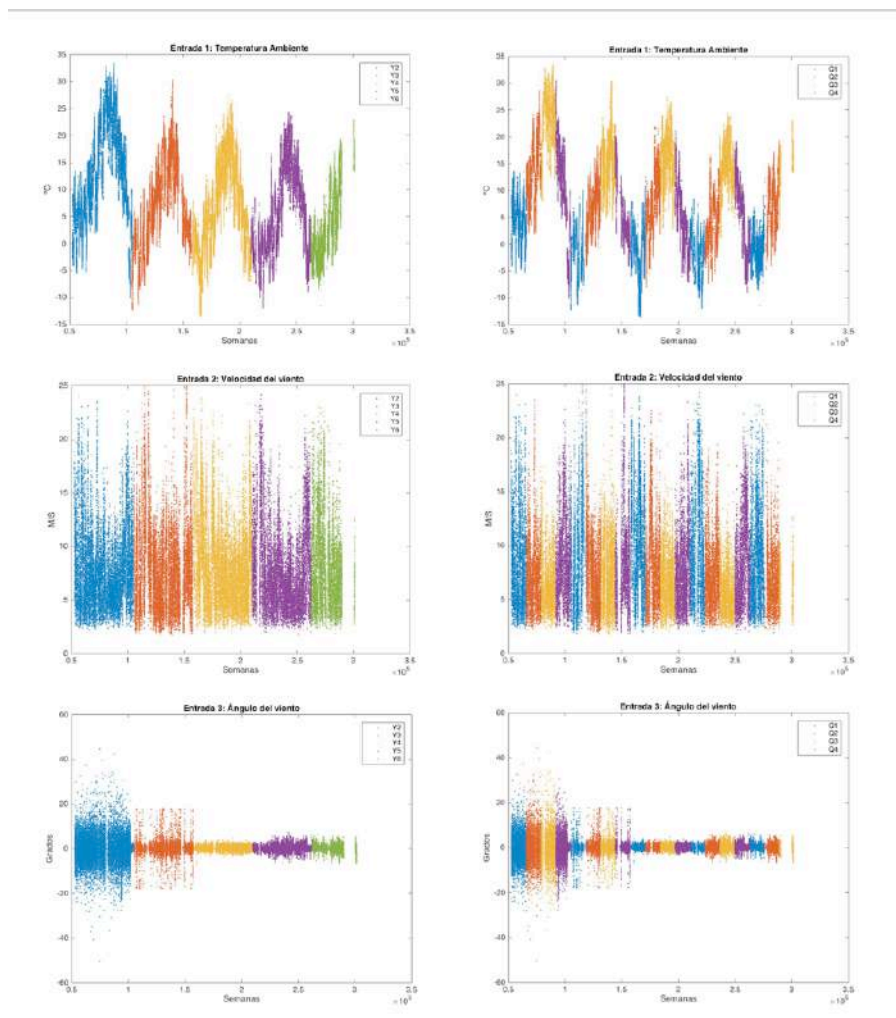


Figura 83. Variables de entrada divididas por años y estaciones

5.3 CURVAS DE ANOMALÍAS

En primer lugar, se muestran en la Figura 84 y la Figura 85 donde se encuentran las anomalías respecto del resto de puntos de cada curva.

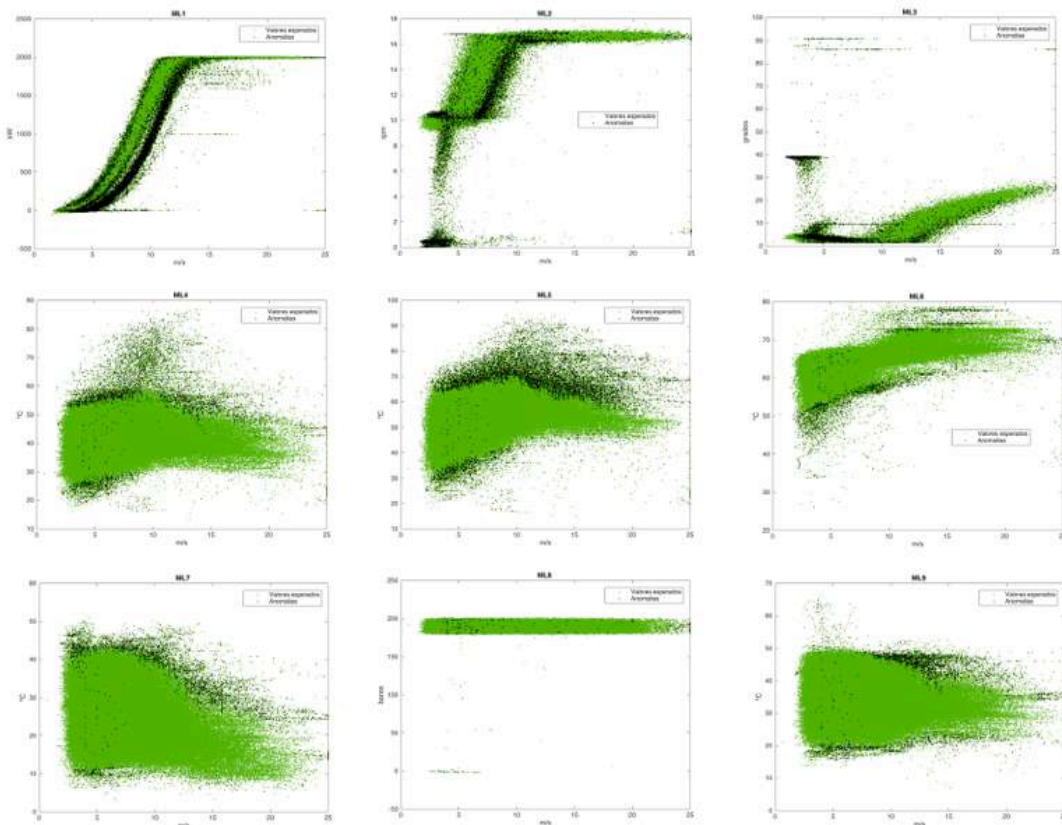


Figura 84. Mapa de puntos de anomalías en los ML

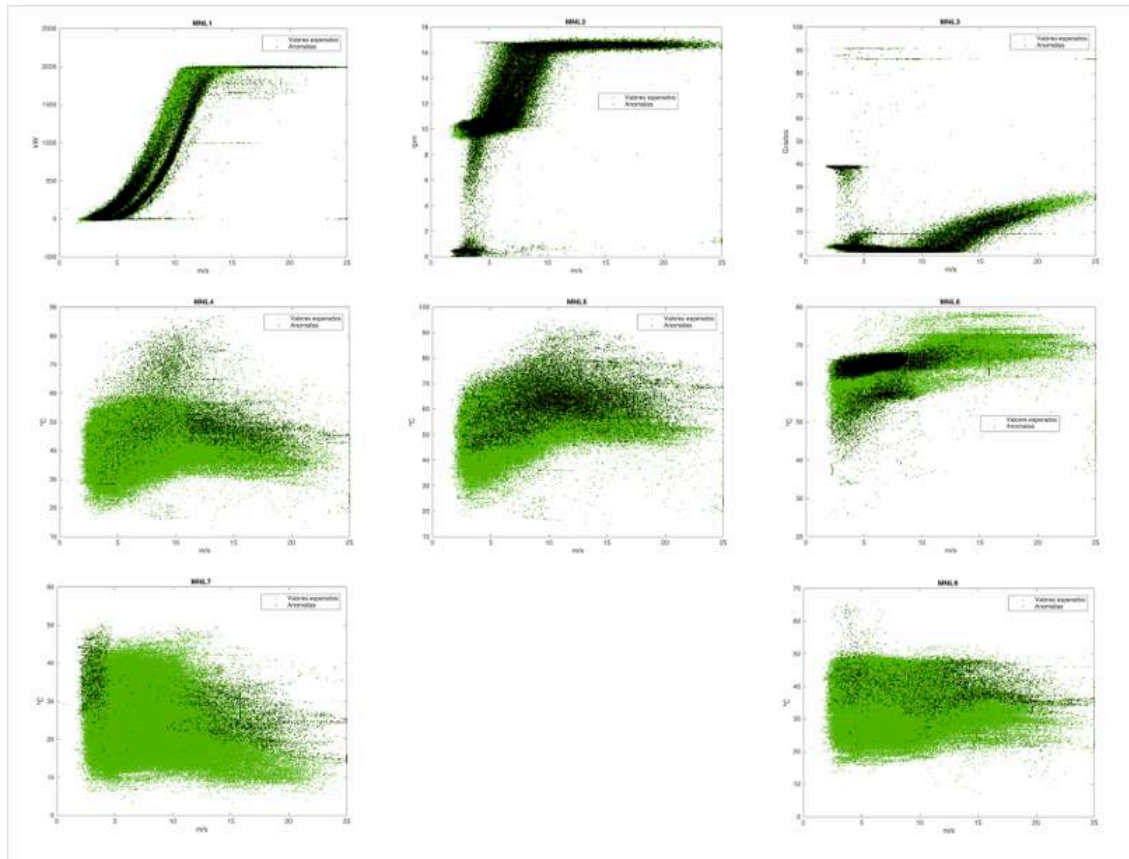


Figura 85. Mapa de puntos de anomalías en los MNL

En estas figuras se observa claramente la desventaja del modelo lineal respecto del no lineal. Mientras que en los ML la distribución de las anomalías se asemeja a una imagen en negativo de los intervalos de confianza del modelo, en los MNL se localizan anomalías entre los puntos que a priori, podría parecer que se encuentran dentro de dichas bandas. Estos puntos no tienen por qué ser normales, por diversos motivos pueden ser anómalos, pero los modelos lineales no los pueden detectar, por ello el número total de anomalías de los MNL es mayor al de los ML en cada curva de estudio.

