COMILLAS
UNIVERSIDAD PONTIFICIA
ICAI

MÁSTER UNIVERSITARIO EN INGENIERÍA INDUSTRIAL

TRABAJO FIN DE MÁSTER

EXPLORACIÓN DE TÉCNICAS DE REDES NEURONALES DE CONVOLUCIÓN EN LA DETECCIÓN DE ANOMALÍAS DE UN AEROGENERADOR

Autor: Mª Teresa Lallana Saldaña

Director: Miguel Angel Sanz Bobi

Madrid

Agosto de 2020

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título

"Exploración de técnicas de redes neuronales de convolución en la detección de anomalías de un aerogenerador"

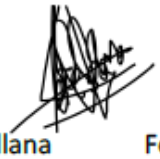
en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el

curso académico 2020 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

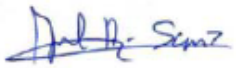
de otros documentos está debidamente referenciada.



Fdo.: Maria Teresa Lallana Fecha: 22/ 08/ 2020

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO



Fdo.: Miguel Ángel Sanz Bobi Fecha: 22 / 08 /2020

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

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

El autor Dña. Maria Teresa Lallana Saldaña

DECLARA ser el titular de los derechos de propiedad intelectual de la obra: "Exploración de técnicas de redes neuronales de convolución en la detección de anomalías de un aerogenerador", 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) Reproducirla 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 exclusive 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 …….. de ……... de …….

ACEPTA

Fdo……

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

Universidad Pontificia de Comillas
Escuela Técnica Superior de Ingeniería (ICAI)
Máster en Ingeniería Industrial

EXPLORACIÓN DE TÉCNICAS DE REDES NEURONALES DE CONVOLUCIÓN EN LA DETECCIÓN DE ANOMALÍAS DE UN AEROGENERADOR

Autor: Lallana Saldaña, Mª Teresa.

Director: Sanz Bobi, Miguel Angel.

Entidad Colaboradora: ICAI – Universidad Pontificia Comillas.

RESUMEN DEL PROYECTO

El objetivo del proyecto es analizar el estado de un aerogenerador, utilizando técnicas de machine learning para predecir los valores que deberían tener las variables objetivo, y comparar las predicciones con los valores realmente medidos. Cuando estas desviaciones sean suficientemente grandes, identificarlas como anomalías y contabilizarlas. A través de un análisis de la contabilización de anomalías, extraer si existe un fallo, el indicador de fallo y el elemento que da fallo, permitiéndonos conocer el estado del aerogenerador.

Se dispone de un histórico de medidas que han sido tomadas durante el transcurso de 6 años. Estas medidas han sido tomadas simultáneamente en distintos puntos del aerogenerador. Se ha hecho un primer barrido recogiendo tan solo las medidas de aquellas variables de interés. Las variables de interés se muestran en la tabla siguiente:

Nombre de la variable		
ANG	fecha	
ANG	aero_tem_aceite_multi	val
ANG	aero_tem_rod_alta_multi	val
ANG	aero_tem_rod_DE_gen	val
ANG	aero_tem_rod_NDE_gen	val
ANG	aero_tem_anillos_sld_gen	val
ANG	aero_tem_interior_nac	val
ANG	aero_tem_ambiente	val
ANG	aero_vel_viento	val
ANG	aero_angulo_viento	val
ANG	aero_pot_total	val

Tabla 1. Variables de entrada y salida al modelo.

El método escogido para predecir el valor de una medida es una técnica de machine learning, las redes neuronales de convolución. Es objeto de este trabajo estudiar el comportamiento de las redes neuronales de convolución utilizándose para la predicción de datos pertenecientes a un muestreo de medidas. Habitualmente, este tipo de redes neuronales se utilizan para tareas de clasificación de imágenes mediante la detección de patrones en ellas.

Esto se traduce en las siguientes tareas principales:

1. Tratamiento de los datos
2. Desarrollo del modelo
3. Construcción del modelo

4. Detección de anomalías
5. Resultados de la detección

Tratamiento de los datos

Para desarrollar un modelo de red neuronal, es necesario un paso previo de tratamiento de los datos. Este paso se ha realizado utilizando el entorno Matlab. Se trata de un paso indispensable para que el modelo opere de forma robusta: por ejemplo, sirve para descartar los fallos del propio sensor.

El tratamiento de los datos se ha divide en varios pasos. El primer paso es la separación de las medidas de las variables de interés, la cual se ha explicado arriba. El segundo paso es la limpieza de los outliers: para cada una de las variables de interés se han eliminado las medidas que se encontraban fuera del intervalo $[\mu-2\sigma; \mu+2\sigma]$. El tercer paso es eliminar la mayor parte de los datos vacíos: se han eliminado las variables con más del 60% de datos vacíos y las observaciones (conjunto de medidas tomadas en un mismo momento) con más de tres medidas vacías.

Desarrollo del modelo

Una vez completado este preprocesado de los datos, las medidas resultantes se exportan a una tabla Excel para poder importarlos posteriormente a R. Este paso es necesario puesto que no se pueden importar los datos a R directamente desde Matlab. R ha sido el lenguaje de programación elegido para desarrollar la red neuronal. Existen dos lenguajes para el desarrollo de modelos de red nuronal: R y Python. Sin embargo, R ofrece mejores capacidades gráficas, requisito indispensable para poder analizar los resultados más adelante. Puesto que trabajar directamente con R no es funcional ni cómodo, se ha empleado un entorno de trabajo llamado Rstudio, más amigable.

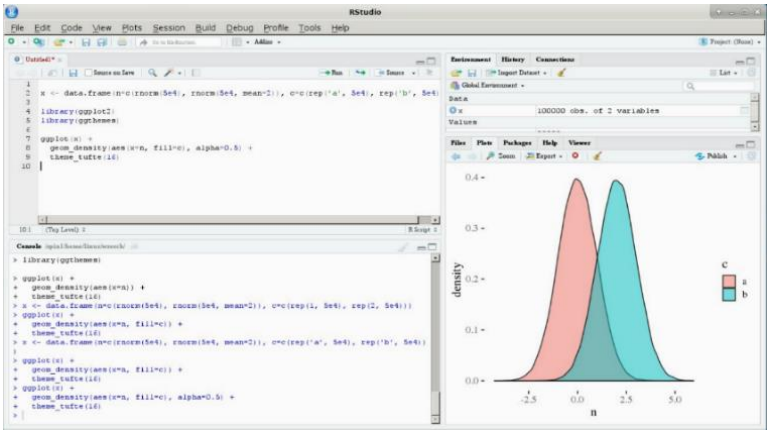


Ilustración 1. Entorno RStudio para la programación en R

Una vez importados los datos en Rstudio, comienza el desarrollo de la red neuronal. El primer paso es cargar los paquetes necesarios para trabajar. Destacamos el paquete Keras, el cual alberga el algoritmo de creación, entrenamiento y uso del modelo de red neuronal. El siguiente

paso será separar del conjunto total de datos los subconjuntos de entrenamiento y test de los datos del primer año, 70% y 30% respectivamente.

Utilizando el conjunto de entrenamiento, se procederá a entrenar el modelo. Con objeto de obtener la red neuronal óptima, se encierra dentro de un conjunto de bucles anidados la creación y el entrenamiento del modelo. Las variables de programación que se iteran en los bucles son los parámetros que caracterizan una capa de convolución: el numero de filtros y el tamaño de los filtros. Habitualmente, las capas de convolución se alternan con capas de pooling. Estas capas tienen como función reducir las dimensiones de las imágenes, en nuestro caso de los datos. Por lo tanto, adicionalmente se itera en los bucles anidados el parámetro que caracteriza una capa de pooling: la cantidad de dimensiones sustraídas.

Una vez completada una iteración, se calculaba el error absoluto medio del modelo y se registraba en una matriz. El error absoluto medio del modelo en una red neuronal es el indicador más habitual de optimalidad del mismo: cuanto más pequeño, mejor. Volviendo a la matriz, albergaba en cada columna tanto los parámetros característicos como el error medio absoluto de una iteración concreta. Por lo tanto, al salir de los bucles anidados, esta matriz ha sido exportada a formato Excel para salvaguardarla de forma permanente y poder extraer de ella la mejor combinación de parámetros, que constituyen la red neuronal óptima.

Año/Year 1		Potencia	Temperatura de los anillos	Temperatura de la multiplicadora	Rodamiento DE	Rodamiento NDE
Capa/Layer 1	f	6	6	2	5	7
	k	2	2	1	1	2
	p	0	0	0	0	0
Capa/Layer 2	f	4	5	7	4	7
	k	1	1	1	1	1
	p	0	0	0	0	0
Capa/Layer 3	f	7	7	3	7	7
	k	2	1	2	2	1
	p	1	0	0	0	0

Tabla 2. Configuraciones óptimas de la red neuronal para cada modelo.

Cabe destacar que en todo momento se ha trabajado con una red neuronal compuesta de 3 capas. 3 capas parecían un valor razonable. De igual manera, un máximo de 7 filtros, de 3 como tamaño de filtro y de 3 como cantidad de dimensiones sustraídas parecían valores razonables. Al aumentar estos parámetros, el tiempo de procesado de la red neuronal aumenta y el tiempo de compleción de los bucles iterativos también. Dadas la cantidad de variables entrantes al modelo (no son muchas por lo que el modelo no precisa mayor complejidad) y los resultados de desempeño del mismo obtenidos (bastante razonables), no ha sido necesario aumentarlos.

El proceso explicado hasta el momento se ha seguido para el desarrollo de más de un modelo de predicción. Los modelos desarrollados han sido:

1. Modelo predictivo de la potencia
2. Modelo predictivo de la temperatura de los anillos
3. Modelo predictivo de la temperatura de la multiplicadora
4. Modelo predictivo de la temperatura del rodamiento DE
5. Modelo predictivo de la temperatura del rodamiento NDE

Construcción de los modelos

Antes de comenzar a analizar el estado del aerogenerador en los años venideros, se desarrolla un script de extracción y análisis de los resultados de predicción. Este script es único y común para todos los modelos de predicción desarrollados. Primeramente, extrae el histograma de residuos, o errores de predicción (desviaciones entre la predicción y el valor medido). Segundo, extrae la nube de puntos y la recta de regresión entre los valores obtenidos del modelo, predichos, y los valores reales, medidos. Por último, plotea en el eje temporal los valores de predicción superpuestos a los valores reales y las desviaciones entre ambos. Adicionalmente, calcula el intervalo de confianza del 95% de la variable de salida del modelo y el número de observaciones fuera de dicho intervalo. Esto lo hace tanto para el conjunto de entrenamiento como para el conjunto de test.

Una vez listos todos los modelos y el script de extracción de los resultados, se analiza su robustez. Para ello se toman los conjuntos de entrenamiento y test de los datos del primer año y se analizan. Si se comprueba, a través de las gráficas y estadísticos extraídos, que el modelo es robusto para el primer año, se podrá utilizar para predecir la variable de salida en los años venideros.

Detección de anomalías

Por lo tanto, para cada modelo se han extraído los resultados del primer año, se ha comprobado que el modelo es robusto con los datos de ese primer año. Se ha desarrollado un script que contabiliza el número de muestras anómalas para en cada mes. Una muestra anómala es una muestra cuya desviación se encuentra fuera del intervalo de confianza del 95% del histograma de residuos. Este intervalo de confianza es el que se ha generado con los datos del primer año, no se calcula cada año con los datos nuevos.

Una vez se han contabilizado las muestras anómalas a nivel mensual, por años (2, 3, 4, 5 y 6), se procede a identificas de entre las muestras anómalas, cuales son anomalías. Se analizarán las muestras sucesivas, anómalas y no anómalas, en grupos de tres. Solo cuando se detecte en el grupo que dos o tres muestras son anómalas, estas muestras anómalas de contabilizarán como anomalías. Para completar esta tarea, se ha creado un vector vacío y se ha rellenado con las muestras anómalas, respetando su posición. Luego, mediante un bucle, se ha repasado el vector y, para aquellos grupos de tres con una sola muestra anómala, esa posición se ha vaciado. De esta forma, se ha obtenido el vector de anomalías.

Al igual que las muestras anómalas, las anomalías se han contabilizado. Se ha calculado el porcentaje mensual de anomalías en función de las muestras tomadas en el mes. Se han registrado todos estos resultados en una tabla resumen, que incorpora los resultados de los 5 modelos.

Resultados de la detección

A continuación, se puede encontrar un extracto de la tabla, extracto que solo incluye los dos primeros años.

Año	-	2	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3
Mes		1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	Potencia	45	45	54	33	12	15	15	9	12	39	39	117	99	111	69	48	36	21	45	51	51	15	51	87
% anomalías con respectoal total mensual de muestras	Potencia	1,56	1,56	1,86	1,14	0,42	0,51	0,51	0,3	0,42	1,35	1,35	4,05	3,21	3,6	2,25	1,56	1,17	0,69	1,47	1,65	1,65	0,48	1,65	2,82
Nº anomalías	Temperatura de los anillos	84	12	6	9	9	12	21	45	66	84	48	135	216	168	165	207	297	270	63	171	90	180	114	174
% anomalías con respectoal total mensual de muestras	Temperatura de los anillos	2,91	0,42	0,21	0,3	0,3	0,42	0,72	1,56	2,28	2,91	1,68	4,68	7,02	5,46	5,37	6,72	9,66	8,76	2,04	5,55	2,91	5,85	3,69	5,64
Nº anomalías	Temperatura de la multiplicadora	135	96	63	21	15	15	9	33	39	9	36	30	21	39	24	15	18	48	120	78	132	81	33	63
% anomalías con respectoal total mensual de muestras	Temperatura de la multiplicadora	4,68	3,33	2,19	0,72	0,51	0,51	0,3	1,14	1,35	0,3	1,26	1,05	0,69	1,26	0,78	0,48	0,57	1,56	3,9	2,52	4,29	2,64	1,08	2,04
Nº anomalías	Rodamiento DE	174	93	69	39	39	45	12	48	54	27	45	72	180	123	156	180	156	123	207	195	165	135	90	114
% anomalías con respectoal total mensual de muestras	Rodamiento DE	6,03	3,24	2,4	1,35	1,35	1,56	0,42	1,68	1,86	0,93	1,56	2,49	5,85	3,99	5,07	5,85	5,07	3,99	6,72	6,33	5,37	4,38	2,91	3,69
Nº anomalías	Rodamiento NDE	207	75	81	39	39	51	54	105	120	75	51	162	243	219	180	237	168	192	255	240	225	168	87	168
% anomalías con respectoal total mensual de muestras	Rodamiento NDE	7,17	2,61	2,82	1,35	1,35	1,77	1,86	3,63	4,17	2,61	1,77	5,61	7,89	7,11	5,85	7,71	5,46	6,24	8,28	7,8	7,32	5,46	2,82	5,46

Tabla 3. Detección de anomalías en los años 2 y 3.

Se ha calculado cada uno de los modelos:

- La media de las proporciones mensuales de anomalía de todos los años
- El tercer cuartil de las proporciones mensuales de anomalía de todos los años

En función de estos resultados, se ha determinado que al superar el valor de la media se indica fallo del elemento del modelo que ha determinado esa proporción, y el momento (mes y año) en el que se está produciendo. Cuando la proporción alcanza el valor del tercer cuartil, se indica que el fallo empieza a ser crítico y que es necesario actuar de inmediato sobre el elemento en cuestión.

Adicionalmente, se han identificado líneas de tendencia, tendencias estacionales y tendencias cíclicas en la visualización gráfica de los resultados.

EXPLORATION OF CONVOLUTIONAL NEURAL NETWORK TECHNIQUES IN THE DETECTION OF ANOMALIES IN AN AEROGENERATOR

The objective of this project is to analyse the condition of an aerogenerator, using machine learning techniques to predict the values the objective variables should have and to compare these predictions with real measured values. When any of these variations are significant enough, they will be identified as anomalies and will be accounted for. Through an analysis of every anomaly, it will discover if an error exists, the indicator of such error and the element which produces such error, allowing us to understand the condition of the aerogenerator.

A measurement record registered throughout the course of the last six years is provided. These measurements have been taken simultaneously at different points of the aerogenerator. An initial sweep has been made, leaving only the measurements of the variables we find of interest. Such variables are shown in the following table:

Name of the variable		
ANG	fecha	
ANG	aero_tem_aceite_multi	val
ANG	aero_tem_rod_alta_multi	val
ANG	aero_tem_rod_DE_gen	val
ANG	aero_tem_rod_NDE_gen	val
ANG	aero_tem_anillos_sld_gen	val
ANG	aero_tem_interior_nac	val
ANG	aero_tem_ambiente	val
ANG	aero_vel_viento	val
ANG	aero_angulo_viento	val
ANG	aero_pot_total	val

Tabla 4. Input and output variables of the model.

The chosen method to predict a measured value is a machine learning technique, convolutional neural networks. Studying the behaviour of these convolutional neural networks is the objective of this task, using for such prediction data provided from a sample of measurements. Normally, these types of neural networks are used for image classification tasks through the detection of patterns they share.

This is translated into the following main tasks:

1. Treatment of data.
2. Development of the model
3. Model building
4. Anomaly detection
5. Results of the detection

Treatment of data

In order to develop a neural network model, it is necessary a previous step regarding data treatment. This step has been performed using Matlab environment. It is considered to be an

essential step to ensure a strong performance: for instance, it helps to dismiss errors appearing in the sensor.

Data treatment has been divided into several steps. The first step is the separation of the measured values which are of interest, as it has been explained above. The second step involves the outliers cleaning: for each of the previous values, the measurements that were not in the range [μ-2σ; μ+2σ] have been eliminated. The third step is to eliminate most of the empty values: values with more than 60% of empty data have been removed, as well as the observations (set of measurements taken at the same time) with more than three empty measures.

Development of the model

Once this data preprocess has been completed, the remaining measurements are included in an Excel table, enabling them to be directed to R in the future. This step is necessary because you cannot direct data from Matlab directly to R. R has been the chosen program to develop the neural network. There are two different languages to develop neural network models: R and Phyton. However, R provides better graphic characteristics, an essential requirement necessary to analyse the results in the future. As working directly on R is neither functional nor comfortable, a different working environment called Rstudio has been used, a friendlier one.

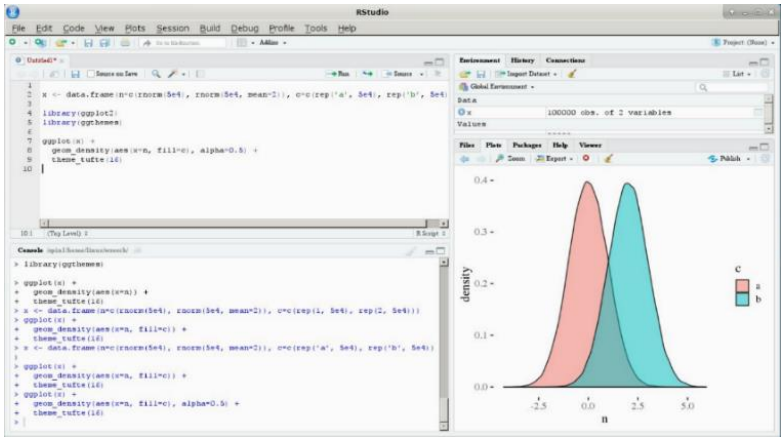


Ilustración 2. RStudio environment for R

Once the data has been directed to Rstudio, the development of the neural network starts. The first step is to load the necessary packages to work. We highlight the Keras package, which includes an algorithm of creation, training, and usage of the neural network model. The next step will be to separate from the total set of data the subsets regarding training and test from the first year data, 70% and 30% respectively.

Using the training subset, we will proceed to train the model. In order to develop an optimal neural network, the model is enclosed inside a set of nested loops the creation and the training of the model. The programming variables that are used in such loops are the parameters that characterise a convolutional layer: the number of filters and the size of the filters. Normally, the convolutional layers are alternated with pooling layers. Pooling layers are designed to reduce

the dimensions of the images, data in our case. Additionally, the parameter that characterises one layer of pooling (the quantity of subtracted dimensions) is iterated in the nested loops.

Once an iteration is completed, the mean absolute error from the model is calculated and registered into a matrix. The mean absolute error from the model in a neural network is the most common indicator for its optimal functionality: the smaller it is, the better. Returning to the matrix, it includes in each column both the characteristic parameters and the mean absolute error of an specific iteration. Therefore, when coming out of the nested loops, this matrix has been exported to Excel file in order to safeguard it permanently and enable it to extract the best possible combination of parameters, that establish the optimal neural network.

Año/Year 1		Potencia	Temperatura de los anillos	Temperatura de la multiplicadora	Rodamiento DE	Rodamiento NDE
Capa/Layer 1	f	6	6	2	5	7
	k	2	2	1	1	2
	p	0	0	0	0	0
Capa/Layer 2	f	4	5	7	4	7
	k	1	1	1	1	1
	p	0	0	0	0	0
Capa/Layer 3	f	7	7	3	7	7
	k	2	1	2	2	1
	p	1	0	0	0	0

Tabla 5. Optimal configurations of the neural network for each model.

It is important to highlight that throughout the task a three-layer neural network has been used. Three layers seemed like a reasonable value. Similarly, a maximum of seven filters, with a maximum size of three for each filter and a quantity of three as the number of subtracted dimensions seemed like reasonable values. If we were to increase these parameters, the time taken by the neural network to process increases, as well as completion time from nested loops. Due to the number of variables entering the model (not many so the model does not need more complexity) and the results provided from the performance (quite reasonable), it has not been necessary to increase them.

The process explained up to this point has been followed for more than one prediction model. The developed models are:

1. Predictive model of power.
2. Predictive model of the temperature of the sleep-ring.
3. Predictive model of the temperature of the gear box.
4. Predictive model of the temperature of the bearing DE.
5. Predictive model of the temperature of the bearing NDE.

Model building

Before starting to analyse the conditions of the aerogenerator in the upcoming years, an extraction and analysis of the predicted results script is developed. This script is unique and common for every predictive model developed. Firstly, extracts from the histogram of residuals, or prediction errors (deviations between the prediction and the measured value). Secondly, generates the point cloud and the regression line between the values registered in the model, the predicted values and the real measured values. Finally, plots in the temporal axis the predicted values overlaying them with the real values and the deviation between them. Additionally, it calculates the confidence interval of a 95% from the output variable from the model and the number of observations outside that range. This is done for both the training set and the test set.

Once all models and the result extraction script are ready, its strength is measured. For such task, the data regarding the training and test sets from the first year are registered and analysed. If, through the graphics and statistics extracted, the strength of the model is verified, it will be available to predict the output variable for the upcoming years.

Anomaly detection

Therefore, for every model the results of the first year have been registered. It has been verified that the model is strong using data from that first year. A script has been developed that counts the number of abnormal samples for each month. An abnormal sample appears when its deviation is outside the 95% confidence range from the histogram of residuals. This confidence range is the one generated with the data registered throughout the first year, it is not recalculated each year with new data.

Once all abnormal samples have been recorded monthly, each year (2,3,4,5 and 6), we proceed to identify which of the abnormal samples are anomalies. In groups of three, sequentially all samples will be analysed, abnormal or not. Only when a group has two or more of the given samples abnormal, it will be considered as an anomaly. In order to complete this task, an empty vector has been created, being filled afterwards with abnormal samples, considering their position. Afterwards, using a loop, the vector has been reviewed and, for those groups in which only one of the three samples was abnormal, that position has been emptied. That way, a vector of anomalies has been obtained.

Like it was done with the abnormal samples, anomalies have been registered. The monthly percentage of anomalies has been calculated, considering the number of samples taken that month. Every result has been registered in a summary table, which includes the results from all five models.

Results of the detection

Underneath, an extract from the table may be appreciated. This extract only includes the two first years.

Año	-	2	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3
Mes		1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	Potencia	45	45	54	33	12	15	15	9	12	39	39	117	99	111	69	48	36	21	45	51	51	15	51	87
% anomalías con respectoal total mensual de muestras		1,56	1,56	1,86	1,14	0,42	0,51	0,51	0,3	0,42	1,35	1,35	4,05	3,21	3,6	2,25	1,56	1,17	0,69	1,47	1,65	1,65	0,48	1,65	2,82
Nº anomalías	Temperatura de los anillos	84	12	6	9	9	12	21	45	66	84	48	135	216	168	165	207	297	270	63	171	90	180	114	174
% anomalías con respectoal total mensual de muestras		2,91	0,42	0,21	0,3	0,3	0,42	0,72	1,56	2,28	2,91	1,68	4,68	7,02	5,46	5,37	6,72	9,66	8,76	2,04	5,55	2,91	5,85	3,69	5,64
Nº anomalías	Temperatura de la multiplicadora	135	96	63	21	15	15	9	33	39	9	36	30	21	39	24	15	18	48	120	78	132	81	33	63
% anomalías con respectoal total mensual de muestras		4,68	3,33	2,19	0,72	0,51	0,51	0,3	1,14	1,35	0,3	1,26	1,05	0,69	1,26	0,78	0,48	0,57	1,56	3,9	2,52	4,29	2,64	1,08	2,04
Nº anomalías	Rodamiento DE	174	93	69	39	39	45	12	48	54	27	45	72	180	123	156	180	156	123	207	195	165	135	90	114
% anomalías con respectoal total mensual de muestras		6,03	3,24	2,4	1,35	1,35	1,56	0,42	1,68	1,86	0,93	1,56	2,49	5,85	3,99	5,07	5,85	5,07	3,99	6,72	6,33	5,37	4,38	2,91	3,69
Nº anomalías	Rodamiento NDE	207	75	81	39	39	51	54	105	120	75	51	162	243	219	180	237	168	192	255	240	225	168	87	168
% anomalías con respectoal total mensual de muestras		7,17	2,61	2,82	1,35	1,35	1,77	1,86	3,63	4,17	2,61	1,77	5,61	7,89	7,11	5,85	7,71	5,46	6,24	8,28	7,8	7,32	5,46	2,82	5,46

Tabla 6. Anomaly detection in years 2 and 3.

For each model, it has been calculated:

- The average of the monthly proportions of anomalies for every year.
- The third quartile of the monthly proportions of abnormal samples for every year.

Considering these results, it has been determined that when the value of the average is surpassed, it is shown error from the element from the model that has established such proportion, and the moment (month and year) it was produced. When the proportion reaches the third quartile, it shows that the error is starting to be severe and that it is necessary to take action as soon as possible in the element affected.

Additionally, trend lines have been identified, as well as seasonal trends and cyclical trends in the graphic visualisation of the results.

Tabla de Contenidos

Índice de Tablas

Índice de Ilustraciones

Capítulo 1.- Introducción y planteamiento del proyecto

1.1.- Introducción

En el parque eólico encontramos un aerogenerador, una máquina de 2 MW de potencia, para la cual se llevará a cabo un estudio de análisis del funcionamiento para detectar posibles anomalías. Esta máquina dispone de múltiples sensores distribuidos en su estructura que han ido tomando medidas de distintas variables a lo largo de un periodo de tiempo. Este periodo de tiempo es un total de 6 años.

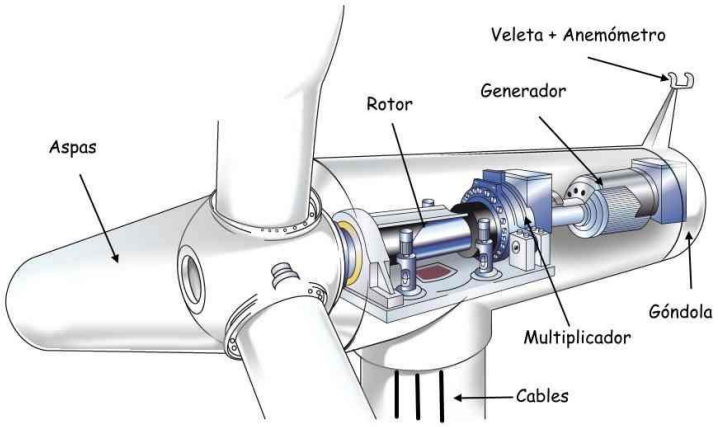


Ilustración 3. Partes principales de un aerogenerador.

El presente proyecto pretende realizar un modelo de análisis de un gran volumen de datos que permita detectar anomalías en el aerogenerador, trabajando con el histórico de medidas recogidas a lo largo de los años y con técnicas de predicción de machine learning.

1.2.- Planteamiento

Como tareas de dirección de proyecto, se hará un listado de objetivos y un planteamiento del proyecto que persiga la consecución de los mismos. Se completará con un estudio del estado de la técnica de aplicación de las redes neuronales a tareas de mantenimiento predictivo.

Para desarrollar el modelo de predicción se ha elegido una técnica llamada redes neuronales de convolución. Esta red neuronal es un tipo de red neuronal habitualmente utilizada para clasificación de imágenes mediante la detección de patrones.

Para desarrollar la red neuronal, es preciso realizar antes un tratamiento de los datos, con objeto de mejorar la robustez del modelo final. Una vez importados los datos limpios, se crea el modelo de red neuronal y se entrena. En medio de este proceso se puede buscar optimizar el modelo mediante la variación de los parámetros que constituyen la red neuronal. Una vez entrenado el modelo con la configuración óptima, se puede empezar a predecir valores.

Los modelos de comportamiento serán creados con objeto de predecir una variable clave para determinar si existe una anomalía en un elemento específico del aerogenerador. Por ello, la

variable de salida del modelo (variable clave) será un indicador habitual del estado del elemento en concreto que se esté inspeccionando. El indicador más evidente, y por lo tanto, el elegido, será la temperatura.

En este caso se va a buscar anomalías en los siguientes elementos del aerogenerador:

1. Anillos
2. Multiplicadora
3. Rodamiento DE
4. Rodamiento NDE

El objetivo es, a través de los modelos de comportamiento, ser capaces de predecir los valores que debería tomar una variable especifica (en este caso se ha elegido la temperatura) en un momento determinado. Luego, analizando la desviación entre el valor real medido y la predicción obtenida con el modelo, determinar si existe una anomalía.

Antes de proceder a calcular estas desviaciones, habrá que construir el modelo. El proceso de construcción del modelo consiste en demostrar la robustez del mismo mediante un análisis estadístico sencillo. Consiste en comprobar las desviaciones son mínimas para el conjunto de datos del primer año.

Una vez determinado que el modelo es robusto, se contabilizarán las anomalías. Para determinar si existe una anomalía, se calculará el intervalo de confianza del 95% para el histograma de residuos. Esto es, se calculará el error, o desviación entre la predicción y el valor real medido, para cada observación. Después, se graficarán en un histograma todos los errores, y aquellos en el 5% superior y el 5% inferior de cada año se considerarán observaciones que presentan una anomalía.

Una vez identificadas estas observaciones que presentan anomalía, se analizarán mes a mes, cuantificándolas, para determinar si el elemento objetivo del modelo precisa una inspección de mantenimiento.

Finalmente, se presentarán los resultados obtenidos que recogerán si se da un modo de fallo y en que elemento se da dicho modo de fallo. Adicionalmente se presentarán las tendencias observadas y un apartado de conclusiones que haga una reflexión global sobre la técnica utilizada, la metodología seguida y los resultados obtenidos.

A continuación, se puede encontrar un diagrama de cajas que albergan los distintos paquetes de trabajo a completar.

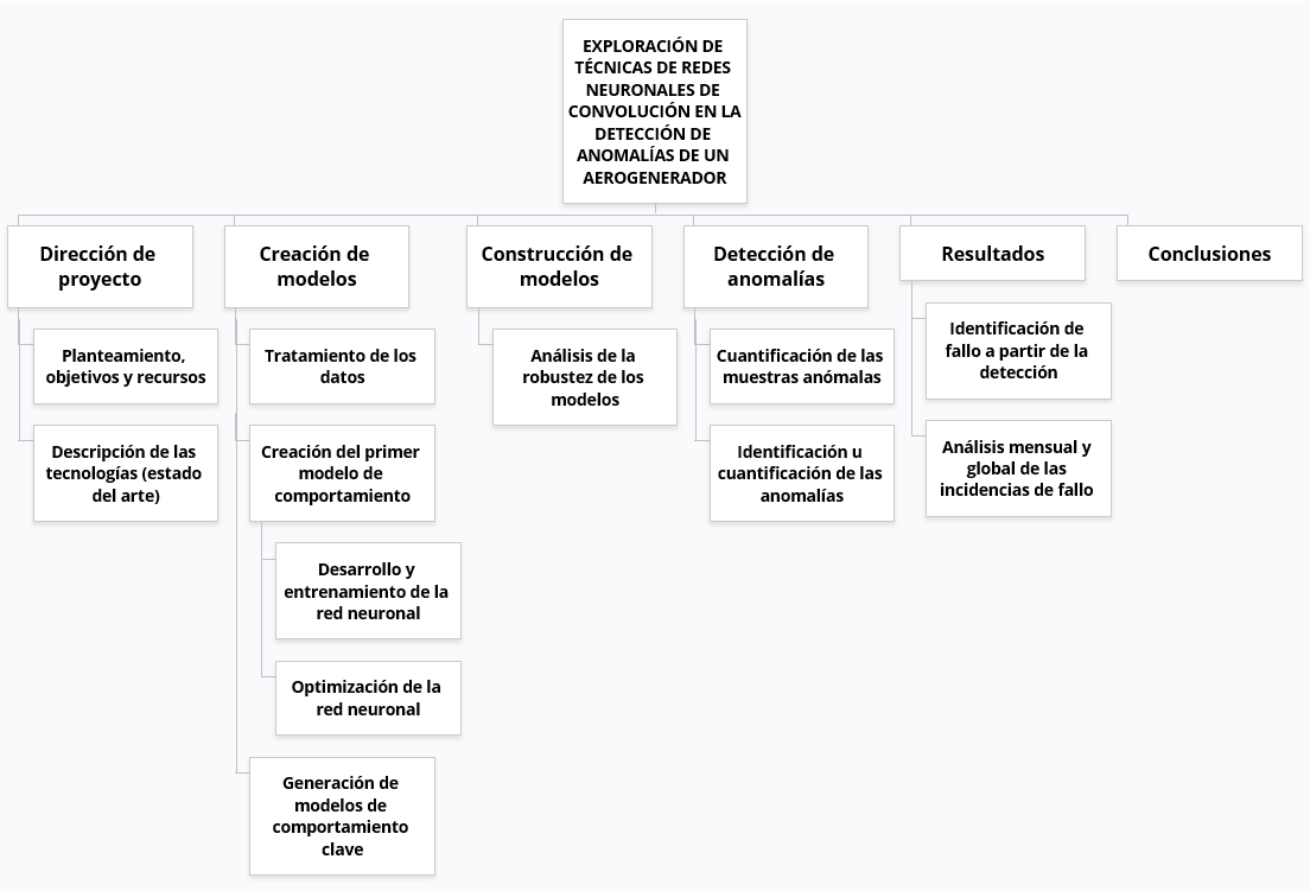


Ilustración 4. Work Breakdown Structure

1.3.- Motivación del proyecto

Las redes neuronales, por sus capacidades predictivas, dominan en la actualidad el campo de la inteligencia computacional. Hoy en día encuentran múltiples campos de aplicación, particularmente en tareas de mantenimiento predictivo. Cabe recordar que el mantenimiento predictivo consiste en monitorizar el estado de los elementos de forma que se puedan reparar o sustituir justo antes de romperse.

En concreto, en un parque eólico de tamaño medio, se consume una media de 20.000 euros por MW en tareas de mantenimiento. Cuando un elemento se rompe, puede afectar a otros elementos de su entorno, por lo que esperar a fallo por rotura para repararlo/sustituirlo puede resultar más costoso que realizar estas tareas de mantenimiento con una anterioridad.

Por lo tanto, desarrollar los modelos de mantenimiento predictivo, compuestos por redes neuronales, es más económico que realizar un mantenimiento correctivo. Especialmente en máquinas de elevado coste y complejas, como puede ser un aerogenerador.

1.4.- Objetivos

A continuación, presentamos los objetivos del trabajo.

1. Emplear redes neuronales de convolución como técnica de previsión de medidas
    a. Realizar un tratamiento previo de los datos
    b. Crear el modelo de comportamiento con una red neuronal de convolución en el lenguaje de programación R
    c. Optimizar el modelo de comportamiento
    d. Adaptar el modelo de comportamiento para generar nuevos modelos de comportamiento para emplear en otros elementos del aerogenerador

2. Detectar anomalías en el aerogenerador
    a. Identificar los elementos en los que se busca identificar anomalías
    b. Establecer la estrategia de detección de anomalías en los elementos
    c. Detectar anomalías en cada elemento mediante la utilización de los modelos creados

3. Identificar fallos en los elementos del aerogenerador
    a. Establecer la estrategia de contabilización de anomalías
    b. Realizar un análisis mensual y anual de las anomalías detectadas
    c. Identificar el tipo de fallo en función de las anomalías contabilizadas

1.5.- Recursos

Para el diseño y puesta a punto del modelo de análisis se utilizará un fichero de datos, proporcionado por el director del proyecto al alumno, que contiene las medidas tomadas por los sensores del aerogenerador durante un periodo de tiempo. El nombre del fichero es "datos.zip".

Para el tratamiento de estos datos las herramientas informáticas utilizadas serán:

- Notepad++
- Matlab
- RStudio
- Microsoft Office: Excel, Word, PowerPoint…

Adicionalmente, se recurrirá a recursos localizados en la red como repositorios o herramientas de modificación de formato:

- Mantenimiento de formato de la programación
- Repositorio GitHub

Finalmente, para la consecución de los objetivos se pretende utilizar la bibliografía:

- (Libro de nociones de estadística, 2013)

Capítulo 2.- Descripción de las tecnologías (estado de la técnica)

En este apartado veremos el estado de la técnica en la actualidad en el campo de las redes neuronales, en concreto las redes neuronales artificiales (ANN) y las redes neuronales de convolución (CNN).

Existen una gran variedad de tipos de red neuronal. Como uno de los objetivos de nuestro trabajo es estudiar la aplicación de las redes neuronales de convolución para la predicción de valores numéricos, haremos una comparativa entre:

1. Redes neuronales artificiales (ANN), redes neuronales clásicas utilizadas con más asiduidad, y las redes neuronales.
2. Redes neuronales de convolución (RNN), redes neuronales utilizadas en este proyecto en concreto.

Primeramente, introduciremos la técnica de red neuronal. A continuación, explicaremos el funcionamiento y características de las redes neuronales enumeradas, incluyendo referencias de proyectos que las han incorporado. Por último, realizaremos una tabla comparativa que muestre las características que ha incluido cada uno de los proyectos referenciados y situaremos en la misma nuestro proyecto.

2.1.- Introducción a las redes neuronales

Las redes neuronales dominan en la actualidad el campo de la inteligencia computacional. Están basadas en algoritmos matemáticos de machine learning que permiten realizar tareas de previsión.

Mediante el gran número de conexiones entre neuronas, son capaces de aprender relaciones entre entradas y salidas a partir de un conjunto de entrenamiento. Para que el entrenamiento sea completo es necesario una gran cantidad de ejemplos, y por tanto, disponer de un histórico con un gran volumen de datos.

Las redes neuronales pueden utilizarse para predecir (una sola neurona de salida) o para clasificar (más de una neurona de salida). En ambos casos se trabaja asociando la salida a un porcentaje de éxito.

A. Las tareas de clasificación consisten en determinar si las entradas al modelo pertenecen a un grupo o a otro (una salida u otra). Son comunes, por ejemplo, en aplicaciones de detección de patrones en imágenes para identificar lo que aparece en ellas.
B. Las tareas se predicción consisten en determinar el valor numérico de una salida en función de las entradas. Son más comunes en tareas de inferencia estadística, por ejemplo, en estimación del valor de un activo económico.

2.2.- Redes neuronales artificiales (ANN)

Las redes neuronales artificiales, comúnmente conocidas como perceptrones multicapa (MLP), son las de uso más habitual. Debido a su sencillez de formato y a la cantidad de ejemplos disponibles, son las redes neuronales de partida a utilizar en la iniciación en el mundo de machine learning.

Funcionamiento

Las redes neuronales clásicas están compuestas por una capa de neuronas de entrada y una capa de neuronas de salida. Entre estas capas se encuentran otras capas "ocultas" de neuronas. Las capas de neuronas están compuestas por un número de neuronas simples, y cada una de estas neuronas está conectada individualmente a cada una de las neuronas de la siguiente capa.

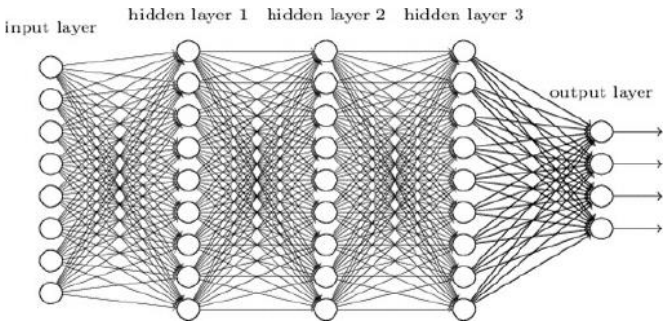


Ilustración 5. Red neuronal artificial

Cada neurona está conectada a la entrada con varias neuronas de la capa anterior, es decir tiene varias entradas. A través de un modelo lineal (net) seguido por un modelo no lineal (σ), obtiene una única salida. Esta salida, a su vez, se conectará con cada una de las neuronas de la siguiente capa.

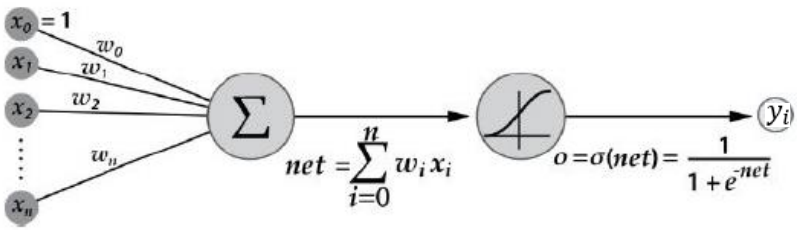


Ilustración 6. Detalle matemático de la conexión entre dos capas.

Cada conexión entre dos neuronas tiene un peso wi. El modelo lineal net se computa utilizando los valores de activación xi de las neuronas anteriores y los pesos wi de las conexiones con las neuronas anteriores.

$$net = \sum_{i=0}^{n} wi\ xi$$

El siguiente paso para obtener el valor de activación de la neurona de salida yi es incorporar una función de activación. La función de activación que aporta el carácter no lineal a la función de transferencia global. La capa de salida suele utilizar funciones de activación lineales mientras que las capas ocultas las neuronas utilizan típicamente una función de activación sigmoidal logística o una función de activación RELU.

$$y_i = \sigma(net) = \sigma\left(\frac{1}{1 + e^{-net}}\right)$$

Por lo tanto, los parámetros que podemos modificar para ajustar la red neuronal durante el entrenamiento son los valores de activación de las neuronas (xi, yi) y los pesos de las conexiones entre las neuronas (wi). Aplicándolo a la función de transferencia entre dos capas, esto se traduce en que se reforzará progresivamente las neuronas que están más activas (xi, yi) y las neuronas que se quiere que estén más activas (wi).

El aprendizaje se realiza a través del algoritmo de propagación hacia atrás o retopropagación. Durante el entrenamiento, este algoritmo se utiliza para ajustar los pesos wi y la función de activación σ. Cuando un conjunto de entrada ha cruzado el modelo, a la salida aparece un error ej entre la salida dada por el modelo y la salida que se debería haber obtenido: en el valor de activación de la neurona de salida no es el que debería. En ese momento, se calcula a la salida el valor del gradiente que nos posibilita la variación conveniente de los pesos wi.

$$\Delta w_{xi\,yi} = \frac{\partial \varepsilon}{\partial}\ y_i \qquad donde\ \frac{\partial \varepsilon}{\partial} = e_j\ \sigma'$$

$$e_j\ error\ en\ el\ valor\ de\ activación\ de\ la\ neurona\ j$$

$$\sigma'\ derivada\ de\ la\ función\ de\ activación$$

De esta forma, se modifican los pesos wi de la última capa hasta la primera. Esto es lo que se llama retropropagación. Una vez se completa, con un nuevo conjunto de entrada al modelo, como se han modificado los pesos wi, se modificarán también los valores de activación de las neuronas de la red neuronal.

Características

Antes de entrar a enumerar las características de una red neuronal artificial (ANN), es necesario entender algunos conceptos asociados al comportamiento de la red neuronal, para poder valorarlos y definir su utilidad:

A. Compartición de parámetros: esta capacidad la tienen algunos tipos de red neuronal, y consiste en compartir parámetros entre un paso y el siguiente durante el entrenamiento. Es decir, entre un conjunto de observaciones de entrada y el siguiente. El resultado es una disminución en los parámetros a entrenar (pesos wi) y, por lo tanto, una disminución de la carga computacional.
B. Relación espacial: en el caso de trabajar con imágenes, las características espaciales recogen la localización de los píxeles en la imagen y la relación entre ellos. Posibilita identificar objetos, localizarlos, y conocer su relación con otros objetos de la imagen.
C. Desaparición o disparo del gradiente: este es un problema asociado a el algoritmo de retropropagación o propagación hacia atrás. Recordemos que los pesos wi se actualizan mediante el algoritmo de retropropagación utilizando el gradiente.

Ahora sí, se enumeran las características de una red neuronal artificial (ANN):

1. Esta red neuronal es capaz de desarrollar una función no lineal a partir de entradas y salidas a la misma, mediante la utilización de una función de activación.
2. Al trabajar con imágenes, el número de parámetros en las capas ocultas aumentan drásticamente con el aumento de píxeles en la imagen.
3. No es capaz de capturar la información secuencial de los datos.
4. Al trabajar con imágenes, se pierden las características espaciales.
5. Si la red neuronal es muy profunda, existe una alta probabilidad de desaparición o disparo de gradiente.
6. Usada para aplicaciones simples por la sencillez del entrenamiento.
7. Tiene un nivel medio de desempeño en términos de precisión.
8. Permite continuar el trabajo ante información de entrada incompleta.

2.3.- Redes neuronales de convolución (CNN)

Funcionamiento

Las redes neuronales de convolución son un tipo de redes neuronales habitualmente utilizadas para procesamiento de imágenes y para procesamiento de texto. A partir de los píxeles, identifican patrones y los asocian a las salidas, permitiendo clasificar imágenes y distinguir objetos, letras o símbolos.

Cada una de las capas en las redes convolucionales está compuesta por tres dimensiones: el alto, el ancho, y la profundidad. Las redes se forman usando tres tipos de capas:

1. Capas convolucionales

En la capa convolucional se usa un filtro, llamado máscara, que recorre toda la imagen. Este cuadrado deslizante aporta a la salida un píxel combinación lineal de lo píxeles de entrada.

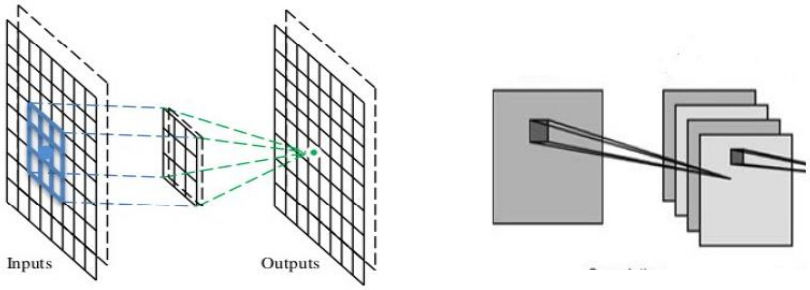


Ilustración 7. Detalle del funcionamiento de una capa convolucional.

Análogamente a las redes neuronales clásicas, los pesos de las conexiones son los valores de la máscara, que representan la conexión entre dos capas sucesivas. Se utilizan varias máscaras para recorrer una sola capa de entrada, obteniendo una capa de salida por cada máscara. Por tanto, el número de capas de salida es n veces el número de capas de entrada, siendo n el número de máscaras. Esto es lo que llamamos profundidad de la capa: el número de capas en paralelo dentro de una misma capa.

2. Capas de pooling

Estas capas se intercalan con las capas de convolución y se utilizan para reducir las dimensiones de alto y ancho de una capa. Básicamente, es un filtro tipo ventana que se pasea por todas las subcapas dentro de una misma capa y reduce localmente el número de píxeles de m a 1, siendo mxm el tamaño del filtro.

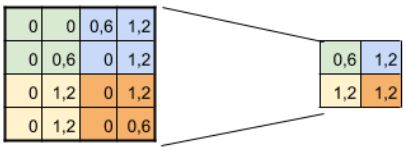


Ilustración 8. Detalle de funcionamiento de una capa de pooling.

Una capa de pooling se suele caracterizar con la opción 'max pooling': de los pixeles recogidos por el filtro, registra en la salida tan solo el de que presenta un mayor valor de activación.

3. Capas totalmente conectadas

Es la capa previa a la salida de la red neuronal. En esta capa se pierde la información espacial.

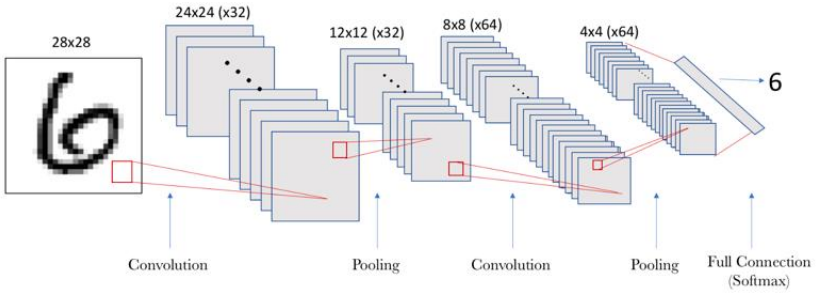


Ilustración 9. Funcionamiento global de una red neuronal de convolución.

Se puede observar en la imagen superior la intercalación de las capas, el funcionamiento de cada una de ellas y el funcionamiento combinado que constituye la red neuronal de convolución.

Características

A continuación, se enumeran las características más interesantes de una red neuronal de convolución (CNN):

1. Esta red neuronal es capaz de desarrollar una función no lineal a partir de entradas y salidas a la misma, mediante la utilización de una función de activación.
2. Al trabajar con imágenes, mediante un filtrado progresivo se capturan las características relevantes de la información de entrada utilizando el proceso de convolución. Esto permite obtener un mapa de características mucho más reducido que la imagen de entrada, disminuyendo la carga computacional.
3. Al trabajar con imágenes, no se pierden las características espaciales.
4. Es capaz de capturar la información secuencial de los datos.
5. Aunque la red neuronal sea muy profunda, no existe la posibilidad de desaparición o disparo de gradiente, que imposibilita la captura de la información secuencial por parte de la red.
6. Comparten parámetros entre un paso del entrenamiento y el siguiente, puesto que un filtro recoge los mismos pixeles limítrofes en una iteración y en la siguiente. Esto les permite capturar la información secuencial.
7. Como estas redes no son muy profundas, existe una baja probabilidad de desaparición o disparo de gradiente.
8. Presenta mayor complejidad en el entrenamiento que otros tipos de redes.
9. Tiene un alto nivel de desempeño en términos de precisión.
10. No permite continuar el trabajo ante información de entrada incompleta.

2.4.- Referencias

En este apartado mostraremos diferentes ejemplos de aplicación de los dos tipos de redes neuronales vistas. Para cada una de ellas, introduciremos el procedimiento utilizado y el fin de las mismas.

El primer ejemplo en el que nos podemos fijar es en un modelo creado para la previsión de consumo de gas de una instalación en un horario determinado. Como entradas utiliza la velocidad del viento, la cantidad de lluvia, la temperatura, el consumo eléctrico y el momento temporal en que se toman las muestras. Como salida predice el consumo de gas. Este modelo utiliza un modelo de red neuronal de convolución (CNN) para la predicción de valores. Utiliza una sola capa de convolución, pero se trata de una capa de convolución de dos dimensiones. Esto es, su vector de entrada tiene 4 dimensiones (3 dimensiones para los colores RGB y una dimensión para la escala de grises).

(Keijzer, GitHub, s.f.)

El siguiente ejemplo tiene como objetivo obtener la vida útil restante de un componente en concreto de un motor turbofán, presente en una flota de 100 aviones del mismo tipo. Busca proveer el tiempo restante antes del fallo. Incorpora una red neuronal convolucional (CNN) que analizará la serie temporal de medidas en formato de imágenes. El conjunto de datos incorpora tres escenarios de operación (motores), cada uno con 21 variables medidas y un ciclo de vida (eje temporal con todas las observaciones medidas). Se trata de una red neuronal de clasificación con tres posibles salidas, cada una de ellas recoge el rango de ciclos restantes antes de fallo en el motor (2=[0,15], 1=[16,45], 0=[46,inf]). El autor utiliza un espectrograma para obtener las imágenes de alimentación del modelo y comenta que el problema de operar así es que puede dar lugar a pérdida de información.

(Cerliani, 23)

Un auto encoder es una técnica de aprendizaje en la que los datos de entrada son comprimidos para obtener un bajo nivel dimensional, y luego descomprimidos utilizando la red neuronal invertida, obteniendo los datos originales. Aunque este método de compresión y descompresión no presenta ventajas frente a la compresión de imágenes tradicional, puede ser útil, por ejemplo, para eliminar el ruido presente en una imagen. El auto encoder variador, una evolución del anterior, no solo aprende a identificar patrones, sino que es capaz de dibujar nuevas imágenes.

El siguiente ejemplo discute como generar un modelo que detecte anomalías utilizando un auto encoder. Las variables de entrada al modelo son un total de 66, que recogen las medidas obtenidas por un sensor de vibración triaxial situado en un soporte. Las variables de salida son también 66, buscando recomponer las medidas de entrada. Por lo tanto, se realizará una tarea de clasificación. Las capas de encoding y decoding son capas de red neuronal artificial (ANN), compuestas por un numero de neuronas que va descendiendo (encoding) para luego ascender (decoding). Con el conjunto de entrenamiento se calcula el error de reconstrucción: diferencia

entre medidas de entrada y medidas de salida. A partir del percentil 85, la medida se considerará una anomalía.

(Amruthnath, 2020)

En este ejemplo se realizan inspecciones de una superficie de cemento. La red neuronal es una red neuronal de convolución (CNN) que analiza las imágenes de entrada, que pueden presentar o no grieta. Por lo tanto, se trata de una tarea de clasificación. Se ajusta la dimensión de las imágenes para no tener en la capa de entrada demasiadas neuronas, y así reducir los parámetros del modelo. La red neuronal está compuesta por tres capas de convolución, seguidas cada una por una capa de pooling, para reducir las dimensiones de la red neuronal.

(Rodrigues, GitHub, 2019)

Este ejemplo propone un modelo de diagnóstico de fallo de un motor de inducción. Utiliza una red neuronal convolucional (CNN) que recibe el detalle de las vibraciones tomadas por un sensor situado sobre el motor. Para una sola señal puede tener una de las tres salidas: operación normal, fallo de motor y fallo del cojinete. Por lo tanto, la tarea que completa es de clasificación. La entrada es la imagen obtenida que muestra el espectro de vibración y, entonces, la capa de entrada recibe un total de 1024 entradas.

(Lee, 2019)

2.5.- Estado de la cuestión

A partir de las referencias encontradas de ejemplos de aplicación de redes neuronales y redes neuronales de convolución, para tareas tanto de clasificación como de predicción, podemos generar una tabla comparativa en la que se resuman las capacidades de cada aplicación y el lugar que ocuparía este proyecto.

A continuación, podemos ver tanto la naturaleza como la tarea de la red neuronal para cada uno de los modelos vistos.

Modelos	ANN	CNN	Clasificación	Predicción
Modelo de previsión de consumo de gas		☑	☑	
Modelo de previsión de la vida útil restante de un motor turbofan		☑	☑	
Modelo de deteccion de anomalías usando un auto encoder	☑			☑
Modelo de detección de grietas en una superficie de cemento		☑	☑	
Modelo de diagnóstico de fallo de un motor de inducción		☑	☑	
Modelo de de anomalías en un aerogenerador		☑		☑

Tabla 7. Naturaleza y tarea de los modelos de referencia.

El fin último de este estudio de estado de la técnica es justificar que la técnica empleada, en términos teóricos, es la óptima para nuestro tipo concreto de aplicación. Y esto estará justificado a través de la composición de esta tabla.

Modelos	Menor complejidad del entrenamiento	Continuidad del trabajo ante información de entrada incompleta	Reducción de carga computacional mediante compatición de parametros	Riesgo bajo de desvanecimiento/disparo del gradiente	Captura de la relacion espacial	Captura de la información secuencial de los datos	Mejor desempeño en términos de precisión	Clasificación	Predicción
Modelo de previsión de consumo de gas			☑	☑	☑	☑	☑	☑	
Modelo de previsión de la vida útil restante de un motor turbofan			☑	☑	☑	☑	☑	☑	
Modelo de deteccion de anomalías usando un auto encoder	☑	☑							☑
Modelo de detección de grietas en una superficie de cemento			☑	☑	☑	☑	☑	☑	
Modelo de diagnóstico de fallo de un motor de inducción			☑	☑	☑	☑	☑	☑	
Modelo de de anomalías en un aerogenerador			☑	☑	☑	☑	☑		☑

Tabla 8. Comparativa de las características aportadas por los modelos de referencia y el modelo de este proyecto.

Como podemos observar, en la última fila se sitúa el proyecto presente. Reúne las capacidades propias, ventajosas, de una red de convolución. Solo que, en este caso, el modelo opera la red empleándola para tareas de predicción de medidas numéricas.

Capítulo 3.- Datos y preprocesado

Para trabajar con los datos y sacar resultados concluyentes, es necesario limpiar los mismos previamente. Se trata de un paso clave para asegurar que el modelo se construye sobre una base de datos sólidos, aportando robustez y fiabilidad al mismo.

La limpieza consistirá en organizar los datos, seleccionar las variables de interés, eliminar outliers y eliminar datos vacíos.

Para esta tarea hemos seleccionado el entorno Matlab, el cual es necesario como intermediario para exportar los paquetes de datos a Excel. Una vez en Excel, con los datos limpios, se podrán importar al entorno R.

3.1.- Organización de los datos

Los datos de partida dados son tres ficheros encriptados de Matlab que incluyen conjuntos de variables obtenidas de los sensores del aerogenerador:

- '_ANG_alarms.mat'
- 'repara_ANG.mat'
- 'ANG_analog_short_2.mat'

Dentro de estos paquetes encontramos dos matrices, 'Alarmas_ANG' y 'repara_ANG', del primer y segundo fichero respectivamente. Estas matrices no aportan información útil y, por lo tanto, no se utilizarán en el estudio. Dentro del tercer fichero encontramos una estructura de datos llamada 'ANG'. Esta estructura contiene la siguiente información útil:

ANG	fecha	
ANG	aero_tem_aceite_multi	val
ANG	aero_tem_rod_alta_multi	val
ANG	aero_tem_rod_DE_gen	val
ANG	aero_tem_rod_NDE_gen	val
ANG	aero_tem_anillos_sld_gen	val
ANG	aero_tem_max_dev_gen	val
ANG	aero_tem_dev_A_gen	val
ANG	aero_tem_dev_B_gen	val
ANG	aero_tem_dev_C_gen	val
ANG	aero_tem_grupo_h	val
ANG	aero_pres_grupo_h	val
ANG	aero_pres_yaw	val
ANG	aero_tem_interior_nac	val
ANG	aero_tem_ambiente	val
ANG	aero_tem_max_trafo	val
ANG	aero_tem_dev_A_trafo	val
ANG	aero_tem_dev_B_trafo	val
ANG	aero_tem_dev_C_trafo	val

ANG	aero_W_eje_principal	val
ANG	aero_w_gen	val
ANG	aero_angulo_palas	val
ANG	aero_vel_viento	val
ANG	aero_angulo_viento	val
ANG	aero_angulo_yaw	val
ANG	aero_V_red	val
ANG	aero_pot_reactiva	val
ANG	aero_factor_potencia	val
ANG	aero_pot_act_est	val
ANG	aero_pot_act_rot	val
ANG	aero_pot_total	val
ANG	aero_Energia	val
ANG	aero_state	avg

Tabla 9. Selección de variables útiles para el modelo.

Cada una de las 33 variables de esta estructura contiene 291312 datos, donde el dato i de 'ANG aero' corresponde a las medidas tomadas en la fecha i de 'ANG fecha'.

Tras importar los datos a Matlab, y guardar cada variable de la estructura en un vector independiente de dimensión 291312x1, se juntan conformando una matriz de 291312x33: 291312 observaciones (filas) de 33 variables (columnas).

3.1.- Reducción de variables

Dentro de las 33 variables, tan solo hay 20 que con las que nos interesa trabajar. Por lo tanto, el primer paso será eliminar las variables intranscendentes, obteniendo una nueva matriz resultado de dimensión 291312x20.

3.2.- Limpieza de los outliers

Los outliers son, en estadística, unas observaciones con valores atípicos, numéricamente distantes del resto de los datos. Estos outliers son frecuentemente engañosos y desvían los resultados de la tendencia real.

Por esta razón se ha decidido retirarlos del conjunto de datos de trabajo. A través de Matlab, se calculan la media aritmética y la desviación típica de cada una de las variables (columnas), omitiendo para su cálculo los datos vacíos 'NaN'.

A continuación, para cada una de las observaciones (filas), se analiza dato a dato si está dentro del intervalo $[\mu-2\sigma, \mu+2\sigma]$. Si todos los datos de la observación (fila) están dentro del intervalo, esta fila se copia en una nueva matriz, la matriz que será la resultante de la limpieza de outliers.

Esto se ha realizado mediante dos bucles for: el exterior va recorriendo las filas una a una y el interior va recorriendo las columnas. Cuando el interior ya ha recorrido todas las columnas de

una fila y ha comprobado, por ejemplo, que no existe ningún outlier, es cuando se almacena la observación (fila) en la matriz resultado. Entonces se sale del bucle interior y se itera en el bucle exterior, desplazándose a la siguiente fila, donde se repite el proceso.

La matriz resultado tiene una dimensión de 210523x20.

3.3.- Limpieza de los datos vacíos 'NaN'

Los datos vacíos pueden ser debidos a fallos en los sensores, a desconexión de los mismos o a valores fuera de su rango de medida. Es necesario retirar los máximos posibles puesto que impiden el entrenamiento del modelo en la red neuronal.

1. Variables (columnas) con más del 60% de datos vacíos 'NaN'

Usando el mismo procedimiento que antes, mediante dos bucles for, se recorre la matriz comuna a columna contabilizando los datos 'NaN' que aparecen. Si son más del 60 por ciento de los datos de la variable (columna) datos vacíos, se elimina la columna.

La matriz obtenida tiene una dimensión de 210523x16, habiéndose eliminado las variables. Se han eliminado las columnas:

6. 'ANG_aero_tem_dev_A_gen_val'
7. 'ANG_aero_tem_dev_B_gen_val'
8. 'ANG_aero_tem_dev_C_gen_val'
19. 'ANG_aero_angulo_yaw_val'

2. Observaciones (filas) con más de 3 datos 'NaN'

De nuevo recorreremos la matriz, ahora fila por fila, para eliminar aquellas con mas de tres datos vacíos 'NaN'. Así, eliminamos las observaciones (filas) con demasiados datos vacíos.

La matriz obtenida tiene una dimensión de 200425x16, habiéndose eliminado las observaciones.

Por último, se representan los resultados para comprobar que la limpieza de los datos de partida es válida. Para ello se representan los datos tras cada uno de los filtrados a continuación.

3.4.- Resultado de la limpieza de datos

Verificamos que el filtrado realizado sea correcto mediante la representación gráfica de cada una de las etapas del proceso.

1. Variables (columnas) con más del 60% de datos vacíos 'NaN'

Los gráficos obtenidos son los siguientes:

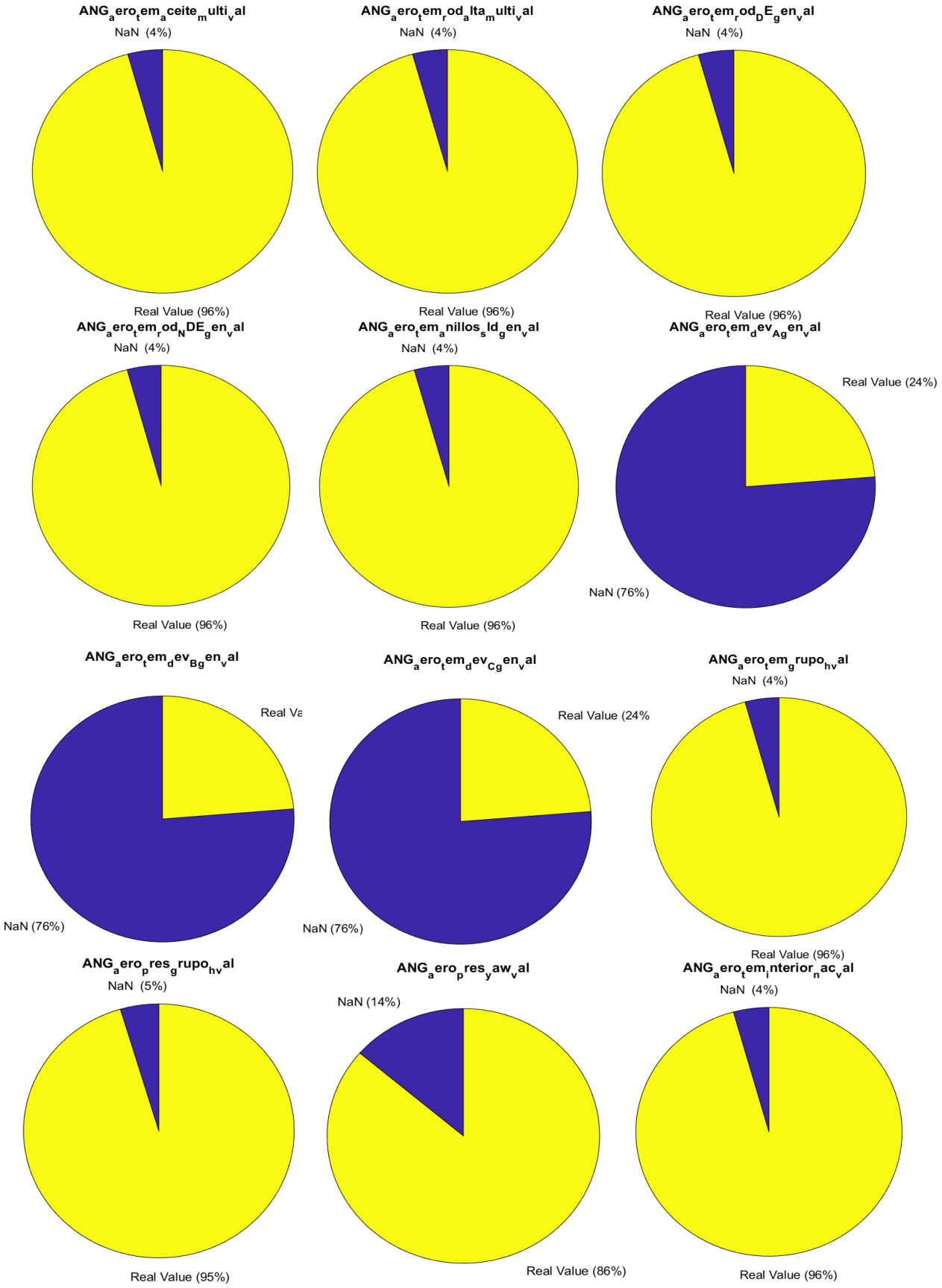


Ilustración 10. Resultados de la limpieza de datos vacíos NaN.

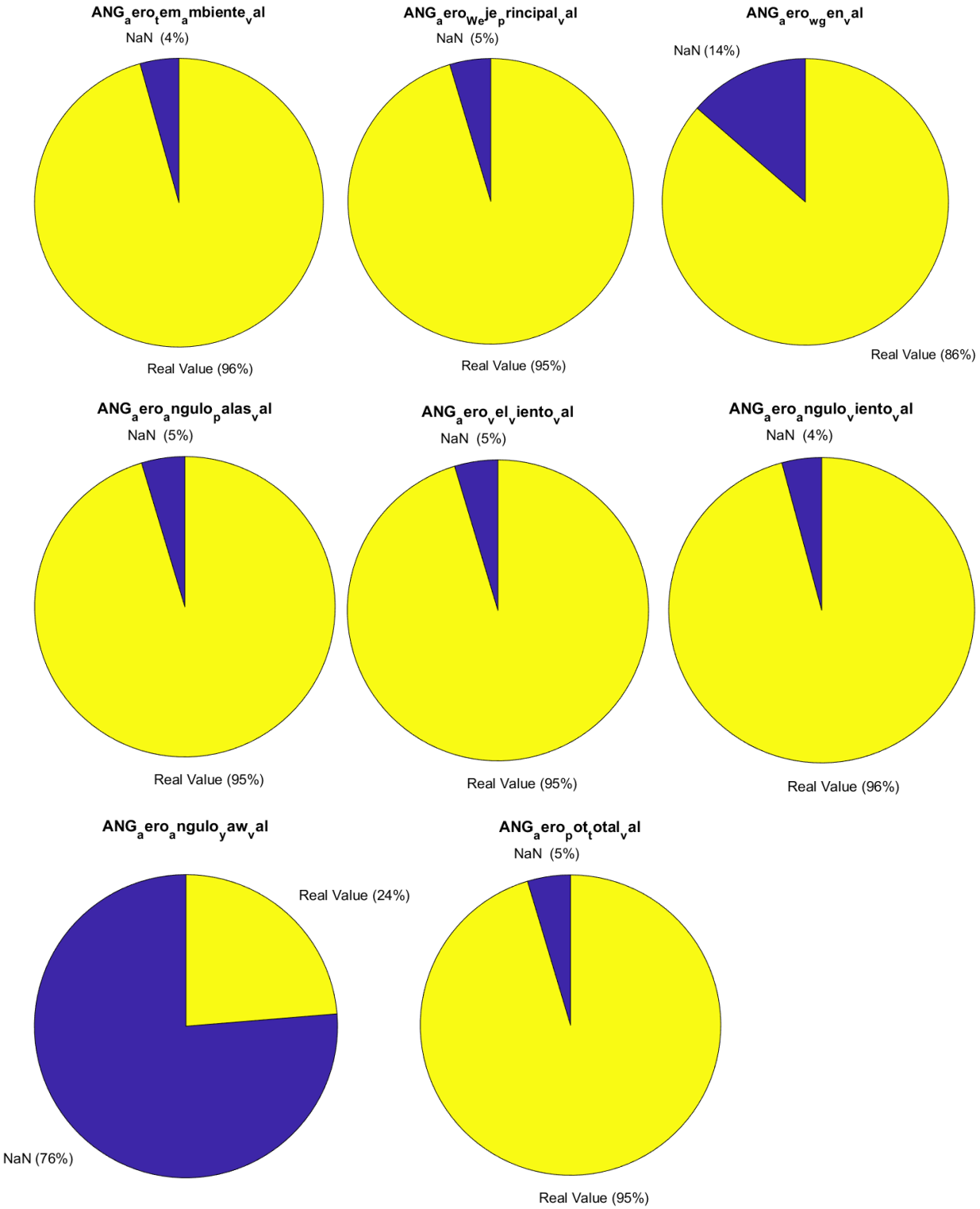


Ilustración 11. Resultados de la limpieza de datos vacíos NaN.

Tal y como se puede comprobar, las columnas que se han eliminado (6, 7, 8, 19) cumplen la condición de más del 60% de datos vacíos 'NaN'.

2. Observaciones (filas) con más de 3 datos 'NaN'

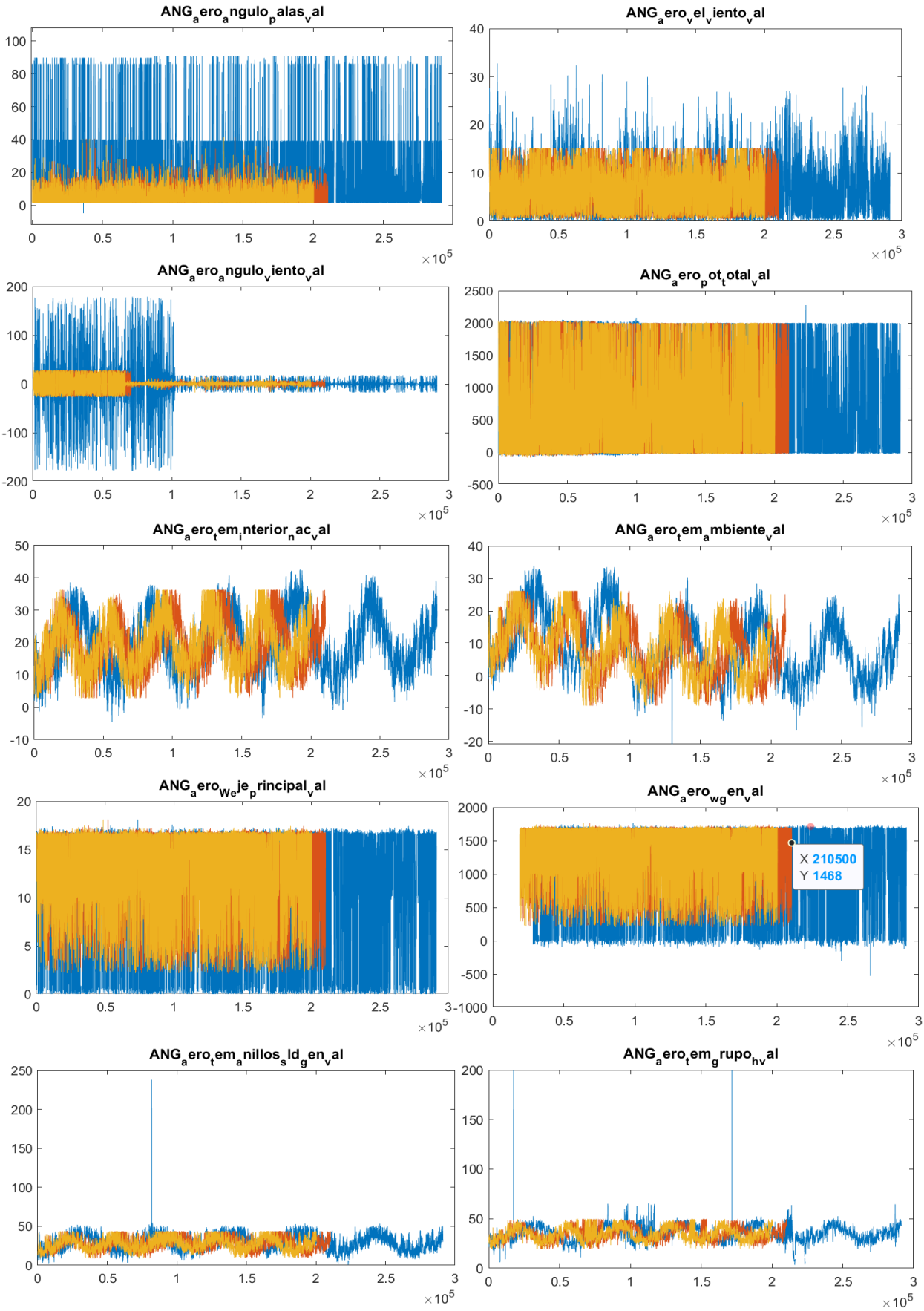


Ilustración 12. Evolución de los datos tras cada fase de limpieza.

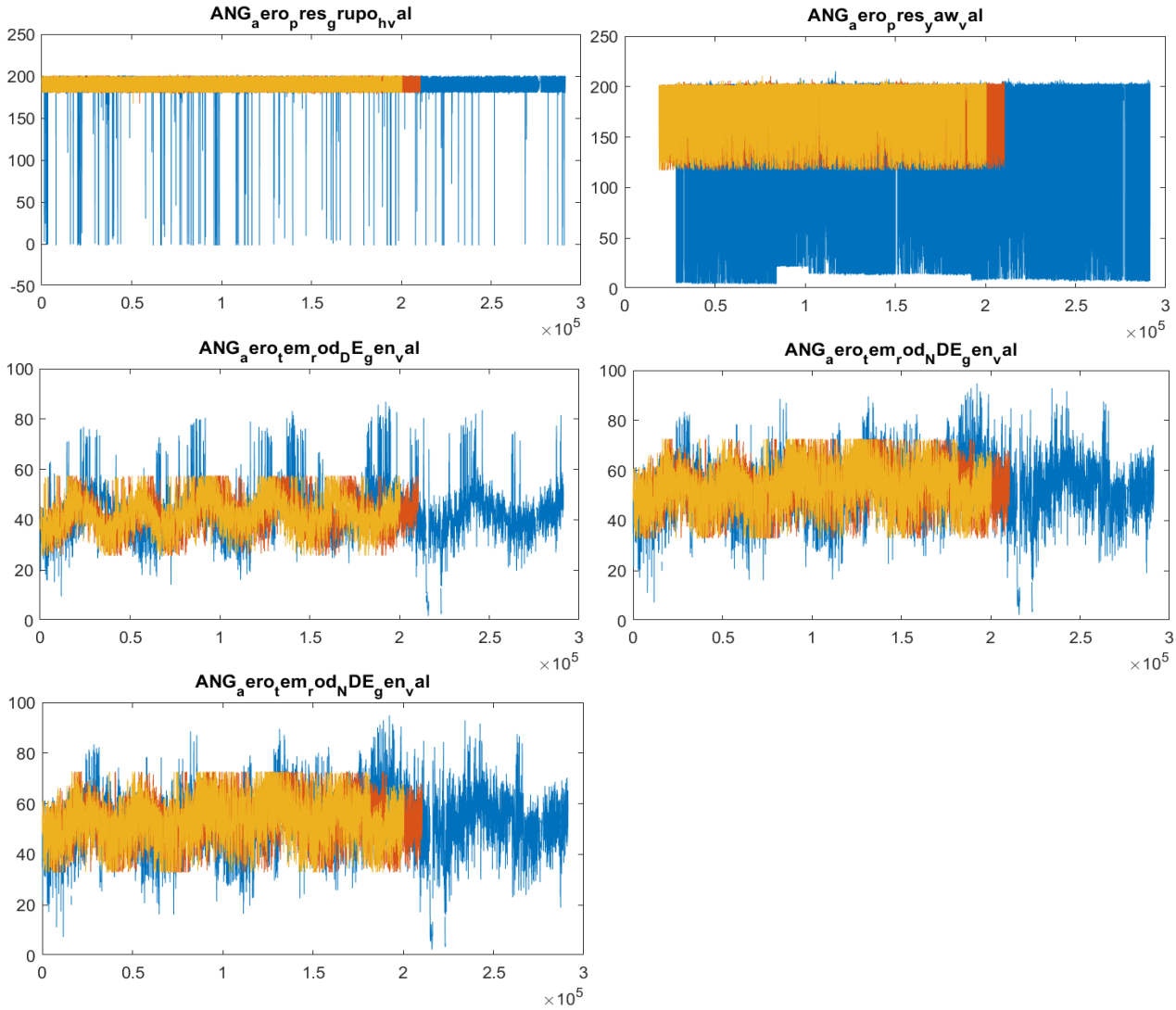


Ilustración 13. Evolución de los datos tras cada fase de limpieza.

En los gráficos sobre el texto se puede observar en azul los datos originales, en rojo los datos tras la limpieza de outliers y en amarillo los datos tras la limpieza de observaciones con mas de 3 datos vacíos. En el eje horizontal se muestra la posición cronológica de la observación y en el eje vertical el valor de la variable que se está graficando en esa observación.

Se puede comprobar visualmente como tanto horizontal (número de observaciones) como verticalmente (valor de la variable concreta en dicha observación) se acotan los valores de trabajo.

A continuación, desde Matlab se exporta el conjunto resultante a Excel, con los títulos de cada variable (columna), para más tarde importar el archivo Excel al entorno RStudio.

Capítulo 4.- Creación de modelos de comportamiento normal

Tras haber limpiado los datos, es el momento de crear los modelos de comportamiento. En este paso reside todo el trabajo generado a través del lenguaje de programación R, es decir, la generación del código que constituye el cuerpo de los modelos de comportamiento.

Para poder crear los modelos hay que conocer las variables de entrada y salida de los mismos. Por lo tanto, el primer paso será identificar las variables que se utilizan para cada uno de ellos y justificar su elección.

Una vez determinadas las variables de entrada y salida a cada uno de los modelos, se generará el cuerpo de código de cada uno de ellos.

4.1.- Conjuntos de datos usados en entrenamiento y test

Para comprender mejor la eleccion de las variables a predecir y su dependencia de otras, resulta necesario hacer una breve introducción a los elementos que componen un aerogenerador. Estos elementos serán monitorizados con objeto de comprobar el estado en el que se encuentran y si precisan algún tipo de mantenimiento.

A continuación podemos econtrar un esquemático de un aerogenerador.

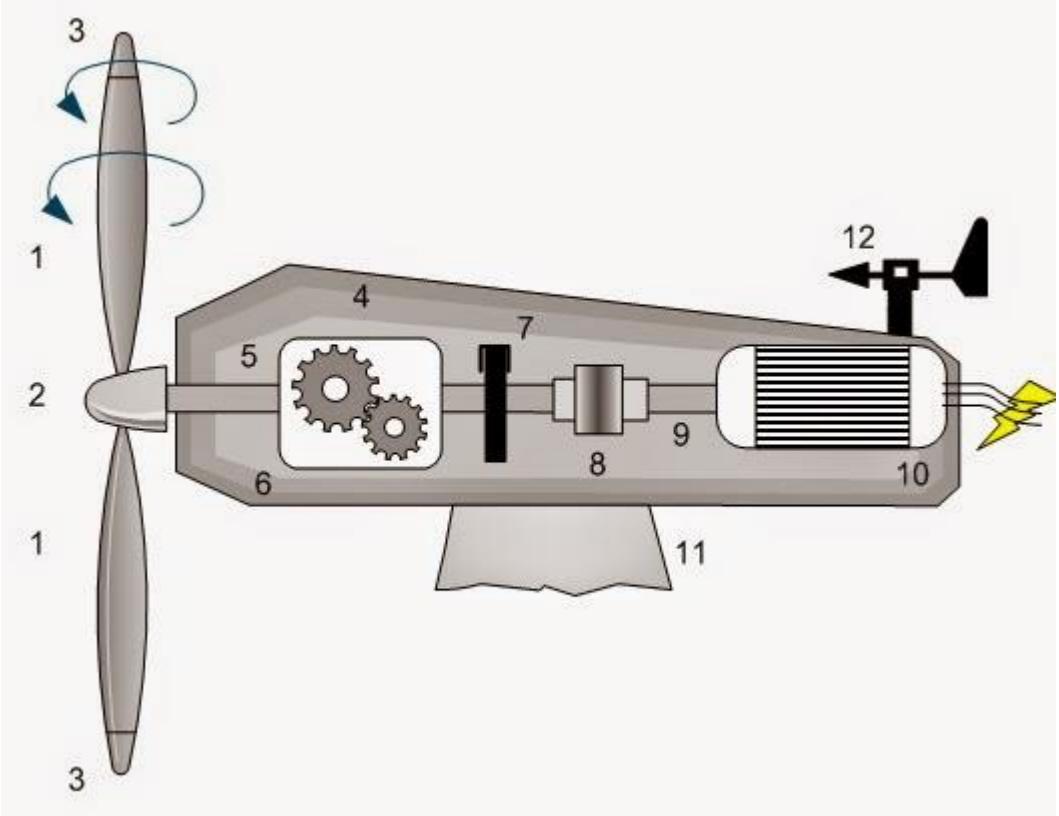


Ilustración 14. Elementos internos del aerogenerador.

Las partes numeradas del aerogenerador son:

1. Pala.

2. Buje.

3. Freno aerodinámico.

4. Góndola.

5. Eje principal.

6. Caja multiplicadora.

7. Freno mecánico.

10. Generador.

11. Torre.

12. Veleta y anemómetro.

A continuación, sobre el esquemático anterior, indicaremos los puntos en donde se tomarán las medidas empleadas en nuestro modelo.

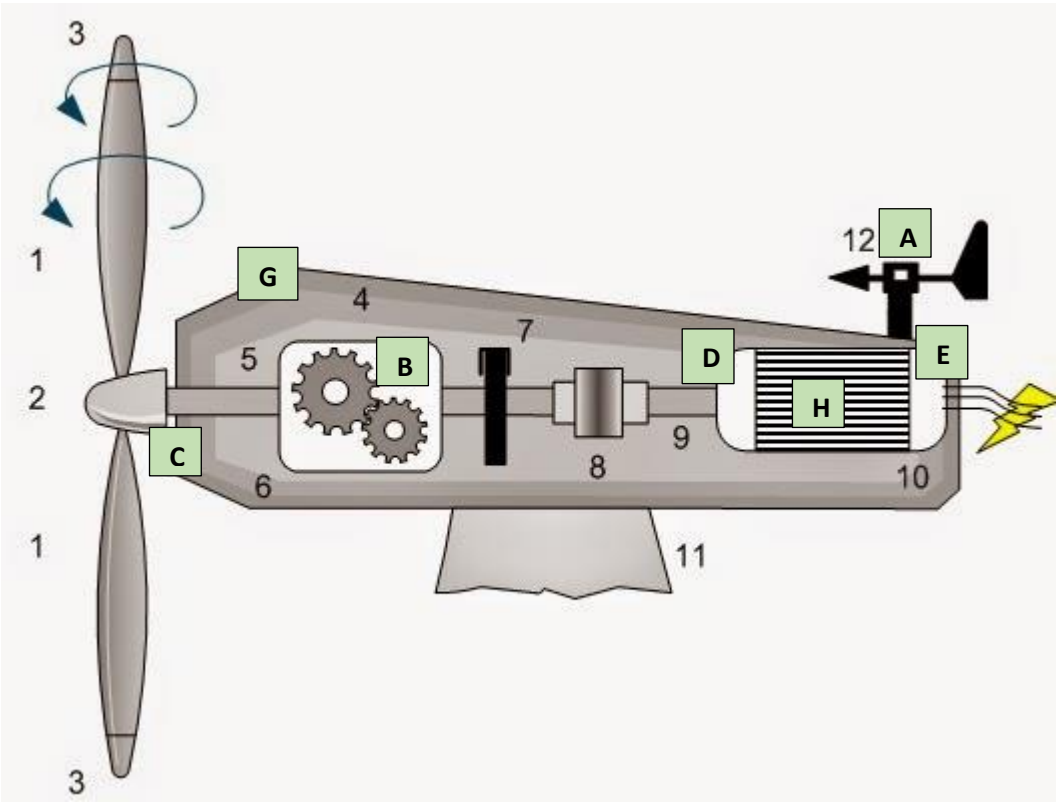


Ilustración 15. Localización de los sensores en el aerogenerador.