En la curva de potencia (C1), ambos modelos detectan los puntos del Y3 como anomalías, se trata de la curva de potencia en negro que se ve desfasada hacia la derecha en ambas gráficas. Como se ha explicado, más del 80% de los datos correspondientes a ese año son valores anómalos.

Para poder analizar de forma gráfica las anomalías de cada modelo, se ha construido para cada modelo una curva que muestre el porcentaje de anomalías que han sucedido en un tiempo determinado durante el periodo de simulación.

Cada curva tiene una memoria de 336 anomalías, lo que equivale a 7 días de información. Desde que terminan las 336 primeras anomalías del Y2, la curva muestra que porcentaje de anomalías ha habido en los últimos 7 días de información.

Se observa que, durante el Y3, el porcentaje de anomalías de los modelos de las curvas C1 C2 y C3 aumenta notablemente y que el modelo lineal de la C8 en ningún momento supera el 15% de anomalías.

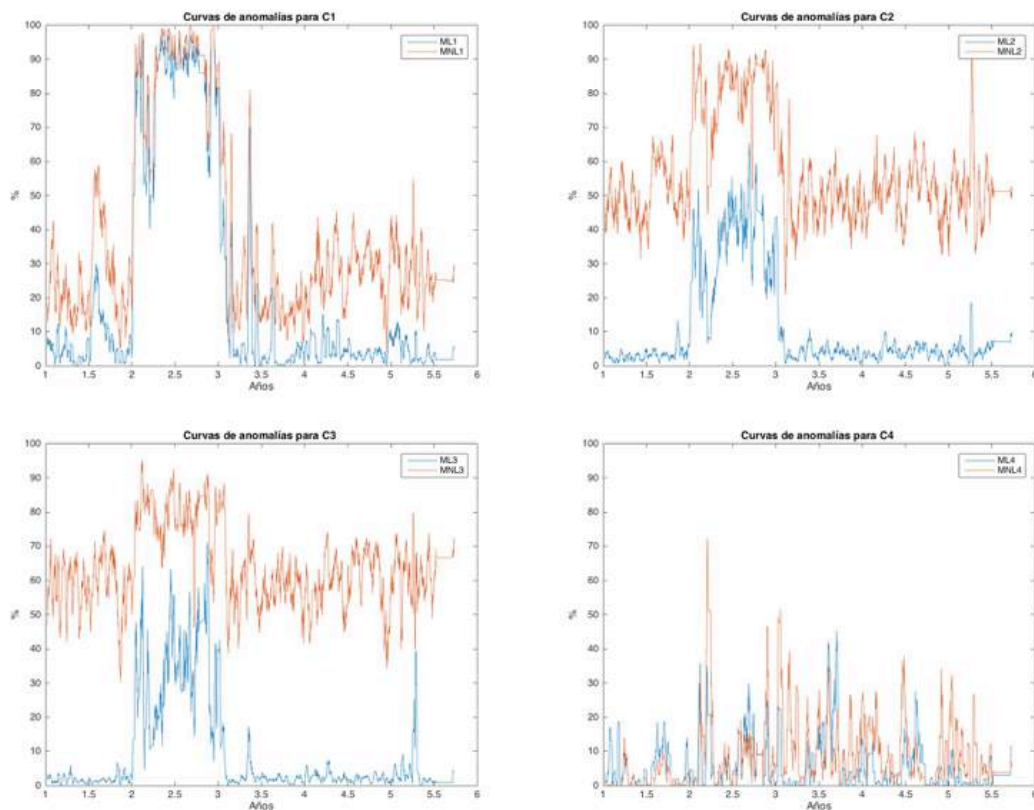


Figura 86. Curvas de anomalías de las curvas C1, C2, C3 y C4

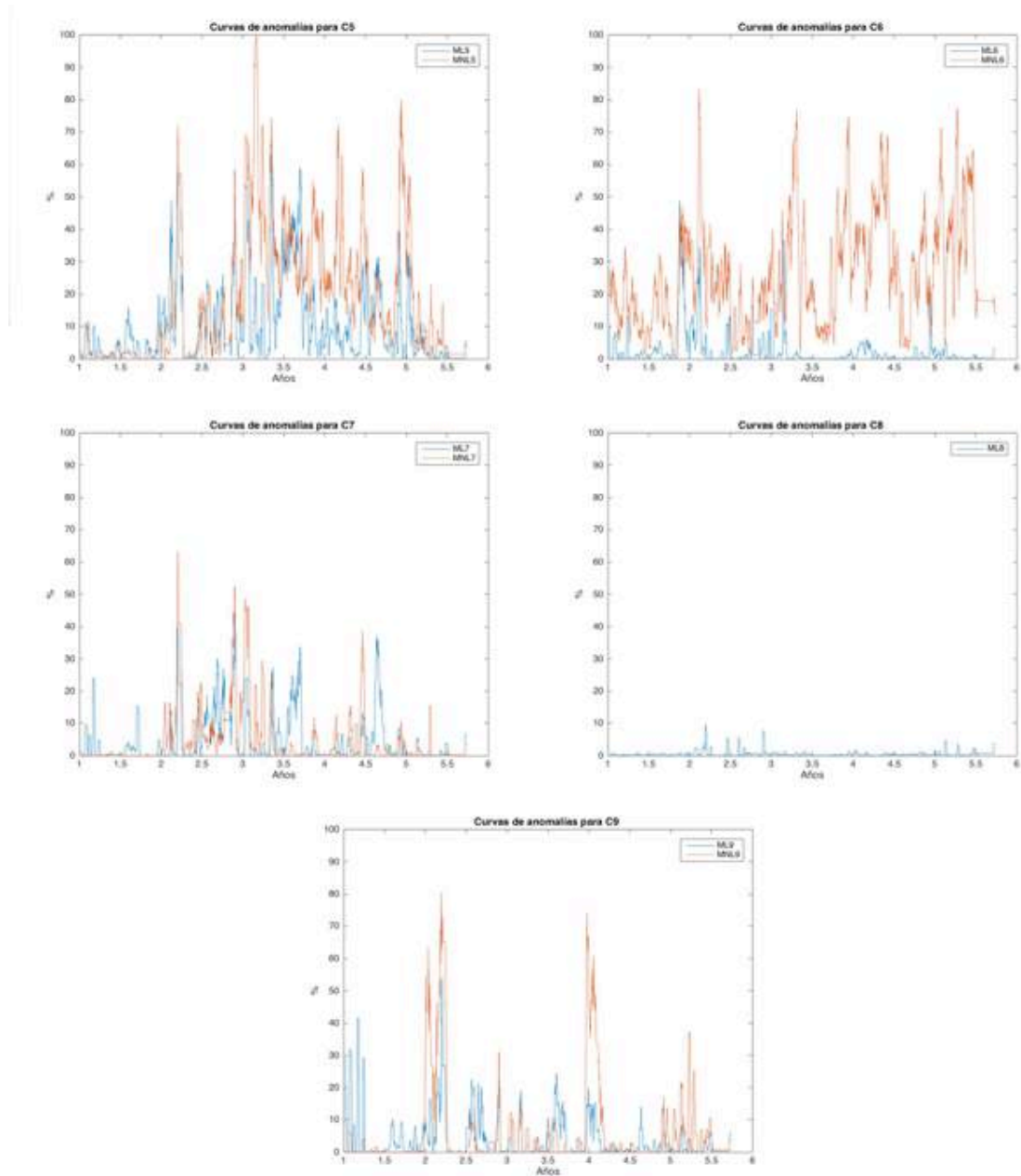


Figura 87. Curvas de anomalías de las curvas C5, C6, C7, C8 y C9

Conocidas las curvas de anomalías, sólo queda comprobar si hay una relación entre las reparaciones que se han hecho en el aerogenerador y los fallos en su comportamiento que han simulado los modelos de comportamiento

5.4 REPARACIONES

Durante el periodo de estudio se realizaron 163 reparaciones, de las cuales, 56 fueron correcciones o sustituciones de componentes. Estas reparaciones son las más interesantes, pues solo se dan si se ha detectado un fallo no previsto en el sistema. De las 56 reparaciones realizadas, 20 se efectuaron durante el Y1, por lo que están fuera del periodo de simulación y por lo tanto no se utilizan. Las 36 reparaciones restantes se recogen en la Tabla 27, ordenadas cronológicamente. Todas ellas tienen un impacto indirecto en todos los modelos, pero sólo algunas tienen un impacto directo en los componentes que se han simulado en los modelos de comportamiento. Se señalan las reparaciones que han tenido un impacto directo los modelos y se estudia cada caso en detenimiento, así como todas las reparaciones que sucedieron durante Y3, periodo de porcentaje de anomalías más elevado.