En la siguiente lista se puede encontrar, junto al nombre de la variable, el posicionamiento del sensor que entrega dicha medida al modelo.

Nombre de la variable a medir			Localizador
ANG	fecha		-
ANG	aero_tem_aceite_multi	val	B
ANG	aero_tem_rod_alta_multi	val	C
ANG	aero_tem_rod_DE_gen	val	D
ANG	aero_tem_rod_NDE_gen	val	E
ANG	aero_tem_anillos_sld_gen	val	F
ANG	aero_tem_interior_nac	val	G
ANG	aero_tem_ambiente	val	A
ANG	aero_vel_viento	val	A
ANG	aero_angulo_viento	val	A
ANG	aero_pot_total	val	H

Tabla 10. Variables de entrada y salida del modelo.

Para los siguientes modelos de predicción se emplearán variables de la lista, tanto las variables de entrada como la variable de salida del modelo. La variable de salida del modelo será la variable a predecir y, por lo tanto, deberá tener una relación de significancia con las variables de entrada al mismo.

La elección de estas variables de salida se basa en los posibles modos de fallo que se desean detectar. Como cada modelo tiene una única variable de salida, cada modelo de seguimiento del comportamiento se construirá para detectar un tipo particular de fallo:

A. Modelo de previsión de la potencia
    a. Modelo base para verificar el estado del aerogenerador, es decir, que se está operando con regularidad
B. Modelo de previsión de la temperatura de los anillos
    a. Fallo de los anillos con alta temperatura
C. Modelo de previsión de la temperatura de la multiplicadora
    a. Fallo de la multiplicadora con alta temperatura
D. Modelo de previsión de la temperatura de del rodamiento DE
    a. Fallo del rodamiento DE con alta temperatura
E. Modelo de previsión de la temperatura de del rodamiento NDE
    a. Fallo del rodamiento NDE con alta temperatura

Modelo de previsión de la potencia generada por el aerogenerador

Este modelo tiene como objetivo evaluar si la potencia generada se corresponde con la predicha, en función de las condiciones de viento observadas. En el caso en que no se hallen anomalías (desviaciones altas entre la predicción y el valor real medido), se conocerá que el aerogenerador opera con normalidad.

Este modelo es el caso base, a partir del cual se realizarán pequeñas modificaciones para adaptarlo a los distintos posibles detectores de anomalías a construir. Así, la explicación del

código de programación utilizado, que se puede ver en la siguiente sección, está ejemplarizada con este modelo, y dicho código es el que conforma el modelo predictivo de la potencia.

Como caso de partida vamos a predecir la potencia:

- ANG_aero_pot_total_val

a partir de las variables:

- ANG_aero_tem_ambiente_val
- ANG_aero_vel_viento_val
- ANG_aero_angulo_viento_val
- ANG_aero_tem_rod_DE_gen_val
- ANG_aero_tem_rod_NDE_gen_val

Modelo de previsión de la temperatura de los anillos

Este modelo tiene como objetivo evaluar si la temperatura medida en los anillos se corresponde con la predicha, en función de las condiciones refrigeración y operación. En el caso en que no se hallen anomalías (desviaciones altas entre la predicción y el valor real medido), se conocerá que no existe fallo por alta temperatura en los anillos, o en otras palabras, que los anillos mantienen un correcto funcionamiento.

A continuación, vamos a predecir la temperatura de los anillos:

- ANG_aero_tem_anillos_sld_gen_val

a partir de las variables:

- ANG_aero_tem_interior_nac_val
- ANG_aero_pot_total_val
- ANG_aero_tem_ambiente_val
- ANG_aero_vel_viento_val

Modelo de previsión de la temperatura de la multiplicadora

Este modelo tiene como objetivo evaluar si la temperatura medida en la multiplicadora se corresponde con la predicha, en función de las condiciones refrigeración y operación. En el caso en que no se hallen anomalías (desviaciones altas entre la predicción y el valor real medido), se conocerá que no existe fallo por alta temperatura en la multiplicadora, o en otras palabras, que la multiplicadora mantiene un correcto funcionamiento.

A continuación, vamos a predecir la temperatura de la multiplicadora:

- ANG_aero_tem_rod_alta_multi_val

a partir de las variables:

- ANG_aero_pot_total_val

- ANG_aero_tem_interior_nac_val
- ANG_aero_vel_viento_val
- ANG_aero_tem_aceite_multi_val

Modelo de previsión de la temperatura del rodamiento DE

Este modelo tiene como objetivo evaluar si la temperatura medida en el rodamiento DE se corresponde con la predicha, en función de las condiciones refrigeración, carga de trabajo y operación. En el caso en que no se hallen anomalías (desviaciones altas entre la predicción y el valor real medido), se conocerá que no existe fallo por alta temperatura en el rodamiento DE, o en otras palabras, que el rodamiento DE mantiene un correcto funcionamiento.

A continuación, vamos a predecir la temperatura del rodamiento DE, "Drive end", el rodamiento en el lado del accionamiento:

- ANG_aero_tem_rod_DE_gen_val

a partir de las variables:

- ANG_aero_tem_rod_NDE_gen_val
- ANG_aero_pot_total_val
- ANG_aero_tem_ambiente_val
- ANG_aero_vel_viento_val

Modelo de previsión de la temperatura del rodamiento NDE

Este modelo tiene como objetivo evaluar si la temperatura medida en el rodamiento NDE se corresponde con la predicha, en función de las condiciones refrigeración, carga de trabajo y operación. En el caso en que no se hallen anomalías (desviaciones altas entre la predicción y el valor real medido), se conocerá que no existe fallo por alta temperatura en el rodamiento NDE, o en otras palabras, que el rodamiento NDE mantiene un correcto funcionamiento.

A continuación, vamos a predecir la temperatura del rodamiento NDE, "Non drive end", el rodamiento en el lado opuesto al del accionamiento:

- ANG_aero_tem_rod_NDE_gen_val

a partir de las variables:

- ANG_aero_tem_rod_DE_gen_val
- ANG_aero_pot_total_val
- ANG_aero_tem_ambiente_val
- ANG_aero_vel_viento_val

4.2.- Modelos de comportamiento normal

Tras haber entendido la utilidad de cada modelo y haber identificado las variables que deben entrar y salir de cada uno de ellos, podemos generar el código que albergará los algoritmos de red neuronal. Estos algoritmos permitirán generar las predicciones de las variables salida de los mismos, variables objetivo a monitorizar para poder identificar los posibles fallos.

Objetivos y especificación

Recordemos que para nuestro estudio vamos a desarrollar una red neuronal, cuyos resultados posteriormente serán analizados para determinar si el estado del aerogenerador es bueno, empleando un conjunto de datos numérico. Este conjunto de datos numéricos son las medidas recopiladas durante un periodo de 6 años desde los distintos sensores del aerogenerador.

La estructura genérica de código es la siguiente:

1. Creación, entrenamiento y optimización del modelo

Estos scripts contienen la línea general de desarrollo de la red neuronal. Se trata de 5 scripts, uno por modelo, y todos siguen el mismo desarrollo. Esta etapa comprende las secciones de código:

   a. Importar datos desde Excel
   b. Seleccionar variables
   c. Dividir los datos
   d. Establecer el formato de los datos
   e. Normalizar los datos
   f. Ajustar la dimensión
   g. Definir parámetros base
   h. Abrir bucles iterativos
   i. Crear y compilar modelo
   j. Construir modelo
   k. Entrenar modelo
   l. Evaluar el entrenamiento del modelo
   m. Cerrar bucle iterativo
   n. Exportar resultados numéricos del bucle iterativo

   Redefinir el conjunto de test en función del año.

2. Construcción del modelo

Este script contiene los análisis estadísticos necesarios para construir un modelo. Construir el modelo consiste en justificar que se trata de un modelo robusto, poco sensible a pequeñas variaciones causadas por impurezas (como podría ser un outlier) y fiable: que para un mismo set de entrada siempre se obtenga la misma salida. Se trata de 1 script solo, el cual se puede reutilizar para los distintos modelos. Esta etapa comprende las secciones de código:

   a. Calcular las predicciones del conjunto de entrenamiento
   b. Dibujar los gráficos de residuos del conjunto de entrenamiento
   c. Dibujar las predicciones frente a los valores reales del conjunto de entrenamiento

d. Calcular las predicciones del conjunto de test
e. Dibujar los gráficos de residuos del conjunto de test
f. Dibujar las predicciones frente a los valores reales del conjunto de test
g. Calcular los intervalos de confianza del modelo
h. Exportar los resultados gráficos del modelo

3. Detección y registro de anomalías

Este script contiene el código necesario para generar las predicciones de manera anual y contabilizar las predicciones de manera mensual. Habrá que identificar las muestras anómalas, determinar si se trata de anomalías y registrarlas de forma ordenada para poder exportar los resultados después. También se trata de un script único, adaptado a todos los modelos. Esta etapa comprende las siguientes secciones de código:

a. Calcular las predicciones del conjunto de test
b. Dibujar los gráficos de residuos del conjunto de test
c. Identificación y registro de las anomalías del conjunto de las muestras anómalas
d. Dibujar las predicciones frente a los valores reales del conjunto de test
e. Contabilización de anomalías
f. Exportar resultados numéricos del bucle iterativo

A continuación, se enumeran sus componentes de los scripts en orden cronológico. Bajo cada título, se puede encontrar su desglose y el código utilizado, con aclaraciones en el mismo.

Limpiar entorno y establecer directorio

```
# ESTABLECER DIRECTORIO
# Comprueba el directorio de trabajo
getwd()
# Si no es correcto, actualiza el directorio de trabajo
setwd("E:/TFM/scripts")

# LIMPIAR ENTORNO
# Limpiar consola->Ctrl+L
# Limpiar variables
rm(list=ls())
```

En R se carga por defecto el último entorno sobre el que se ha trabajado. Para evitar complicaciones, al comienzo de cada algoritmo reinicializaremos tanto la consola como la memoria de variables del entorno, y comprobaremos que nos encontramos en el directorio de trabajo apropiado.

Cargar paquetes

Para trabajar con los paquetes en R es necesario haberlos previamente instalado. Habitualmente la instalación de los paquetes se realiza con el comando:

```
install.packages("NOMBRE")
```

También existe una pestaña en la interfaz de RStudio que te permite buscar e instalar un paquete sin necesidad de conocer el comando.

En particular, para el paquete Keras, no se puede usar directamente el comando en RStudio, puesto que Keras está pensado para R, y RStudio es un entorno más amigable para trabajar la programación en R, pero no es el entorno origen de la programación en R. Por ello, es necesario correr el siguiente programa, de manera independiente, para que el ordenador pueda trabajar con Keras bajo el entorno RStudio.

```
#Instalation Keras
#The steps to install Keras in RStudio is very simple. Just follow the below steps and you would be good to make your first Neural Network Model in R.
install.packages("devtools")
devtools::install_github("rstudio/keras")
#The above step will load the keras library from the GitHub repository. Now it is time to load keras into R and install tensorflow.
library(keras)
#By default RStudio loads the CPU version of tensorflow. Use the below command to download the CPU version of tensorflow.
install_tensorflow()
#To install the tensorflow version with GPU support for a single user/desktop system, use the below command.
install_tensorflow(gpu=TRUE)
#For multi-user installation, refer this installation guide.
#Now that we have keras and tensorflow installed inside RStudio, let us start and build our first neural network in R.
```

También, para completar la instalación, será necesario tener algún programa compilador de Python instalado en nuestro ordenador. En nuestro caso, vamos a instalar el recomendado por su buena compatibilidad con RStudio, cuyo link es: https://www.anaconda.com/

Por último, procederemos a cargar los paquetes con el siguiente comando:

```
12. # CARGAR PAQUETES
13. # Cargar el paquete Keras
14. library(keras)
15. # Cargar el paquete para normalizar los datos
16. library(RSNNS)
17. # Cargar el paquete para procesar las fechas
18. library(lubridate)
```

La instalación de los paquetes solo es necesaria una vez. Una vez se ha instalado en un directorio, las siguientes veces tan solo será necesario cargar el paquete al inicializar el entorno de trabajo, siempre y cuando se siga trabajando en el mismo directorio.

Cargar variables

Como el procedimiento de importar y preparar el conjunto de datos en el entorno de trabajo.

El primer paso es importar los datos al entorno. Para importar los datos, se han filtrado previamente con Matlab y se han exportado desde Matlab a formato Excel. En el mismo

RStudio existe una pestaña que te permite ubicar el archivo tipo Excel y seleccionar las características de importación del mismo, así como el comando asociado a esta operación. Inicialmente se opta por esta opción, y una vez obtenido el código, se incluirá en la programación para correrlo directamente en las inicializaciones posteriores.

Importar datos desde Excel

```
# CARGAR CONJUNTO DE DATOS
# Importar datos desde excel
library(readxl)
data_complete <- read_excel("C:/Users/Teresa/Desktop/TFM/datos/ANG_filtrada.xlsx",
                            sheet = "Hoja1")
# Visualizar en tabla de datos
#View(data_complete)
# Visualizar en formato columnas cada variable
#head(data_complete)
# Visualizar en formato filas cada variable
#str(data_complete);
```

Adicionalmente, comentadas, se pueden encontrar distintas opciones de visualización de los datos importados. Estos datos, por defecto, se importan en formato 'dataframe'.

Seleccionar las variables

```
# PREPARAR LOS DATOS
# Extract Specific columns form dataframe to data frame
data <- data.frame(data_complete$ANG_fecha,data_complete$ANG_aero_pot_total_val,data_complete$ANG_aero_tem_ambiente_val,data_complete$ANG_aero_vel_viento_val,data_complete$ANG_aero_angulo_viento_val,data_complete$ANG_aero_tem_rod_DE_gen_val,data_complete$ANG_aero_tem_rod_NDE_gen_val)

# Quito todas las filas que contengan NaN
data<-na.omit(data)

# Quito los títulos
#names(data) <- NULL
```

Preparar los datos para una correcta adaptación de los mismos al modelo es un proceso importante. Como se puede ver más arriba se comienza reduciendo las columnas (variables), eliminando aquellas que no tienen transcendencia en la predicción del modelo que se va a desarrollar.

A continuación, se eliminan las filas (observaciones) de que contienen datos vacíos 'NaN'. Es importante realizar este paso posterior a la eliminación de columnas (variables) intranscendentes. Como las variables son independientes entre ellas (son distintos sensores los que miden cada una de las variables), en una observación (fila) puede haber variables erróneas o vacías junto a variables cuyo valor no se debería despreciar. Por ello, se debe eliminar las filas (observaciones) solo tras haber seleccionado las variables (columnas) a incorporar en el modelo.

Tras ello, comentado, aparece el comando con el que se pueden eliminar los títulos de las columnas (nombres de las variables), si fuese necesario, para poder correr el modelo sin obstáculos.

Dividir los datos

La idea es entrenar el modelo con los datos del primer año y, analizar el estado del aerogenerador utilizando los datos de los años consecutivos. Para ello, se han importado los datos con la fecha de recogida de cada uno de ellos y se van a separar por años a partir de dichas fechas.

```
# Divido por años
year1_end<-match(data$data_complete.ANG_fecha[1]+years(1), data$data_complete.ANG_fecha)
year2_end<-
match(data$data_complete.ANG_fecha[1]+years(2)+hours(2), data$data_complete.ANG_fecha)
year3_end<-match(data$data_complete.ANG_fecha[1]+years(3), data$data_complete.ANG_fecha)
year4_end<-
match(data$data_complete.ANG_fecha[1]+years(4)+hours(3), data$data_complete.ANG_fecha)
year5_end<-match(data$data_complete.ANG_fecha[1]+years(5), data$data_complete.ANG_fecha)
lastdate<-dim(data)[1]
```

En esta ocasión, se utilizará la fecha del primer año para coger del conjunto de datos aquellos que nos interesan para entrenar el modelo.

```
# División de los datos
set.seed(1234)
data_year1<-data[1:year1_end,]
ind <- sample(2, nrow(data_year1), replace = T, prob = c(0.7, 0.3))
training <- data_year1[ind==1,3:7]
trainingtarget <- data_year1[ind==1, 2]
test <- data_year1[ind==2,3:7]
testtarget <- data_year1[ind==2, 2]


```

Una vez se ha reducido la dimensión de los datos de entrada a solo los respectivos al primer año, y antes de empezar a trabajar con ellos, se establece una semilla para que en futuras repeticiones del trabajo se obtengan los mismos resultados.

A continuación, se divide el conjunto de datos en las partes 1 y 2, que se separan horizontalmente, y que incluyen el 70% y el 30% de las observaciones respectivamente:

- La parte 1, con el 70% de las observaciones, es el denominado conjunto de entrenamiento.
- La parte 2, con el 30% de las observaciones, es el denominado conjunto de test.

Recordemos que el propio conjunto de entrenamiento ya incluye dentro de su 70% un 20%, denominado conjunto de validación, que se utiliza para evaluar el desempeño del modelo durante el entrenamiento del mismo.

También se separa, mediante una división vertical:

- Los datos de entrada del modelo, tanto del conjunto de entrenamiento como del conjunto de test.
- Los datos objetivo/de salida del modelo, tanto del conjunto de entrenamiento como del conjunto de test.

Establecer el formato de los datos

```
# Cambiar de dataframe a matriz
training = data.matrix(training)
test = data.matrix(test)
trainingtarget = data.matrix(trainingtarget)
testtarget = data.matrix(testtarget)

```

Para trabajar con un modelo de redes neuronales es necesario que el formato utilizado por los datos sea una matriz. Teniendo en cuenta las dos divisiones que se acaban establecer arriba, obtendremos un total de 4 matrices con las que vamos a trabajar.

Normalizar los datos

```
imputs_norm<-normalizeData(data_year1[,3:7], type="0_1")
target_norm<-normalizeData(data_year1[,2], type="0_1")

# Normalizar datos del entrenamiento y test
training_norm<-normalizeData(training, type="0_1")
trainingtarget_norm<-normalizeData(trainingtarget, type="0_1")
test_norm<-normalizeData(test, type="0_1")
testtarget_norm<-normalizeData(testtarget, type="0_1")
# Visualizar la primera observación del conjunto de entrenamiento, normalizada
#training[1, ]
```

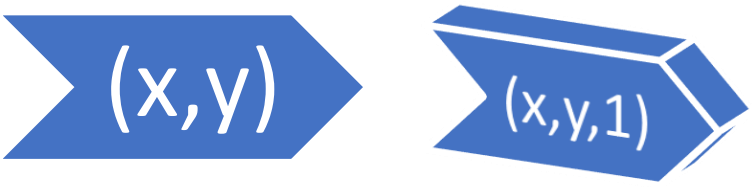
Para evitar que los diferentes rangos de los valores de las variables empleadas puedan sesgar los resultados y se obtengan modelos inútiles, se han de normalizar los datos del conjunto de entrenamiento y test. Así mismo, para comprobar que el conjunto está correctamente normalizado se ha incluido un comando a continuación, comentado, que permite visualizar la primera fila (observación) del conjunto de entrenamiento.

Adicionalmente, en la primeras líneas podemos encontrar la normalización del conjunto de entradas y del conjunto de salidas, por si las necesitásemos posteriormente.

Ajustar la dimensión

```
# De 2 a 3 dimensiones
training_norm=array_reshape(training_norm,c(dim(training_norm),1))
training=array_reshape(training,c(dim(training),1))
test_norm=array_reshape(test_norm, c(dim(test_norm),1))
test=array_reshape(test, c(dim(test),1))
```

Un asunto adicional a tener en cuenta es que las redes neuronales de convolución como ya se dijo nacieron en el mundo de la imagen y por tanto utilizan conjuntos de entrada de tres dimensiones, correspondientes a las tres gamas de color RGB . Nuestros datos de entrada al modelo, tanto el conjunto de entrenamiento como el conjunto de test, son matrices: tienen 2 dimensiones. Por ello, se hace necesario añadir una dimensión a nuestros conjuntos de entrada al modelo, utilizando las funciones predefinidas en R que aparecen más arriba.



Definir parámetros base

A continuación, se define el valor de cada uno de los parámetros que determinan la dimensión de nuestra red neuronal de convolución.

```
# CARGAR PARAMETROS
# Define los parametros base
# Numero de filtros maximo
filter<-7
# Tamaño del filtro maximo
kernel<-2
# Tamaño de pool maximo
pool<-3
# Numero de epocas
epochs <- 600
res<-data.frame(1:11)
```

Estos parámetros definen los límites de las posibles combinaciones que vamos a estudiar. Si a partir de dichas combinaciones no encontrásemos una que diese un resultado lo suficientemente bueno en el entrenamiento, aumentaríamos estos límites. Estos parámetros límite son:

- filter: en una capa convolucional, sería el número de filtros/máscaras a pasar por la imagen.
- kernel: en una capa convolucional, sería el tamaño del filtro/mascara a pasar por la imagen.
- pool: en una capa de pooling, sería el número de dimensiones a reducir.
- epochs: es el número de veces que se pasa el conjunto de datos de entrada por la red neuronal con el objetivo de entrenarla.
- res: variable vacía, en formato ‘dataframe’, para almacenar los resultados de las iteraciones de los bucles for.

Abrir bucles iterativos

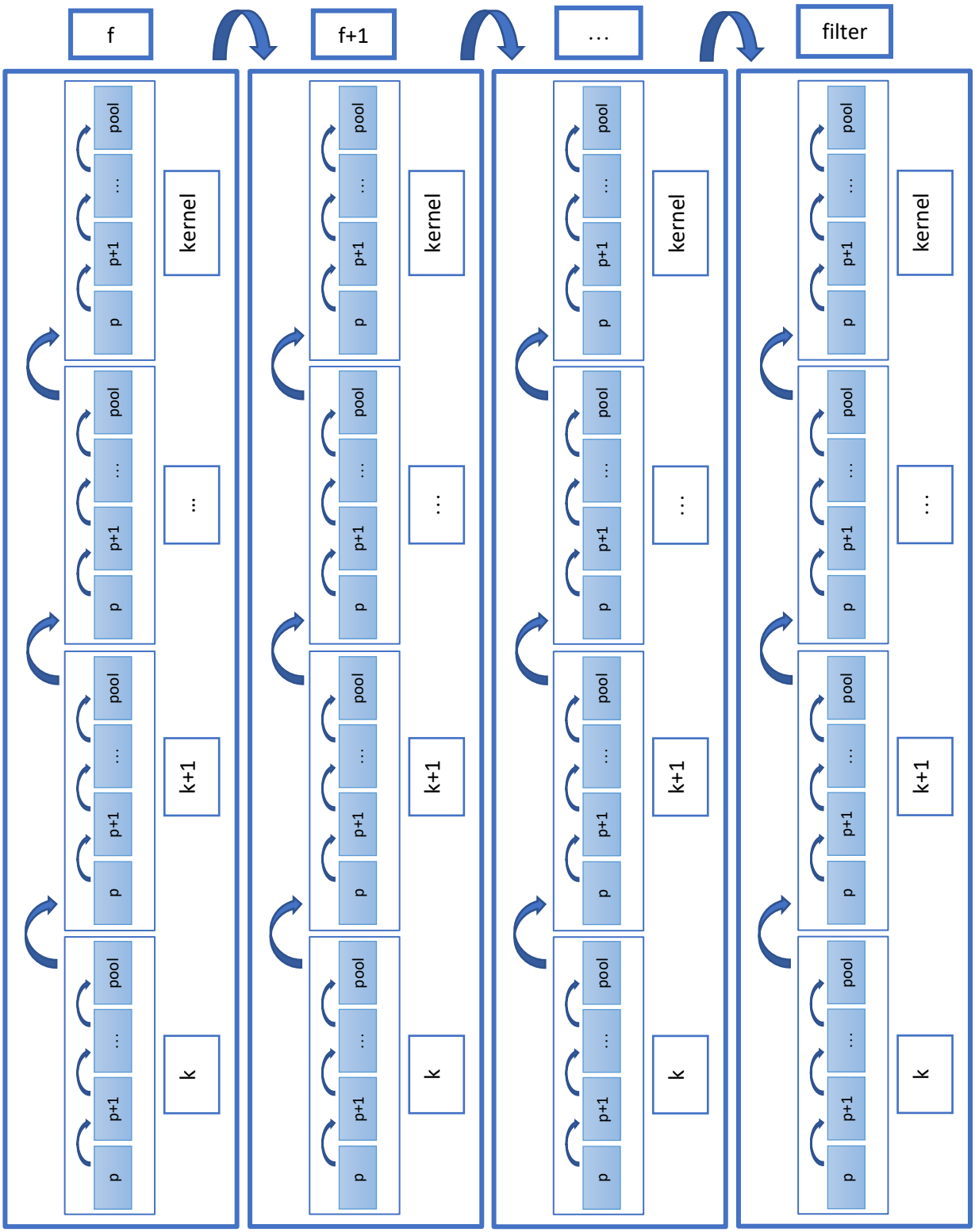


Ilustración 16. Descriptivo del desarrollo del bucle iterativo en el tiempo.

Como se puede ver en la imagen más arriba, se abren varios bucles for, anidados, que pretenden recorrer todas las posibilidades para poder analizarlas. Así, en cada iteración se entrenará una red de convolución diferente: con un valor de k (kernel), un valor de p (pool) y un valor de f (filter), parámetros que caracterizan la red. Se podrá pasar por todas las combinaciones posibles y entrenarlas, para averiguar que combinación es la óptima para componer la red.

```
for(layer1_f in 1:filter){
 for(layer1_k in 1:kernel){
  for(layer1_p in 1:(pool+1-layer1_k)){
   for(layer2_f in 1:filter){
    for(layer2_k in 1:kernel){
     for(layer2_p in 1:(pool+1-layer2_k)){
      for(layer3_f in 1:filter){
       for(layer3_k in 1:kernel){
        for(layer3_p in 1:(pool+1-k)){
```

Como ya se ha comentado en la definición de los parámetros, los límites de los bucles se han elegido en función del tiempo de procesado del conjunto de los bucles (aproximadamente 12h). Si no se obtuviese algún resultado de red neuronal lo suficientemente bueno, entonces se ampliarían los límites de los bucles. Para el caso del límite de pool, es necesario sustraer del mismo la dimensión kernel (tamaño de la máscara): se produce un error si el tamaño de la ventana de pool es más grande que el tamaño del resultado después de haber convolucionado la imagen.

Crear y compilar modelo

En este apartado se explica cómo se crea la función build_model. Por lo tanto, se genera esta función que, con tan solo llamarla, se crea un nuevo modelo.

```
# Crear modelo
# Una unica neurona de salida obtiene un valor de variable continua

build_model <- function() {

 model <- keras_model_sequential()%>%
  layer_conv_1d(filters=layer1_f, kernel_size=layer1_k,  activation = "relu", input_shape = c(dim(traini
ng_norm)[2],dim(training_norm)[3])) %>%
  layer_max_pooling_1d(pool_size = layer1_p)%>%
  layer_conv_1d(filters=layer2_f, kernel_size=layer2_k,  activation = "relu", input_shape = c(dim(traini
ng_norm)[2],dim(training_norm)[3])) %>%
  layer_max_pooling_1d(pool_size = layer2_p)%>%
  layer_conv_1d(filters=layer3_f, kernel_size=layer3_k,  activation = "relu", input_shape = c(dim(traini
ng_norm)[2],dim(training_norm)[3])) %>%
  layer_max_pooling_1d(pool_size = layer3_p)%>%
  layer_flatten() %>%
  layer_dense(units = 64, activation = "relu") %>%
  layer_dense(units = 1)


 model %>% compile(
  loss = "mse",
  optimizer = optimizer_rmsprop(),
```

```
    metrics = list("mean_absolute_error")

  )

  model
}
```

En la definición del modelo, el número de capas debe cambiarse manualmente, no puede cambiarse de manera automática mediante programación. Por lo tanto, se decide realizar el estudio con tres capas. Tres capas es un buen punto de partida dados los rangos de dimensión de las variables de entrada a los modelos. Si con tres capas no se obtuviese un entrenamiento lo suficientemente bueno, se procedería con una nueva iteración, empleando cuatro capas.

Para cada capa añadida de convolución, es posible añadir una capa de pooling posterior. En nuestro caso, podemos ver como se intercala la convolución con el pooling. Podemos ver como cada capa está determinada por los parámetros dados por los bucles anidados: "layerX_Y", siendo "X" la capa e "Y" el parámetro. Tras la sucesión de capas de convolución y pooling, encontramos una capa "flatten", o capa de aplanado. Esta capa hace el trabajo de quitar las dimensiones creadas en la red neuronal, de forma que la salida sea unidimensional. Por último, siguiéndola, encontramos dos capas sucesivas de red neuronal clásica, la primera con 64 neuronas (elegido al azar) y la segunda con 1 neuronal (número de salidas esperadas del modelo).

A pesar de haber incluido ambos términos dentro de la función, la compilación del modelo es independiente a la creación del modelo. El compilador sirve para traducir el lenguaje de programación a un lenguaje de bajo nivel, que habitualmente es lenguaje máquina.

Los parámetros característicos de la compilación del modelo son:

- Loss: función de pérdidas/función de coste empleada.
- Optimizer: optimizador de los pesos de las conexiones entre neuronas.
- Metrics: criterio de evaluación para el modelo.

De esta forma, cada vez que se escriba en la consola "build_model", se correrá el código interno de la función, creando y compilando el modelo que esté definido dentro.

Construir modelo

```
model <- build_model()
# Los pasos de creacion del modelo se envuleven en la funcion built_model, util
para poder crear mas modelos posteriormente
model %>% summary()

```

Una vez creada la función built_model, se llama a la misma, construyendo el modelo.

Entrenar modelo

Por razones meramente estéticas, vamos a mostrar un punto por cada época transcurrida en el entrenamiento del modelo. Por defecto se muestra de forma numérica. A continuación, se muestra el código que posibilita esta sustitución.

```
# Mostrar el progreso del entrenamiento mediante imprimir un solo punto por epoca completada
print_dot_callback <- callback_lambda(
  on_epoch_end = function(epoch, logs) {
    if (epoch %% 80 == 0) cat("\n")
    cat(".")
  }
)

```

Una vez determinado como se mostrará en pantalla la evolución del entrenamiento, incorporamos el código para entrenar el modelo, y especificamos los parámetros de dicho entrenamiento.

```
# Entrenar (fit) el modelo y guardar las estadísticas de entrenamiento
history <- model %>% fit(
  training,
  trainingtarget,
  epochs = epochs,
  validation_split = 0.2,
  verbose = 0,
  callbacks = list(print_dot_callback)
)
```

Los parámetros especificados son:

- Training: datos de entrada al modelo.
- Trainingtarget: datos objetivo del modelo, un objetivo para cada observación de los datos de entrada.
- Batch_size: número de observaciones por cada actualización del gradiente. Si no se define, se toma el valor por defecto que es 32.
- Epochs: una época es una iteración completa de todos los datos de entrada, con sus datos objetivo. Una vez alcanzado el número de épocas especificado por este parámetro, el entrenamiento del modelo finaliza.
- Validation_split: Fracción de los datos de entrenamiento que se utiliza para validar los resultados del entrenamiento. Esta fracción se mantiene al margen durante el entrenamiento y, al finalizar cada época, se evalúa el error de la misma. Esta fracción se selecciona comenzando por el final, tras el proceso de shuffle (mezclado).
- Verbose: indicar o no el progreso del entrenamiento. En nuestro caso, 0 indica en silencio.
- Callbacks: lista funciones que se pueden llamar durante el entrenamiento, para ver el progreso o estadísticas del modelo. Nosotros incorporamos el imprimir puntos para cada época transcurrida.
- Shuffle: valor booleano (True/False) que indica si se producirá o no un mezclado de los datos de entrenamiento antes de empezar cada época. Por defecto es toma el valor True.

Evaluar el entrenamiento del modelo

```
# EVALUAR EL MODELO
# Se evalua el desempeño del modelo con los datos del conjunto de test
c(loss, mae) %<-% (model %>% evaluate(test, testtarget, verbose = 0))
# Se muestra en pantalla el valor del error absoluto medio
paste0("Mean absolute error on test set: $", sprintf("%.2f", mae * 1000))

# Predicciones y matriz de confusion, de los datos del conjunto de test
#test_predictions <- model %>% predict(test)
#test_predictions[ , 1]
# Mostrar las predicciones del conjunto de test vs los objetivos del conjunto de test
#cbind(test_probability,test_predictions,testtarget)

```

Para cada iteración es posible conocer el resultado del entrenamiento en términos de:

1. Función de pérdidas o función de coste
2. Error medio absoluto del modelo

A continuación de registrar estos indicadores ya mencionados, se muestra a través de la consola el error absoluto medio obtenido, para facilitar su conocimiento al operador.

Adicionalmente, comentadas, se pueden encontrar las alternativas de visualizado de los resultados del entrenamiento. Entre ellas, mostrar en pantalla los resultados de predicciones a partir del conjunto de test y el display de una matriz cuyas columnas sean la probabilidad de acierto, el valor predecido y el valor objetivo.

```
# VISUALIZAR EL PROGRESO
# Se visualiza la evolucion del entrenamiento, una vez finalizado
library(ggplot2)
plot(history)

```

Otra forma de evaluar el entenamiento del modelo es comprobar gráficamente que la caída del error medio absoluto es exponencial a medida que se van completando las épocas en el entrenamiento. En las líneas de código se puede ver que el paso previo es cargar el paquete que alberga las funciones de grafiado en R. A continuación, se grafica la evolución del entrenamiento.

Más abajo se puede observar un entrenamiento con un número pequeño de épocas, en el que es posible percibir más fácilmente la caída exponencial ya mencionada.

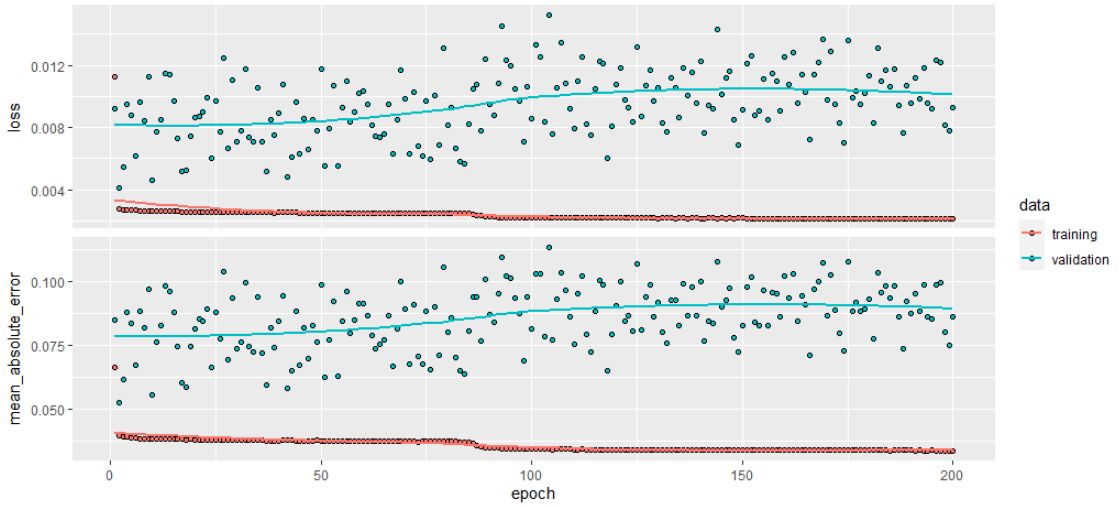


Tabla 11. Entrenamiento que aún no ha alcanzado el régimen permanente.

La elaboración de un modelo durante su entrenamiento sigue un patrón muy característico si realmente se está aprendiendo. Inicialmente, la caída del error medio es exponencial, y es precedida por una ligera subida. Estas oscilaciones se perpetúan a lo largo del tiempo. Con el paso de las épocas, las oscilaciones se amortiguan y se alargan en el tiempo, acercándose al régimen permanente más lentamente.

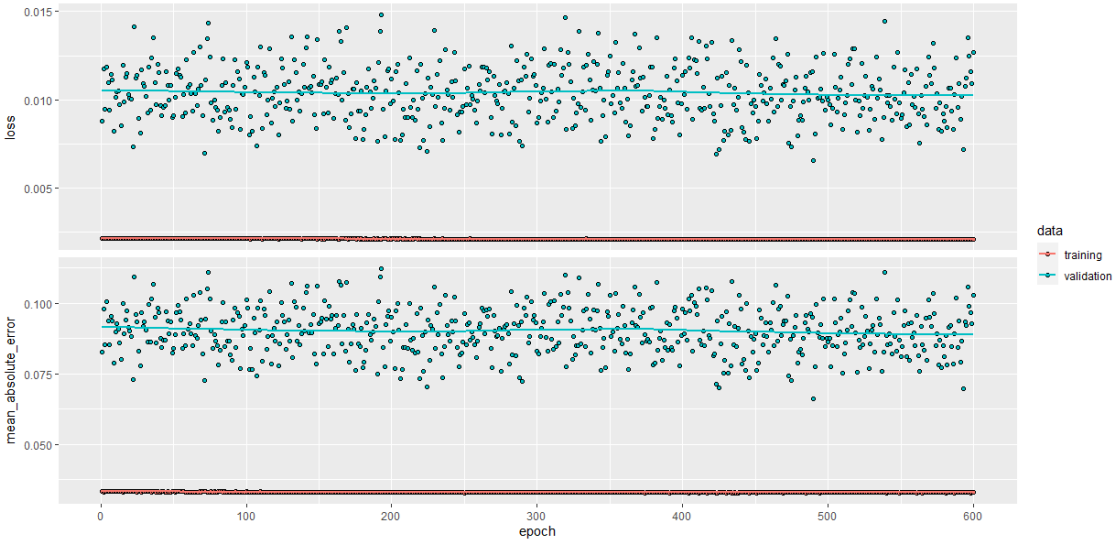


Tabla 12. Entrenamiento que ya ha alcanzado el regimen permanente.

Cerrar bucle iterativo

```
# Cerrar bucle iterativo, almacenando los resultados en la variable res

   vec<-data.frame(c(layer1_f, layer1_k, layer1_p,layer2_f, layer2_k, layer2_p, layer3_f, layer3_k, layer3_p,mae,loss))
   res<-cbind(res,vec)

             }

           }

         }

        }

       }

     }

   }

  }

}
```

Antes de cerrar cada iteración del bucle iterativo, es necesario registrar los indicadores ya mencionados (mae y loss), así como los parámetros utilizados para componer la red neuronal de esa iteración. Todos estos datos se registran en un vector "vec", el cual luego se une a un dataframe "res" en el que se recogen los resultados de todas las iteraciones.

Una vez registrados los datos, se puede cerrar el bucle respectivo del conjunto de bucles anidados para proceder con la siguiente iteración.

Exportar resultados numéricos del bucle iterativo

Una vez completadas todas las vueltas del bucle iterativo, lo cual puede tardar en torno a unas 12h, se procede a exportar los resultados almacenados en el dataframe "res".

```
# Exportar resultados de todas las iteraciones
setwd("E:/TFM/scripts/Estudio Potencia")
library(writexl)
write_xlsx(x = res, path = "W eje_Capa 3_Resultados_1.xlsx", col_names = TRUE)
setwd("E:/TFM/scripts")
```

En este caso, como ya hemos comentado, se está ejemplificando el código con el modelo de predicción de la potencia. Para registrar de forma permanente los resultados, se exportan a un documento Excel. Para llevar un procedimiento ordenado, se ha creado una carpeta independiente para almacenar los resultados de cada uno de los modelos. Por lo tanto, el primer paso consiste en cambiar el directorio de trabajo a la carpeta pertinente. Una vez hecho esto, se carga el paquete que contiene la lógica de exportación de una variable a un documento Excel.

A continuación, se exporta la variable "res" y se retorna al directorio de trabajo inicialmente definido.

Año/Year 1		Potencia	Temperatura de los anillos	Temperatura de la multiplicadora	Rodamiento DE	Rodamiento NDE
Capa/Layer 1	f	6	6	2	5	7
	k	2	2	1	1	2
	p	0	0	0	0	0
Capa/Layer 2	f	4	5	7	4	7
	k	1	1	1	1	1
	p	0	0	0	0	0
Capa/Layer 3	f	7	7	3	7	7
	k	2	1	2	2	1
	p	1	0	0	0	0

Tabla 13. Configuraciones óptimas de la red neuronal para cada modelo.

En la tabla se muestra un resumen de los resultados exportados del bucle iterativo. Contiene la combinación de parámetros óptima para cada modelo, parámetros que compondrán su red neuronal.

Redefinir el conjunto de test en función del año

Para analizar el comportamiento del aerogenerador en cada año, de forma independiente, es necesario redefinir el conjunto de test: del antiguo conjunto de datos a un nuevo conjunto de datos (que serán los datos correspondientes al año en que se va a emplear el modelo para analizar el aerogenerador).

```
# Dividir los datos
test <- data[year1_end: year2_end,3:7]
testtarget <- data[year1_end: year2_end, 2]

# Establecer el formato de los datos
test = data.matrix(test)
testtarget = data.matrix(testtarget)


# Normalizar los datos
test_norm<-normalizeData(test, type="0_1")
testtarget_norm<-normalizeData(testtarget, type="0_1")


# Ajustar la dimension
test_norm=array_reshape(test_norm, c(dim(test_norm),1))
test=array_reshape(test, c(dim(test),1))
```

Se sigue el mismo procedimiento empleado anteriormente para preparar el conjunto de datos, solo que en esta ocasión solo se preparará el conjunto de test: dividir los datos, establecer el formato de los datos, normalizar los datos y ajusta la dimensión. Para conocer en detalle cada uno de estos pasos, se recomienda recapitular a los apartados precedentes, más detallados.

Calcular las predicciones del conjunto de entrenamiento

Una vez el modelo está entrenado, habiendo utilizado la red neuronal optima, podemos empezar a utilizar el modelo para predecir.

```
# PREDICCIONES DEL CONJUNTO DE ENTRENAMIENTO
predictions_training_norm <- model %>% predict(training_norm)
#predictions_training_norm[ , 1]
# Mostrar las predicciones frente a los valores objetivo
#cbind(predictions_test_norm,testtarget_norm)

# Resultados de la NN están normalizados. Se convierten a unidades reales
predictions_training_denorm <-denormalizeData(predictions_training_norm,
                              getNormParameters(target_norm))

target_training_denorm<- denormalizeData(trainingtarget_norm,
                               getNormParameters(target_norm))

# plot(pred_tr_denorm)
```

En esta ocasión, mostramos primero la línea de código utilizada para predecir los valores objetivo del modelo. Comentadas a continuación, se pueden encontrar dos formas de comprobar los resultados de la predicción: mostrar en pantalla las predicciones o mostrar en pantalla las predicciones frente a los valores objetivo, en dos columnas adyacentes.