Nº	Fecha de inicio	Año	Duración	Descripción	Componente de interés afectado	Modelos
1	20/5/10 16:33	Y2	0:07:48	Plc		
2	25/5/10 9:07	Y2	2:00:00	Fusibles		
3	3/8/10 10:33	Y2	2:00:00	Armario top control sistema dtc g8x		
4	17/8/10 6:57	Y2	6:04:12	Ups002-ups opcional		
5	9/9/10 23:02	Y2	11:48:00	Ups001-cargador de baterías		
6	16/9/10 13:26	Y2	3:00:00	Fg008-estator del generador	Generador eléctrico	C5 C7
7	19/9/10 13:12	Y2	4:46:48	Escobilla descarga l.o.a.	Generador eléctrico	C5 C7
8	4/11/10 11:02	Y2	0:00:00	Tramo 1		
9	27/2/11 17:31	Y3	19:18:00	Km280-bba grupo hidráulico	G. Hidráulico	C8 C9
10	28/2/11 13:40	Y3	2:21:00	Módulo 6.9-salida analógicas 1 cab.6		
11	5/8/11 8:24	Y3	8:00:00	Baliza tipo a de media intensidad		
12	8/8/11 15:50	Y3	2:58:12	Acumulador 20 l (pos. 87)		
13	16/8/11 7:40	Y3	55:15:00	Rodamiento loa	Generador eléctrico	C5 C7
14	29/9/11 20:24	Y3	16:21:00	F15-fusible de precarga		
15	4/11/11 11:31	Y3	1:10:12	Anemómetro ultrasónico	Medida de velocidad del viento	Todos
16	16/11/11 9:07	Y3	1:50:59	Circuitería hidráulica sistema grupo hidráulico	G. Hidráulico	C8 C9
17	31/1/12 8:24	Y4	0:00:00	Alumbrado nacelle		
18	22/2/12 9:21	Y4	8:40:12	Rodamiento eje rápido-lado rotor	Multiplicadora	C4 C6

19	22/2/12 11:02	Y4	3:00:00	Microinterruptores final de carrera		
20	22/2/12 15:36	Y4	2:30:00	Buje		
21	23/4/12 9:36	Y4	1:34:48	Anemómetro ultrasónico	Medida velocidad viento	Todos
22	29/5/12 11:31	Y4	1:00:00	Ups002-ups opcional		
23	23/6/12 23:31	Y4	11:15:00	Escobilla descarga l.o.a.	Generador eléctrico	C5 C7
24	7/10/12 9:07	Y4	26:40:12	Válvula proporcional (pos. 83)		
25	16/10/12 5:31	Y4	5:35:59	Sensor de enrollamiento de cables		
26	16/10/12 11:02	Y4	1:55:12	Válvula proporcional (pos. 83)		
27	14/11/12 14:24	Y4	0:25:12	Pantalla táctil		
28	14/2/13 8:24	Y5	5:30:00	Batería 001		
29	14/4/13 20:24	Y5	6:43:48	Sonda pt100 aceite multiplicadora	Multiplicadora	C4 C6
30	4/6/13 1:12	Y5	9:58:12	Km280-bba grupo hidráulico	G. Hidráulico	C8 C9
31	19/6/13 13:26	Y5	0:06:00	Tt001-690/230-130v		
32	4/7/13 1:12	Y5	8:52:12	Módulo 6.5-de temperatura 8 cab.6		
33	7/7/13 4:48	Y5	8:30:00	Motor eléctrico (pos. 8)		
34	18/7/13 8:38	Y5	5:28:48	Filtros		
35	30/7/13 3:21	Y5	6:10:47	Ta001-f.a. 230vac/24vcc		
36	25/9/13 16:33	Y5	0:30:00	Rejillas de aireación		

Tabla 27. Listado de reparaciones tipo “correc” en el periodo de entrenamiento

En los gráficos Figura 88, Figura 89, Figura 90, Figura 91, Figura 92 las líneas verticales corresponden con el final de una reparación. El objetivo no es tanto una disminución inmediata del porcentaje de anomalías tras una reparación, sino un descenso notable en la pendiente de la curva de anomalías.

- Las reparaciones 9, 16 y 30 corresponden a componentes del sistema hidráulico. En la Figura 88 se observa como tras el final de las reparaciones 9 y 16 disminuye notablemente el porcentaje de anomalías en los 3 modelos, mientras que la reparación 30 parece no tener efecto en los modelos. Cabe destacar que en ML8 el porcentaje de anomalías es siempre muy bajo, lo que indica que los errores en el

comportamiento del grupo hidráulico no se ven reflejados en demasía en la variable de presión (C8).

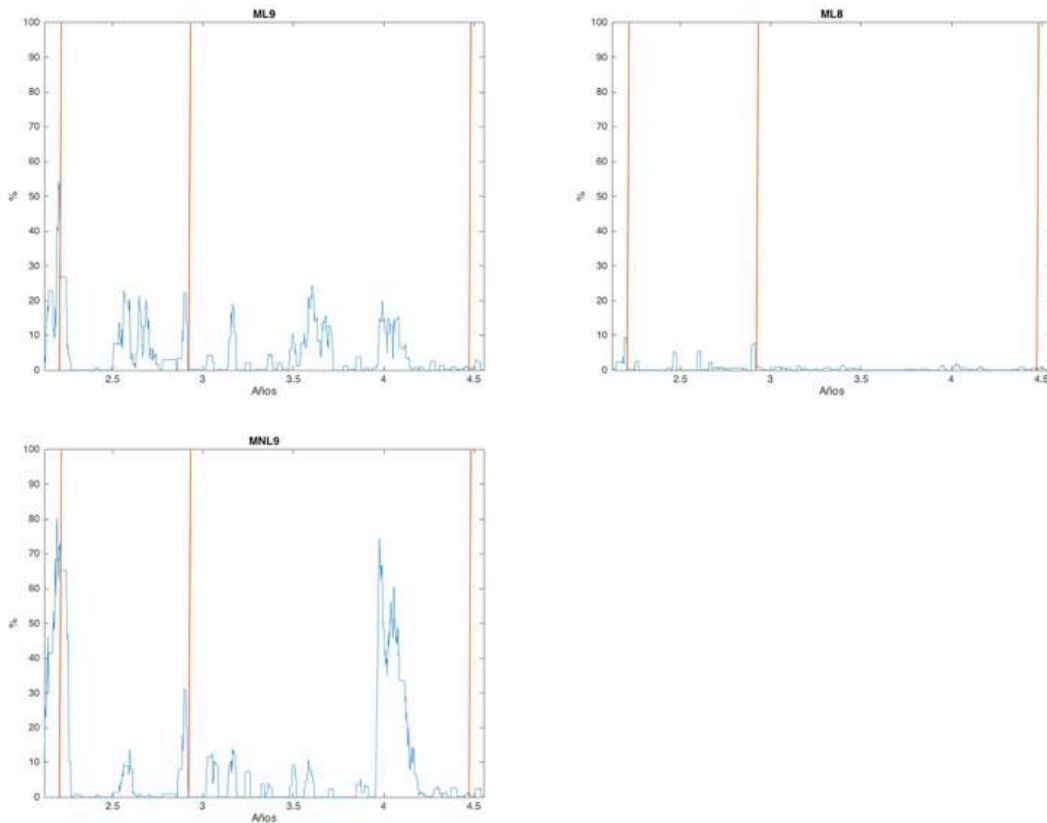


Figura 88. Curva de anomalías de C8 y C9 para reparaciones del grupo hidráulico

- Las reparaciones 6, 7, 13 y 23 corresponden a componentes del generador. Las reparaciones 6 y 7 suceden al mismo tiempo, por ello su marca prácticamente se superpone en la Figura 89. En este caso la relación entre la reparación y el descenso de anomalías no se ve tan claramente como en el caso anterior, pero se observa que en los modelos lineales la pendiente de la curva de anomalías desciende tras producirse cada reparación.

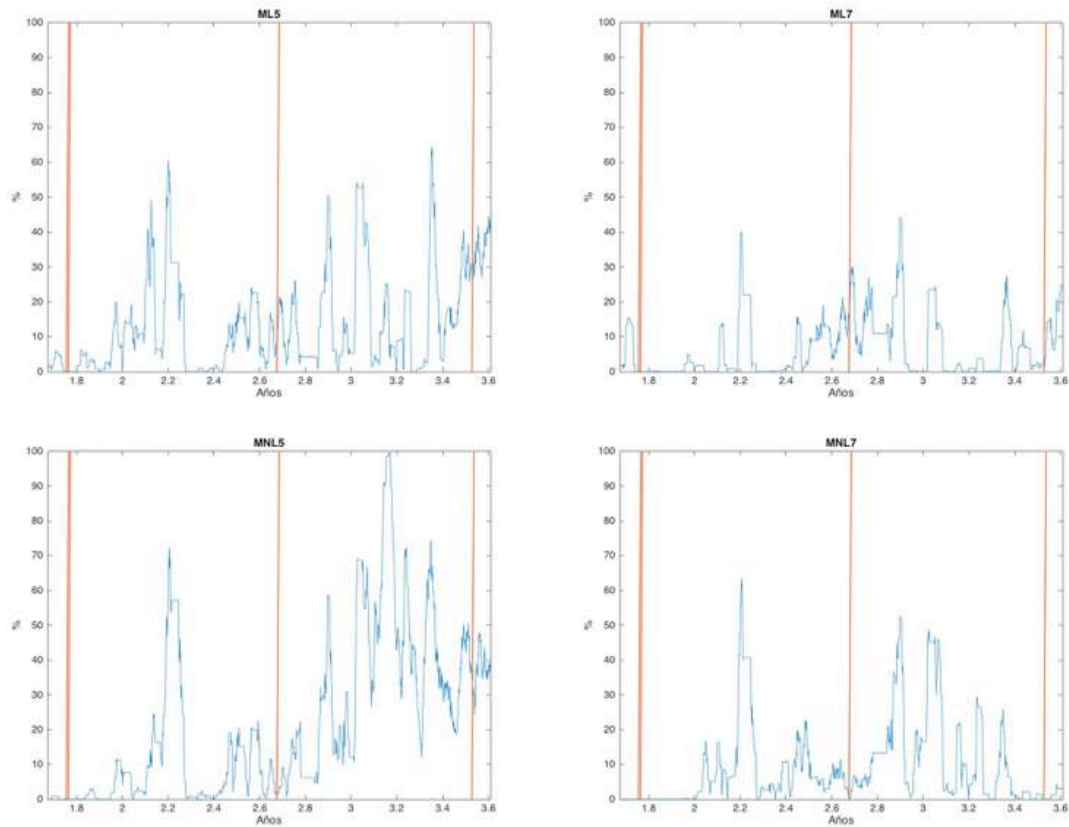


Figura 89. Curva de anomalías de C5 y C7 para reparaciones del generador

- Las 18 y 29 corresponden a componentes de la multiplicadora. En este caso, análogamente al caso anterior, son los modelos no lineales los que leen más acertadamente como mejora el rendimiento del componente tras las reparaciones.

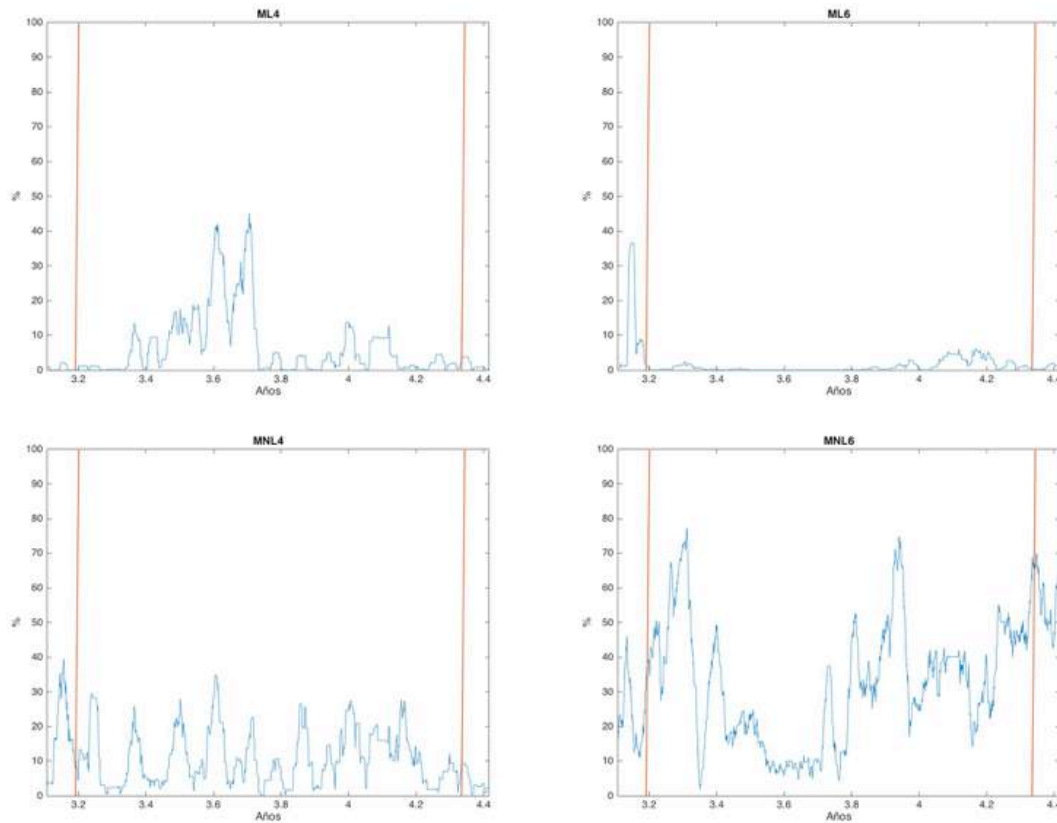


Figura 90. Curva de anomalías de C4 y C6 para reparaciones de la multiplicadora

- Las 15 y 21 corresponden muestran un fallo en la medida de la velocidad del viento. Esta es una variable con una relación directa con la creación de los 17 modelos, por lo que en todos se aprecia en mayor o menor medida cuando se arregla. La Figura 91 muestra los modelos en los que es más claro el descenso de la pendiente de anomalías.