A continuación, se convierte de unidades normalizadas a unidades reales tanto las predicciones como los valores objetivo. También, comentada, aparece la posibilidad de graficar estas predicciones en unidades reales.

Dibujar los gráficos de residuos del conjunto de entrenamiento

Una vez obtenidas las predicciones y haberlas escalado, se procede a dibujar los gráficos necesarios para analizar los resultados.

```
#Errores de prediccion del conjunto de entrenamiento
error_tr<-target_training_denorm - predictions_training_denorm

#Histograma de residuos del conjunto de entrenamiento
ggplot(data=as.data.frame(error_tr), mapping= aes(x=error_tr))+
  geom_histogram(binwidth=0.5, col=c('blue'))
```

Se empieza calculando los residuos de la regresión, es decir, la diferencia entre el valor predicho y el valor objetivo para cada muestra. Una vez calculados, se procede a graficarlos, esperando

que sigan una distribución normal centrada en 0 (lo cual validaría la hipótesis de normalidad de las variables del modelo).

Dibujar la regresión del conjunto de entrenamiento

Una vez obtenidas las predicciones y haberlas escalado, se procede a dibujar los gráficos necesarios para analizar los resultados.

```
#REGRESIÓN
#Estimación de R^2 para el conjunto de entrenamiento
num<-sum((error_tr)^2)
den<-sum((target_training_denorm-mean(trainingtarget))^2)

R2_tr<-1-(num/den)
R2_tr_adjust<-(1-(1-R2_tr)*(nrow(target_training_denorm)-1)/(nrow(target_training_denorm)
                                    -ncol(target_training_denorm)-1))
```

El primer paso es calcular los indicadores del ajuste de la regresión. Una vez hecho esto, se procede a dibujar la regresión.

```
# Comparar las predicciones vs los valores reales
plot(target_training_denorm,predictions_training_denorm,col='red',
    main='Training: Real vs predicted NN',pch=18,cex=0.7)
abline(0,1,lwd=2)

#Plot del error de la regresion
plotRegressionError(trainingtarget_norm,predictions_training_norm,pch=3)
```

Para el conjunto de entrenamiento se hace un doble ejercicio: se dibuja la recta de ajuste real sobre la nube de puntos y se dibuja la recta de ajuste óptima sobre la nube de puntos. El objetivo es comparar visualmente la diferencia de inclinación entre ambas.

Dibujar las predicciones frente a los valores reales del conjunto de entrenamiento

Una vez obtenidas las predicciones y haberlas escalado, se procede a dibujar los gráficos necesarios para analizar los resultados.

```
# Graficar valores reales y valores predichos
plot(target_training_denorm,type = "p",col = "red", xlab = "Sample",
    ylab = "ANG_aero_pot_total_val Value",
    main = "Training: Real (red) - Predicted (blue)")

## lines(test$pred, type = "p", col = "blue") #  opcion para añadir las predicciones
lines(predictions_training_denorm, type = "p", col = "blue")
```

En este caso, se dibujan en el eje temporal tanto los valores reales como las predicciones, en rojo y verde respectivamente. El dibujo se hace sobre el mismo gráfico, con ambas variables superpuestas. Es una buena manera de identificar visualmente, por ejemplo, la naturaleza cíclica de los datos o la existencia de outliers.

Calcular las predicciones del conjunto de test

Una vez el modelo está entrenado, habiendo utilizado la red neuronal optima, podemos empezar a utilizar el modelo para predecir. Cabe recordar que antes se debe haber redefinido el conjunto de test (véase el apartado pertinente).

```
# PREDICCIONES DEL CONJUNTO DE TEST
predictions_test_norm <- model %>% predict(test_norm)

# Resultados de la NN están normalizados
# Se escalan a unidades reales.
predictions_test_denorm <-denormalizeData(predictions_test_norm,
                          getNormParameters(target_norm))

target_test_denorm<- denormalizeData(testtarget_norm,
                           getNormParameters(target_norm))
```

En esta ocasión, mostramos primero la línea de código utilizada para predecir los valores objetivo del modelo. A continuación, se convierte de unidades normalizadas a unidades reales tanto las predicciones como los valores objetivo.

Dibujar los gráficos de residuos del conjunto de test

Una vez obtenidas las predicciones y haberlas escalado, se procede a dibujar los gráficos necesarios para analizar los resultados.

```
# Errores de prediccion del conjunto de test
error_ts<-target_test_denorm-predictions_test_denorm
plot(error_ts)

# Histograma de residuos del conjunto de test
ggplot(data=as.data.frame(error_ts), mapping= aes(x=error_ts))+
 geom_histogram(binwidth=0.5, col=c('blue'))
```

Se empieza calculando los residuos de la regresión, es decir, la diferencia entre el valor predicho y el valor objetivo para cada muestra. Una vez calculados, se procede a graficarlos:

1. Histograma: esperando que los residuos sigan una distribución normal centrada en 0 (lo cual validaría la hipótesis de normalidad de las variables del modelo).
2. Eje temporal: esperando inferir patrones que nos ayuden en el análisis del estado del aerogenerador.

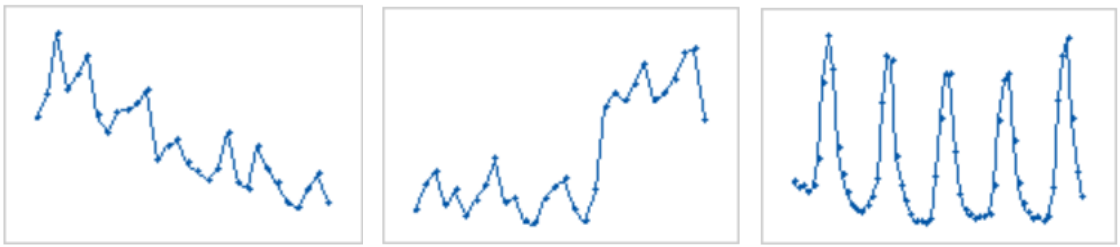


Ilustración 17. De izquierda a derecha: línea de tendencia, línea de cambio y línea cíclica.

Para una visualización mas clara de los resultados, es posible dibujar las gráficas junto con el intervalo de confianza superpuesto. Resulta más intuitivo para ver como las desviaciones se salen del rango del intervalo. Se puede encontrar el código a continuación.

```
#Histograma de residuos del conjunto de test
hist(error_ts)
abline(v=min_ts,col="red",lwd=3, lty=2)
abline(v=max_ts,col="red", lwd=3, lty=2)

#Errores de prediccion del conjunto de test
plot(error_ts)

abline(h=min_ts,col="red", lwd=3, lty=2)
abline(h=max_ts,col="red", lwd=3, lty=2)
```

Dibujar la regresión del conjunto de test

Una vez obtenidas las predicciones y haberlas escalado, se procede a dibujar los gráficos necesarios para analizar los resultados.

```
#REGRESIÓN
# Estimación de R^2 para el conjunto de test
num<-sum((error_ts)^2)
den<-sum((target_test_denorm-mean(testtarget))^2)

R2_ts<-1-(num/den)
R2_ts_adjust<-(1-(1-R2_ts)*(nrow(target_test_denorm)-1)/(nrow(target_test_denorm)
                                    -ncol(target_test_denorm)-1))
```

El primer paso es calcular los indicadores del ajuste de la regresión. Una vez hecho esto, se procede a dibujar la regresión.

```
# Comparar las predicciones vs los valores reales
plot(target_test_denorm,predictions_test_denorm,col='red',
     main='Test Real vs predicted NN',pch=18,cex=0.7)
abline(0,1,lwd=2)
```

Para el conjunto de test se dibuja la recta de ajuste real sobre la nube de puntos. El objetivo es comparar visualmente el posicionamiento de la recta de ajuste sobre la nube de puntos, y la dispersión de la misma.

Dibujar las predicciones frente a los valores reales del conjunto de test

Una vez obtenidas las predicciones y haberlas escalado, se procede a dibujar los gráficos necesarios para analizar los resultados.

```
# Graficar valores reales y valores predichos
plot(target_test_denorm,type = "p",col = "red", xlab = "Sample",
    ylab = "ANG_aero_pot_total_val Value",
    main = "Test: Real (red) - Predicted (blue)")

## lines(test$pred, type = "p", col = "blue") #  opcion para añadir las predicciones
lines(predictions_test_denorm, type = "p", col = "blue")
```

En este caso, se dibujan en el eje temporal tanto los valores reales como las predicciones, en rojo y verde respectivamente. El dibujo se hace sobre el mismo gráfico, con ambas variables superpuestas. Es una buena manera de identificar visualmente, por ejemplo, la naturaleza cíclica de los datos o la existencia de outliers.

En el caso que se esté usando esta función para la visualización de anomalías, una vez creado el vector de anomalías, es posible superponerlas a las predicciones y valores reales, en el eje temporal. Resulta un método bastante intuitivo para conocer además de su localización en el tiempo, el valor aproximado de las mismas.

```
# Graficar valores reales, valores predichos y anomalias
plot(target_test_denorm,type = "p",col = "red", xlab = "Sample",
    ylab = "Value",
    main = "Real (red) - Predicted (blue) - Anomaly (green)")
## lines(test$pred, type = "p", col = "blue") #  opcion para añadir las predicciones y anomalias
lines(predictions_test_denorm, type = "p", col = "blue")
lines(anom,type = "p", col = "green")
```

Calcular los intervalos de confianza del modelo

Una vez dibujados los gráficos necesarios para analizar los resultados, se finaliza calculando el intervalo de confianza de los resultados del modelo (véase más información en el apartado pertinente). Se calculará tanto el intervalo de confianza del conjunto de entrenamiento como el intervalo de confianza del conjunto de test.

```
#Calculo de estadisticos del conjunto de entrenamiento
sigma_tr<-sd(error_tr, na.rm = FALSE)
var(error_tr, na.rm = FALSE)
media_tr<-mean(error_tr)
#Calculo de limites del intervalo de confianza del conjunto de entrenamiento
min_tr<-(-1.96*sd(error_tr, na.rm = FALSE))+mean(error_tr)
max_tr<-(1.96*sd(error_tr, na.rm = FALSE))+mean(error_tr)

#Numero de muestras fuera del intervalo de confianza del conjunto de entrenamiento
count_tr<-0
for(t in 1:nrow(error_tr)){
```

```
    if(error_tr[t]<min_tr){
      count_tr<-count_tr+1

    }else if(error_tr[t]>max_tr){
      count_tr<-count_tr+1

    }
 }

percent_tr<-(count_tr/nrow(error_ts))*100

#Calculo de estadisticos del conjunto de test
sigma_ts<-sd(error_ts, na.rm = FALSE)
var(error_ts, na.rm = FALSE)
media_ts<-mean(error_ts)
#Calculo de limites del intervalo de confianza del conjunto de test
min_ts<-(-1.96*sd(error_ts, na.rm = FALSE))+mean(error_ts)
max_ts<-(1.96*sd(error_ts, na.rm = FALSE))+mean(error_ts)


#Numero de muestras fuera del intervalo de confianza del conjunto de test
count_ts<-0
for(t in 1:nrow(error_ts)){
 if(error_ts[t]<min_ts){
  count_ts<-count_ts+1

 }else if(error_ts[t]>max_ts){
  count_ts<-count_ts+1

 }
}

percent_ts<-(count_ts/nrow(error_ts))*100
```

Como podemos ver, para los dos casos se sigue el mismo procedimiento:

1. Calcular los estadísticos de la muestra, en este caso el conjunto de entrenamiento o el conjunto de test.
2. Calcular los límites del intervalo de confianza, con un nivel de confianza del 95%. Véase el apartado pertinente para una explicación mas completa.
3. Contar el número de observaciones que se encuentran fuera del intervalo de confianza.
4. Calcular el porcentaje de las observaciones de la muestra que se encuentran fuera de la misma.

Exportar los resultados gráficos del modelo

Por último, es necesario almacenar de forma permanente el análisis del comportamiento del modelo. En este caso, utilizando el entorno de trabajo R, se puede exportar en formato imagen los gráficos obtenidos.

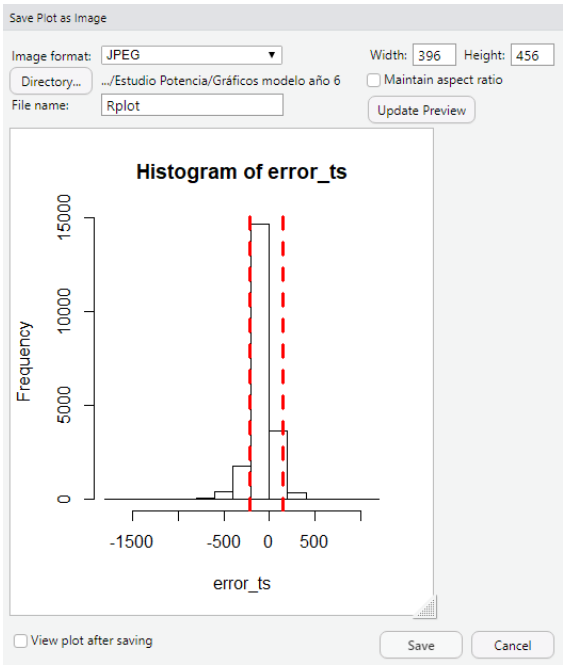


Ilustración 18. Detalle de la pestaña en Rstudio para exportación de gráficas.

Esta es una opción que provee el propio entorno de Rstudio a través de una pestaña situada sobre el visor de gráficos. Permite seleccionar la carpeta donde se quiere salvaguardar la imagen, el nombre del archivo, el tipo de archivo imagen que se va a generar y la dimensión de píxeles de la imagen a generar.

Identificación y registro de las anomalías del conjunto de las muestras anómalas

```
#Creación del vector anomalias
#Se registran los valores de las muestras anómalas en un vector vacio
anom<- numeric(0)
for(t in 1:nrow(error_ts)){
  if(error_ts[t]<min_ts){
    anom[t]<-target_test_denorm[t]

  }else if(error_ts[t]>max_ts){
    anom[t]<-target_test_denorm[t]

  }
}

#Se eliminan las muestras anómalas que no se consideran anomalías
for(t in 1:nrow(error_ts)){
  if( !is.na(anom[t]) && (!is.na(anom[t+1]) || !is.na(anom[t+2]))){

  }else{
    is.na(anom)<-t
  }
}
```

No todas las muestras anómalas se consideran anomalías. Para considerar anomalía una muestra anómala se intercepta un grupo de tres muestras seguidas, y si al menos dos de ellas son anómalas, entonces esas muestras anómalas se considerarán anomalías.

Como se puede ver en el código, se ha creado un vector vacío, y para los valores de error (desviación entre la predicción y el valor real medido), se registra en el vector vacío el valor de la medida anómala. Se registra en la misma posición de vector en que se sitúan el error y la medida desviada en sus respectivos vectores. Este será el vector que incluya las medidas anómalas.

A continuación, se chequea de tres en tres que al menos dos de cada tres espacios dentro del vector de medidas anómalas estén ocupados. En este caso, se considera que las tres medidas anómalas son anomalías, por lo que no se actúa sobre el vector de anomalías. En el caso que solo se encuentre un espacio ocupado en un grupo de tres espacios seguidos, se considera que la medida anómala que ocupa dicho espacio no es una anomalía, y por lo tanto se vacía.

Contabilización de anomalías

Se procede a contabilizar las anomalías registradas y computar el porcentaje de anomalías que presenta cada mes del año.

```
#Indicar el año en el que estas trabajando
year<-2
month<-0
b<-match(data$data_complete.ANG_fecha[1]+months(month)+years(year)
, data$data_complete.ANG_fecha)-year2_end

#Bucle que correr 12 veces, 1 por cada mes del año
a<-b+1
month<-month+1
b<-match(data$data_complete.ANG_fecha[1]+months(month)+years(year)
, data$data_complete.ANG_fecha)-year2_end
count_ts_anom<-0
count_ts_obs<-0
for(t in a:b){
 if(!is.na(anom[t])){
  count_ts_anom<-count_ts_anom+1
 }

 if(error_ts[t]<min_ts){
  count_ts_obs<-count_ts_obs+1

 }else if(error_ts[t]>max_ts){
  count_ts_obs<-count_ts_obs+1

 }


}


percent_ts_obs<-(count_ts_obs/t)*100
percent_ts_anom<-(count_ts_anom/t)*100

vec<-data.frame(c(month,year,count_ts_obs,percent_ts_obs,count_ts_anom,percent_ts_anom))
res<-cbind(res,vec)
```

Como podemos ver, las primeras cuatro líneas de código corresponden a la inicialización de variables. Había que modificar el año cada vez que se trabaje con ello. Recordemos que la obtención predicciones y análisis de anomalías se realiza de manera anual.

También habría que modificar el 'yearX_end' en la actualización de la variable b. Para comprender mejor esta línea de código, véase el apartado 'Redefinir en conjunto de test en función del año'.

A continuación, tanto el bucle como el registro de resultados en la variable 'res' deben correrse 12 veces seguidas, tantas como meses tiene el año. Esto es necesario porque para cada mes, se computan las anomalías y el porcentaje de anomalías mensual. El bucle está diseñado para que él solo se reajuste automáticamente en cada iteración.

Una vez registrados los resultados mensuales, para un año determinado, en la variable 'res', se puede exportar a una hoja Excel. Véase el apartado 'Exportar resultados numéricos del bucle iterativo'

Capítulo 5.- Construcción del modelo

Antes de proceder a detectar anomalías, es necesario demostrar que el modelo es robusto. Este paso previo es indispensable y se realizará utilizando los datos pertenecientes al primer año.

Para ello, es esta sección, comenzaremos explicando la estrategia seguida para determinar si un modelo es robusto. A continuación, se hará una introducción de cada modelo explicando su objetivo, como se construye y para qué sirve. Después, se demostrará que ese modelo en concreto es robusto.

Una vez completada la justificación de la construcción de cada modelo y demostrada su robustez, se podrá proceder a identificar y cuantificar las anomalías que se encuentran en los elementos a los que hace referencia cada modelo.

5.1.- Estrategia de determinación de la robustez del modelo

Para determinar la robustez del modelo se hace una división de los datos pertenecientes al primer año: los datos que se emplearán para el entrenamiento son un 70% de los datos del primer año, y los datos que se emplearán para comprobar la robustez del modelo son el 30% restante.

El número de épocas empleadas para el entrenamiento varía en función del modelo y su respuesta al entrenamiento: se emplearán tantas épocas como sean necesarias hasta poder comprobar visualmente que el gráfico que dibuja los errores de entrenamiento para cada una de las épocas se ha estabilizado. Aumentar el número de épocas no necesariamente mejora los resultados del modelo: una vez el modelo está suficientemente ajustado, seguir aumentando el número de épocas da lugar a 'overfitting': efecto que provoca que el modelo se ajuste tanto a la nube de puntos que, en lugar de obtener la recta que mejor se ajusta a los puntos, se obtiene la recta que pasa por todos los puntos. Cabe recordar que dentro del propio set de entrenamiento se hace una partición para validar cada época de entrenamiento y así ir ajustando el modelo: un 20% de los datos empleados en el entrenamiento se emplean para validación.

Una vez el modelo está suficientemente entrenado, se comprueba su robustez. Para ellos, se utilizarán los métodos enumerados más abajo.

La robustez se validará con respecto a los dos conjuntos de datos: el conjunto de datos de entrenamiento (70% de los datos del primer año) y el conjunto de datos de test (30% de los datos del primer año). Una vez se haya demostrado que el modelo es robusto, se podrá utilizar para predecir los valores esperados en los años sucesivos.

Los métodos de análisis de la robustez de un modelo de comportamiento son que se explican a continuación, y que se emplearán en este apartado:

1. Histograma de errores

Para verificar que el modelo es robusto se grafica un histograma de errores. Estos errores tienen que distribuirse siguiendo una distribución normal y dicha estar centrada en 0. Se verificará visualmente.

A continuación, se calculará un intervalo de confianza que recoja un porcentaje específico del total de los datos (los datos serán los errores), comprendido entre dos valores simétricos con respecto a la media. De esta forma se puede asegurar que el porcentaje elegido de datos se encuentra entre esos valores, y se puede cuantificar la cantidad de muestras que se encuentran fuera de los límites del intervalo. Esto es útil para conocer la imperfección del propio modelo. Así, cuando el modelo se emplee utilizando un nuevo set de datos de entrada, la desviación o error entre predicción y valor real es la suma de la desviación debida a la imperfección del modelo y la desviación debida a problemas en el aerogenerador.

Para ello se utilizará la normal estándar o tipificada N(0,1). Esta es una distribución normal que tiene media 0 y varianza 1, y que se denota mediante la variable Z. La ventaja de utilización de esta es que función de distribución $\Phi(z)$ está tabulada.

Estableciendo un intervalo de confianza de un nivel de confianza del $\delta = 95\% \rightarrow \alpha = 0{,}05$

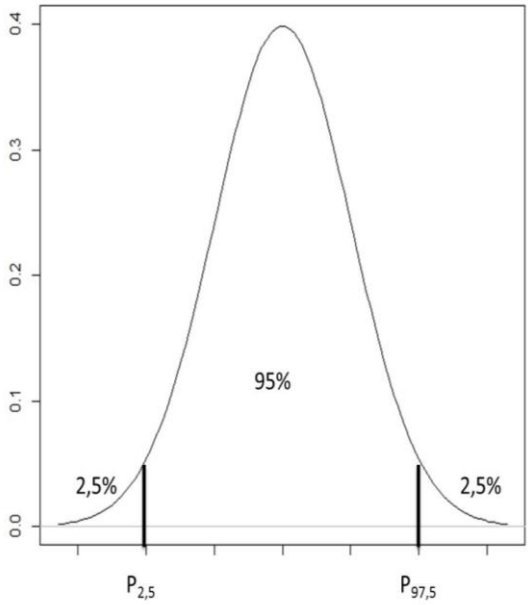


Tabla de la distribución normal N(0,1) para probabilidad acumulada inferior

μ = Media

σ = Desviación típica

$P(z \le z_0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_0} e^{\frac{-z^2}{2}} dz$

Tipificación: $z_0 = \frac{x-\mu}{\sigma}$

z_0	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	z_0
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359	0,0
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753	0,1
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141	0,2
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517	0,3
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879	0,4
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224	0,5
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549	0,6
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852	0,7
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133	0,8
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389	0,9
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621	1,0
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830	1,1
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015	1,2
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177	1,3
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319	1,4
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441	1,5
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545	1,6
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633	1,7
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706	1,8
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767	1,9

Tabla 14. Tabla detalle de cálculo del intervalo de confianza del 95 %.

$$P(Z \le z_0) = \Phi(z) \rightarrow Para\ \Phi(z) = 0{,}975 \rightarrow\ z_0 = z_{\alpha/2} =\ 1{,}96$$

$$z_{\alpha/2} =\ 1{,}96\ ;\ -z_{\alpha/2} =\ -1{,}96$$

Una vez obtenidos los valores para la normal estándar N(0,1), se lleva a cabo la destipificacion de la variable, que consiste en devolverla nuestra distribución normal efectuando un cambio de variable.

$$Z \sim N(0,1)\ ;\ Z = \frac{X-\mu}{\sigma} \rightarrow X \sim N(\mu, \sigma)$$

Así, en función de los valores μ y σ que obtengamos del conjunto de los errores que conforman el histograma, obtendremos unos límites para el intervalo de confianza del 95%. Y con estos límites, podremos cuantificar la cantidad de muestras que quedan fuera de los mismos.

2. Regresión

Las redes neuronales al final son modelos aditivos de regresión no lineales. Por ello, se pueden utilizar los indicadores clásicos de regresión para controlar la robustez del modelo. Si la regresión es robusta, el modelo lo será también.

La línea de regresión es la línea que mejor se ajusta a la nube de puntos, explicando la relación entre las entradas o variables independientes, el eje x, y las salidas o variables dependientes, el eje y.

El coeficiente de determinación R^2 mide la dispersión de la nube de puntos alrededor de la línea de regresión. Su valor se encuentra entre 0 y 1, 0 indicando una dispersión muy elevada y 1 indicando que la línea se ajusta perfectamente a los puntos. En una regresión múltiple, R^2 representa la variabilidad de la salida que explican las variables de entrada: el porcentaje de variabilidad de la salida explicada por el modelo de regresión. El coeficiente de determinación ajustado $\overline{R^2}$ tiene las mismas propiedades que R^2 pero penaliza la inclusión de variables de entrada que no son significativas, es decir, que se trata se una medida más representativa de la bondad del ajuste real.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

$$\overline{R^2} = 1 - (1 - R^2) * \frac{n-1}{n-k-1}$$

$$e_i \; error \; depredicción \rightarrow e_i = valor \; real - predicción$$

$$y_i \; valor \; real$$

$$y_i \; media \; de \; los \; valores \; reales$$

$$n \; número \; total \; de \; valores \; reales$$

$$k \; número \; de \; variables \; que \; predecimos$$

Se graficarán la predicción frente al valor real para cada una de las muestras. Se hará con los valores a escala, por una parte, y con los valores normalizados por otra. En este último caso, para representar junto a la recta de ajuste real (en rojo), para la nube de puntos, la que sería la recta de ajuste óptima (en negro). La recta de ajuste real se genera para la nube de puntos, mientras que la recta de ajuste óptima es la recta que pasaría por 0 con pendiente 1.

De esta forma, se podrá verificar visualmente la diferencia entre valores reales y predicciones, la existencia de outliers y el desvío de la recta de regresión.

5.3.- Modelo de previsión de la potencia

Para el modelo de previsión de la potencia, se va a proceder a describir objetivo, utilidad y construcción del mismo. A continuación, se procederá a demostrar su robustez, para poder emplearlo de manera segura y certera en los años venideros.

Objetivos y especificación

Este modelo es el caso base, a partir del cual se realizarán pequeñas modificaciones para adaptarlo a los distintos posibles detectores de anomalías a construir.

El objetivo de este modelo es determinar la potencia que debería tener el aerogenerador en un determinado momento. La potencia es un claro indicador del estado de operación del aerogenerador, si opera con regularidad o no. Por lo tanto, monitorizando su potencia real y siendo capaces de predecir la potencia que debería tener, podemos calcular la desviación entre ambas.

La desviación se considerará grande cuando se encuentre en fuera del intervalo de confianza del 95% del error creado a partir de los datos del primer año. Si la desviación es suficientemente grande y se suceden tres de ellas seguidas, se podrá considerar que la observación es una anomalía. Estas anomalías se podrán graficar y cuantificar.

Esto servirá para:

- Monitorizar el estado del aerogenerador en tiempo real, a través de su potencia.
- Identificar que una anomalía ha aparecido durante la operación del aerogenerador.
- Tener una visión global de las anomalías: a nivel mensual, anual…
- Dar soporte a los demás modelos de cara a identificar un posible tipo de fallo.

El modelo se construye en función de las condiciones de viento observadas. También se incluyen las condiciones de operación en las que se encuentra el aerogenerador en ese momento. Se va a predecir la potencia:

- ANG_aero_pot_total_val

a partir de las variables:

- ANG_aero_tem_ambiente_val
- ANG_aero_vel_viento_val
- ANG_aero_angulo_viento_val
- ANG_aero_tem_rod_DE_gen_val
- ANG_aero_tem_rod_NDE_gen_val

Robustez del modelo

Para el caso base de análisis de la potencia, se debe recordar que para entrenar el modelo se utilizarán 23897 observaciones (el 70% de los datos del primer año), y que para verificar la robustez del modelo se utilizarán 10197 observaciones (el 30% restante de los datos del primer año). Cabe destacar que la selección de los datos del primer año para entrenamiento o para test se realiza de forma aleatoria y atemporal.

Para entrenar el modelo, se observa que es preciso utilizar un total de 600 épocas. A continuación, se encuentra el gráfico que muestra la estabilización del error medio absoluto y la caída de la función de pérdida.

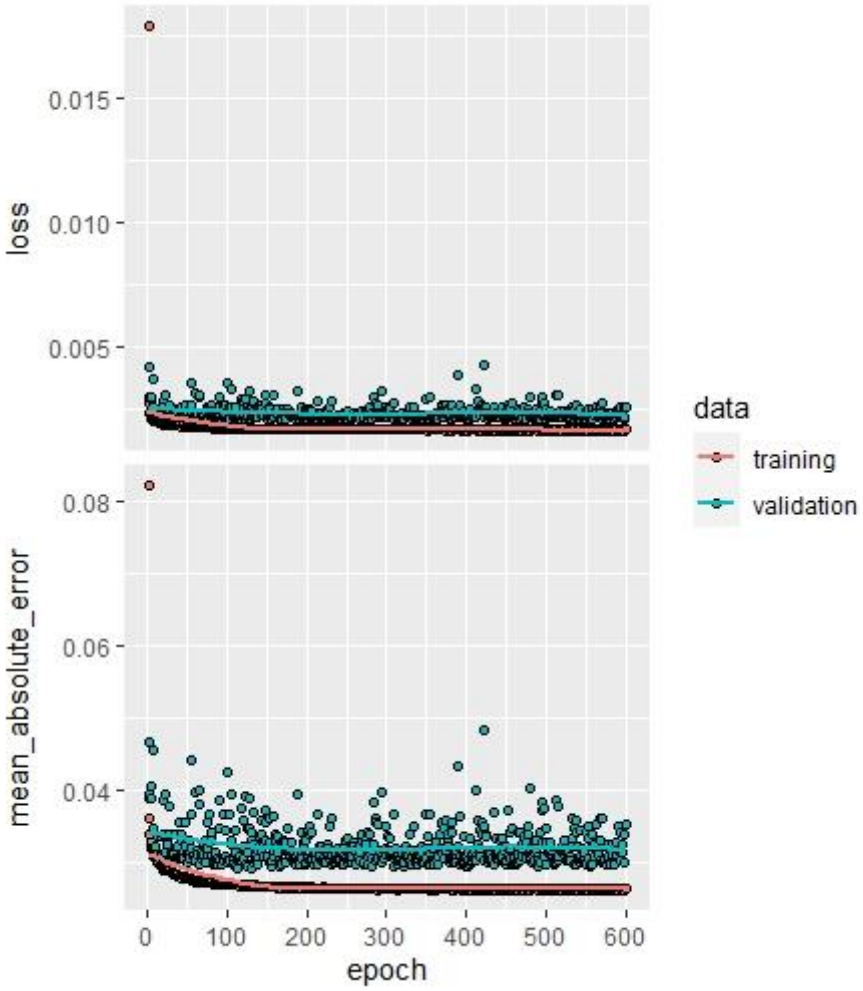


Ilustración 19. Función de coste y función de pérdidas del modelo de previsión de la potencia.

Tras el entrenamiento, se analiza en profundidad el modelo utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos parámetros del entorno de trabajo se pueden encontrar a continuación, y así como las gráficas obtenidas del modelo de regresión utilizado con los dos conjuntos de observaciones:

- Gráficos del conjunto de datos de entrenamiento

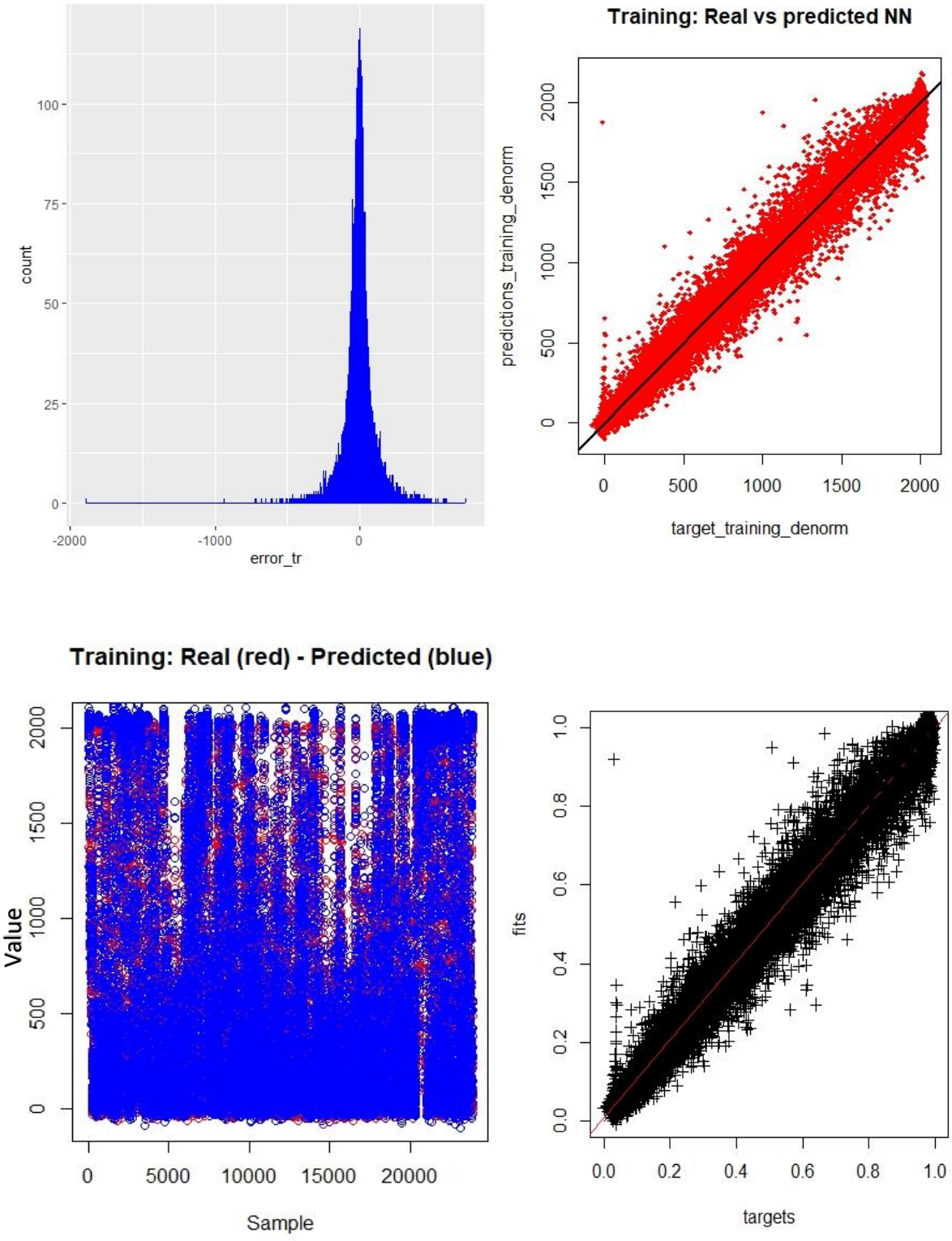


Ilustración 20. Resultado de robustez del entrenamiento del modelo de previsión de la potencia.

- Gráficos del conjunto de datos de test

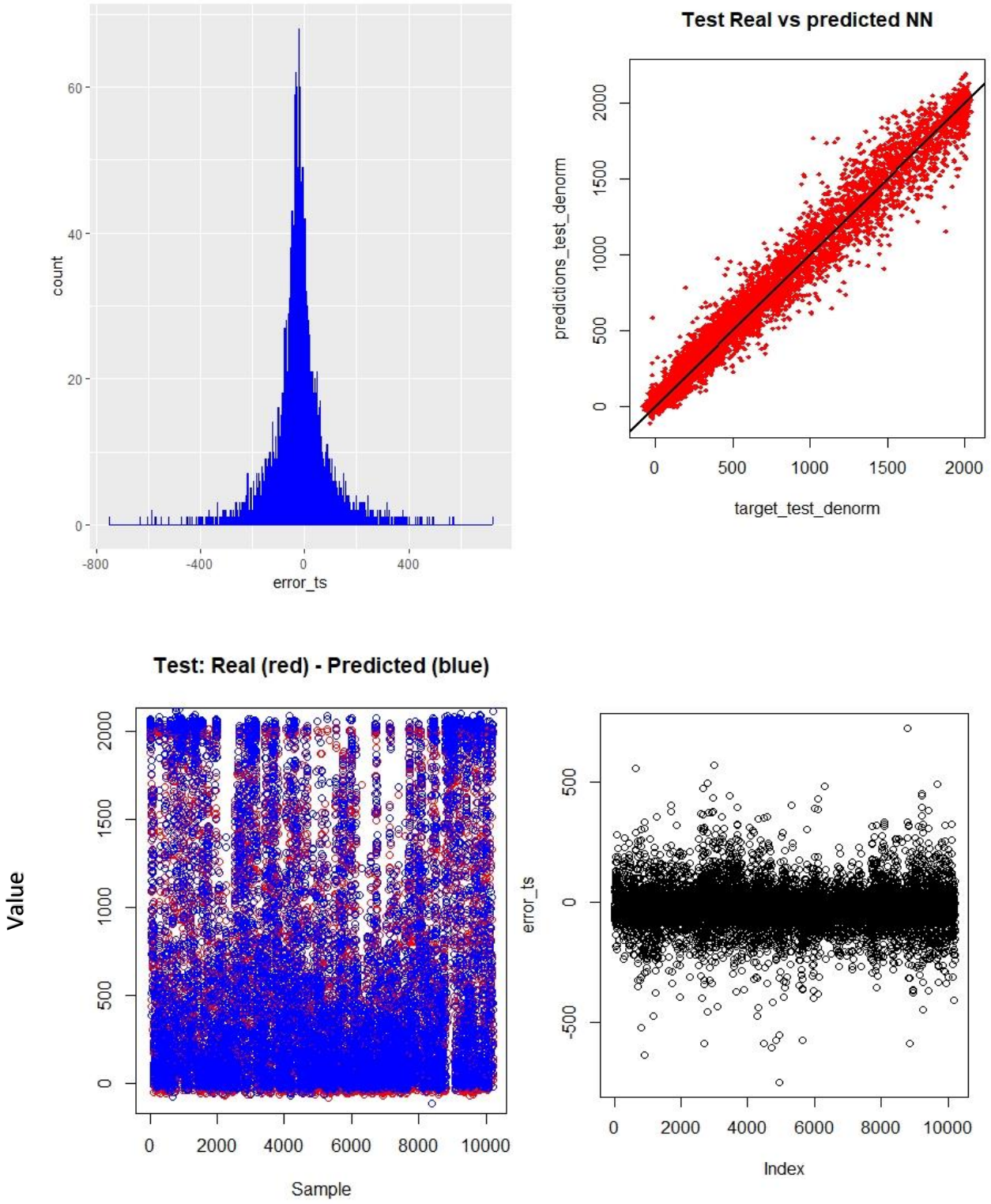


Ilustración 21. Resultado de robustez del test del modelo de previsión de la potencia.

1. Histograma

Recordando que el primer año se ha dividido en dos subconjuntos, entrenamiento y test, el número de muestras que consideraremos que el modelo ha predicho mal será igual a la suma de ambos. Concretamente, se calcularán los intervalos de confianza del 95% para cada subconjunto, se cuantificarán las muestras fuera de el intervalo para cada subconjunto, y se sumarán. Estas muestras fuera de los intervalos se consideran predicciones incorrectas. De esta forma, la desviación que aporta el modelo de forma adicional por su imperfección se podrá medir mediante el número de predicciones incorrectas para el primer año.

- Conjunto de entrenamiento

La media obtenida para este conjunto es μ = -27,807

La desviación típica obtenida para este conjunto es σ = 89,166

$$Z \sim N(0,1)\ ;\ Z = \frac{X-\mu}{\sigma} \rightarrow X \sim N(\mu,\sigma)$$

Destipificando, obtenemos que

$$Z = \frac{X-\mu}{\sigma} \rightarrow 1{,}96 = \frac{X-(-27{,}807)}{89{,}166} \rightarrow\ X = 146{,}959$$

$$Z = \frac{X-\mu}{\sigma} \rightarrow -1{,}96 = \frac{X-(-27{,}807)}{89{,}166} \rightarrow\ X = -202{,}574$$

$$P[-202{,}574 \leq X \leq 146{,}959] = 0{,}95$$

El número de muestras fuera del intervalo en este subconjunto es de 1517 muestras.

- Conjunto de test

La media obtenida para este conjunto es μ = -42,311

La desviación típica obtenida para este conjunto es σ = 86,2590

$$Z \sim N(0,1)\ ;\ Z = \frac{X-\mu}{\sigma} \rightarrow X \sim N(\mu,\sigma)$$

Destipificando, obtenemos que

$$Z = \frac{X-\mu}{\sigma} \rightarrow 1{,}96 = \frac{X-(-42{,}311)}{86{,}2590} \rightarrow\ X = 126{,}756$$

$$Z = \frac{X-\mu}{\sigma} \rightarrow -1{,}96 = \frac{X-(-42{,}311)}{86{,}2590} \rightarrow\ X = -211{,}378$$

$$P[-211{,}378 \leq X \leq 126{,}756] = 0{,}95$$

El número de muestras fuera del intervalo en este subconjunto es de 653 muestras.

2. Regresión

- Conjunto de entrenamiento

Se ha graficado la predicción frente al valor real para cada una de las muestras. Se ha hecho con los valores a escala, por una parte, y con los valores normalizados por otra. En este último caso, para representar junto a la recta de ajuste real (en rojo) la que sería la recta de ajuste óptima (en negro). Se puede comprobar visualmente que en ambos casos:

a. La existencia de outliers es reducida, lo que confirma que la mayoría de las predicciones son muy parecidas al valor real medido.
b. La recta de ajuste real (roja) obtenida prácticamente se superpone a la recta de ajuste óptima (negra). Por su similitud en inclinación deducimos que el modelo está bien optimizado.

Adicionalmente, se obtienen los siguientes indicadores:

R^2	0,9801
$\overline{R^2}$	0,9801

Gracias a ellos, podemos sacar las siguientes conclusiones:

a. Al ser R^2 muy próximo a 1, las variables de entrada del modelo están muy relacionadas con las variables de salida. Es decir, la variabilidad de la salida está muy explicada por las variables de entrada, o lo que es lo mismo, por el modelo de regresión.
b. Al ser R^2 muy próximo a 1, la recta de regresión tiene un ajuste muy bueno a la nube de puntos; o lo que es lo mismo, la nube de puntos no tiene una gran dispersión entre los puntos.
c. Al ser ambos parámetros R^2 y $\overline{R^2}$ muy parecidos, podemos afirmar que todas las variables de entrada al modelo son significativas.

- Conjunto de test

Se ha graficado la predicción frente al valor real para cada una de las muestras. Se ha hecho con los valores a escala. Se puede comprobar visualmente que en ambos casos:

a. La existencia de outliers es reducida, lo que confirma que la mayoría de las predicciones son muy parecidas al valor real medido.
b. La recta de ajuste real obtenida prácticamente se ajusta de manera precisa a la nube de puntos, lo que indica que el grado de entrenamiento del modelo ha sido bueno.

Adicionalmente, se obtienen los siguientes indicadores:

R^2	0,9793
$\overline{R^2}$	0,9793

Gracias a ellos, podemos sacar las siguientes conclusiones:

a. Al ser R^2 muy próximo a 1, las variables de entrada del modelo están muy relacionadas con las variables de salida. Es decir, la variabilidad de la salida está muy explicada por las variables de entrada, o lo que es lo mismo, por el modelo de regresión.
b. Al ser R^2 muy próximo a 1, la recta de regresión tiene un ajuste muy bueno a la nube de puntos; o lo que es lo mismo, la nube de puntos no tiene una gran dispersión entre los puntos.
c. Al ser ambos parámetros R^2 y $\overline{R^2}$ muy parecidos, podemos afirmar que todas las variables de entrada al modelo son significativas.

5.3.- Modelo de previsión de la temperatura de los anillos

Para el modelo de previsión de la temperatura de los anillos, se va a proceder a describir objetivo, utilidad y construcción del mismo. A continuación, se procederá a demostrar su robustez, para poder emplearlo de manera segura y certera en los años venideros.

Objetivos y especificación

El objetivo de este modelo es determinar la temperatura que deberían tener los anillos en un determinado momento. La temperatura es un claro indicador del estado de los anillos. Por lo tanto, monitorizando su temperatura real y siendo capaces de predecir la temperatura que debería tener, podemos calcular la desviación entre ambas.

La desviación se considerará grande cuando se encuentre en fuera del intervalo de confianza del 95% del error creado a partir de los datos del primer año. Si la desviación es suficientemente grande y se suceden tres de ellas seguidas, se podrá considerar que la observación es una anomalía. Estas anomalías se podrán graficar y cuantificar.

Esto servirá para:

- Monitorizar el estado de los anillos en tiempo real, a través de su temperatura.
- Identificar que una anomalía ha aparecido en los anillos.
- Tener una visión global de las anomalías: a nivel mensual, anual…
- Permitir la planificación de las tareas de mantenimiento de los anillos.

El modelo se construye en función de las condiciones de refrigeración observadas. También se incluyen las condiciones de operación en las que se encuentran los anillos en ese momento. Se va a predecir la temperatura de los anillos:

- ANG_aero_tem_anillos_sld_gen_val

a partir de las variables:

- ANG_aero_tem_interior_nac_val
- ANG_aero_pot_total_val
- ANG_aero_tem_ambiente_val
- ANG_aero_vel_viento_val

Robustez del modelo

Para el caso base de análisis de la temperatura de los anillos, se debe recordar que para entrenar el modelo se utilizarán 23897 observaciones (el 70% de los datos del primer año), y que para verificar la robustez del modelo se utilizarán 10197 observaciones (el 30% restante de los datos del primer año). Cabe destacar que la selección de los datos del primer año para entrenamiento o para test se realiza de forma aleatoria y atemporal.

Para entrenar el modelo, se observa que es preciso utilizar un total de 1200 épocas. A continuación, se encuentra el gráfico que muestra la estabilización del error medio absoluto y la caída de la función de pérdida.

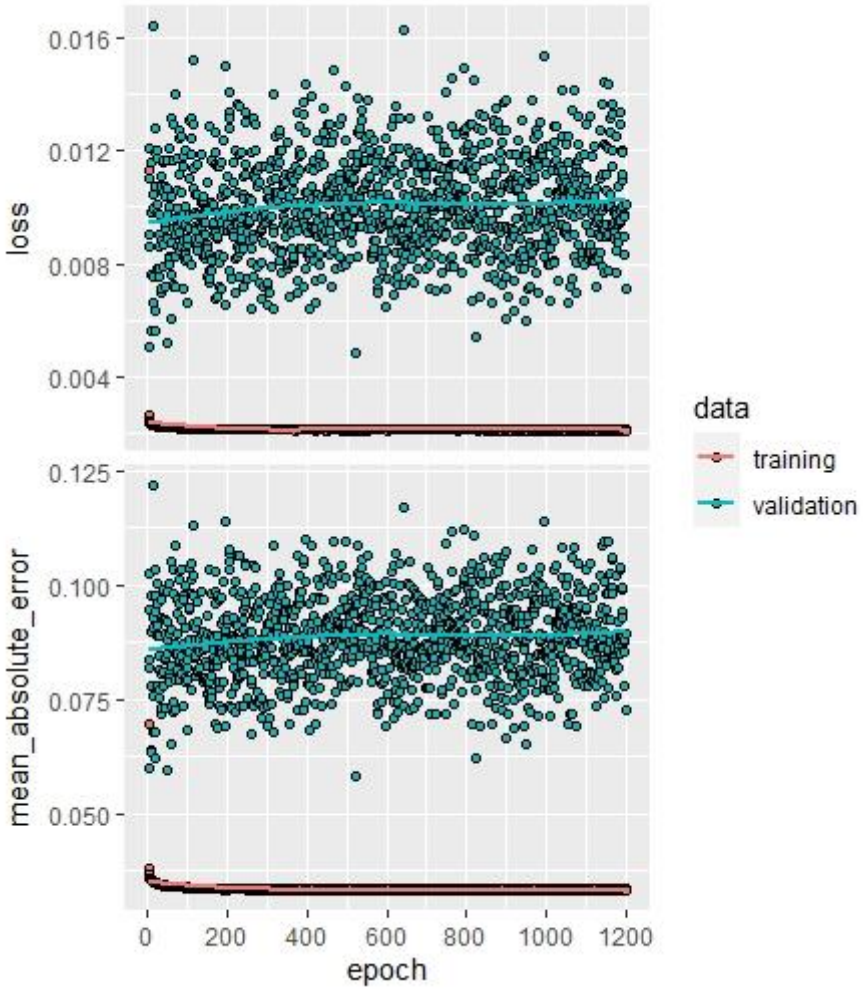


Ilustración 22. Función de coste y función de pérdidas del modelo de previsión de la temperatura de los anillos.

Tras el entrenamiento, se analiza en profundidad el modelo utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos parámetros del entorno de trabajo se pueden encontrar a continuación, y así como las gráficas obtenidas del modelo de regresión utilizado con los dos conjuntos de observaciones:

- Gráficos del conjunto de datos de entrenamiento

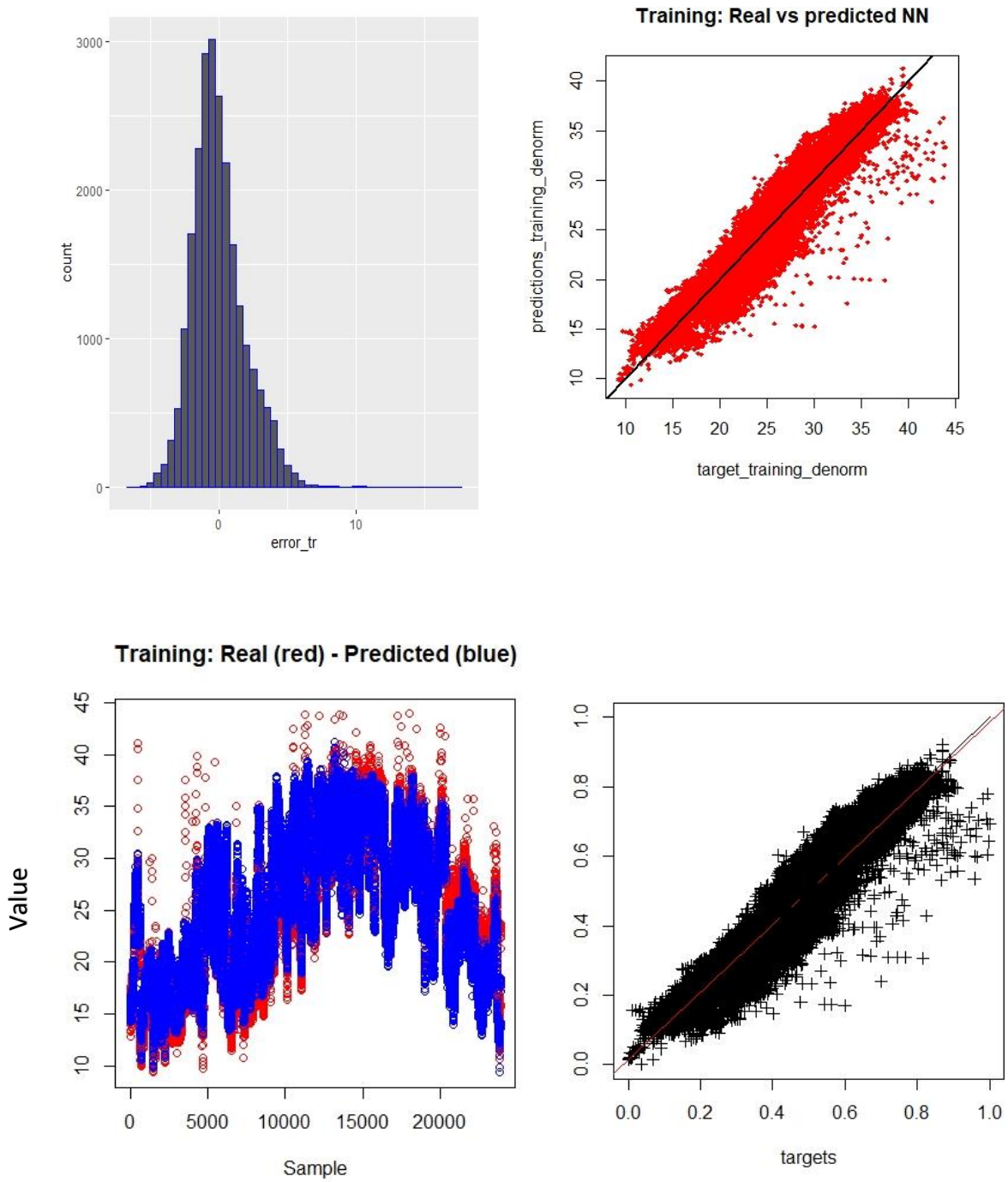


Ilustración 23. Resultado de robustez del entrenamiento del modelo de previsión de la temperatura de los anillos.

- Gráficos del conjunto de datos de test

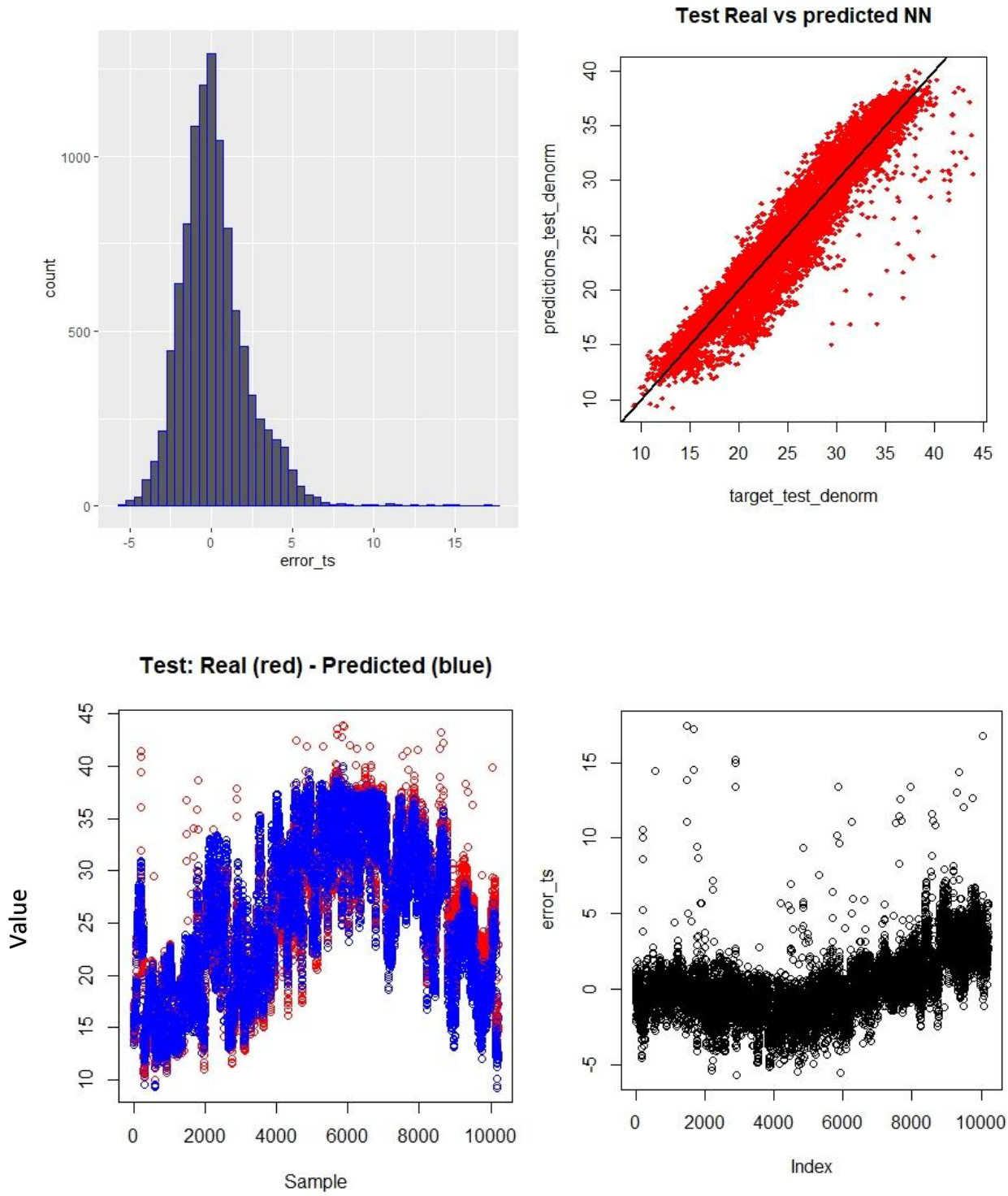


Ilustración 24. Resultado de robustez del test del modelo de previsión de la temperatura de los anillos.

1. Histograma

Recordando que el primer año se ha dividido en dos subconjuntos, entrenamiento y test, el numero de muestras que consideraremos que el modelo ha predicho mal será igual a la suma de ambos. Concretamente, se calcularán los intervalos de confianza del 95% para cada subconjunto, se cuantificarán las muestras fuera de el intervalo para cada subconjunto, y se sumarán. Estas muestras fuera de los intervalos se consideran predicciones incorrectas. De esta forma, la desviación que aporta el modelo de forma adicional por su imperfección se podrá medir mediante el número de predicciones incorrectas para el primer año.

- Conjunto de entrenamiento

La media obtenida para este conjunto es μ = 0,003

La desviación típica obtenida para este conjunto es σ = 1,985

$$Z \sim N(0,1)\ ;\ Z = \frac{X-\mu}{\sigma} \rightarrow X \sim N(\mu,\sigma)$$

Destipificando, obtenemos que

$$Z = \frac{X-\mu}{\sigma} \rightarrow 1{,}96 = \frac{X-(0{,}003)}{1{,}985} \rightarrow\ X = 3{,}894$$

$$Z = \frac{X-\mu}{\sigma} \rightarrow -1{,}96 = \frac{X-(0{,}003)}{1{,}985} \rightarrow\ X = -3{,}887$$

$$P[-3{,}887 \leq X \leq 3{,}894] = 0{,}95$$

El número de muestras fuera del intervalo en este subconjunto es de 1220 muestras.

- Conjunto de test

La media obtenida para este conjunto es μ = 0,039

La desviación típica obtenida para este conjunto es σ = 2,023

$$Z \sim N(0,1)\ ;\ Z = \frac{X-\mu}{\sigma} \rightarrow X \sim N(\mu,\sigma)$$

Destipificando, obtenemos que

$$Z = \frac{X-\mu}{\sigma} \rightarrow 1{,}96 = \frac{X-(0{,}039)}{2{,}023} \rightarrow\ X = 4{,}006$$

$$Z = \frac{X-\mu}{\sigma} \rightarrow -1{,}96 = \frac{X-(0{,}039)}{2{,}023} \rightarrow\ X = -3{,}927$$

$$P[-3{,}927 \leq X \leq 4{,}006] = 0{,}95$$

El número de muestras fuera del intervalo en este subconjunto es de 483 muestras.

2. Regresión

- Conjunto de entrenamiento

Se ha graficado la predicción frente al valor real para cada una de las muestras. Se ha hecho con los valores a escala, por una parte, y con los valores normalizados por otra. En este último caso, para representar junto a la recta de ajuste real (en rojo) la que sería la recta de ajuste óptima (en negro). Se puede comprobar visualmente que en ambos casos:

a. La existencia de outliers es reducida, lo que confirma que la mayoría de las predicciones son muy parecidas al valor real medido.
b. La recta de ajuste real (roja) obtenida prácticamente se superpone a la recta de ajuste óptima (negra). Por su similitud en inclinación deducimos que el modelo está bien optimizado.

Adicionalmente, se obtienen los siguientes indicadores:

R^2	0,9134
$\overline{R^2}$	0,9134

Gracias a ellos, podemos sacar las siguientes conclusiones:

d. Al ser R^2 muy próximo a 1, las variables de entrada del modelo están muy relacionadas con las variables de salida. Es decir, la variabilidad de la salida está muy explicada por las variables de entrada, o lo que es lo mismo, por el modelo de regresión.
e. Al ser R^2 muy próximo a 1, la recta de regresión tiene un ajuste muy bueno a la nube de puntos; o lo que es lo mismo, la nube de puntos no tiene una gran dispersión entre los puntos.
f. Al ser ambos parámetros R^2 y $\overline{R^2}$ muy parecidos, podemos afirmar que todas las variables de entrada al modelo son significativas.

- Conjunto de test

Se ha graficado la predicción frente al valor real para cada una de las muestras. Se ha hecho con los valores a escala. Se puede comprobar visualmente que en ambos casos:

a. La existencia de outliers es reducida, lo que confirma que la mayoría de las predicciones son muy parecidas al valor real medido.
b. La recta de ajuste real obtenida prácticamente se ajusta de manera precisa a la nube de puntos, lo que indica que el grado de entrenamiento del modelo ha sido bueno.

Adicionalmente, se obtienen los siguientes indicadores:

R^2	0,9111
$\overline{R^2}$	0,9111

Gracias a ellos, podemos sacar las siguientes conclusiones:

d. Al ser R^2 muy próximo a 1, las variables de entrada del modelo están muy relacionadas con las variables de salida. Es decir, la variabilidad de la salida está muy explicada por las variables de entrada, o lo que es lo mismo, por el modelo de regresión.
e. Al ser R^2 muy próximo a 1, la recta de regresión tiene un ajuste muy bueno a la nube de puntos; o lo que es lo mismo, la nube de puntos no tiene una gran dispersión entre los puntos.
f. Al ser ambos parámetros R^2 y $\overline{R^2}$ muy parecidos, podemos afirmar que todas las variables de entrada al modelo son significativas.

5.4.- Modelo de previsión de la temperatura de la multiplicadora

Para el modelo de previsión de la temperatura de la multiplicadora, se va a proceder a describir objetivo, utilidad y construcción del mismo. A continuación, se procederá a demostrar su robustez, para poder emplearlo de manera segura y certera en los años venideros.

Objetivos y especificación

El objetivo de este modelo es determinar la temperatura que debería tener la multiplicadora en un determinado momento. La temperatura es un claro indicador del estado de la multiplicadora. Por lo tanto, monitorizando su temperatura real y siendo capaces de predecir la temperatura que debería tener, podemos calcular la desviación entre ambas.

La desviación se considerará grande cuando se encuentre en fuera del intervalo de confianza del 95% del error creado a partir de los datos del primer año. Si la desviación es suficientemente grande y se suceden tres de ellas seguidas, se podrá considerar que la observación es una anomalía. Estas anomalías se podrán graficar y cuantificar.

Esto servirá para:

- Monitorizar el estado de la multiplicadora en tiempo real, a través de su temperatura.
- Identificar que una anomalía ha aparecido en la multiplicadora.
- Tener una visión global de las anomalías: a nivel mensual, anual…
- Permitir la planificación de las tareas de mantenimiento de la multiplicadora.

El modelo se construye en función de las condiciones de lubricación y refrigeración observadas. También se incluyen las condiciones de operación en las que se encuentra la multiplicadora en ese momento. Se va a predecir la temperatura de la multiplicadora:

- ANG_aero_tem_rod_alta_multi_val

a partir de las variables:

- ANG_aero_pot_total_val
- ANG_aero_tem_interior_nac_val
- ANG_aero_vel_viento_val
- ANG_aero_tem_aceite_multi_val

Robustez del modelo

Para el caso base de análisis de la temperatura de la multiplicadora, se debe recordar que para entrenar el modelo se utilizarán 23897 observaciones (el 70% de los datos del primer año), y que para verificar la robustez del modelo se utilizarán 10197 observaciones (el 30% restante de los datos del primer año). Cabe destacar que la selección de los datos del primer año para entrenamiento o para test se realiza de forma aleatoria y atemporal.

Para entrenar el modelo, se observa que es preciso utilizar un total de 1000 épocas. A continuación, se encuentra el gráfico que muestra la estabilización del error medio absoluto y la caída de la función de pérdida.

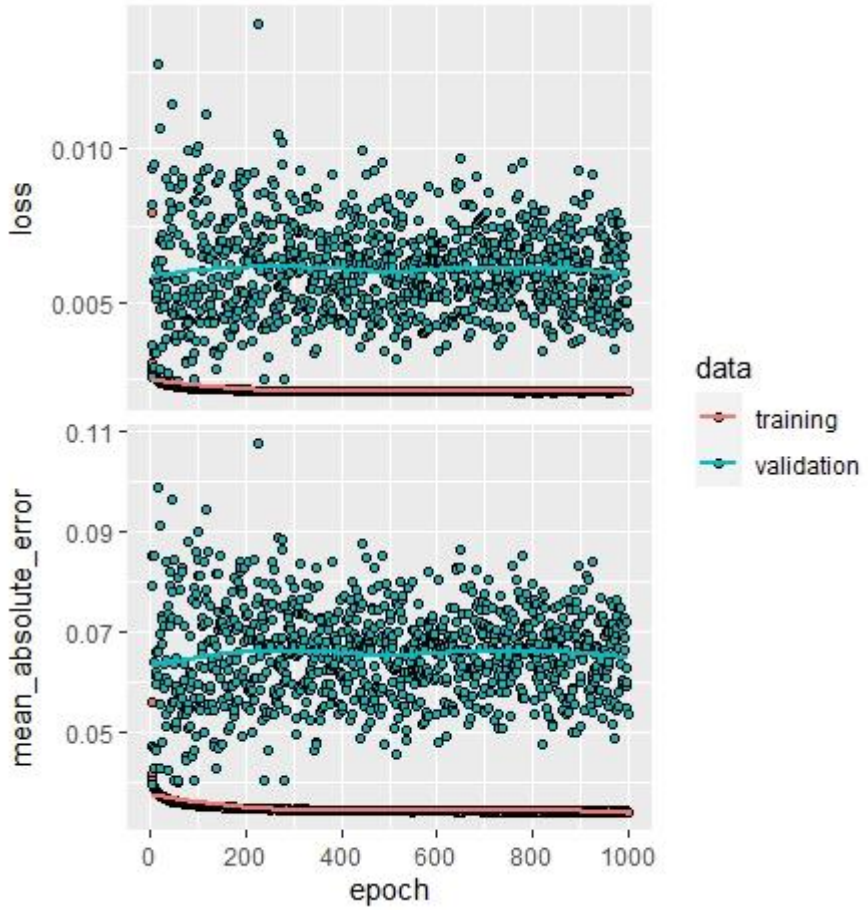


Ilustración 25. Función de coste y función de pérdidas del modelo de previsión de la temperatura de la multiplicadora.

Tras el entrenamiento, se analiza en profundidad el modelo utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos parámetros del entorno de trabajo se pueden encontrar a continuación, y así como las gráficas obtenidas del modelo de regresión utilizado con los dos conjuntos de observaciones:

- Gráficos del conjunto de datos de entrenamiento

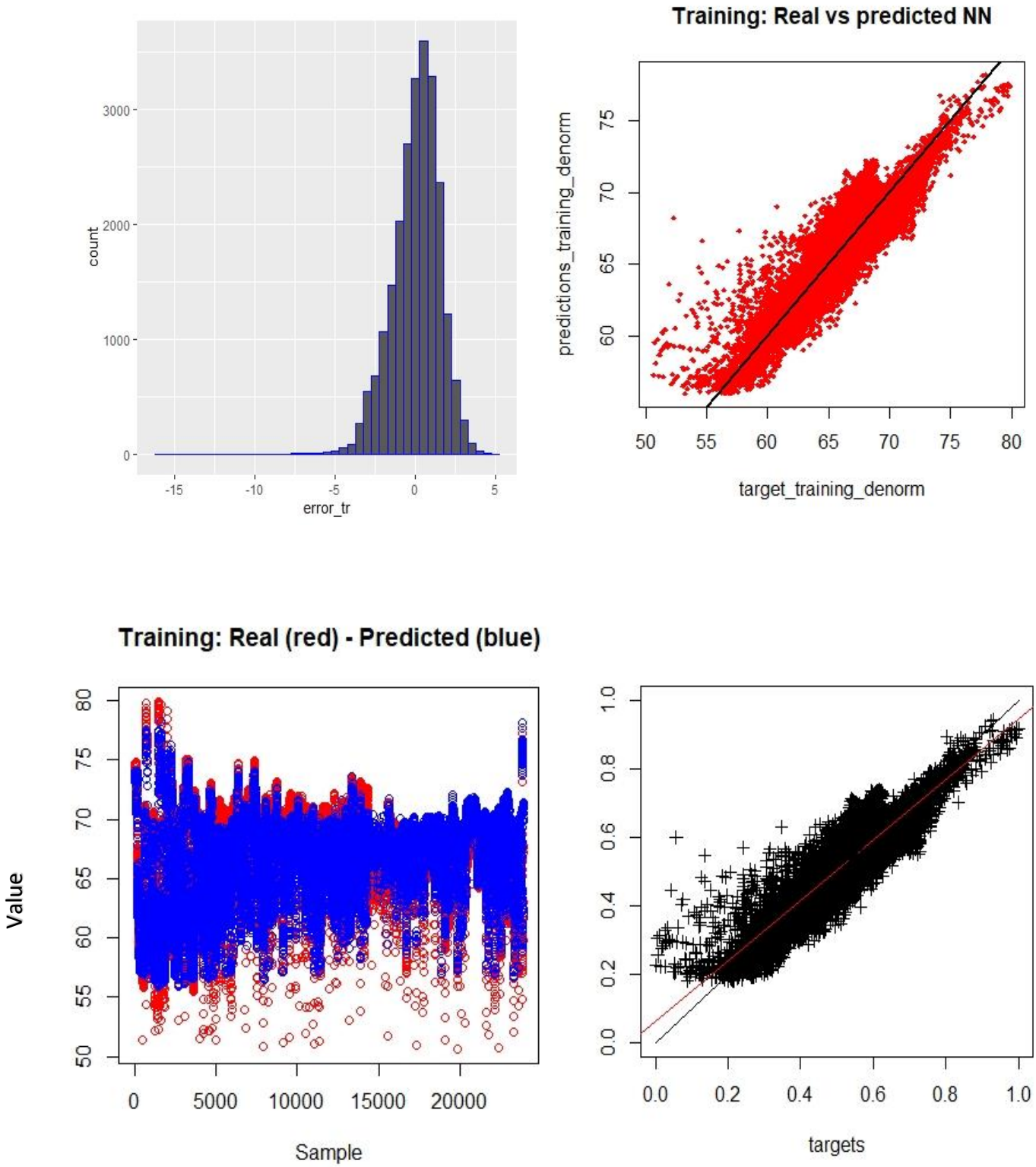


Ilustración 26. Resultado de robustez del entrenamiento del modelo de previsión de la temperatura de la multiplicadora.

- Gráficos del conjunto de datos de test

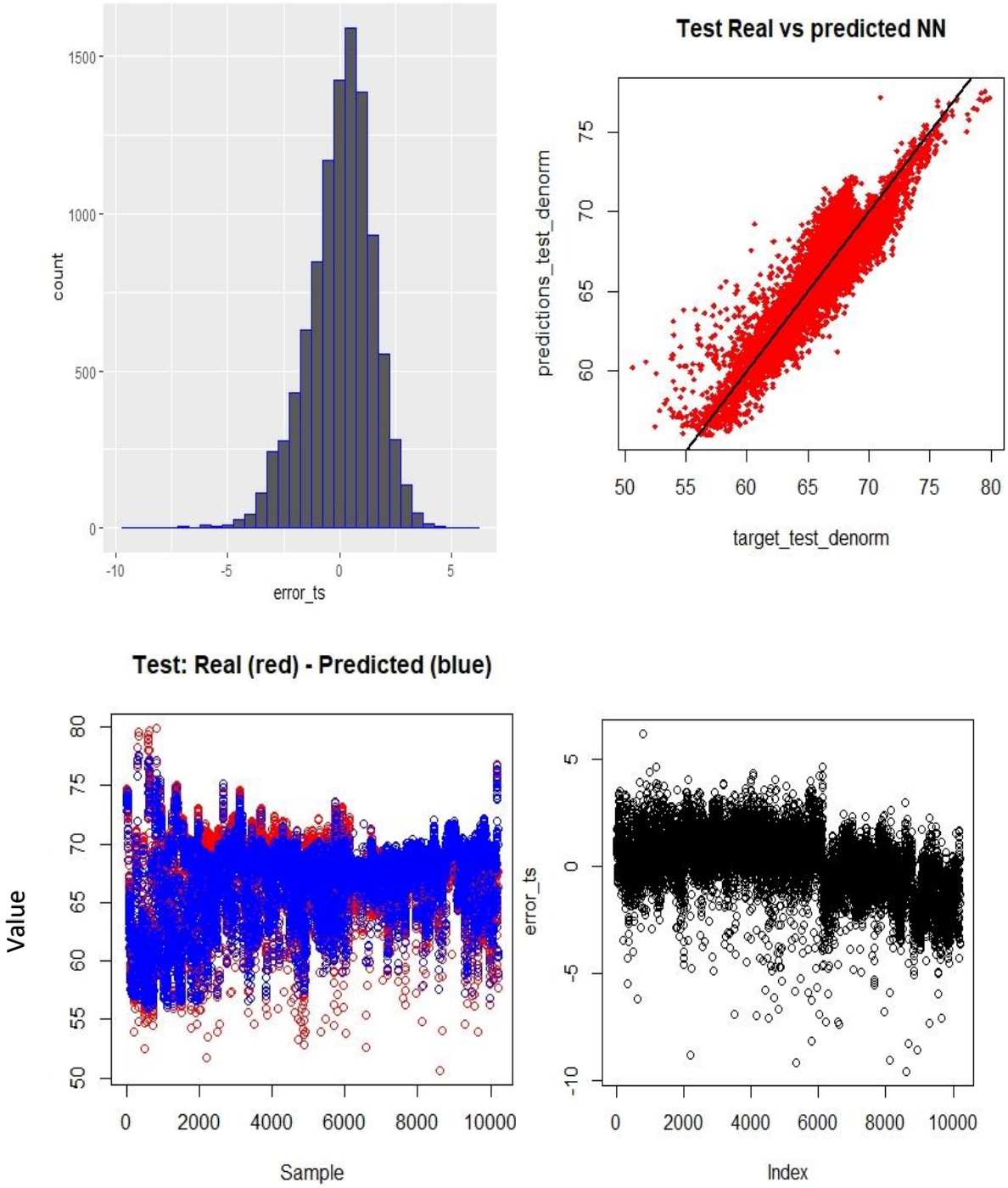


Ilustración 27. Resultado de robustez del test del modelo de previsión de la temperatura de la multiplicadora.

3. Histograma

Recordando que el primer año se ha dividido en dos subconjuntos, entrenamiento y test, el numero de muestras que consideraremos que el modelo ha predicho mal será igual a la suma de ambos. Concretamente, se calcularán los intervalos de confianza del 95% para cada subconjunto, se cuantificarán las muestras fuera de el intervalo para cada subconjunto, y se sumarán. Estas muestras fuera de los intervalos se consideran predicciones incorrectas. De esta forma, la desviación que aporta el modelo de forma adicional por su imperfección se podrá medir mediante el número de predicciones incorrectas para el primer año.

- Conjunto de entrenamiento

La media obtenida para este conjunto es μ = 0,03

La desviación típica obtenida para este conjunto es σ = 1,502

$$Z{\sim}N(0,1)\ ;\ Z=\frac{X-\mu}{\sigma}\rightarrow X{\sim}N(\mu,\sigma)$$

Destipificando, obtenemos que

$$Z=\frac{X-\mu}{\sigma}\rightarrow 1{,}96=\frac{X-(0{,}03)}{1{,}502}\rightarrow\ X=2{,}974$$

$$Z=\frac{X-\mu}{\sigma}\rightarrow -1{,}96=\frac{X-(0{,}03)}{1{,}502}\rightarrow\ X=-2{,}914$$

$$P[-2{,}914\leq X\leq 2{,}974]=0{,}95$$

El número de muestras fuera del intervalo en este subconjunto es de1186 muestras.

- Conjunto de test

La media obtenida para este conjunto es μ = 0,042

La desviación típica obtenida para este conjunto es σ = 1,845

$$Z{\sim}N(0,1)\ ;\ Z=\frac{X-\mu}{\sigma}\rightarrow X{\sim}N(\mu,\sigma)$$

Destipificando, obtenemos que

$$Z=\frac{X-\mu}{\sigma}\rightarrow 1{,}96=\frac{X-(0{,}042)}{1{,}845}\rightarrow\ X=2{,}954$$

$$Z=\frac{X-\mu}{\sigma}\rightarrow -1{,}96=\frac{X-(0{,}042)}{1{,}845}\rightarrow\ X=-2{,}869$$

$$P[-2{,}869\leq X\leq 2{,}954]=0{,}95$$

El número de muestras fuera del intervalo en este subconjunto es de 540 muestras.

4. Regresión

- Conjunto de entrenamiento

Se ha graficado la predicción frente al valor real para cada una de las muestras. Se ha hecho con los valores a escala, por una parte, y con los valores normalizados por otra. En este último caso, para representar junto a la recta de ajuste real (en rojo) la que sería la recta de ajuste óptima (en negro). Se puede comprobar visualmente que en ambos casos:

a. La existencia de outliers es moderada, lo que confirma que la muchas de las predicciones son muy parecidas al valor real medido.
b. La recta de ajuste real (roja) obtenida tiene una inclinación ligeramente menor que la recta de ajuste óptima (negra). A pesar de ello, debido a la alta variabilidad de rangos de trabajo de la temperatura de la multiplicadora durante el año, deducimos que el modelo está bien optimizado.

Adicionalmente, se obtienen los siguientes indicadores:

R^2	0,835
$\overline{R^2}$	0,835

Gracias a ellos, podemos sacar las siguientes conclusiones:

g. Al ser R^2 cercano a 1, las variables de entrada del modelo están suficientemente relacionadas con las variables de salida. Es decir, la variabilidad de la salida está explicada por las variables de entrada, o lo que es lo mismo, por el modelo de regresión.
h. Al ser R^2 cercano a 1, la recta de regresión tiene un ajuste suficientemente bueno a la nube de puntos; o lo que es lo mismo, la nube de puntos no tiene una gran dispersión entre los puntos.
i. Al ser ambos parámetros R^2 y $\overline{R^2}$ muy parecidos, podemos afirmar que todas las variables de entrada al modelo son significativas.

- Conjunto de test

Se ha graficado la predicción frente al valor real para cada una de las muestras. Se ha hecho con los valores a escala. Se puede comprobar visualmente que en ambos casos:

a. La existencia de outliers es moderada, lo que confirma que la muchas de las predicciones son muy parecidas al valor real medido
b. La recta de ajuste real obtenida se ajusta de manera bastante precisa a la nube de puntos, lo que indica que el grado de entrenamiento del modelo ha sido bueno.

Adicionalmente, se obtienen los siguientes indicadores:

R^2	0,834
$\overline{R^2}$	0,834

Gracias a ellos, podemos sacar las siguientes conclusiones:

g. Al ser R^2 cercano a 1, las variables de entrada del modelo están suficientemente relacionadas con las variables de salida. Es decir, la variabilidad de la salida está explicada por las variables de entrada, o lo que es lo mismo, por el modelo de regresión.
h. Al ser R^2 cercano a 1, la recta de regresión tiene un ajuste suficientemente bueno a la nube de puntos; o lo que es lo mismo, la nube de puntos no tiene una gran dispersión entre los puntos.
i. Al ser ambos parámetros R^2 y $\overline{R^2}$ muy parecidos, podemos afirmar que todas las variables de entrada al modelo son significativas.

5.4.- Modelo de previsión de la temperatura del rodamiento DE

Para el modelo de previsión de la temperatura del rodamiento DE, se va a proceder a describir objetivo, utilidad y construcción del mismo. A continuación, se procederá a demostrar su robustez, para poder emplearlo de manera segura y certera en los años venideros.

Objetivos y especificación

El objetivo de este modelo es determinar la temperatura que debería tener el rodamiento DE en un determinado momento. La temperatura es un claro indicador del estado del rodamiento DE. Por lo tanto, monitorizando su temperatura real y siendo capaces de predecir la temperatura que debería tener, podemos calcular la desviación entre ambas.

La desviación se considerará grande cuando se encuentre en fuera del intervalo de confianza del 95% del error creado a partir de los datos del primer año. Si la desviación es suficientemente grande y se suceden tres de ellas seguidas, se podrá considerar que la observación es una anomalía. Estas anomalías se podrán graficar y cuantificar.

Esto servirá para:

- Monitorizar el estado del rodamiento DE en tiempo real, a través de su temperatura.
- Identificar que una anomalía ha aparecido en el rodamiento DE.
- Tener una visión global de las anomalías: a nivel mensual, anual…
- Permitir la planificación de las tareas de mantenimiento del rodamiento DE.

El modelo se construye en función de las condiciones de refrigeración y carga de trabajo observadas. También se incluyen las condiciones de operación en las que se encuentra el rodamiento DE en ese momento. Se va a predecir la temperatura del rodamiento DE, "Drive end", el rodamiento en el lado del accionamiento:

- ANG_aero_tem_rod_DE_gen_val

a partir de las variables:

- ANG_aero_tem_rod_NDE_gen_val
- ANG_aero_pot_total_val
- ANG_aero_tem_ambiente_val
- ANG_aero_vel_viento_val

Robustez del modelo

Para el caso base de análisis de la temperatura del rodamiento DE, se debe recordar que para entrenar el modelo se utilizarán 23897 observaciones (el 70% de los datos del primer año), y que para verificar la robustez del modelo se utilizarán 10197 observaciones (el 30% restante de los datos del primer año). Cabe destacar que la selección de los datos del primer año para entrenamiento o para test se realiza de forma aleatoria y atemporal.

Para entrenar el modelo, se observa que es preciso utilizar un total de 800 épocas. A continuación, se encuentra el gráfico que muestra la estabilización del error medio absoluto y la caída de la función de pérdida.

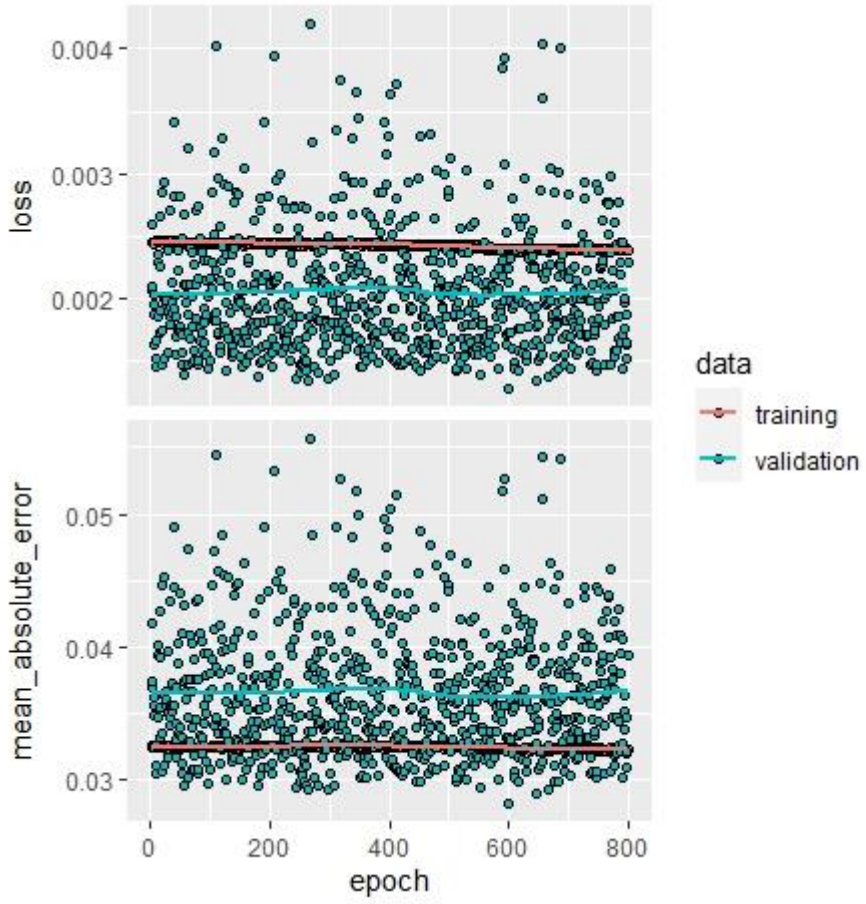


Ilustración 28. Función de coste y función de pérdidas del modelo de previsión de la temperatura del rodamiento DE.

Tras el entrenamiento, se analiza en profundidad el modelo utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos parámetros del entorno de trabajo se pueden encontrar a continuación, y así como las gráficas obtenidas del modelo de regresión utilizado con los dos conjuntos de observaciones:

- Gráficos del conjunto de datos de entrenamiento

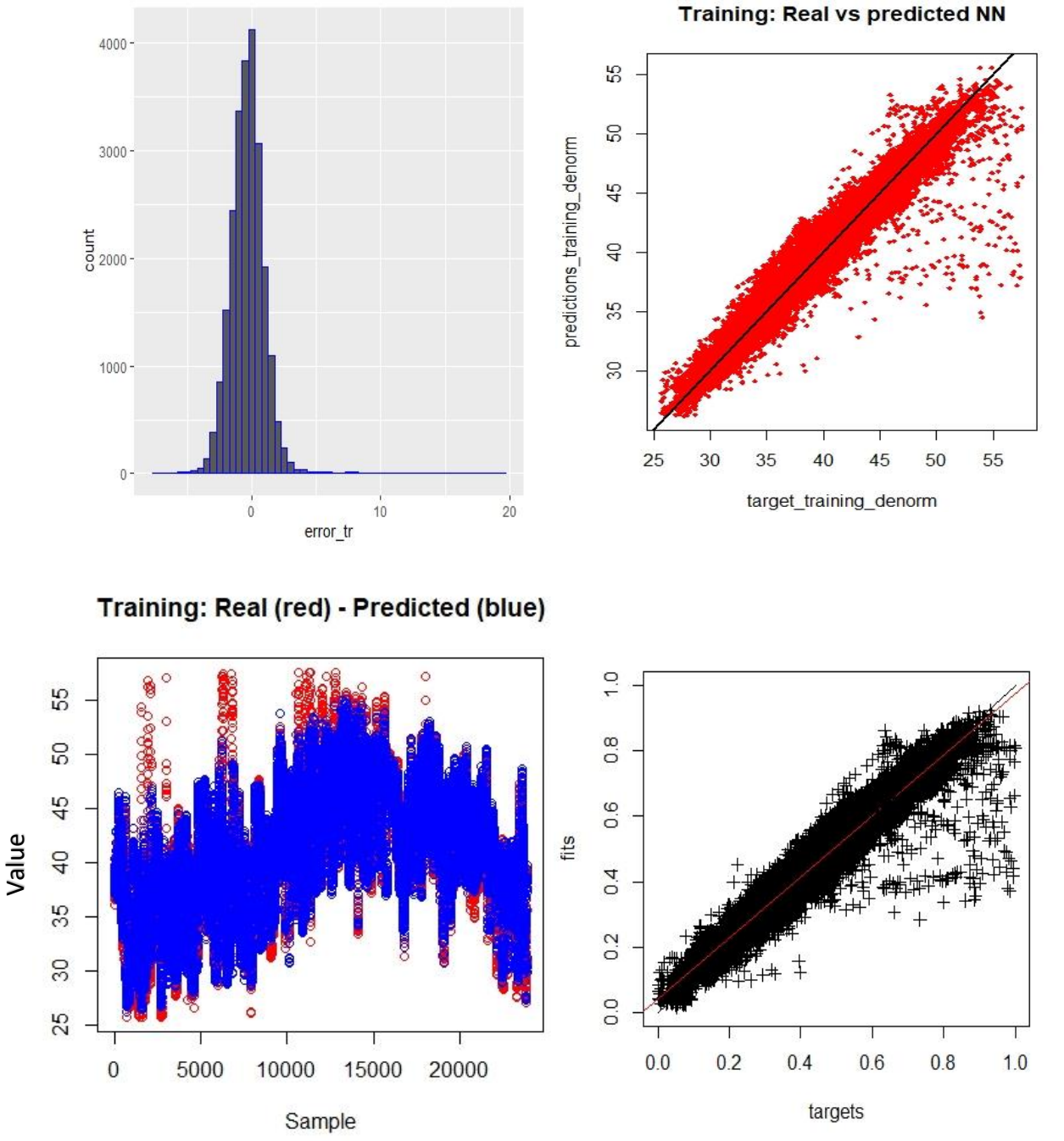


Ilustración 29. Resultado de robustez del entrenamiento del modelo de previsión de la temperatura del rodamiento DE.