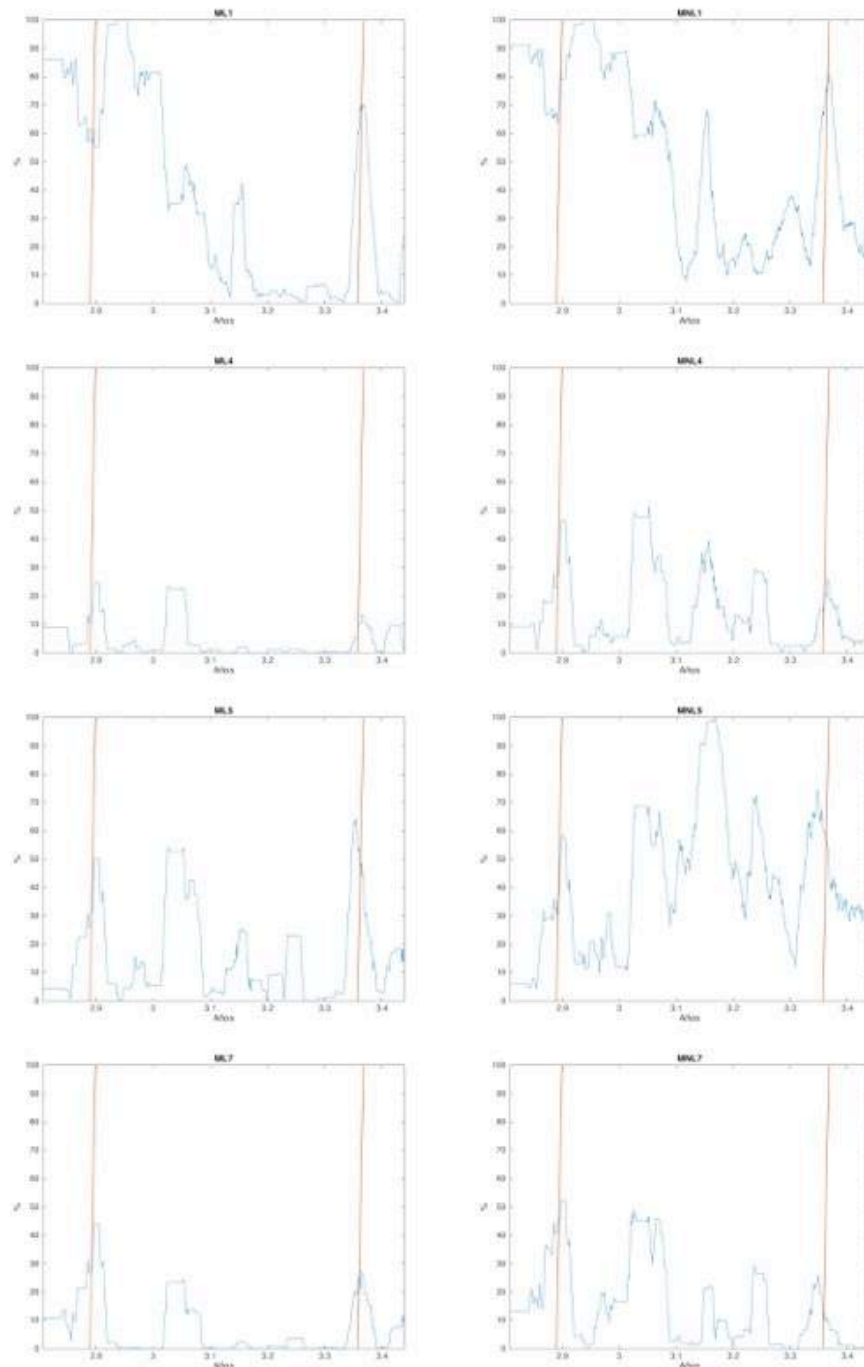


Figura 91. Curva de anomalías de C1, C4, C5 y C7 para reparaciones en el sistema de medida de la velocidad del viento

Finalmente se muestran las anomalías 8, 9, ..., 16 y 17; todas suceden durante Y3. Este es el periodo en el que peor es el rendimiento general de la instalación. Durante este periodo solo se producen reparaciones del tipo corrección o sustitución y de tipo rearme del sistema, que no tienen efecto en el rendimiento de la máquina. Por ello, estas son las reparaciones que hay que estudiar para comprender como se ha arreglado el aerogenerador. A excepción de la reparación 14 (fusible de precarga), que sucede en el año 2,8 (30 de septiembre de 2011), el resto de las reparaciones tienen un efecto notable en la curva de anomalías de los modelos de C1, C2, y C3, los más afectados durante este periodo. Puesto que el valor de regresión lineal de MNL3 era de un 16% durante el Y3 no se ha mostrado la curva de anomalías de ese modelo.

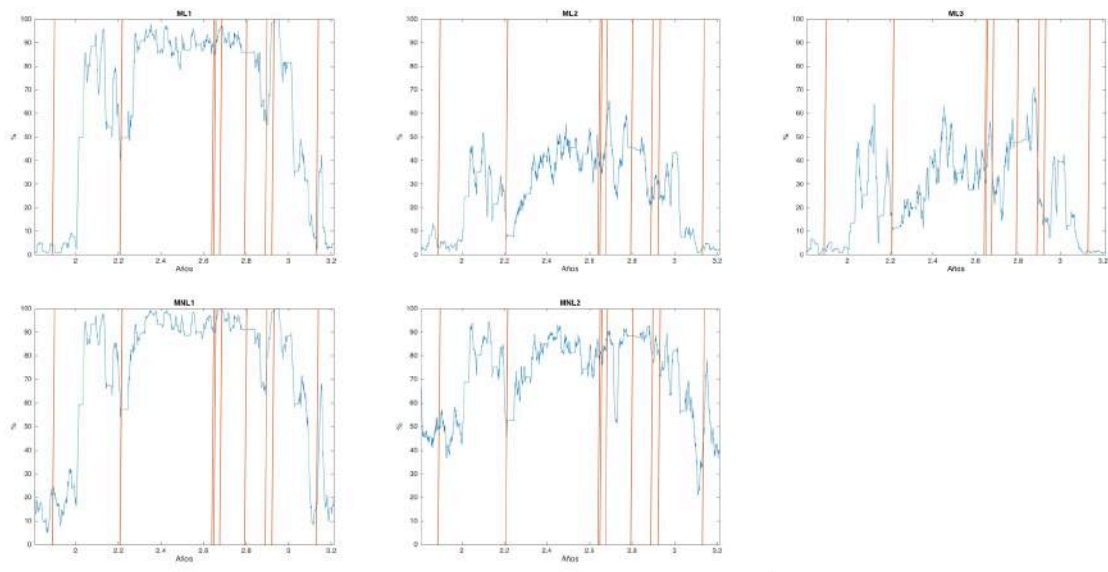


Figura 92. Curva de anomalías de C1, C2 y C3 para reparaciones durante Y3

Capítulo 6. RESULTADOS Y CONCLUSIONES

En este capítulo se concluye la validez del proyecto en base a los resultados que se han obtenido para cumplir los objetivos marcados en el apartado 1.2. Al final del mismo se dan las conclusiones generales que se han obtenido,

El primer objetivo consistía en importar las mediciones del SCADA del aerogenerador y seleccionar las variables con las que mejor se pudiese monitorizar el funcionamiento tanto del aerogenerador en general, como de algunos de sus componentes principales en particular. En el Capítulo 1. se ha expuesto la metodología seguida para importar las mediciones y se han analizado las posibles variables a utilizar. En el Capítulo 2. se expuso que se dispone de mediciones suficientes como para crear modelos de comportamiento para 11 variables (potencia generada y 10 variables intermedias), de las cuales se tomaron 9, únicamente descartando la var2 y var13 por los motivos allí citados.

El segundo objetivo consistía en construir modelos de comportamiento lineales (ML) y no lineales (MNL) de las curvas de estado creadas a raíz de las variables previamente seleccionadas. El Capítulo 3. Muestra las tareas previas a la creación de los modelos de comportamiento. En el Capítulo 4. se explica la metodología utilizada en cada caso para la construcción de los modelos de comportamiento, así como los resultados de cada modelo. Para los modelos lineales ha creado una media móvil de intervalo fijo, y para los modelos no lineales se han creado redes neuronales de tipo perceptrón multicapa (MLP). Se ha observado la precisión de cada modelo y se ha descartado aquel modelo cuyos resultados no fuesen satisfactorios, como en el caso del MNL8, cuya precisión es mínima y por ello no una vez creado se ha decidido no seguir trabajando con él. También se señaló que el MNL3 durante Y3 no opera de manera adecuada, pero no se debe a un problema de modelado, sino a que el comportamiento de C3 durante ese año es notablemente atípico, pues en el resto de años la regresión lineal de MNL3 es muy buena (superior al 92%).

El tercer objetivo se ha tratado en el Capítulo 5. Se trataba de buscar anomalías en el periodo de simulación y buscar relaciones con las reparaciones que se produzcan al mismo

tiempo. En el apartado 5.2 se expone lo que se ha considerado como una anomalía, y en el apartado 5.4 se muestra la relación entre las reparaciones que afectan directamente a cada componente principal estudiado y en consecuencia a cada curva de estudio, y la curva de anomalías de dicho modelo. Se ha demostrado que los todos modelos de las curvas de estado relacionadas con el grupo hidráulico (C8 y C9) predicen con exactitud cuándo se produjo una reparación de en dicho grupo. En el caso de la multiplicadora (C4 y C6), los modelos no lineales (MNL) muestran una mayor precisión, contrariamente a las modelos del generador eléctrico (C5 y C7), siendo el resultado de ambos menos claro de el de los modelos del grupo hidráulico. Las reparaciones del sistema de medición de la velocidad del viento afectan a todos los modelos, viéndose esto más claro que en los modelos de las curvas C1, C4, C5 y C7. Finalmente se observa la relación que hay entre reparaciones y los modelos de las curvas C1, C2 y C3 durante el segundo año del periodo de simulación (Y3). Se concluye que efectivamente se han creado modelos de comportamiento a partir de los cuales se han extraído anomalías en el funcionamiento de los componentes que cada uno representa y se ha comprobado que, en mayor o menor medida, habrían predicho cuando era necesario realizar una reparación no prevista. Sin embargo, hay mucho ruido en las curvas de anomalías de todos los modelos, por lo que el resultado final no es el ideal. Se espera que futuros desarrollos permitan mejorar este aspecto del proyecto.

Finalmente, la conclusión general es que se han abordado con éxito los 3 objetivos principales que se habían citado en el apartado 1.2, obteniendo los resultados deseados en los 2 primeros, y resultados aceptables para el tercer objetivo. Este objetivo se ha cumplido parcialmente debido a que con los métodos utilizados no se ha podido aislar completamente la variable de estudio de cada curva del resto de la instalación, por lo que al crear la curva de anomalías de cada modelo se encuentra una componente importante de ruido.

Capítulo 7. FUTUROS DESARROLLOS

Como se ha expuesto en el capítulo anterior, se ha cumplido completamente con los 2 primeros objetivos principales del proyecto, y parcialmente con el tercero. Por lo tanto, los futuros desarrollos relacionados con este proyecto deberían apuntar en la dirección de mejorar el sistema de captación de anomalías. A continuación, se nombran soluciones que se cree que podrían aportar unos mejores resultados:

- Utilizar otro tipo de algoritmos predictivos para crear los modelos lineales. Se valoran los modelos de análisis de varianza (ANOVA).
- Utilizando redes neuronales profundas (Deep Learning) para los modelos no lineales puede resultar en una mayor precisión, se recomienda el uso de redes neuronales convolucionales (CNN).
- Utilizar algoritmos de aprendizaje de refuerzo en las redes neuronales como Deep Q-Network o Advantage Actor Critic (A2C).
- Utilizar técnicas de Clustering para analizar las anomalías que se han extraído de los modelos para establecer una diferenciación entre las que se generan directamente por fallos en el funcionamiento del componente al que hace referencia el modelo de comportamiento y las que se dan por fallos de otros componentes.
- Utilizar un sistema híbrido con una combinación de SOM y MLP para mejorar la precisión de este último.

También se valoraría positivamente el estudio de otras variables para modelar otros componentes. En el caso de este proyecto no se disponían de medidas suficientes de variables relacionadas directamente con el transformador o el YAW.

Capítulo 8. BIBLIOGRAFÍA

- [1] Peyman Mazidi, Student Member, IEEE, Lina Bertling Tjernberg, Senior Member, IEEE, and Miguel A. Sanz-Bobi, Senior Member, IEEE (2017). *Performance Analysis and Anomaly Detection in Wind Turbines based on Neural Networks and Principal Component Analysis*. Conference: 12th Workshop on Industrial Systems and Energy Technologies in Madrid, Spain (May 2017).
- [2] Mian Du, Lina Bertling Tjernberg, Shicong Ma, Qing He¹, Lin Cheng and Jianbo Guo (2016). *A SOM based Anomaly Detection Method for Wind Turbines Health Management through SCADA Data*. Article in the Energies Journal, published on February 20, 2017. DOI: 10.3390/en10020253.
- [3] Nassim Laouti, Nida Sheibat-Othman, Sami Othman (2011). *Support Vector Machines for Fault Detection in Wind Turbines*. Conference: 18th IFAC World Congress in Milano, Italy (August 28 – September 2, 2011).
- [4] Peyman Mazidi, Mian Du, Lina Bertling Tjernberg, and Miguel A. Sanz-Bobi (2017). *A Health condition model for wind turbine monitoring through neural networks and proportional hazard models*. Article in the SAGE Journal, published on May 4, 2011. DOI: 10.1177/1748006X17707902.
- [5] Yonglong Yan, Jian Li, and David Wenzhong Gao (2014). *Condition Parameter Modeling for Anomaly Detection in Wind Turbines*. Article in the Energies Journal, published on May 6, 2014. DOI: 10.3390/en7053104.
- [6] Meng Li ¹, Shuangxin Wang ¹, Shanxiang Fang ^{1,2} and Juchao Zhao (February 12, 2020). *Anomaly Detection of Wind Turbines Based on Deep Small-World Neural Network*. Article in the Applied Sciences Journal, published on February 12, 2020. DOI: 10.3390/app10041243

Parte II ESTUDIO

ECONÓMICO

Capítulo 1. PRESUPUESTO

A continuación, se desglosan los costes asociados a la realización del proyecto con objeto de poder estudiar la rentabilidad y viabilidad del mismo, teniendo en cuenta los resultados que se han obtenido.

1.1 MEDICIONES

En este apartado se recogen las horas dedicadas a cada parte del proyecto y los componentes utilizados.

1.1.1 EQUIPO

Elemento	Cantidad	Horas de proyecto
Ordenador	1	480

Tabla 28. Mediciones de equipo

1.1.2 SOFTWARE Y LICENCIAS

Elemento	Cantidad	Horas de proyecto
Matlab 2015b	1	400
Statistics and Machine Learning Toolbox	1	360
Curve Fitting toolbox	1	20

Tabla 29. Mediciones de softwares y licencias

1.1.3 MANO DE OBRA

Elemento	Horas
Estudio de la tecnología	80
Diseño de modelos de comportamiento	100
Programación	200
Comprobación de resultados	60
Documentación	120
Total	560

Tabla 30. Mediciones de mano de obra directa

1.2 PRECIOS UNITARIOS

Es este apartado se muestran los precios de los elementos enumerados en el apartado anterior.

1.2.1 EQUIPO

Elemento	Cantidad	Precio (€/ud.)
Ordenador	1	1200

Tabla 31. Precios unitarios de equipos

1.2.2 SOFTWARE Y LICENCIAS

Elemento	Cantidad	Precio (€/ud.)
Matlab 2015b	1	119
Statistics and Machine Learning Toolbox	1	35
Curve Fitting toolbox	1	35

Tabla 32. Precios unitarios de softwares y licencias

1.2.3 MANO DE OBRA

Elemento	Precio (€/hora)
Estudio de la tecnología	30
Diseño de modelos de comportamiento	44
Programación	44
Comprobación de resultados	40
Documentación	35

Tabla 33. Precios unitarios de mano de obra directa

1.3 SUMAS PARCIALES

En este apartado se muestran los importes totales que cuesta cada componente y cada parte del proyecto.

1.3.1 EQUIPO

Elemento	Cantidad	Precio (€/ud.)	Precio total (€)
Ordenador	1	1200	1.200,00
Total			1.200,00

Tabla 34. Suma parcial de equipos

1.3.2 SOFTWARE Y LICENCIAS

Elemento	Cantidad	Precio (€/ud.)	Precio total (€)
Matlab 2015b	1	119	119
Statistics and Machine Learning Toolbox	1	35	35
Curve Fitting toolbox	1	35	35
Total			189

Tabla 35. Suma parcial de softwares y licencias

1.3.3 MANO DE OBRA

Elemento	Horas	Precio (€/hora)	Precio Total (€)
Estudio de la tecnología	80	30	2.400,00
Diseño de modelos de comportamiento	100	44	4.400,00
Programación	200	44	8.800,00
Comprobación de resultados	60	40	2.400,00
Documentación	120	35	4.200,00
Total			22.200,00

Tabla 36. Suma parcial de mano de obra directa

1.4 PRESUPUESTO GENERAL

Concepto	Coste (€)
Equipo	1.200,00
Software y licencias	189
Mano de obra	22.200,00
Total	23.589,00

Tabla 37. Presupuesto general

Parte III APÉNDICE I

Capítulo 1. LISTA DE VARIABLES

La siguiente tabla recoge las variables originales del aerogenerador y la numeración que se les ha dado durante la programación.