- Gráficos del conjunto de datos de test

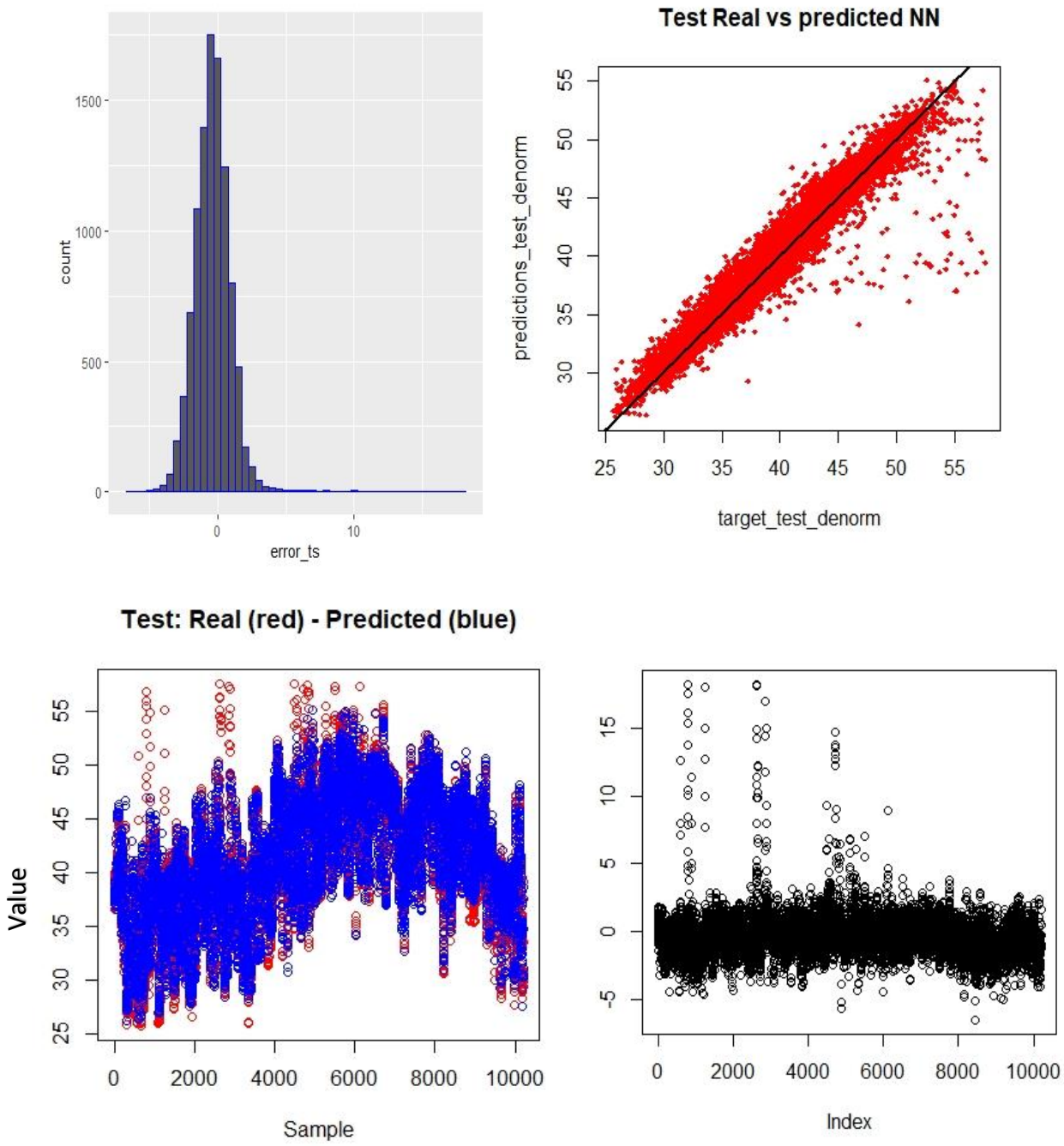


Ilustración 30. Resultado de robustez del test del modelo de previsión de la temperatura del rodamiento DE.

5. Histograma

Recordando que el primer año se ha dividido en dos subconjuntos, entrenamiento y test, el numero de muestras que consideraremos que el modelo ha predicho mal será igual a la suma de ambos. Concretamente, se calcularán los intervalos de confianza del 95% para cada subconjunto, se cuantificarán las muestras fuera de el intervalo para cada subconjunto, y se sumarán. Estas muestras fuera de los intervalos se consideran predicciones incorrectas. De esta forma, la desviación que aporta el modelo de forma adicional por su imperfección se podrá medir mediante el número de predicciones incorrectas para el primer año.

- Conjunto de entrenamiento

La media obtenida para este conjunto es μ = -0,309

La desviación típica obtenida para este conjunto es σ = 1,5039

$$Z \sim N(0,1)\ ;\ Z = \frac{X-\mu}{\sigma} \rightarrow X \sim N(\mu,\sigma)$$

Destipificando, obtenemos que

$$Z = \frac{X-\mu}{\sigma} \rightarrow 1{,}96 = \frac{X-(-0{,}309)}{1{,}5039} \rightarrow\ X = 2{,}638$$

$$Z = \frac{X-\mu}{\sigma} \rightarrow -1{,}96 = \frac{X-(-0{,}309)}{1{,}5039} \rightarrow\ X = -3{,}257$$

$$P[-3{,}257 \leq X \leq 2{,}638] = 0{,}95$$

El número de muestras fuera del intervalo en este subconjunto es de 644 muestras.

- Conjunto de test

La media obtenida para este conjunto es μ = -0,335

La desviación típica obtenida para este conjunto es σ = 2,689

$$Z \sim N(0,1)\ ;\ Z = \frac{X-\mu}{\sigma} \rightarrow X \sim N(\mu,\sigma)$$

Destipificando, obtenemos que

$$Z = \frac{X-\mu}{\sigma} \rightarrow 1{,}96 = \frac{X-(-0{,}335)}{2{,}689} \rightarrow\ X = 2{,}689$$

$$Z = \frac{X-\mu}{\sigma} \rightarrow -1{,}96 = \frac{X-(-0{,}335)}{2{,}689} \rightarrow\ X = -3{,}359$$

$$P[-3{,}359 \leq X \leq 2{,}689] = 0{,}95$$

El número de muestras fuera del intervalo en este subconjunto es de 273 muestras.

6. Regresión

- Conjunto de entrenamiento

Se ha graficado la predicción frente al valor real para cada una de las muestras. Se ha hecho con los valores a escala, por una parte, y con los valores normalizados por otra. En este último caso, para representar junto a la recta de ajuste real (en rojo) la que sería la recta de ajuste óptima (en negro). Se puede comprobar visualmente que en ambos casos:

a. La existencia de outliers es reducida, lo que confirma que la mayoría de las predicciones son muy parecidas al valor real medido.
b. La recta de ajuste real (roja) obtenida prácticamente se superpone a la recta de ajuste óptima (negra). Por su similitud en inclinación deducimos que el modelo está bien optimizado.

Adicionalmente, se obtienen los siguientes indicadores:

R^2	0,927
$\overline{R^2}$	0,927

Gracias a ellos, podemos sacar las siguientes conclusiones:

j. Al ser R^2 muy próximo a 1, las variables de entrada del modelo están muy relacionadas con las variables de salida. Es decir, la variabilidad de la salida está muy explicada por las variables de entrada, o lo que es lo mismo, por el modelo de regresión.
k. Al ser R^2 muy próximo a 1, la recta de regresión tiene un ajuste muy bueno a la nube de puntos; o lo que es lo mismo, la nube de puntos no tiene una gran dispersión entre los puntos.
l. Al ser ambos parámetros R^2 y $\overline{R^2}$ muy parecidos, podemos afirmar que todas las variables de entrada al modelo son significativas.

- Conjunto de test

Se ha graficado la predicción frente al valor real para cada una de las muestras. Se ha hecho con los valores a escala. Se puede comprobar visualmente que en ambos casos:

a. La existencia de outliers es reducida, lo que confirma que la mayoría de las predicciones son muy parecidas al valor real medido.
b. La recta de ajuste real obtenida prácticamente se ajusta de manera precisa a la nube de puntos, lo que indica que el grado de entrenamiento del modelo ha sido bueno.

Adicionalmente, se obtienen los siguientes indicadores:

R^2	0,922
$\overline{R^2}$	0,922

Gracias a ellos, podemos sacar las siguientes conclusiones:

j. Al ser R^2 muy próximo a 1, las variables de entrada del modelo están muy relacionadas con las variables de salida. Es decir, la variabilidad de la salida está muy explicada por las variables de entrada, o lo que es lo mismo, por el modelo de regresión.
k. Al ser R^2 muy próximo a 1, la recta de regresión tiene un ajuste muy bueno a la nube de puntos; o lo que es lo mismo, la nube de puntos no tiene una gran dispersión entre los puntos.
l. Al ser ambos parámetros R^2 y $\overline{R^2}$ muy parecidos, podemos afirmar que todas las variables de entrada al modelo son significativas.

5.4.- Modelo de previsión de la temperatura del rodamiento NDE

Para el modelo de previsión de la temperatura del rodamiento NDE, se va a proceder a describir objetivo, utilidad y construcción del mismo. A continuación, se procederá a demostrar su robustez, para poder emplearlo de manera segura y certera en los años venideros.

Objetivos y especificación

El objetivo de este modelo es determinar la temperatura que debería tener el rodamiento NDE en un determinado momento. La temperatura es un claro indicador del estado del rodamiento NDE. Por lo tanto, monitorizando su temperatura real y siendo capaces de predecir la temperatura que debería tener, podemos calcular la desviación entre ambas.

La desviación se considerará grande cuando se encuentre en fuera del intervalo de confianza del 95% del error creado a partir de los datos del primer año. Si la desviación es suficientemente grande y se suceden tres de ellas seguidas, se podrá considerar que la observación es una anomalía. Estas anomalías se podrán graficar y cuantificar.

Esto servirá para:

- Monitorizar el estado del rodamiento NDE en tiempo real, a través de su temperatura.
- Identificar que una anomalía ha aparecido en el rodamiento NDE.
- Tener una visión global de las anomalías: a nivel mensual, anual…
- Permitir la planificación de las tareas de mantenimiento del rodamiento NDE.

El modelo se construye en función de las condiciones de refrigeración y carga de trabajo observadas. También se incluyen las condiciones de operación en las que se encuentra el rodamiento NDE en ese momento. Se va a predecir la temperatura del rodamiento NDE, "Non drive end", el rodamiento en el lado opuesto al del accionamiento:

- ANG_aero_tem_rod_NDE_gen_val

a partir de las variables:

- ANG_aero_tem_rod_DE_gen_val
- ANG_aero_pot_total_val
- ANG_aero_tem_ambiente_val
- ANG_aero_vel_viento_val

Robustez del modelo

Para el caso base de análisis de la temperatura del rodamiento NDE, se debe recordar que para entrenar el modelo se utilizarán 23897 observaciones (el 70% de los datos del primer año), y que para verificar la robustez del modelo se utilizarán 10197 observaciones (el 30% restante de los datos del primer año). Cabe destacar que la selección de los datos del primer año para entrenamiento o para test se realiza de forma aleatoria y atemporal.

Para entrenar el modelo, se observa que es preciso utilizar un total de 1000 épocas. A continuación, se encuentra el gráfico que muestra la estabilización del error medio absoluto y la caída de la función de pérdida.

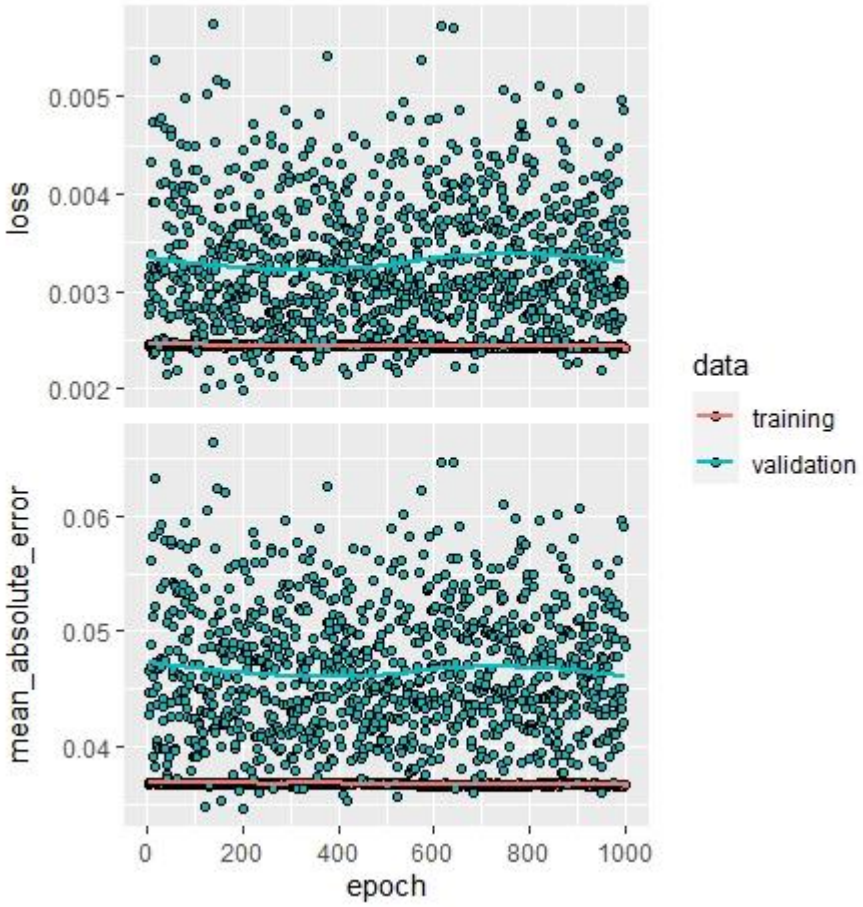


Ilustración 31. Función de coste y función de pérdidas del modelo de previsión de la temperatura del rodamiento NDE.

Tras el entrenamiento, se analiza en profundidad el modelo utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos parámetros del entorno de trabajo se pueden encontrar a continuación, y así como las gráficas obtenidas del modelo de regresión utilizado con los dos conjuntos de observaciones:

- Gráficos del conjunto de datos de entrenamiento

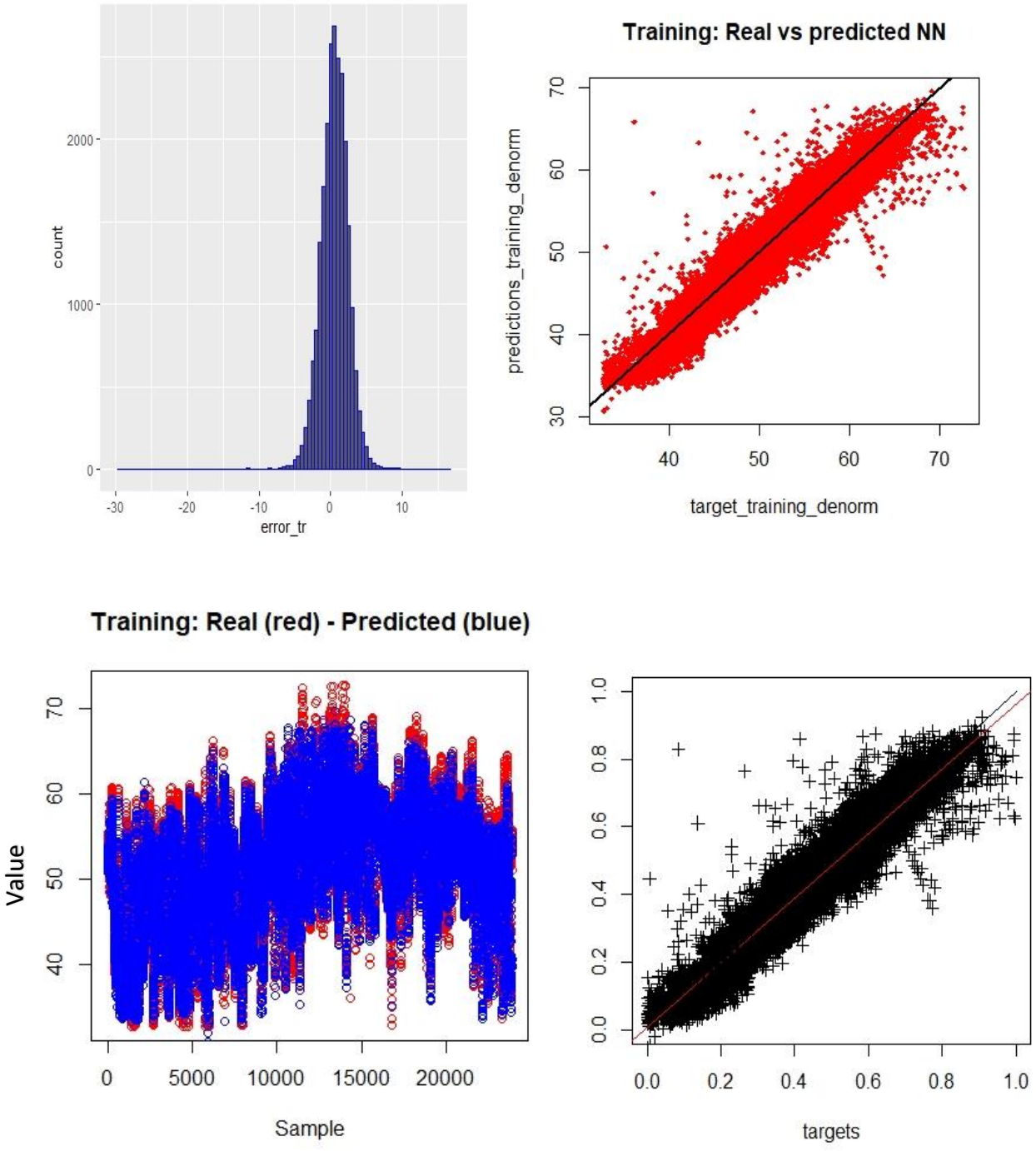


Ilustración 32. Resultado de robustez del entrenamiento del modelo de previsión de la temperatura del rodamiento NDE.

- Gráficos del conjunto de datos de test

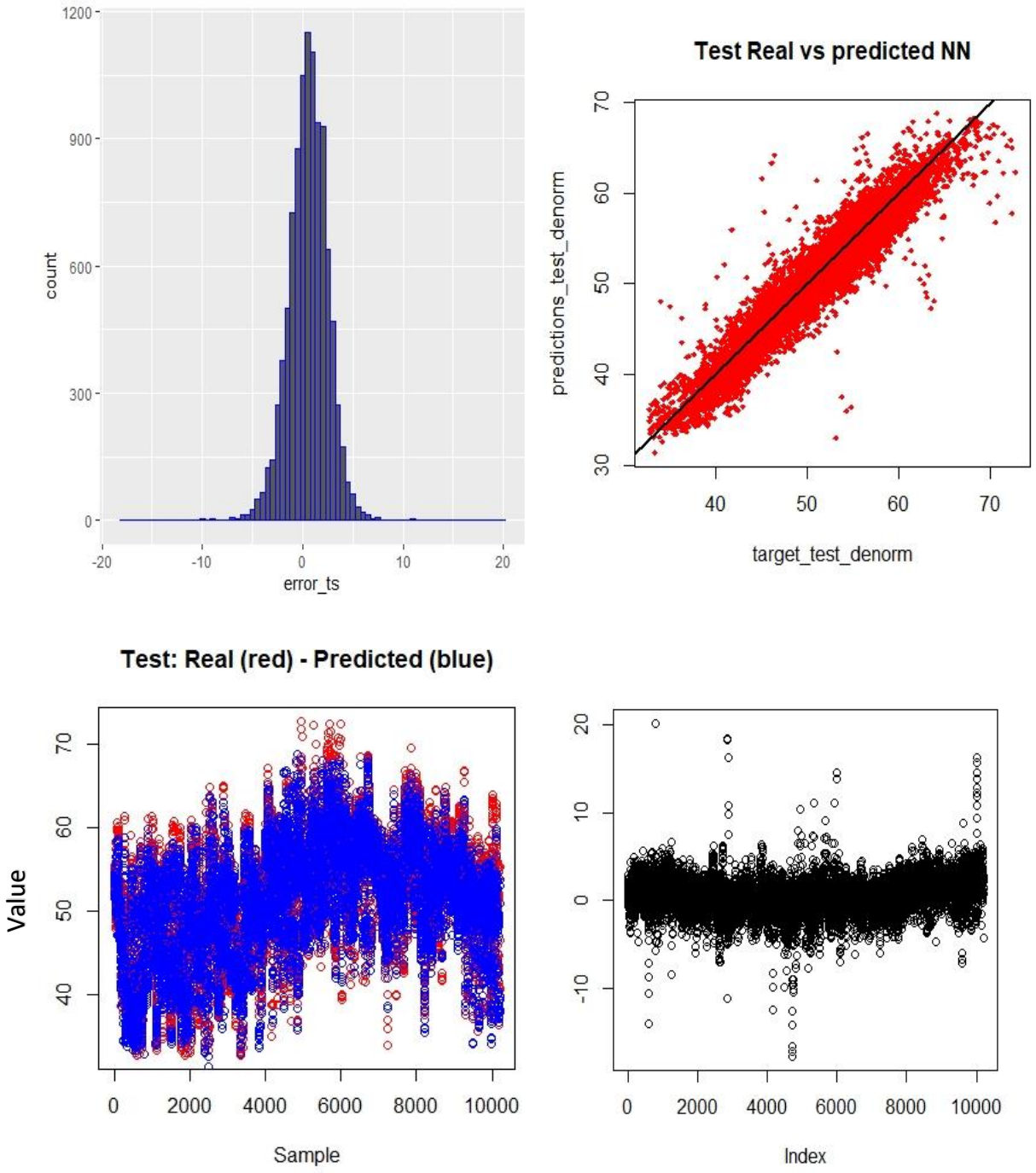


Ilustración 33. Resultado de robustez del test del modelo de previsión de la temperatura del rodamiento NDE.

7. Histograma

Recordando que el primer año se ha dividido en dos subconjuntos, entrenamiento y test, el numero de muestras que consideraremos que el modelo ha predicho mal será igual a la suma de ambos. Concretamente, se calcularán los intervalos de confianza del 95% para cada subconjunto, se cuantificarán las muestras fuera de el intervalo para cada subconjunto, y se sumarán. Estas muestras fuera de los intervalos se consideran predicciones incorrectas. De esta forma, la desviación que aporta el modelo de forma adicional por su imperfección se podrá medir mediante el número de predicciones incorrectas para el primer año.

- Conjunto de entrenamiento

La media obtenida para este conjunto es μ = 0,506

La desviación típica obtenida para este conjunto es σ = 1,994

$$Z \sim N(0,1)\ ;\ Z = \frac{X-\mu}{\sigma} \rightarrow X \sim N(\mu,\sigma)$$

Destipificando, obtenemos que

$$Z = \frac{X-\mu}{\sigma} \rightarrow 1{,}96 = \frac{X-(0{,}506)}{1{,}994} \rightarrow\ X = 4{,}416$$

$$Z = \frac{X-\mu}{\sigma} \rightarrow -1{,}96 = \frac{X-(0{,}506)}{1{,}994} \rightarrow\ X = -3{,}403$$

$$P[-3{,}403 \leq X \leq 4{,}416] = 0{,}95$$

El número de muestras fuera del intervalo en este subconjunto es de 1058 muestras.

- Conjunto de test

La media obtenida para este conjunto es μ = 0,563

La desviación típica obtenida para este conjunto es σ = 2,063

$$Z \sim N(0,1)\ ;\ Z = \frac{X-\mu}{\sigma} \rightarrow X \sim N(\mu,\sigma)$$

Destipificando, obtenemos que

$$Z = \frac{X-\mu}{\sigma} \rightarrow 1{,}96 = \frac{X-(0{,}563)}{2{,}063} \rightarrow\ X = 4{,}608$$

$$Z = \frac{X-\mu}{\sigma} \rightarrow -1{,}96 = \frac{X-(0{,}563)}{2{,}063} \rightarrow\ X = -3{,}481$$

$$P[-3{,}481 \leq X \leq 4{,}608] = 0{,}95$$

El número de muestras fuera del intervalo en este subconjunto es de 428 muestras.

8. Regresión

- Conjunto de entrenamiento

Se ha graficado la predicción frente al valor real para cada una de las muestras. Se ha hecho con los valores a escala, por una parte, y con los valores normalizados por otra. En este último caso, para representar junto a la recta de ajuste real (en rojo) la que sería la recta de ajuste óptima (en negro). Se puede comprobar visualmente que en ambos casos:

a. La existencia de outliers es reducida, lo que confirma que la mayoría de las predicciones son muy parecidas al valor real medido.
b. La recta de ajuste real (roja) obtenida prácticamente se superpone a la recta de ajuste óptima (negra). Por su similitud en inclinación deducimos que el modelo está bien optimizado.

Adicionalmente, se obtienen los siguientes indicadores:

R^2	0,906
$\overline{R^2}$	0,906

Gracias a ellos, podemos sacar las siguientes conclusiones:

m. Al ser R^2 muy próximo a 1, las variables de entrada del modelo están muy relacionadas con las variables de salida. Es decir, la variabilidad de la salida está muy explicada por las variables de entrada, o lo que es lo mismo, por el modelo de regresión.
n. Al ser R^2 muy próximo a 1, la recta de regresión tiene un ajuste muy bueno a la nube de puntos; o lo que es lo mismo, la nube de puntos no tiene una gran dispersión entre los puntos.
o. Al ser ambos parámetros R^2 y $\overline{R^2}$ muy parecidos, podemos afirmar que todas las variables de entrada al modelo son significativas.

- Conjunto de test

Se ha graficado la predicción frente al valor real para cada una de las muestras. Se ha hecho con los valores a escala. Se puede comprobar visualmente que en ambos casos:

a. La existencia de outliers es reducida, lo que confirma que la mayoría de las predicciones son muy parecidas al valor real medido.
b. La recta de ajuste real obtenida prácticamente se ajusta de manera precisa a la nube de puntos, lo que indica que el grado de entrenamiento del modelo ha sido bueno.

Adicionalmente, se obtienen los siguientes indicadores:

R^2	0,896
$\overline{R^2}$	0,896

Gracias a ellos, podemos sacar las siguientes conclusiones:

m. Al ser R^2 muy próximo a 1, las variables de entrada del modelo están muy relacionadas con las variables de salida. Es decir, la variabilidad de la salida está muy explicada por las variables de entrada, o lo que es lo mismo, por el modelo de regresión.
n. Al ser R^2 muy próximo a 1, la recta de regresión tiene un ajuste muy bueno a la nube de puntos; o lo que es lo mismo, la nube de puntos no tiene una gran dispersión entre los puntos.
o. Al ser ambos parámetros R^2 y $\overline{R^2}$ muy parecidos, podemos afirmar que todas las variables de entrada al modelo son significativas.

Capítulo 6.- Detección de anomalías

Tras haber construido un modelo robusto, se puede proceder a detectar anomalías. El proceso de detección de anomalías se operará utilizando los datos a partir del segundo año. Anteriormente, en el primer año, estos datos se habían dividido en un 70% para el conjunto de entrenamiento y el 30% restante para el conjunto de test. Para los años distintos del primer año, el 100% de los datos de cada año pertenecerán al conjunto de test.

6.1.- Estrategia de detección de anomalías

Se van a considerar como posibles anomalías las observaciones medidas que más difieren de su predicción. Por ello, no se trabajará con las predicciones, sino fundamentalmente con los errores, siendo estos la diferencia entre el valor medido y el valor predicho. El ejercicio de detección de anomalías consistirá en los siguientes pasos.

A. Para detectar las anomalías, se graficará el histograma de residuos. Para esté gráfico se determinará un intervalo de confianza del 95% que incluirá las observaciones que se considera que no presentan anomalía. Este intervalo de confianza será el que se calculó para el primer año, y por lo tanto será fijo para los siguientes años.
B. Se graficarán los errores en el eje temporal, y se trazará los límites superior e inferior que separan los valores considerados dentro de los límites de la normalidad de los valores considerados anómalos.
C. Adicionalmente, se graficarán, superpuestos a las predicciones y a los valores medidos en el eje temporal, las observaciones consideradas anómalas, para conocer su valor real también.
D. Se contabilizarán mensualmente los conjuntos de predicciones que determinan que existe una anomalía, para conocer la proporción de las mismas en un determinado mes en comparación con el total anual.

Es importante destacar que una anomalía es una sucesión de tres desviaciones lo suficientemente grandes (observaciones que se encuentran fuera del intervalo de confianza). Por lo tanto, aunque se grafiquen las observaciones que presentan anomalía (más de tres observaciones seguidas cuya desviación está fuera del intervalo de confianza), el conjunto de las mismas tan solo se contabilizará como una anomalía.

Para calcular el intervalo de confianza del 95%, calve en este apartado, se empleará el mismo procedimiento que es utilizó para definir la robustez del modelo (véase el apartado de construcción del modelo. Recordamos como se definía el intervalo de confianza:

1. Histograma de errores

Para contabilizar las observaciones que pueden presentar anomalía se grafica un histograma de errores. Estos errores tienen que distribuirse siguiendo una distribución normal y dicha estar centrada en 0. Se verificará visualmente.

A continuación, se calculará un intervalo de confianza que recoja un porcentaje específico del total de los datos (los datos serán los errores), comprendido entre dos valores simétricos con respecto a la media. De esta forma se puede asegurar que el porcentaje elegido de datos se encuentra entre esos valores, y se puede cuantificar la cantidad de muestras que se encuentran fuera de los límites del intervalo.

Esto también es útil para conocer la imperfección del propio modelo. Así, cuando el modelo se emplee utilizando un nuevo set de datos de entrada, la desviación o error entre predicción y valor real es la suma de la desviación debida a la imperfección del modelo y la desviación debida a problemas en el aerogenerador.

Para ello se utilizará la normal estándar o tipificada N(0,1). Esta es una distribución normal que tiene media 0 y varianza 1, y que se denota mediante la variable Z. La ventaja de utilización de esta es que función de distribución $\Phi(z)$ está tabulada.

Estableciendo un intervalo de confianza de un nivel de confianza del $\delta = 95\% \rightarrow \alpha = 0{,}05$

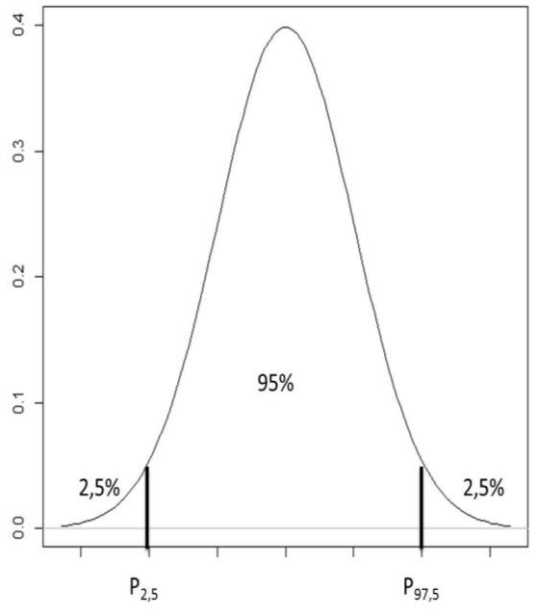


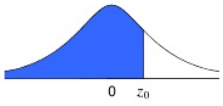
Tabla de la distribución normal N(0,1) para probabilidad acumulada inferior

μ = Media

σ = Desviación típica

$P(z \le z_0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_0} e^{\frac{-z^2}{2}} dz$

Tipificación: $z_0 = \frac{x-\mu}{\sigma}$



z_0	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	z_0
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359	0,0
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753	0,1
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141	0,2
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517	0,3
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879	0,4
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224	0,5
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549	0,6
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852	0,7
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133	0,8
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389	0,9
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621	1,0
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830	1,1
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015	1,2
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177	1,3
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319	1,4
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441	1,5
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545	1,6
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633	1,7
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706	1,8
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767	1,9

Tabla 15. Tabla detalle de cálculo del intervalo de confianza del 95 %.

$$P(Z \le z_0) = \Phi(z) \rightarrow Para\ \Phi(z) = 0{,}975 \rightarrow\ z_0 = z_{\alpha/2} =\ 1{,}96$$

$$z_{\alpha/2} =\ 1{,}96\ ;\ -z_{\alpha/2} =\ -1{,}96$$

Una vez obtenidos los valores para la normal estándar N(0,1), se lleva a cabo la destipificación de la variable, que consiste en devolverla nuestra distribución normal efectuando un cambio de variable.

$$Z \sim N(0,1)\ ;\ Z = \frac{X-\mu}{\sigma} \rightarrow X \sim N(\mu, \sigma)$$

Así, en función de los valores μ y σ que obtengamos del conjunto de los errores que conforman el histograma, obtendremos unos límites para el intervalo de confianza del 95%. Y con estos límites, podremos cuantificar la cantidad de muestras que quedan fuera de los mismos.

6.2.- Modelo de previsión de la potencia

A partir de los intervalos de confianza calculados para el primer año, establecemos el intervalo de confianza que se empleará para los años sucesivos.

$$[-202{,}574 \leq X \leq 146{,}959]\ U\ [-211{,}378 \leq X \leq 126{,}756] = [-211{,}378 \leq X \leq 146{,}959]$$

Recordamos que el intervalo de confianza del 95% para el modelo ha sido establecido con el intervalo de confianza del 95% del conjunto de entrenamiento unido al intervalo de intervalo de confianza del 95% del conjunto de test de los datos del primer año. Esto se ha hecho cogiendo la opción menos restringida, para evitar incluir errores aportados por el propio modelo.

Año 2

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

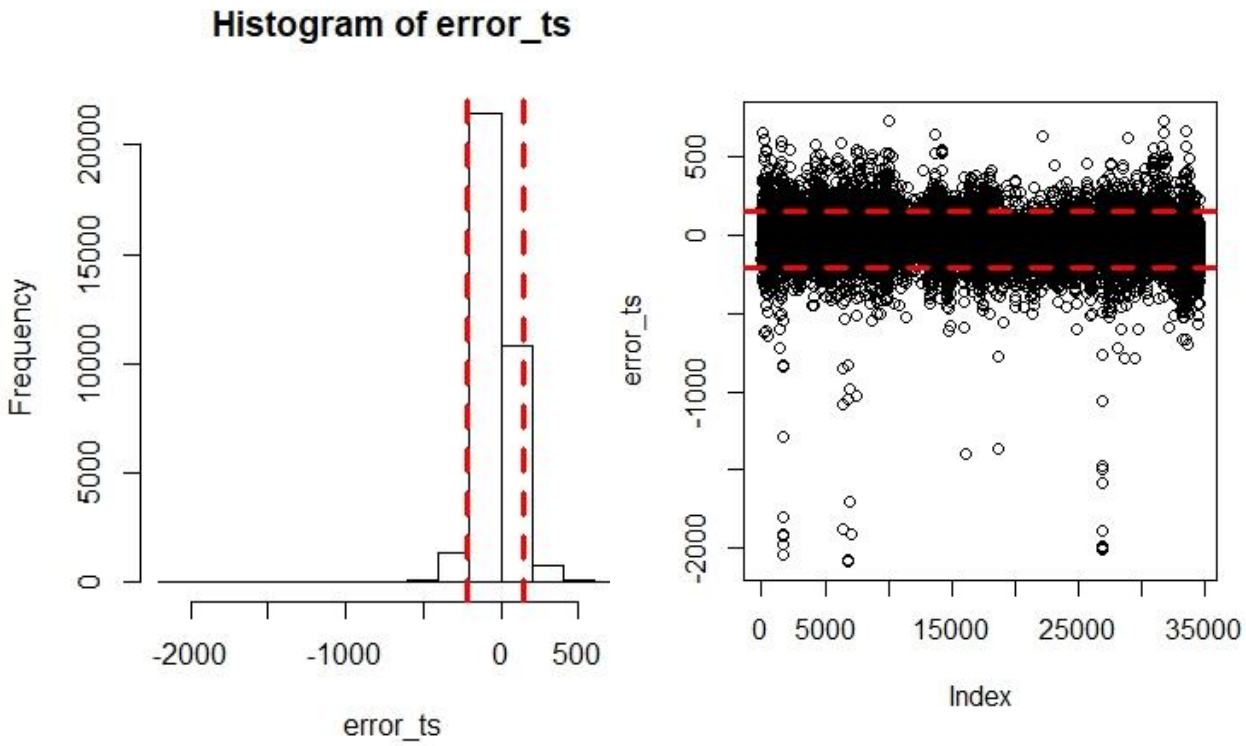


Ilustración 34. Histograma de residuos y errores de predicción en potencia. Año 2.

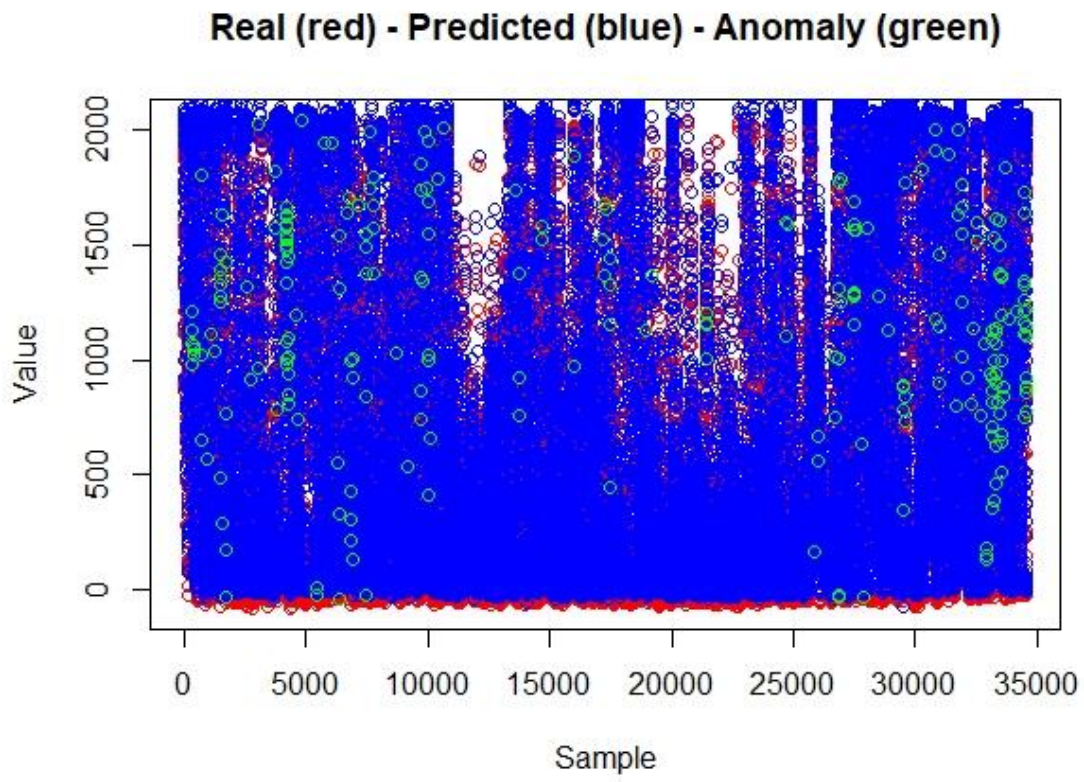


Ilustración 35. Predicciones, valores reales y anomalías en potencia. Año 2.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-211{,}378 \leq X \leq 146{,}959]$, generado con los datos del primer año, encontramos 2911 muestras. Esto es un 8,42% de las muestras, de un total de 34592 muestras.

Año	**2**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	331	268	326	292	142	178	179	90	158	258	260	429
% muestras anómalas con respecto al total mensual de muestras	11,48	9,30	11,31	10,13	4,93	6,17	6,21	3,12	5,48	8,95	9,02	14,88

Tabla 16. Cuantificación muestras anómalas en potencia. Año 2

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas se distribuyen equitativamente durante el año.