Nombre	Parámetro representado	Descripción	Rango 1%-99%	Udm
var1	ANG.aero.tem_aceite_multi	Temperatura del aceite de la multiplicadora	38,90 – 72,96	°C
var2	ANG.aero.tem_rod_alta_multi	Temperatura de los rodamientos de la multiplicadora en el lado de generación	33,18 – 73,94	°C
var3	ANG.aero.tem_rod_DE_gen	Temperatura de los rodamientos DE en el lado de generación	21,79 – 63,22	°C
var4	ANG.aero.tem_rod_NDE_gen	Temperatura de los rodamientos NDE en el lado de generación	25,10 – 76,23	°C
var5	ANG.aero.tem_anillos_sld_gen	Temperatura de los anillos sld en el lado de generación	9,332 – 45,34	°C
var6	ANG.aero.tem_max_dev_gen	Temperatura máxima de los devanados en el lado de generación	31,27 – 102,8	°C
var7	ANG.aero.tem_dev_A_gen	Temperatura máxima del devanado A en el lado de generación	32,83 – 110,5	°C
var8	ANG.aero.tem_dev_B_gen	Temperatura máxima del devanado B en el lado de generación	32,59 – 111,5	°C
var9	ANG.aero.tem_dev_C_gen	Temperatura máxima del devanado C en el lado de generación	33,02 – 111,8	°C
var10	ANG.aero.tem_grupo_h	Temperatura máxima del grupo electrógeno	18,20 – 48,61	°C
var11	ANG.aero.pres_grupo_h	Presión máxima del grupo electrógeno	180,6 – 199,3	Bar
var12	ANG.aero.pres_yaw	Presión del mecanismo de orientación	14,24 – 202,9	Bar
var13	ANG.aero.tem_interior_nac	Temperatura del interior de la góndola	1,966 – 36,88	°C
var14	ANG.aero.tem_ambiente	Temperatura ambiental	-8,02 – 29,21	°C
var15	ANG.aero.tem_max_trafo	Temperatura máxima del transformador	45,53 – 90,87	°C
var16	ANG.aero.tem_dev_A_trafo	Temperatura máxima del devanado A del transformador	46,62 – 95,55	°C
var17	ANG.aero.tem_dev_B_trafo	Temperatura máxima del devanado B del transformador	49,97 – 95,46	°C
var18	ANG.aero.tem_dev_C_trafo	Temperatura máxima del devanado C del transformador	46,97 – 93,76	°C
var19	ANG.aero.W_eje_principal	Giro del rotor	0,120 – 16,80	Rpm
var20	ANG.aero.w_gen	Giro del eje de generación	-12,5 – 1696	Rpm
var21	ANG.aero.angulo_palas	Ángulo de pitch (Ángulo de las palas respecto a la vertical)	1,509 – 90,75	°
var22	ANG.aero.vel_viento/avg	Velocidad del viento	0,738 – 19,62	m/s

var23	ANG.aero.angulo_viento	Ángulo de incidencia del viento	-33,6 – 45,96	°
var24	ANG.aero.angulo_yaw	Ángulo del mecanismo de orientación	2,000 – 358,7	°
var25	ANG.aero.V_red	Tensión de salida de red	670,8 – 714,0	V
var26	ANG.aero.pot_reactiva	Potencia reactiva	-253 – 602,4	KVA
var27	ANG.aero.factor_potencia	Factor de potencia	-1,00 – 1,000	
var28	ANG.aero.pot_act_est	Potencia activa del estator	-5,18 – 1869	KW
var29	ANG.aero.pot_act_rot	Potencia activa del rotor	-116 – 179,0	KW
var30	ANG.aero.pot_total/avg	Potencia generada total	-15,9 – 2001	KW
var31	ANG.aero.Energia	Energía generada	3,722 – 32,49	MWh
var32	ANG.aero.state	Estado	0,000 – 100,0	---

Tabla 38. Variables del aerogenerador

La siguiente tabla recoge los principales vectores y matrices de fechas que se han utilizado durante la programación.

fecha	291312x1	Fecha original de mediciones
fecha_on	216949x1	Fechas de mediciones para sistema en operación
fecha_anom	178308x1	Fechas de mediciones para sistema en operación y durante el periodo de simulación
registro	56813x4	Fechas de instantes de 30 minutos consecutivos susceptibles de ser anomalías. Primera columna para ordenandos (1,2,3...56813) y columnas 2 a 4 con las fechas pertinentes

Tabla 39. Variables temporales

En la siguiente tabla se encuentran las variables creadas relativas a los listados de alarmas y reparaciones.

C_rep	153x6	Matriz original de reparaciones con los eventos que suceden durante el periodo de estudio. Columnas 1 y 2 para fechas (inicio y final) de reparaciones en formato datenum. Columnas 3 a 6 para detalles de la reparación
C_alm	1407x4	Matriz original de alarmas con los eventos que suceden durante el periodo de estudio. Columnas 1 y 2 para detalles de la alarmas y columnas 3 y 4 para fechas (inicio y final) de alarmas en formato datenum
T_rep	153x2	Fechas de reparaciones en formato de programación. Primera columna para el inicio y segunda columna para el fin
T_alm	1407x2	Fechas de alarmas en formato de programación. Primera columna para el inicio y segunda columna para el fin

Tabla 40. Variables de alarmas y reparaciones

Capítulo 2. NOMENCLATURA

En la siguiente tabla se recogen algunos de los acrónimos más utilizados en la memoria.

Acrónimo	Nombre
MLP	Perceptrón multicapa
SCADA	Software de Supervisión, Control y adquisición de Datos
MSE	Error cuadrático medio
E/V/T	Entrenamiento-Validación-Test
LM	Algoritmo de entrenamiento Levenberg-Marquardt
BR	Algoritmo de entrenamiento Regularización Bayesiana

Tabla 41. Acrónimos

La siguiente tabla muestra los símbolos que se han nombrado en este documento.

Símbolo	Nombre
$v_n, n=1, 2, 3, \dots, 32$	Variable
$Y_n, n=1, 2, 3, \dots, 6$	Año del periodo de estudio
$ML_n, n=1, 2, 3, \dots, 9$	Modelo lineal
$MNL_n, n=1, 2, 3, \dots, 9$	Modelo no lineal
$C_n, n=1, 2, 3, \dots, 9$	Curva de estado
$Pred_n, n=1, 2, 3, \dots, 9$	Predicción del modelo no lineal correspondiente
$Q_n, n=1, 2, 3, 4$	Aproximadamente estación del año (de invierno a otoño)

Tabla 42. Símbolos

Parte IV CÓDIGO FUENTE

Capítulo 1. DESCRIPCIÓN DE LA PROGRAMACIÓN

La programación se ha realizado íntegramente en Matlab y se ha dividido en 5 etapas, tal y como se muestra en la Figura 93. Cada etapa cuenta con un script principal en el que se ejecutan los comandos oportunos, y se ha guardado un conjunto de datos correspondiente a las variables del programa tras ejecutar cada uno de los scripts principales. Se les ha dado el nombre de ws_n , referente a workspace, donde “ n ” es el número de la etapa correspondiente. La ventaja que supone tener estos conjuntos de datos radica en que no hace falta volver a ejecutar un script entero para centrarse en un apartado en concreto del código, sino que directamente se carga el workspace correspondiente.

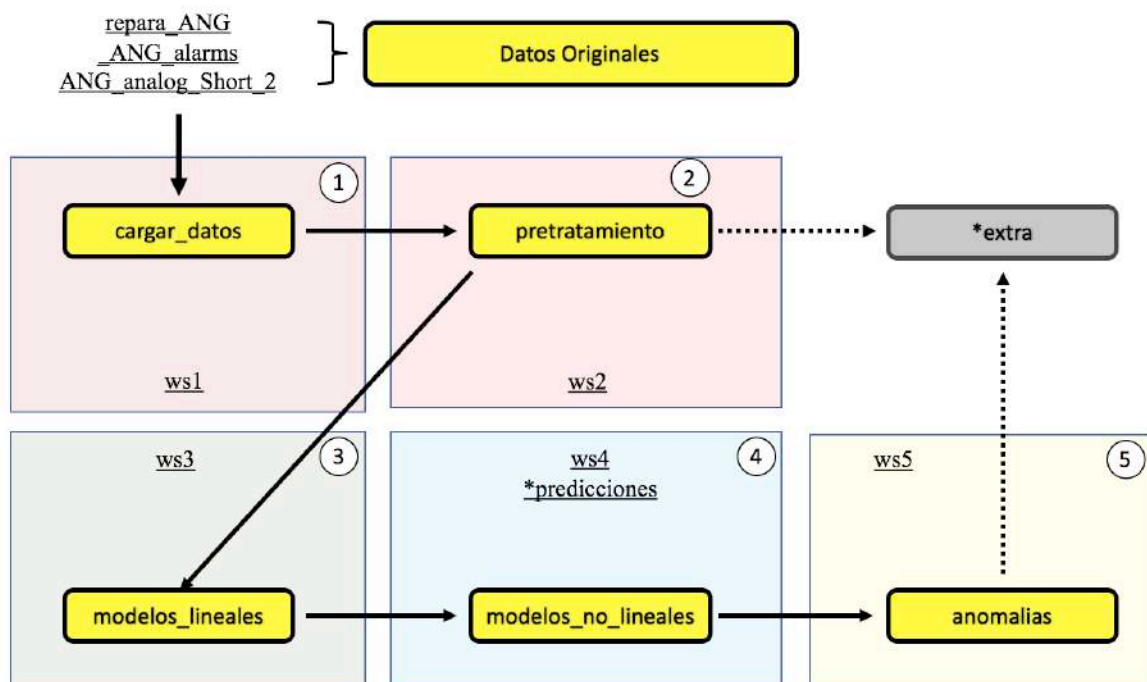


Figura 93. Esquema de programación

Además de los scripts principales, se ha contado con otro script llamado extra, donde se realizan acciones necesarias para la redacción de la memoria y el desarrollo del proyecto. Dado que estas acciones corresponden a diferentes etapas (2 y 5), se utilizan los conjuntos de datos correspondientes (ws2 y ws5) para cargar las variables en el punto adecuado a cada operación.