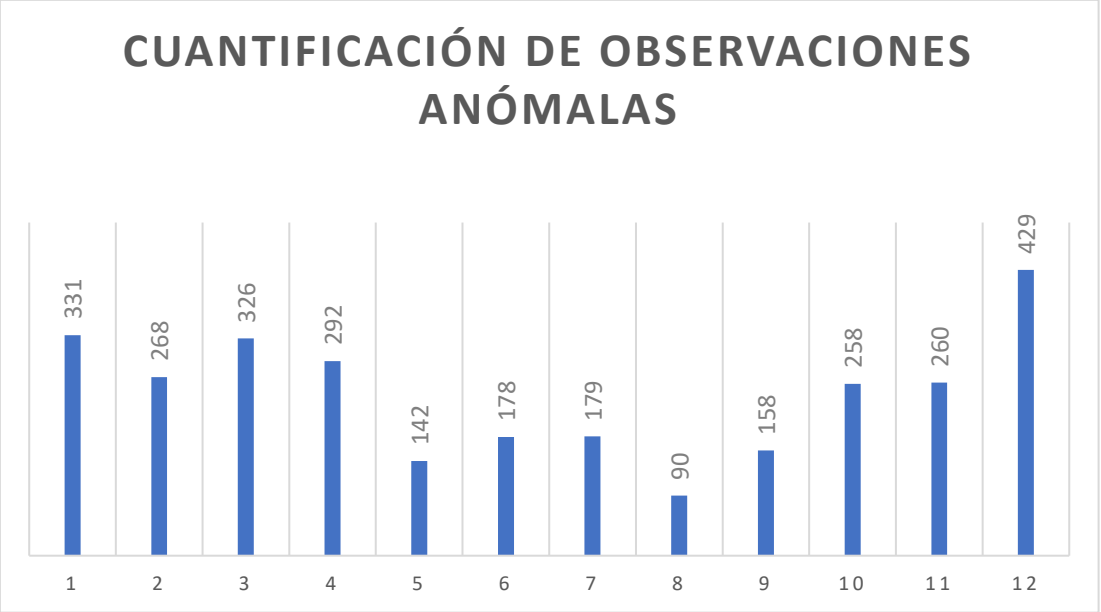


Ilustración 36. Cuantificación de observaciones anómalas en potencia. Año 2.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o mas desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**2**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	45	45	54	33	12	15	15	9	12	39	39	117
% anomalías con respecto al total mensual de muestras	1,56	1,56	1,86	1,14	0,42	0,51	0,51	0,3	0,42	1,35	1,35	4,05

Tabla 17. Cuantificación anomalías en potencia. Año 2.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías se concentran en el último mes del año. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

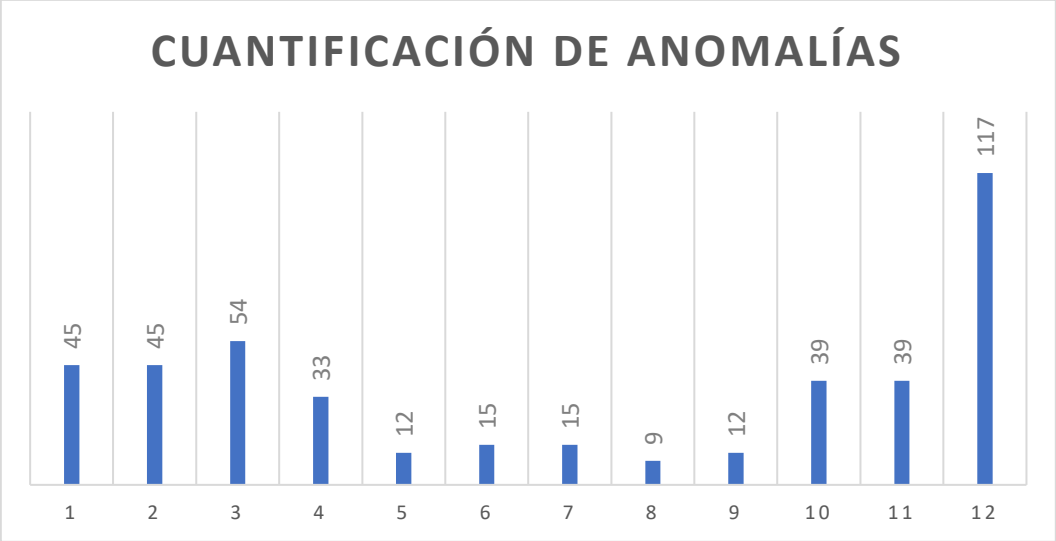


Ilustración 37. Cuantificación de anomalías en potencia. Año 2.

Año 3

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

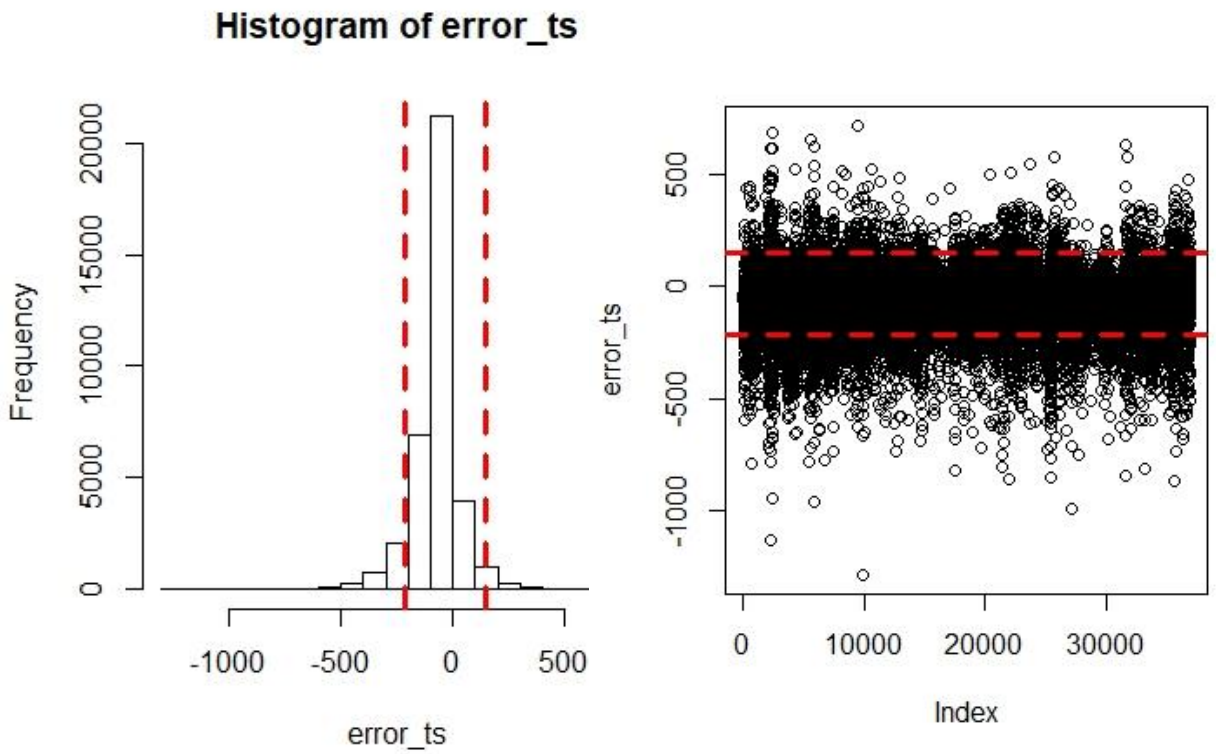


Ilustración 38. Histograma de residuos y errores de predicción en potencia. Año 3.

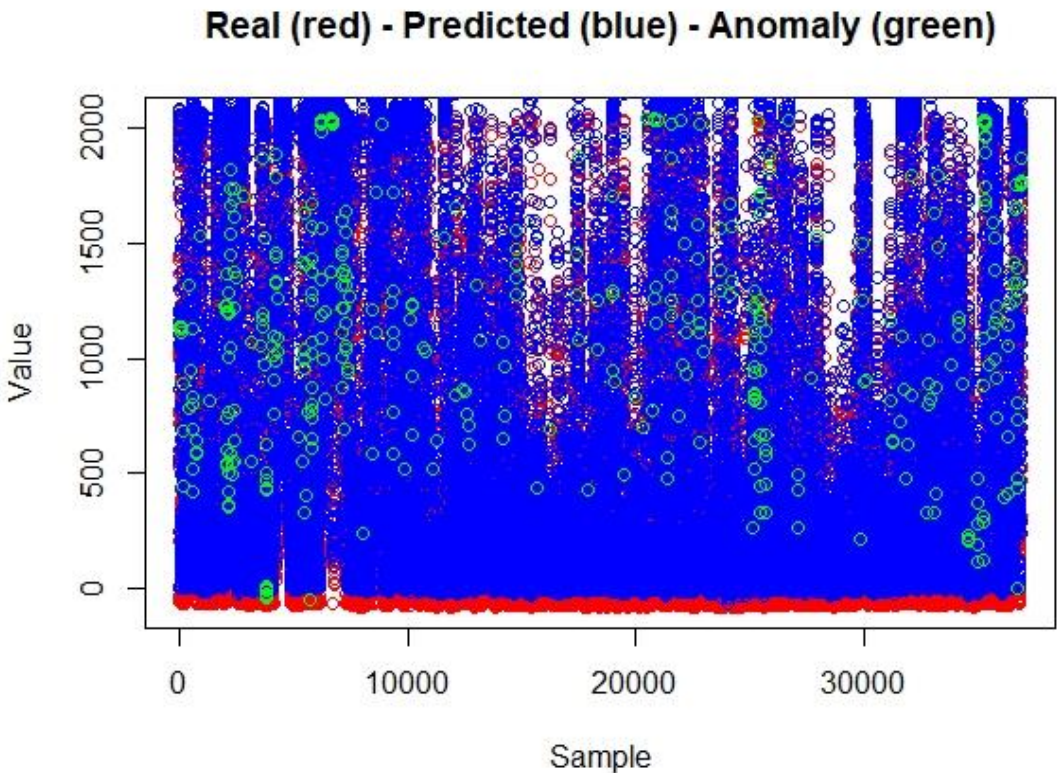


Ilustración 39. Predicciones, valores reales y anomalías en potencia. Año 3.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-211{,}378 \leq X \leq 146{,}959]$, generado con los datos del primer año, encontramos 3754 muestras. Esto es un 10,17% de las muestras, de un total de 36929 muestras.

Año	**3**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	452	359	403	332	294	158	262	280	305	163	345	401
% muestras anómalas con respecto al total mensual de muestras	14,69	11,67	13,10	10,79	9,55	5,13	8,51	9,10	9,91	5,30	11,21	13,03

Tabla 18. Cuantificación muestras anómalas en potencia. Año 3.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas son mayoritarias a principios de año, aunque son abundantes durante todo el año.

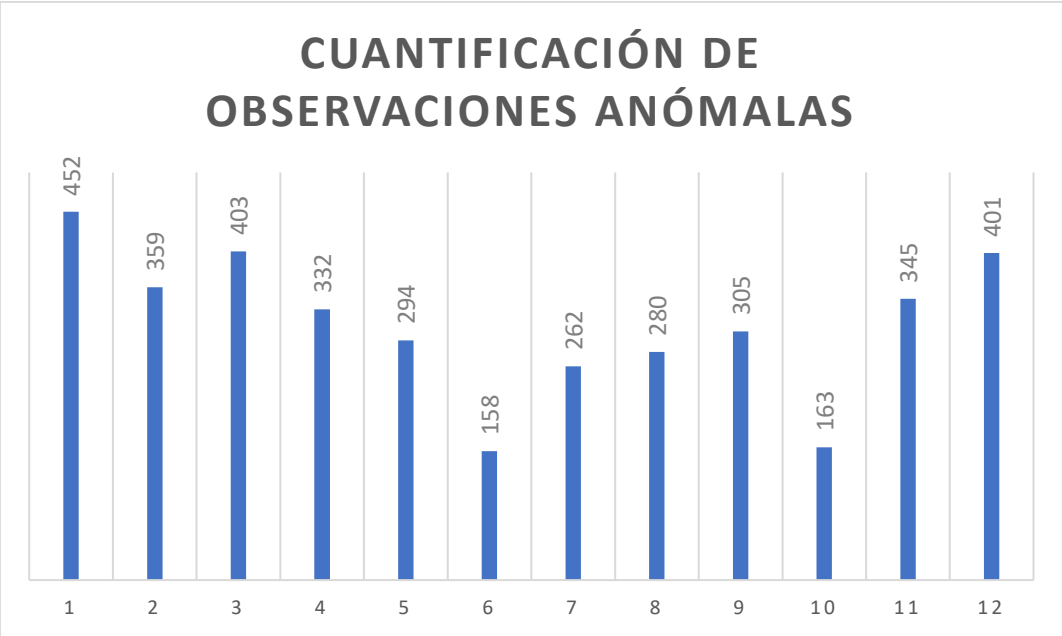


Ilustración 40. Cuantificación de observaciones anómalas en potencia. Año 3.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o mas desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**3**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	99	111	69	48	36	21	45	51	51	15	51	87
% anomalías con respecto al total mensual de muestras	3,21	3,6	2,25	1,56	1,17	0,69	1,47	1,65	1,65	0,48	1,65	2,82

Tabla 19. Cuantificación anomalías en potencia. Año 3.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías son más abundantes a principios de año, aunque aparecen a lo largo de todo el año. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

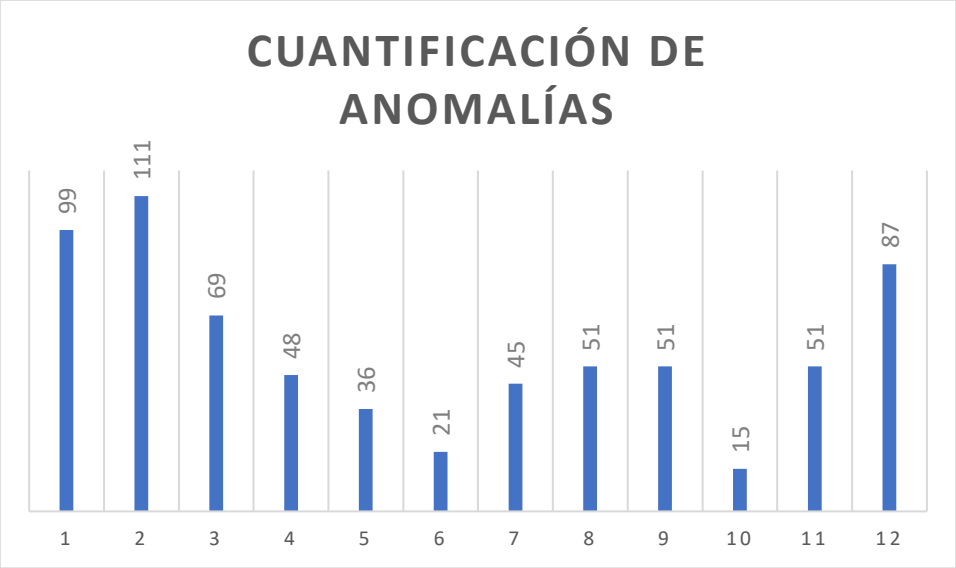


Ilustración 41. Cuantificación de anomalías en potencia. Año 3.

Año 4

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

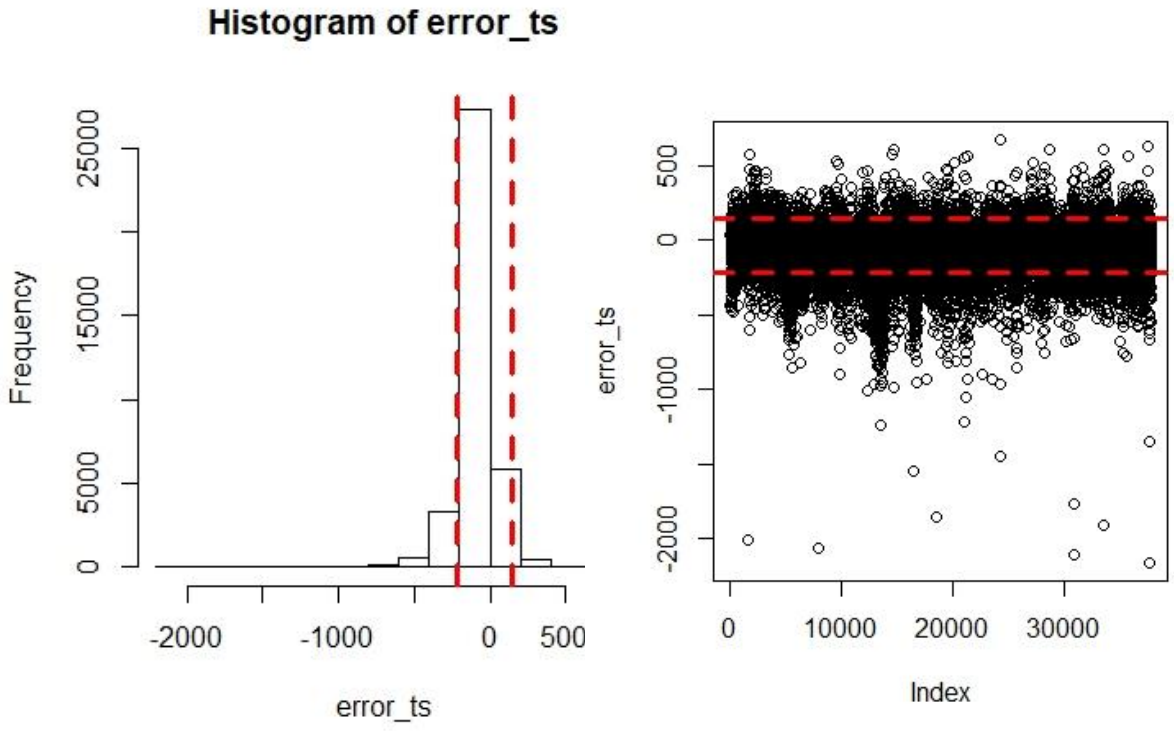


Ilustración 42. Histograma de residuos y errores de predicción en potencia. Año 4.

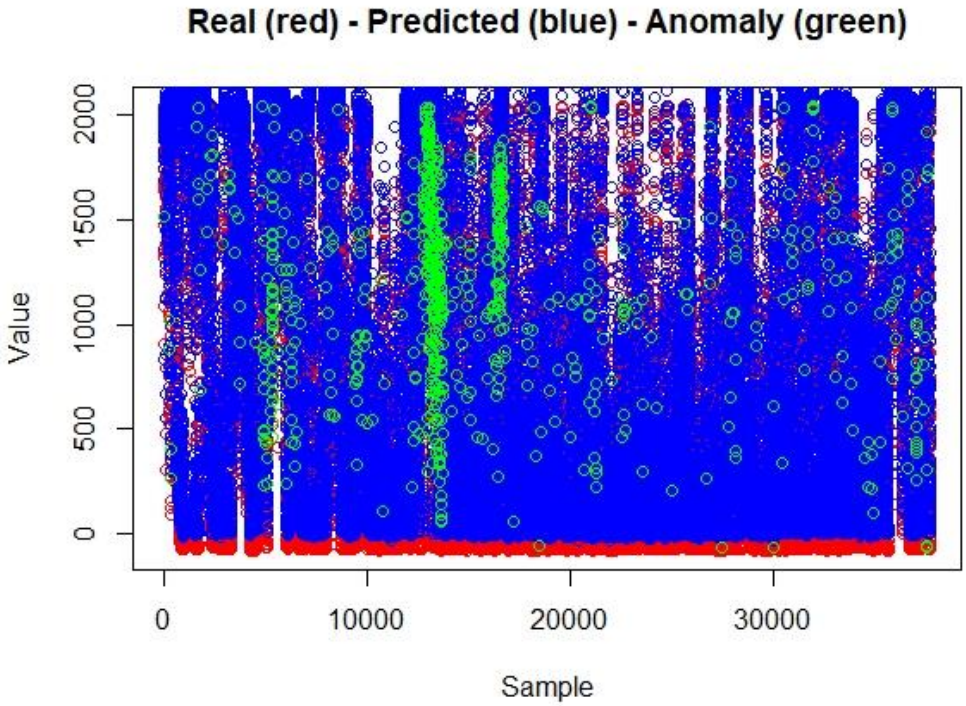


Ilustración 43. Predicciones, valores reales y anomalías en potencia. Año 4.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-211{,}378 \leq X \leq 146{,}959]$, generado con los datos del primer año, encontramos 4622 muestras. Esto es un 12,25% de las muestras, de un total de 37721 muestras.

Año	**4**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	246	483	315	394	781	460	354	218	336	261	384	390
% muestras anómalas con respecto al total mensual de muestras	7,83	15,37	10,02	12,53	24,85	14,63	11,26	6,94	10,69	8,30	12,22	12,41

Tabla 20. Cuantificación muestras anómalas en potencia. Año 4.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas son mas abundantes a mitad de año.

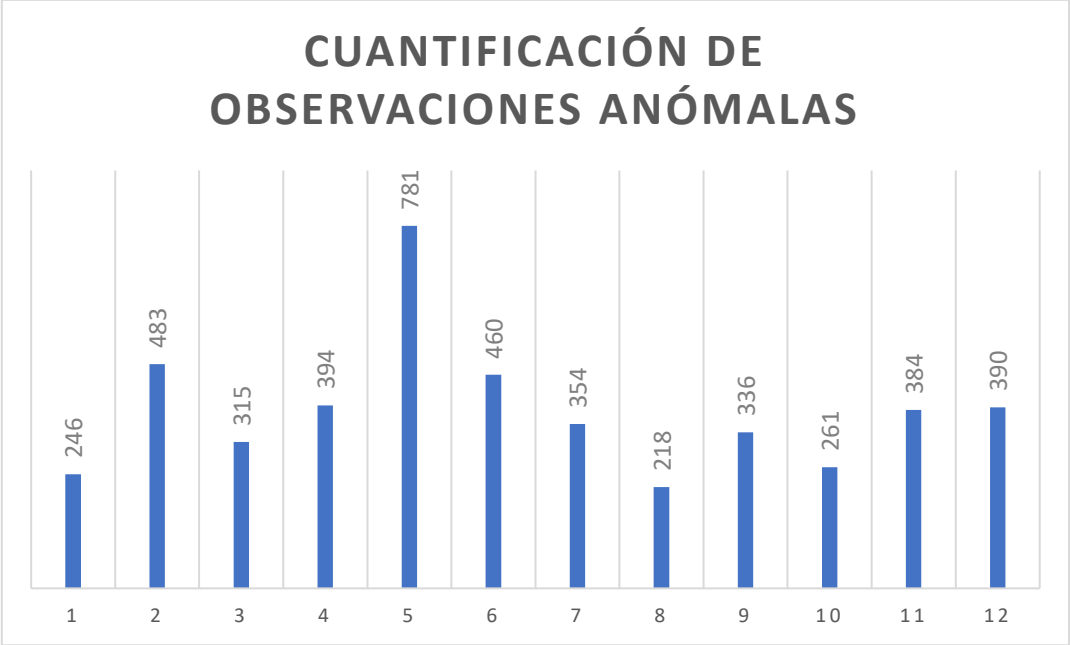


Ilustración 44. Cuantificación de observaciones anómalas en potencia. Año 4.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o mas desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**4**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	45	111	66	87	282	123	66	39	51	24	48	84
% anomalías con respecto al total mensual de muestras	1,44	3,54	2,1	2,76	8,97	3,9	2,1	1,23	1,62	0,75	1,53	2,67

Tabla 21. Cuantificación anomalías en potencia. Año 4.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías se especialmente en el mes de mayo, donde aparece un pico. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

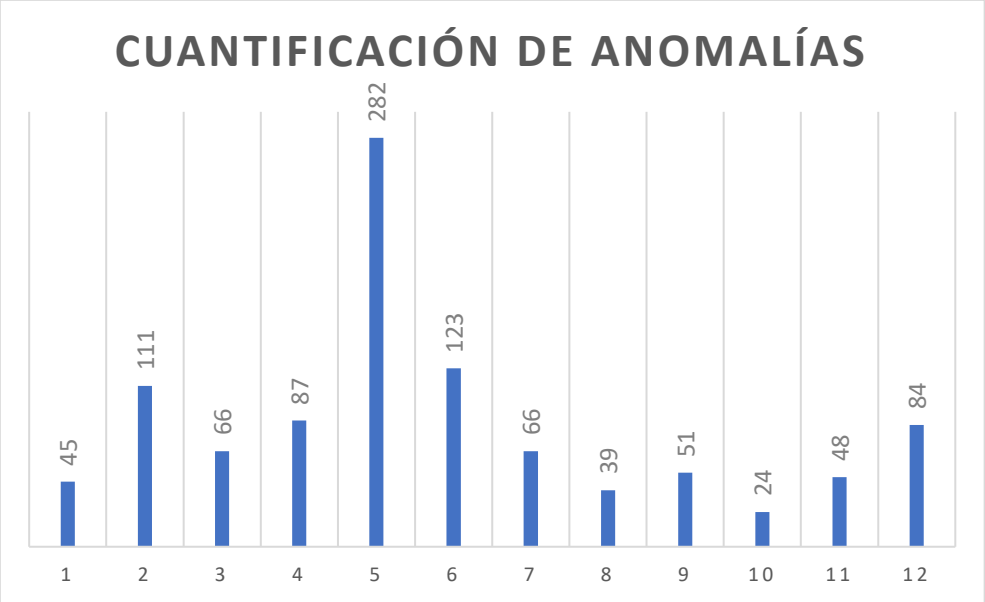


Ilustración 45. Cuantificación de anomalías en potencia. Año 4.

Año 5

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

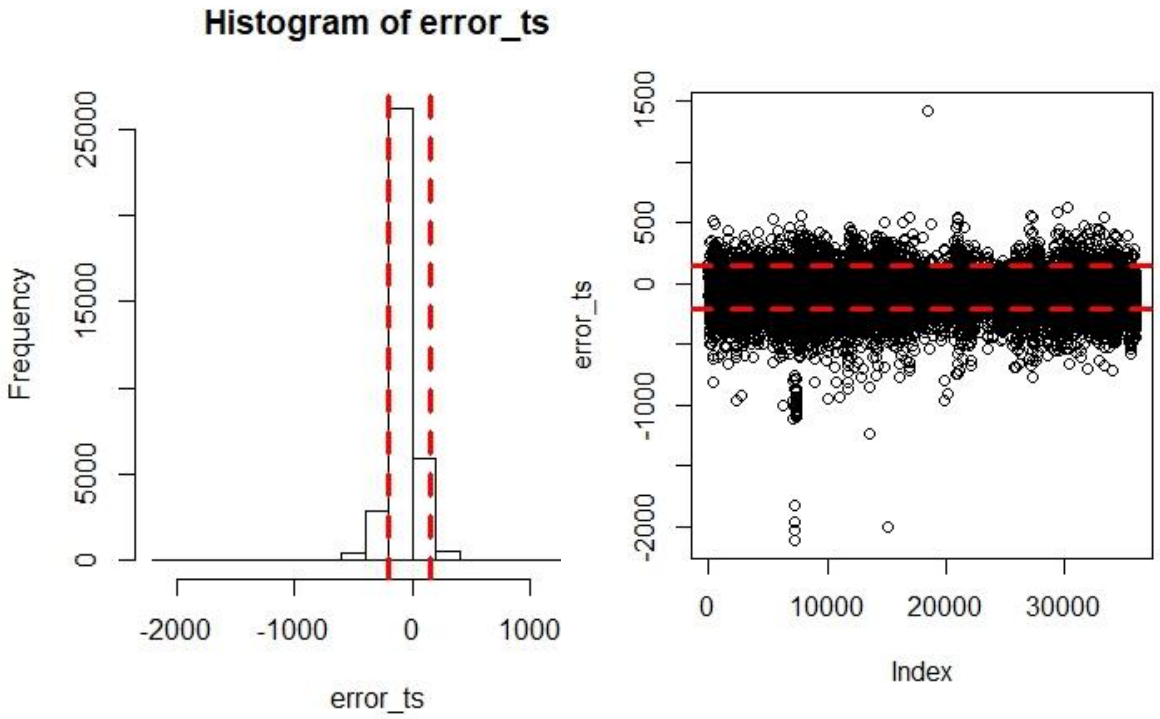


Ilustración 46. Histograma de residuos y errores de predicción en potencia. Año 5.

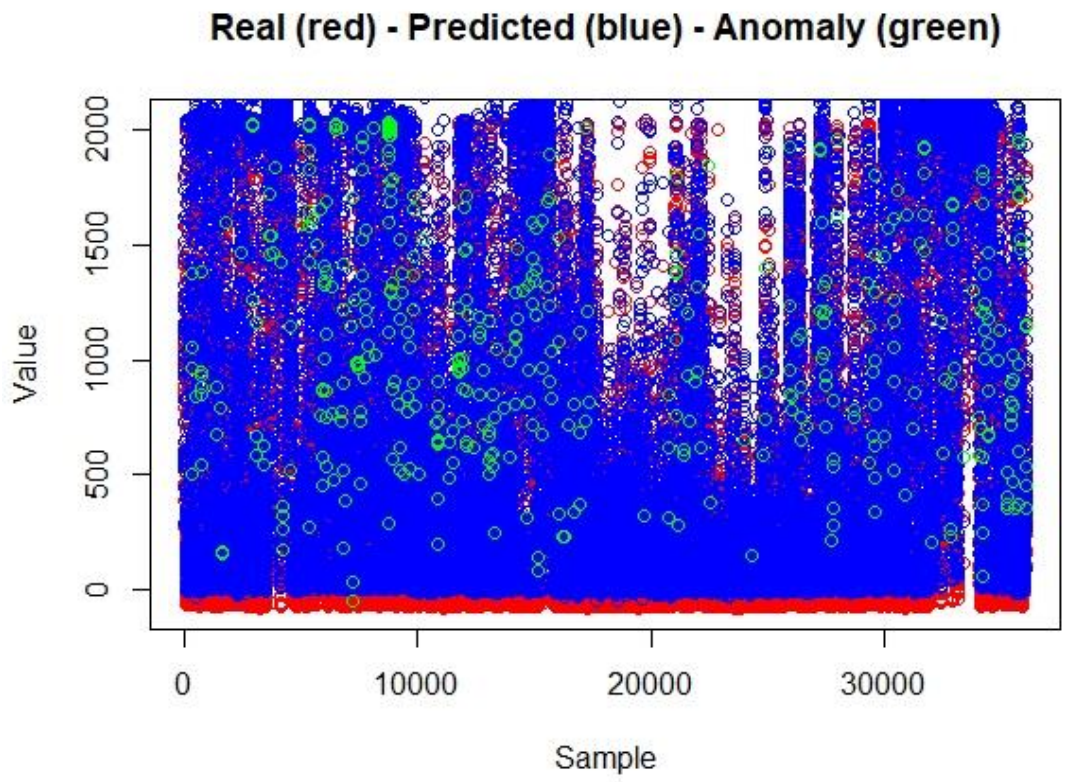


Ilustración 47. Predicciones, valores reales y anomalías en potencia. Año 5.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-211{,}378 \leq X \leq 146{,}959]$, generado con los datos del primer año, encontramos 3993 muestras. Esto es un 11,07% de las muestras, de un total de 36083 muestras.

Año	**5**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	331	343	532	420	390	333	142	230	179	352	375	362
% muestras anómalas con respecto al total mensual de muestras	11,01	11,41	17,69	13,97	12,97	11,07	4,72	7,65	5,95	11,71	12,47	12,04

Tabla 22. Cuantificación muestras anómalas en potencia. Año 5.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas se distribuyen equilibradamente a lo largo de todo el año.

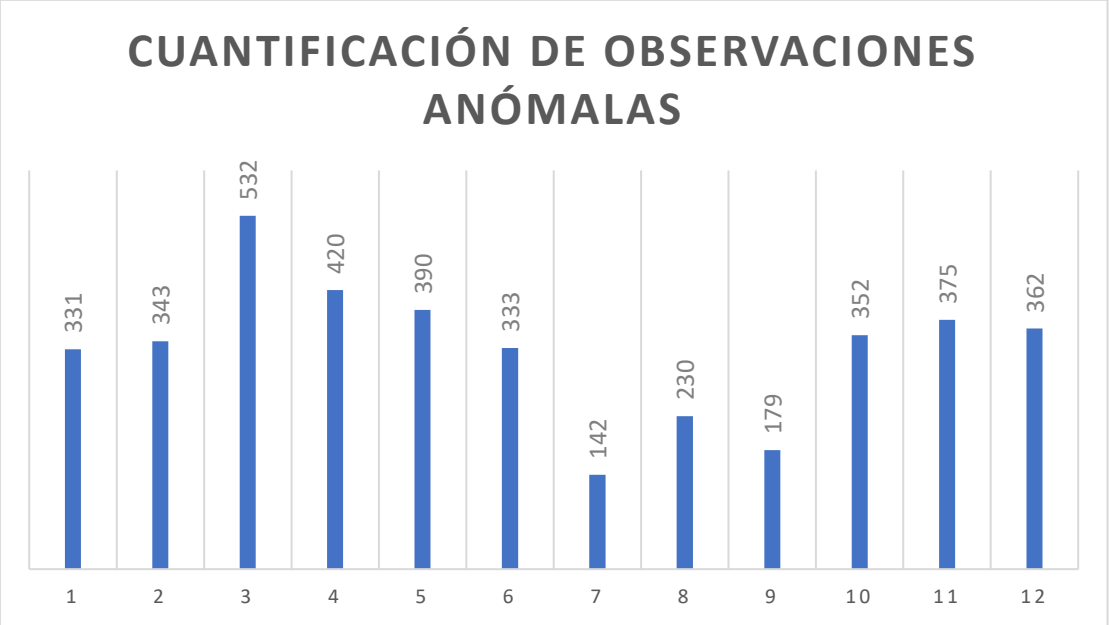


Ilustración 48. Cuantificación de observaciones anómalas en potencia. Año 5.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o mas desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**5**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	36	84	114	84	93	78	6	48	42	60	66	75
% anomalías con respecto al total mensual de muestras	1,2	2,79	3,78	2,79	3,09	2,58	0,21	1,59	1,41	2,01	2,19	2,49

Tabla 23. Cuantificación anomalías en potencia. Año 5.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías se concentran especialmente en el primer semestre. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

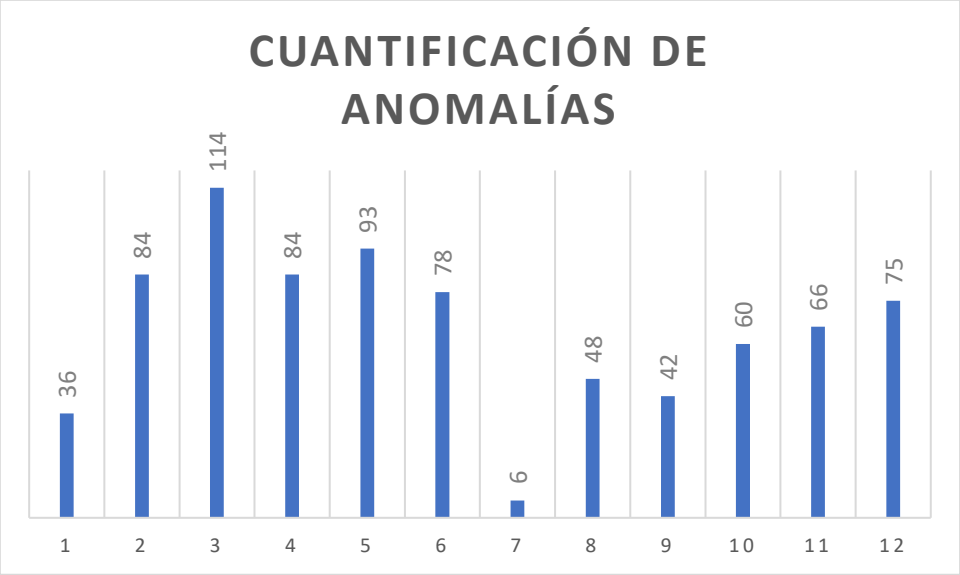


Ilustración 49. Cuantificación de anomalías en potencia. Año 5.

Año 6

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

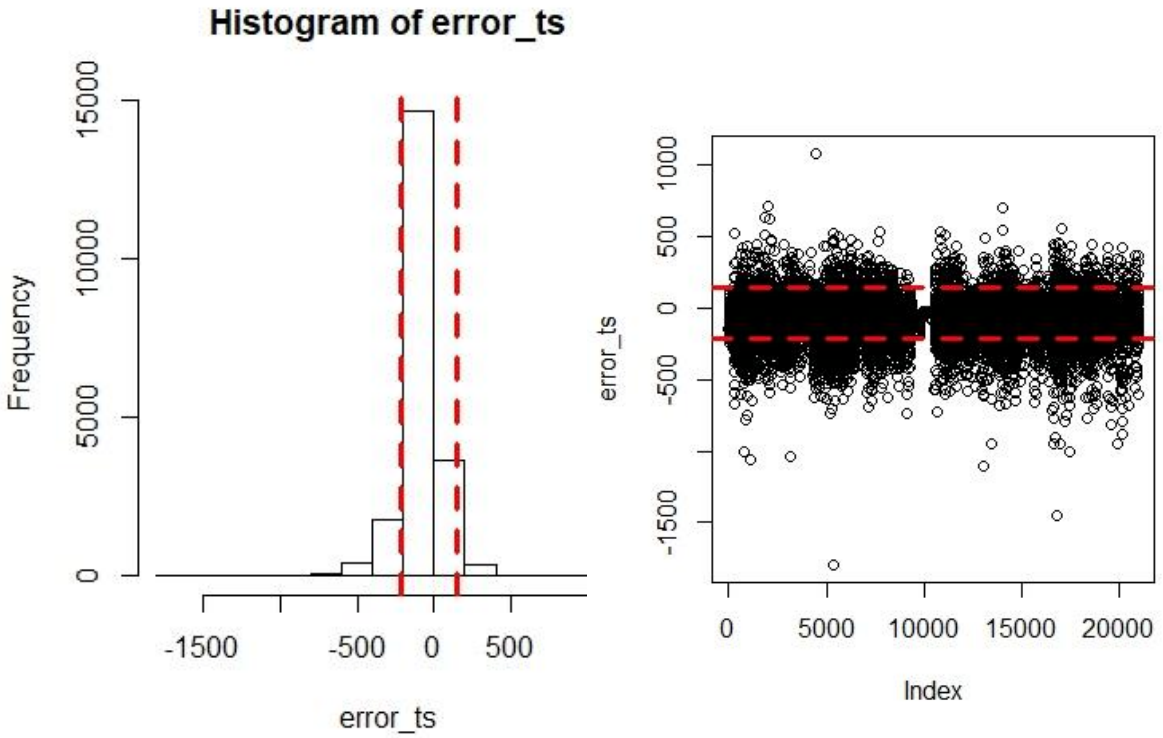


Ilustración 50. Histograma de residuos y errores de predicción en potencia. Año 6.

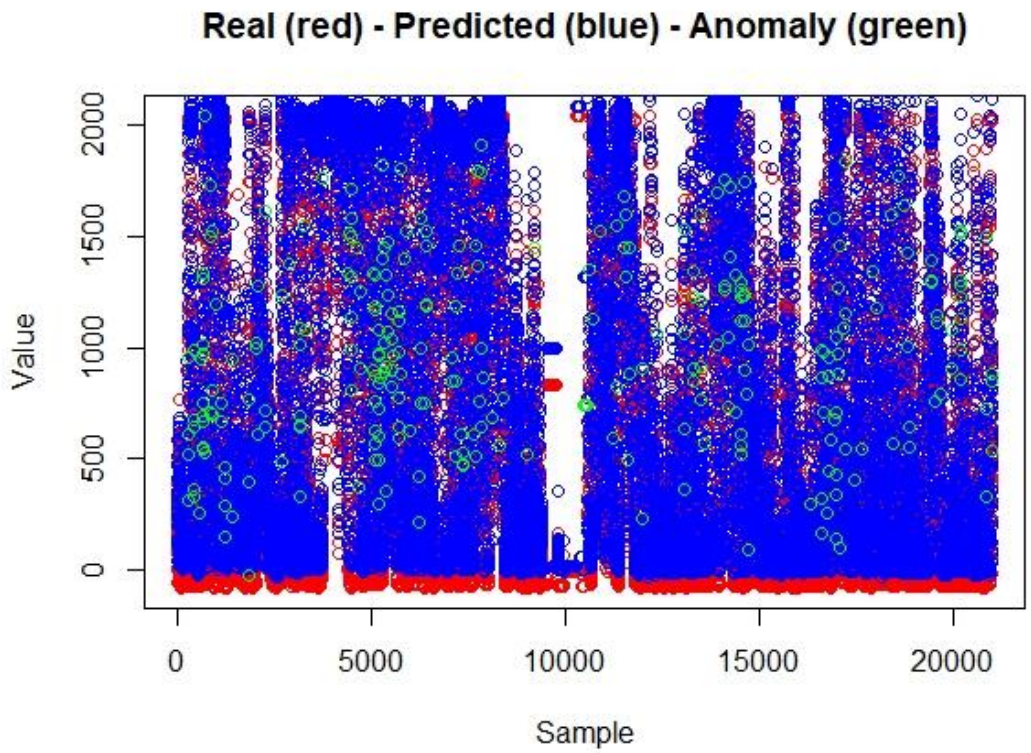


Ilustración 51. Predicciones, valores reales y anomalías en potencia. Año 6.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-211{,}378 \leq X \leq 146{,}959]$, generado con los datos del primer año, encontramos 2803 muestras. Esto es un 13,34% de las muestras, de un total de 21006 muestras.

Año	**6**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	266	262	237	281	264	66	333	174	239	279	216	185
% muestras anómalas con respecto al total mensual de muestras	15,20	14,97	13,54	16,05	15,08	3,77	19,02	9,94	13,65	15,94	12,34	10,57

Tabla 24. Cuantificación muestras anómalas en potencia. Año 6.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas se distribuyen equilibradamente a lo largo del año.

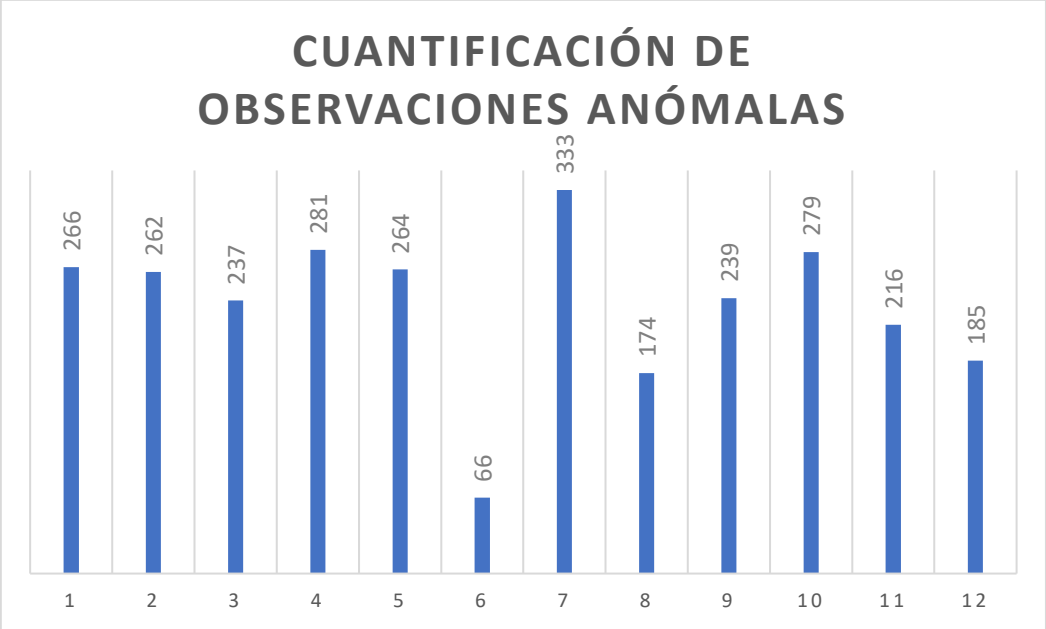


Ilustración 52. Cuantificación de observaciones anómalas en potencia. Año 6.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o mas desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**5**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	57	54	36	66	48	6	42	30	39	66	30	39
% anomalías con respecto al total mensual de muestras	3,26	3,08	2,06	3,77	2,74	0,34	2,40	1,71	2,23	3,77	1,71	2,23

Tabla 25. Cuantificación anomalías en potencia. Año 6.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías se distribuyen equilibradamente a lo largo del año. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

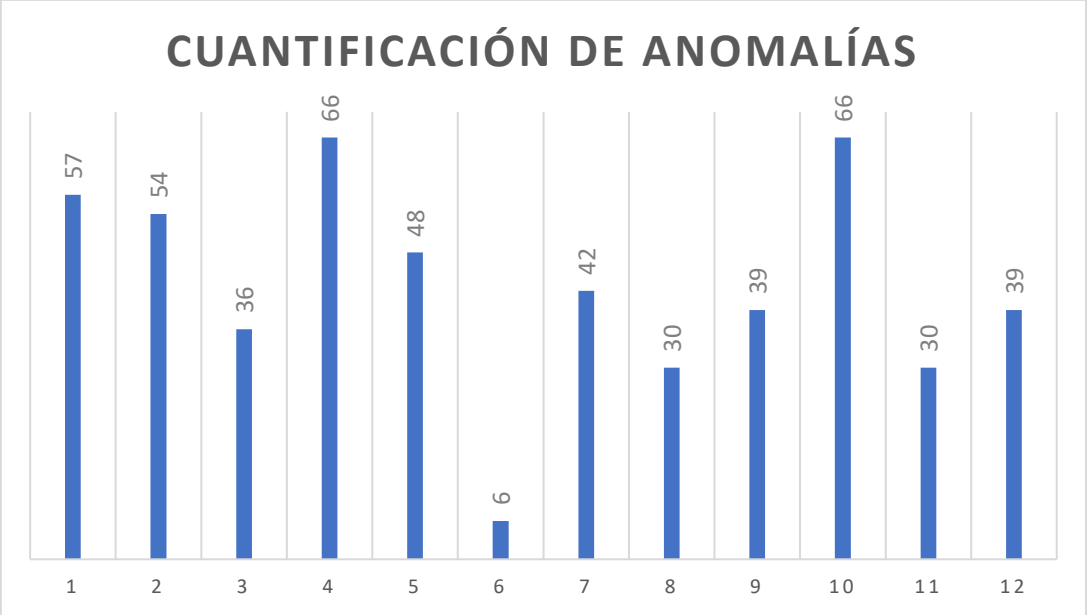


Ilustración 53. Cuantificación de anomalías en potencia. Año 6.