1.1 ETAPA 1

Se toman los datos provenientes de las matrices originales correspondientes a variables, fecha, reparaciones y alarmas, y se copian como vectores individuales de longitud igual a la amplitud del conjunto original de datos (291312 valores). En el caso de las alarmas y las reparaciones directamente se eliminan las que suceden fuera del periodo de datos disponibles, pues había 10 reparaciones anteriores al instante 1 del periodo de estudio (15 de diciembre de 2008 00:00) y 7 alarmas posteriores al instante 291312 o instante final del periodo de estudio (7 de septiembre de 2014 23:50). En el Apéndice I se encuentra las tablas que recoge la información referente a cada variable tomada de los datos originales.

Una vez finalizado y corregido el código, se ha guardado el espacio de trabajo como ws1.mat.

1.2 ETAPA 2

Se modifican las variables creadas en el script anterior para ajustarlas a los requisitos del proyecto.

En primer lugar, se descartan los datos para los que el aerogenerador se encuentra fuera de operación, pasando así de 291312 valores a 216949 valores por vector, y se crea un nuevo vector fecha_on con la nueva amplitud de las variables de estudio, pero con las fechas originales. Se crea la Figura 14 con todas las variables implicadas en la creación de los modelos antes de realizar el siguiente paso.

Se corrige el rango y el offset negativo que se ha encontrado en la variable velocidad del viento (var22) tal y como se ha explicado en el Parte ICapítulo 3. Se crean los gráficos correspondientes a las curvas de estado de cada año de estudio.

En esta etapa se han utilizado 2 de las funciones creadas para el programa, a saber:

- NaNoff: con esta función se eliminan los valores en los que el aerogenerador esta fuera de servicio, y se sustituyen los valores NaN. Estos valores han de ser sustituidos porque imposibilitan el trabajo con redes neuronales.
- rango: con esta función se corrige el rango de la variable deseada. Se ha utilizado para limitar el estudio a unos márgenes de velocidad de viento de entre 0 y 25 m/s.

Una vez finalizado y corregido el código, se ha guardado el espacio de trabajo como ws2.mat.

En esta etapa se hace uso del script secundario extra para realizar 2 acciones. En primer lugar, se carga ws2.mat para crear los histogramas correspondientes a la Figura 8 y la Figura 9 con los valores absolutos de velocidad de viento y potencia anuales. En segundo lugar, se muestra el resultado de la construcción de los modelos lineales de la curva de potencia (ML1) como los creados en la etapa 3 para diferentes anchos de intervalo. Para ello se carga previamente ws2.mat y se hace uso de la función mod_lin1 que se explica en el Parte ICapítulo 3.

1.3 ETAPA 3

Se construyen los modelos lineales de las curvas de estado (ML).

Primeramente, se quitan los valores más irregulares del mapa de puntos de cada una de las 9 curvas de estado para que no afecten negativamente al modelo lineal. Después se construyen los modelos lineales de un intervalo de 0,1 m/s (250 puntos) y se extienden a la amplitud de los vectores de variables de estado en operación (216949 valores). Finalmente

se extrae el error de entrenamiento y se corrigen los puntos fuera de las bandas de confianza con el error medio de cada modelo en el periodo de entrenamiento.

Una vez finalizado y corregido el código, se ha guardado el espacio de trabajo como ws2.mat.

Las funciones propias utilizadas en este script son:

- `limpia`: sustituye los puntos que se encuentran en unas coordenadas determinadas por NaN.
- `mod_lin1`: crea los modelos lineales a partir de las variables, el ancho de bandas de confianza y el ancho de intervalo deseados.
- `mod_lin2`: extiende el modelo creado en `mod_lin1` del número de divisiones del rango de la velocidad del viento (25 m/s) entre el intervalo deseado (0,1 m/s en todos los casos) al número total de valores de los vectores de estudio. Para ello analiza la coordenada de velocidad de viento de cada instante y aplica el valor del modelo correspondiente.

1.4 ETAPA 4

Se construyen los modelos no lineales de las curvas de estado (MNL).

Se crean 10 configuraciones de MLP para cada curva variando el número de neuronas por capa oculta y el algoritmo de entrenamiento. Se selecciona la configuración que mejor resultado aporta, basándose en que cuanto menor sea el MSE y que su duración sea menor de 10 segundos (se ha observado que las configuraciones que tardan más de 10 segundos funcionan ligeramente mejor en el periodo de referencia, pero notablemente peor en el periodo de simulación).

Se han guardado los resultados de las configuraciones anteriores en el fichero predicciones.mat, pues el resultado de las redes neuronales es aleatorio y si se tuviesen que

volver a ejecutar los resultados que darían no serían los mismos que los que se utilizan más adelante. Por ello se deja comentada la parte en la que se construyen las diferentes configuraciones de redes neuronales y se carga el fichero directamente.

A continuación, se extraen los datos pertinentes de cada una y se obtienen las bandas de confianza de los modelos. Nuevamente se extrae el error medio del modelo y se suma a las bandas de confianza.

Una vez finalizado y corregido el código, se ha guardado el espacio de trabajo como ws4.mat.

Las funciones propias utilizadas en este script son:

- MLP: para crear el MLP con la configuración y las entradas deseadas.
- mod_lin1: utilizada en la etapa 3. En esta etapa se usa para crear las bandas de confianza de 250 intervalos de los MNL.
- mod_lin2: utilizada en la etapa 3. Arrastra las bandas de confianza generadas en mod_lin1 a todos los puntos de las variables de estado (de 250 puntos de 216949 puntos).

1.5 ETAPA 5

Se extraen anomalías a partir del error de simulación de los ML y MNL en el periodo de simulación (Y2 a Y6) y se comprueba la validez de las mismas observando si anticipan una reparación importante.

Se obtiene el error del periodo de simulación (valor absoluto de la diferencia en tre predicción y valor real) de cada modelo, se agrupa en espacios de 30 minutos (3 valores consecutivos) y se determina en cada modelo en cuales se produce una anomalía (2 de 3 o 3 de 3 valores fuera de las bandas de confianza).

Finalmente se comprueba que tras una reparación de carácter de sustitución o corrección de componentes disminuye el número de anomalías.

Una vez finalizado y corregido el código, se ha guardado el espacio de trabajo como ws5.mat.

Las funciones propias utilizadas en este script son:

- `curvas_anom`: devuelve las curvas de anomalías (porcentaje de anomalías en el rango de tiempo previo deseado) de cada modelo a lo largo del periodo de simulación (Y2 a Y6).
- `rep_anom`: devuelve las curvas de anomalías de cada modelo y el instante en el que terminan las reparaciones que le sean señaladas. En el script principal se nombran las reparaciones más significativas para el estudio de los modelos.

En esta etapa se hace uso del script secundario extra para realizar 2 acciones. Para cada una de ellas se carga previamente el ws4.mat. Estas acciones consisten en 2 funciones que devuelven gráficas necesarias para la comprensión del trabajo realizado. Estas funciones son:

- `mostrar_entradas`: devuelve las entradas a los modelos no lineales (`var14`, `var22`, y `var23`) divididas por años y estaciones.
- `anom_mapa`: devuelve para cada modelo, la nube de puntos de la máquina en operación, y superpone los puntos de anomalía.

Capítulo 2. LISTADO DE ARCHIVOS

- Scripts (Cuerpo de la programación)

- cargar_datos.m
- pretratamiento.m
- modelos_lineales.m
- modelos_no_lineales.m
- anomalías.m
- extra.m

- Funciones

- NaNoff.m
- rango.m
- limpia.m
- mod_lin1.m
- mod_lin2.m
- MLP.m
- mostrar_entradas.m
- rep_anom.m
- anom_mapa.m
- curvas_anom.m

- Conjuntos de datos
 - Originales
 - _ANG_alarms.mat
 - ANG_analog_short_2.mat
 - repara_ANG.mat
 - Propios
 - ws1.mat
 - ws2.mat
 - ws3.mat
 - ws4.mat
 - ws5.mat
 - predicciones.mat

Capítulo 3. ENLACE A ARCHIVOS

https://drive.google.com/drive/folders/1wkRYktzRy9Md_A3eT9AzrYdA26TWapU2?usp=sharing