6.3.- Modelo de previsión de la temperatura de los anillos

A partir de los intervalos de confianza calculados para el primer año, establecemos el intervalo de confianza que se empleará para los años sucesivos.

$$[-3{,}887 \leq X \leq 3{,}894]\ U\ [-3{,}927 \leq X \leq 4{,}006] = [-3{,}927 \leq X \leq 4{,}006]$$

Recordamos que el intervalo de confianza del 95% para el modelo ha sido establecido con el intervalo de confianza del 95% del conjunto de entrenamiento unido al intervalo de intervalo de confianza del 95% del conjunto de test de los datos del primer año. Esto se ha hecho cogiendo la opción menos restringida, para evitar incluir errores aportados por el propio modelo.

Año 2

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

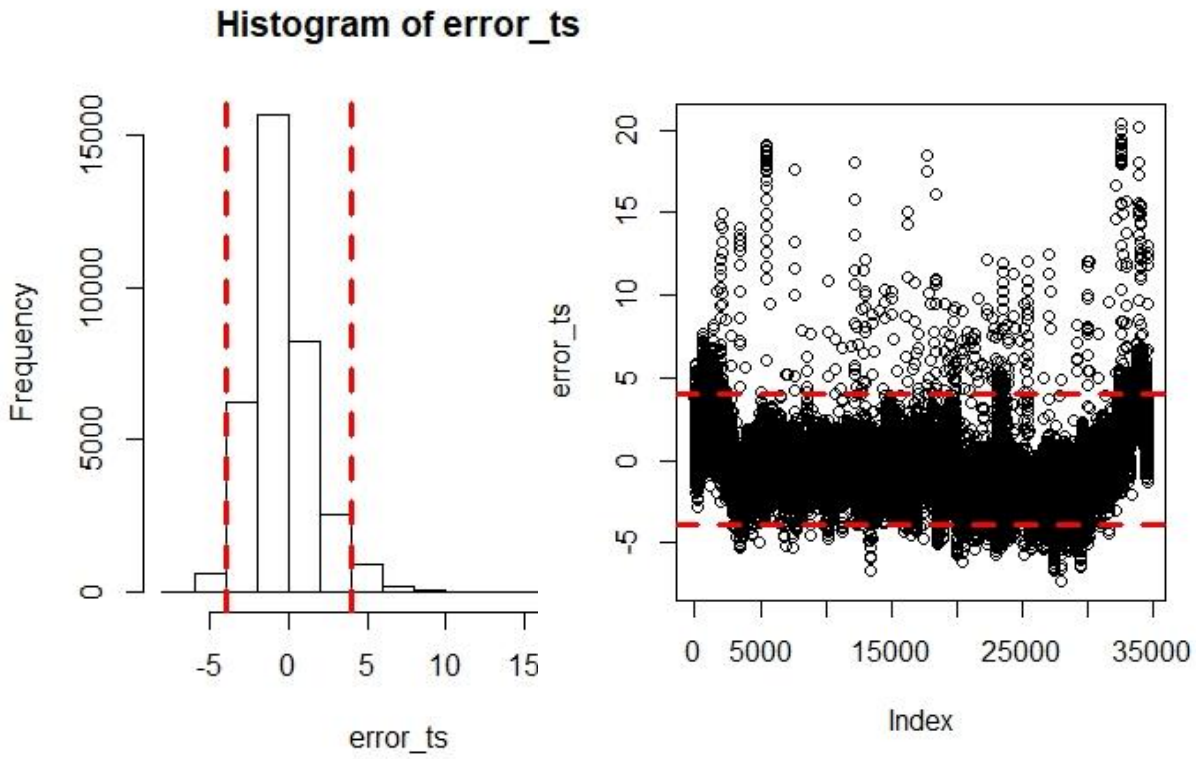


Ilustración 54. Histograma de residuos y errores de predicción en los anillos. Año 2.

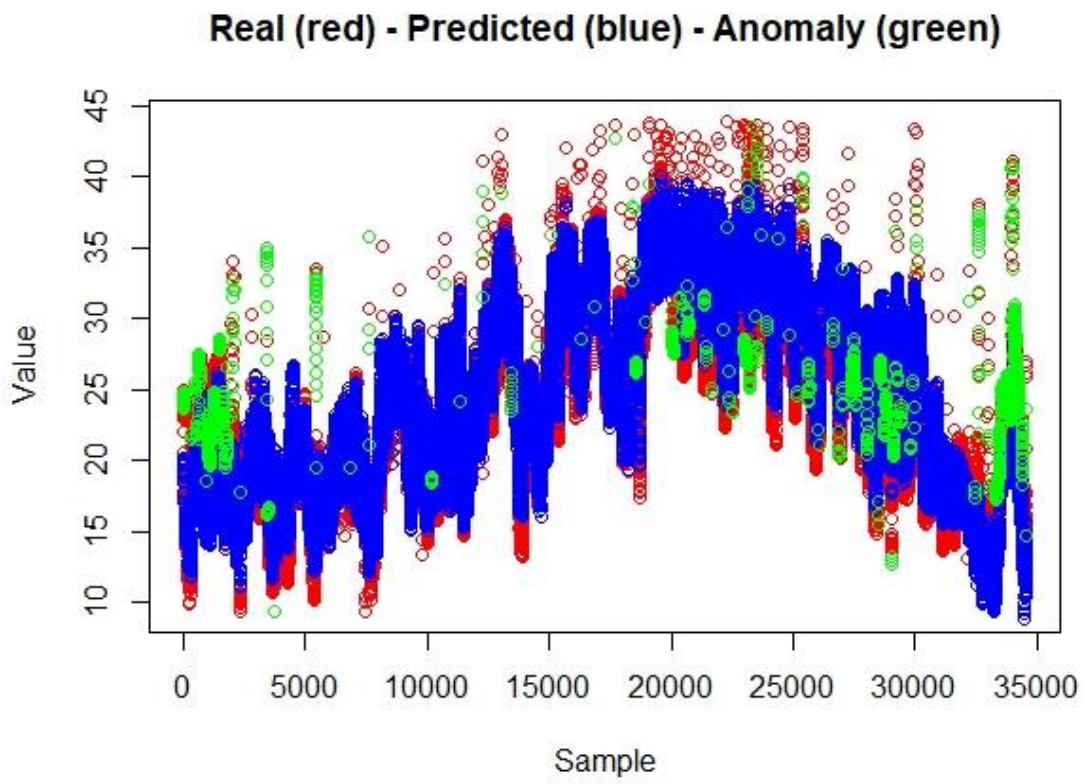


Ilustración 55. Predicciones, valores reales y anomalías en los anillos. Año 2.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-3{,}927 \leq X \leq 4{,}006]$, generado con los datos del primer año, encontramos 2025 muestras. Esto es un 5,58% de las muestras, de un total de 34590 muestras

Año	**2**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	311	75	24	24	44	36	87	160	151	283	137	692
% muestras anómalas con respecto al total mensual de muestras	10,79	2,60	0,83	0,83	1,53	1,25	3,02	5,55	5,24	9,82	4,75	24,01

Tabla 26. Cuantificación muestras anómalas en los anillos. Año 2.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas se concentran el último mes del año.

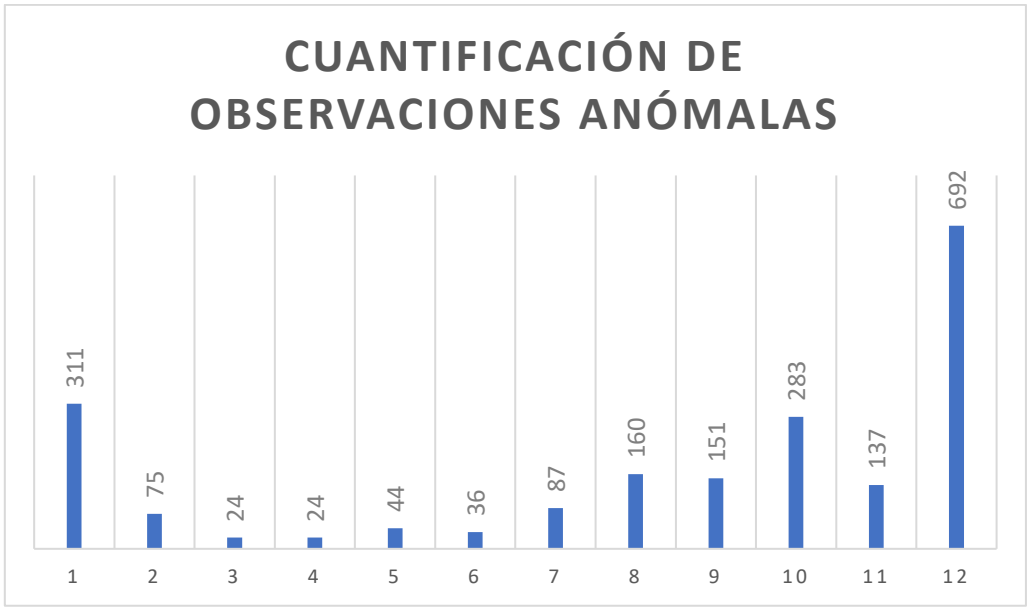


Ilustración 56. Cuantificación de observaciones anómalas en los anillos. Año 2.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o mas desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**2**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	84	12	6	9	9	12	21	45	66	84	48	135
% anomalías con respecto al total mensual de muestras	2,91	0,42	0,21	0,3	0,3	0,42	0,72	1,56	2,28	2,91	1,68	4,68

Tabla 27. Cuantificación anomalías en los anillos. Año 2.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías se concentran en los últimos meses del año. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

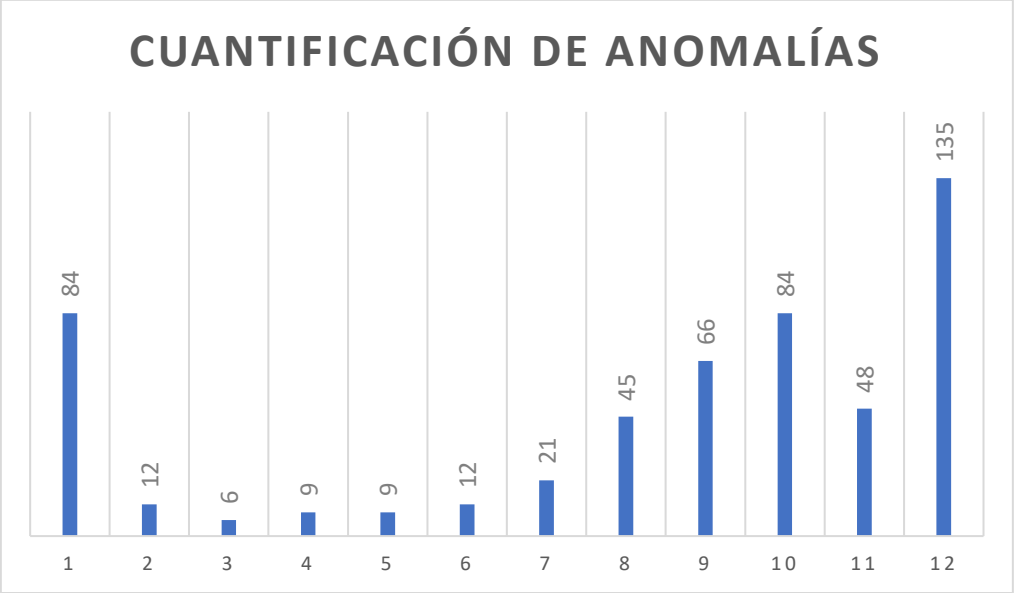


Ilustración 57. Cuantificación de anomalías en los anillos. Año 2.

Año 3

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

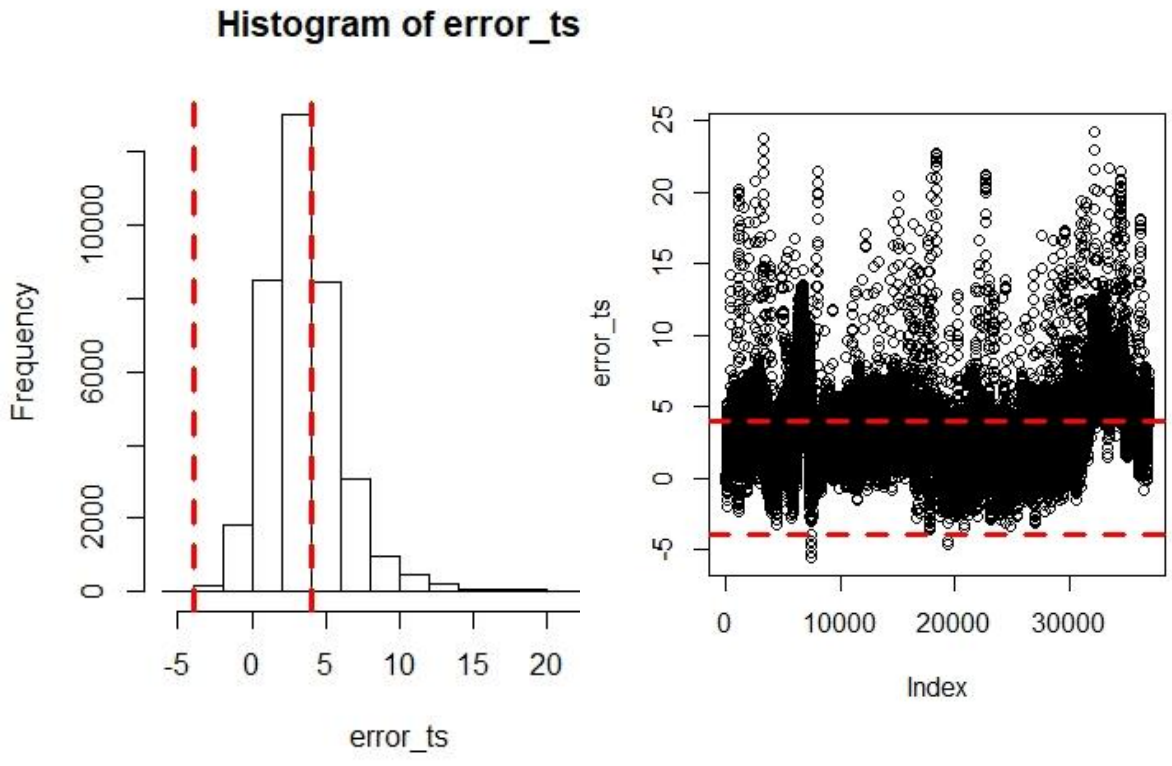


Ilustración 58. Histograma de residuos y errores de predicción en los anillos. Año 3.

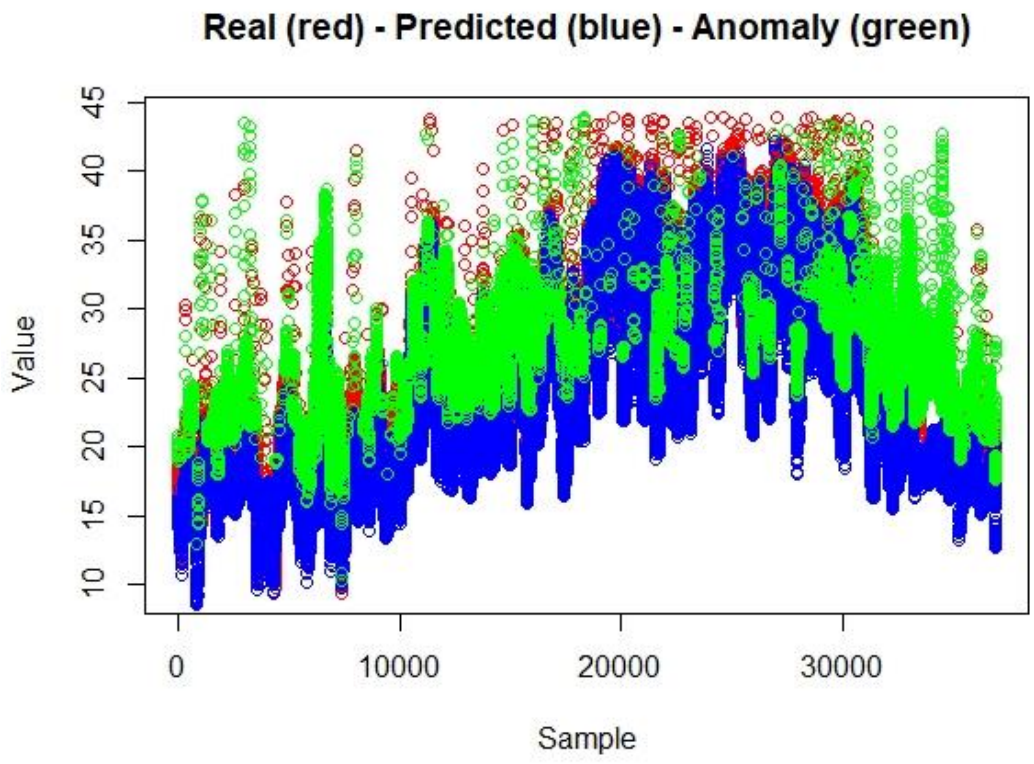


Ilustración 59. Predicciones, valores reales y anomalías en los anillos. Año 3.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-3{,}927 \leq X \leq 4{,}006]$, generado con los datos del primer año, encontramos 13353 muestras. Esto es un 36,16% de las muestras, de un total de 36928 muestras

Año	**3**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	1102	555	1398	1002	1474	1066	154	502	381	978	2710	2030
% muestras anómalas con respecto al total mensual de muestras	35,81	18,04	45,43	32,56	47,90	34,64	5,00	16,31	12,38	31,78	88,06	65,97

Tabla 28. Cuantificación muestras anómalas en los anillos. Año 3.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas se concentran en los últimos del año, aunque son abundantes durante todo el año.

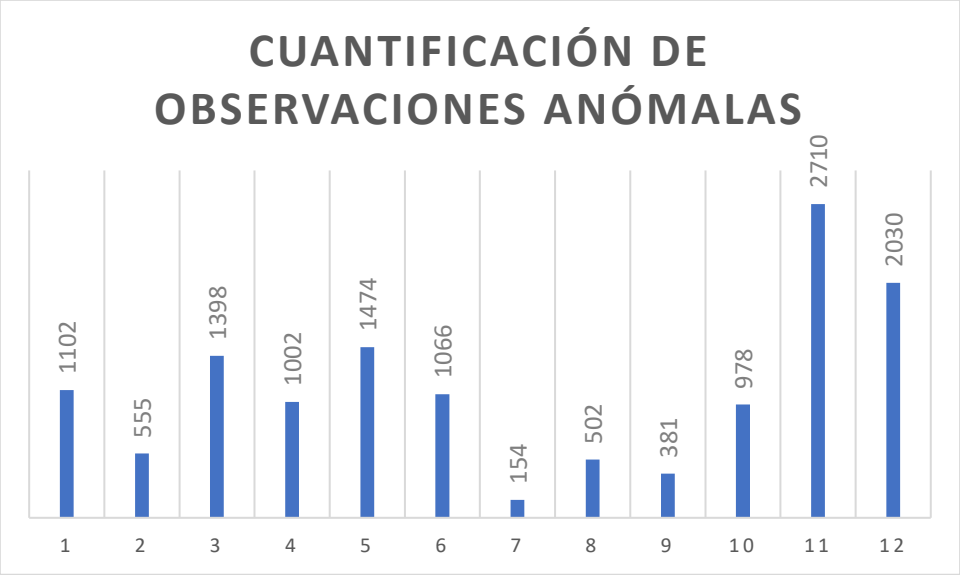


Ilustración 60. Cuantificación de observaciones anómalas en los anillos. Año 3.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o mas desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**3**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	216	168	165	207	297	270	63	171	90	180	114	174
% anomalías con respecto al total mensual de muestras	7,02	5,46	5,37	6,72	9,66	8,76	2,04	5,55	2,91	5,85	3,69	5,64

Tabla 29. Cuantificación anomalías en los anillos. Año 3.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías se concentran a lo largo del año, especialmente en la primera mitad del semestre. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

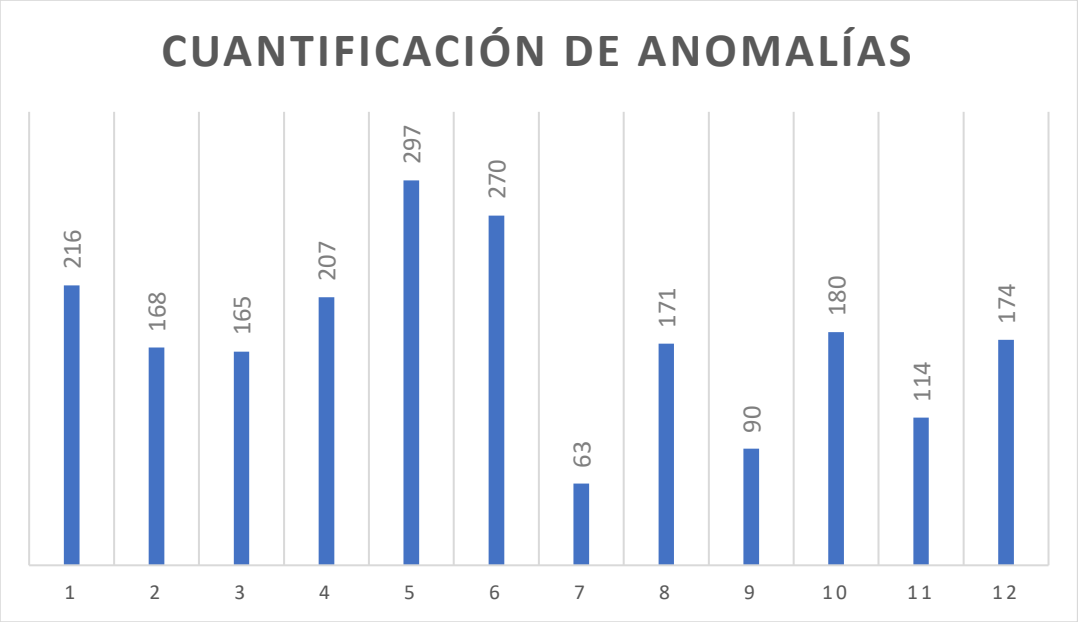


Ilustración 61. Cuantificación de anomalías en los anillos. Año 3.

Año 4

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

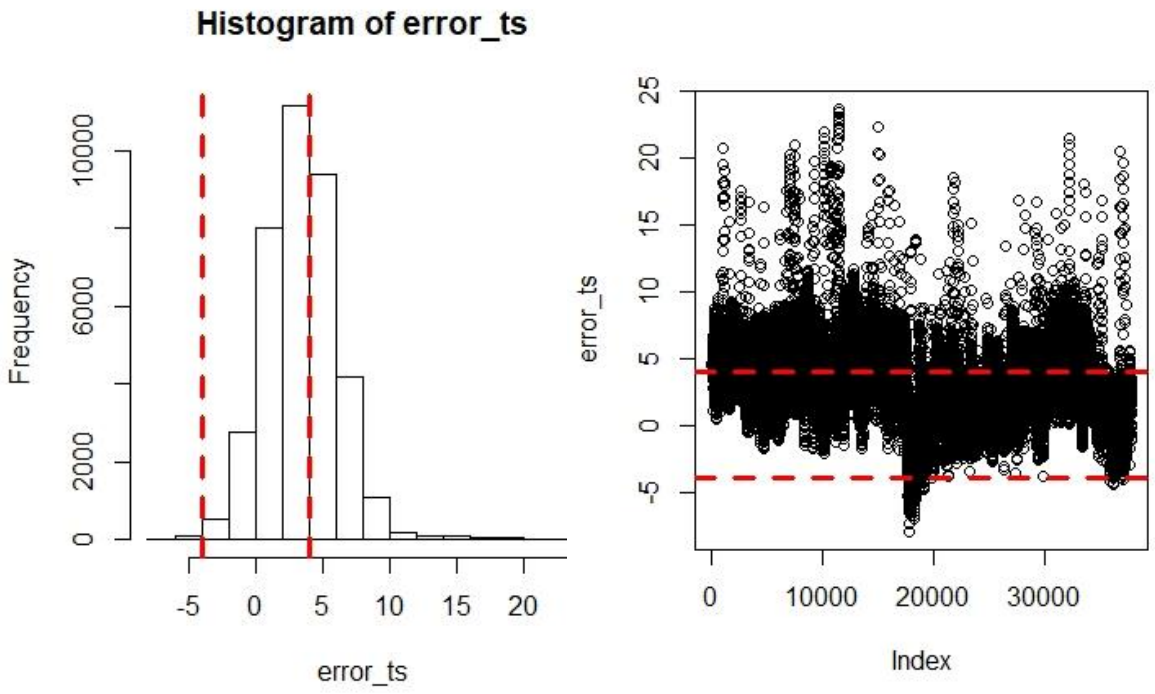


Ilustración 62. Histograma de residuos y errores de predicción en los anillos. Año 4.

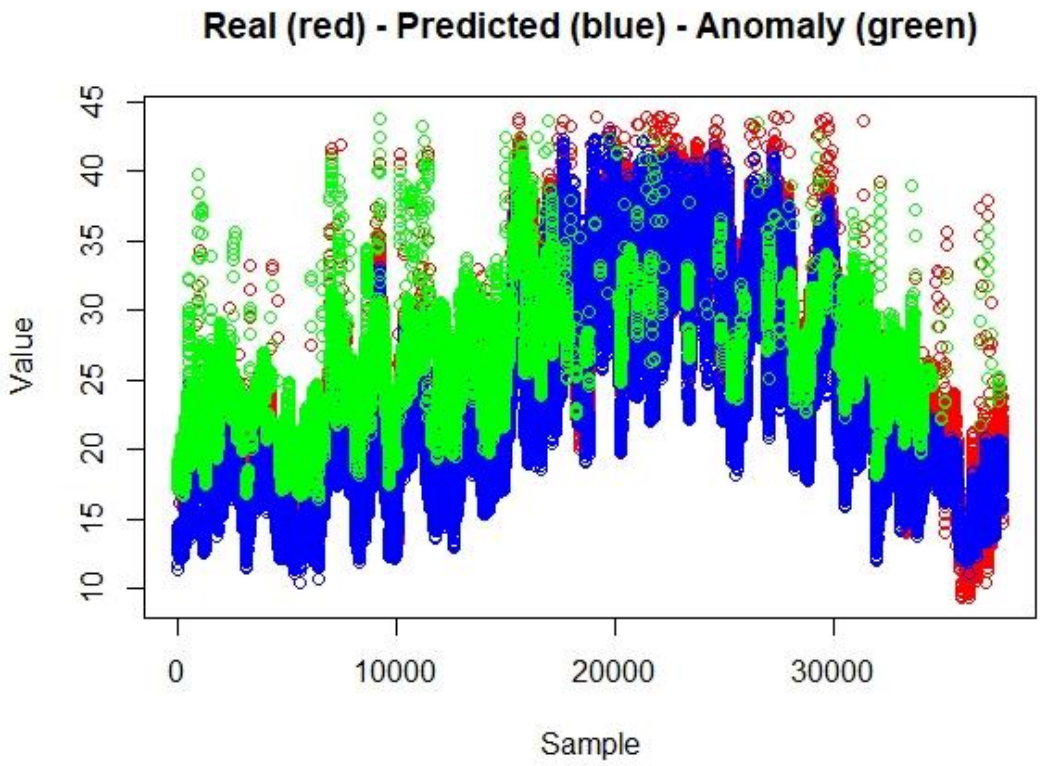


Ilustración 63. Predicciones, valores reales y anomalías en los anillos. Año 4.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-3{,}927 \leq X \leq 4{,}006]$, generado con los datos del primer año, encontramos 15234 muestras. Esto es un 40,38% de las muestras, de un total de 37723 muestras.

Año	**3**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	1102	2150	1800	1864	1838	1993	1201	351	200	589	1314	1829
% muestras anómalas con respecto al total mensual de muestras	68,39	57,26	59,30	58,47	63,40	38,20	11,17	6,36	18,74	41,80	58,18	3,28

Tabla 30. Cuantificación muestras anómalas en los anillos. Año 4.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas se concentran en la primera mitad del semestre.

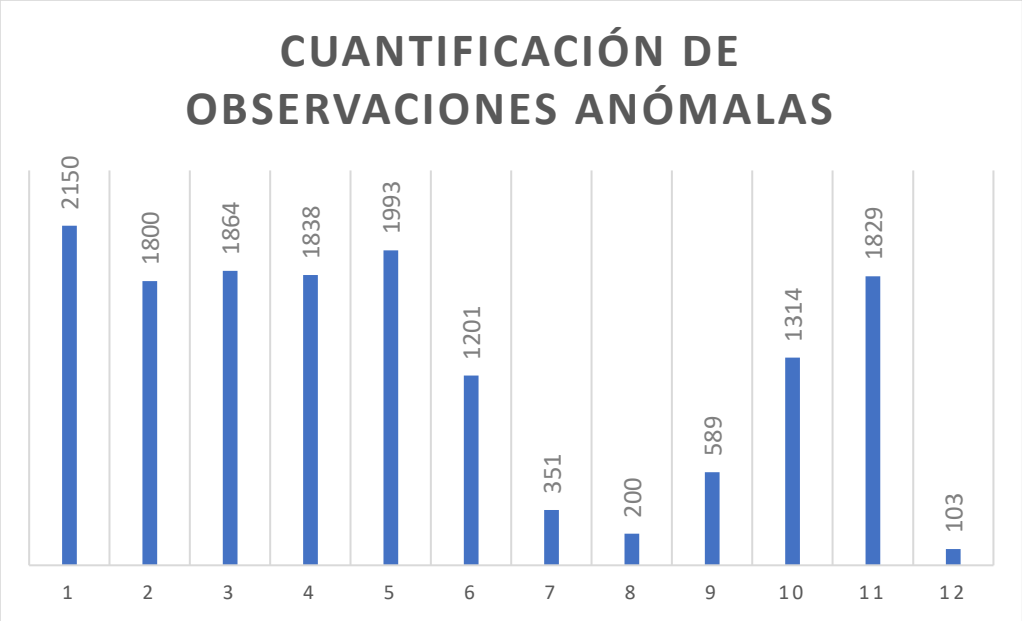


Ilustración 64. Cuantificación de observaciones anómalas en los anillos. Año 4.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o mas desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**2**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	144	144	219	195	177	180	93	51	129	210	138	36
% anomalías con respecto al total mensual de muestras	4,59	4,59	6,96	6,21	5,64	5,73	2,97	1,62	4,11	6,69	4,38	1,14

Tabla 31. Cuantificación anomalías en los anillos. Año 4.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías se concentran a lo largo del año, aunque especialmente la primera mitad del semestre. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

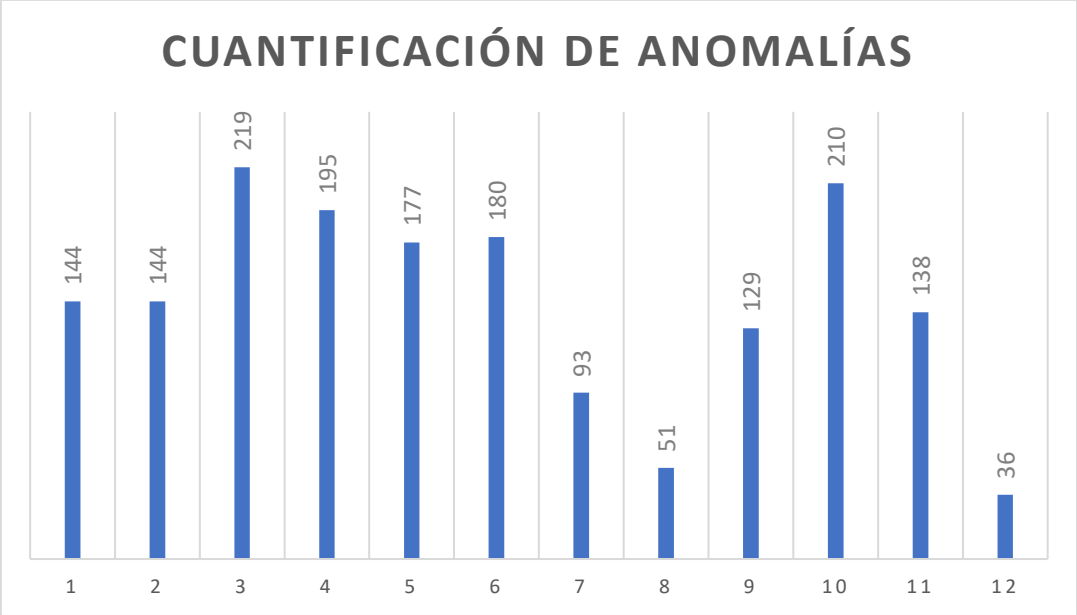


Ilustración 65. Cuantificación de anomalías en los anillos. Año 4.

Año 5

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

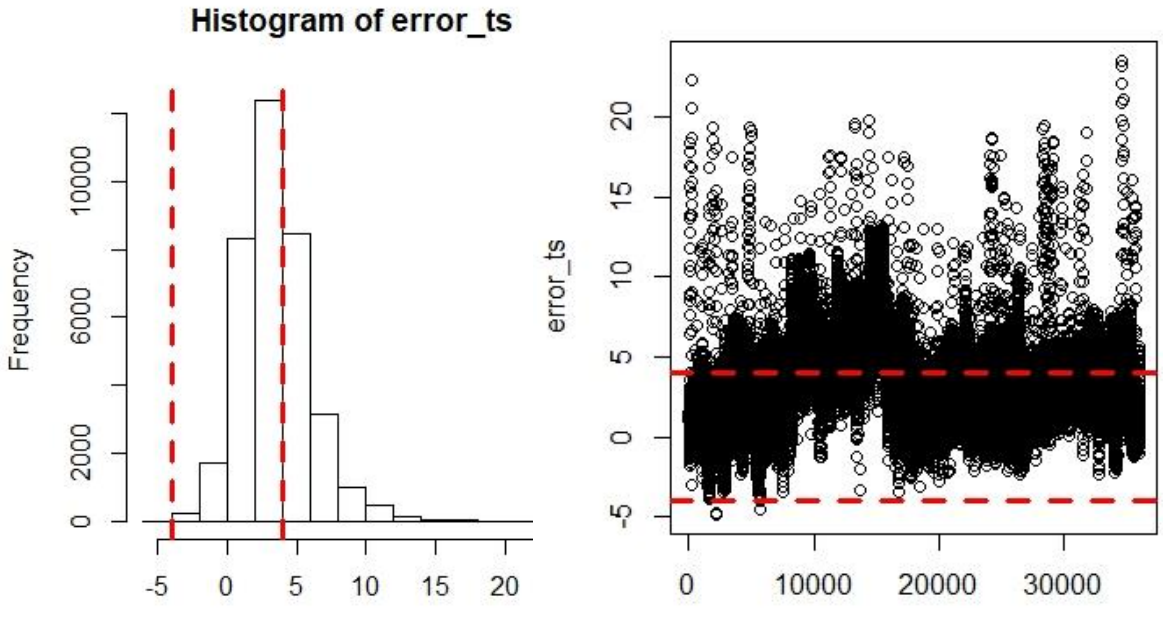


Ilustración 66. Histograma de residuos y errores de predicción en los anillos. Año 5.

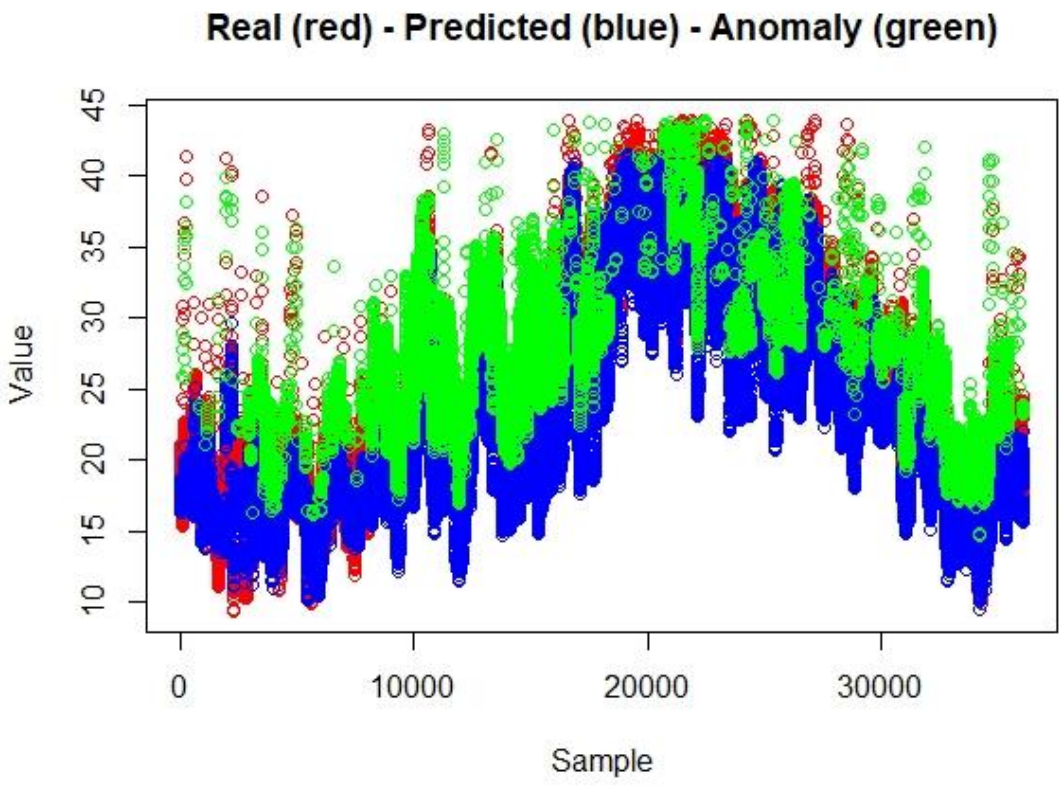


Ilustración 67. Predicciones, valores reales y anomalías en los anillos. Año 5.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-3{,}927 \leq X \leq 4{,}006]$, generado con los datos del primer año, encontramos 13401 muestras. Esto es un 37,14% de las muestras, de un total de 36083 muestras

Año	**5**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	178	653	1043	2319	2626	1710	277	733	958	533	1242	1125
% muestras anómalas con respecto al total mensual de muestras	8,42	5,92	21,72	34,69	77,12	87,33	56,87	9,21	24,38	31,86	17,73	41,30

Tabla 32. Cuantificación muestras anómalas en los anillos. Año 5.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas se concentran en primavera.

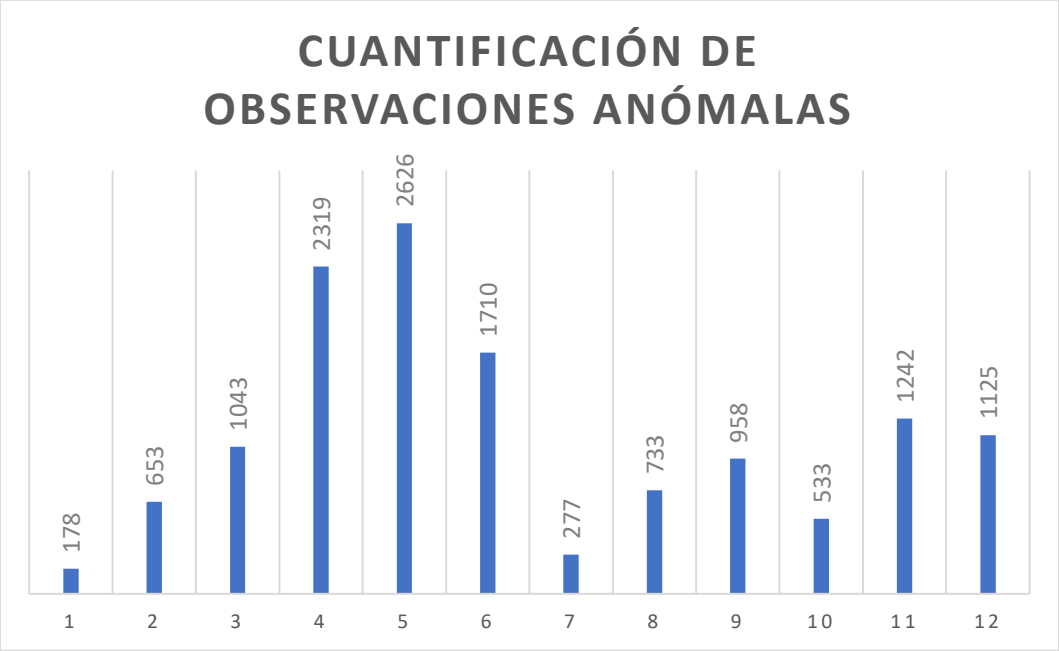


Ilustración 68. Cuantificación de observaciones anómalas en los anillos. Año 5.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o mas desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**5**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	66	177	153	201	141	150	93	198	201	192	213	195
% anomalías con respecto al total mensual de muestras	2,19	5,88	5,1	6,69	4,68	4,98	3,09	6,57	6,69	6,39	7,08	6,48

Tabla 33. Cuantificación anomalías en los anillos. Año 5.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías se distribuyen equilibradamente a lo largo del año. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

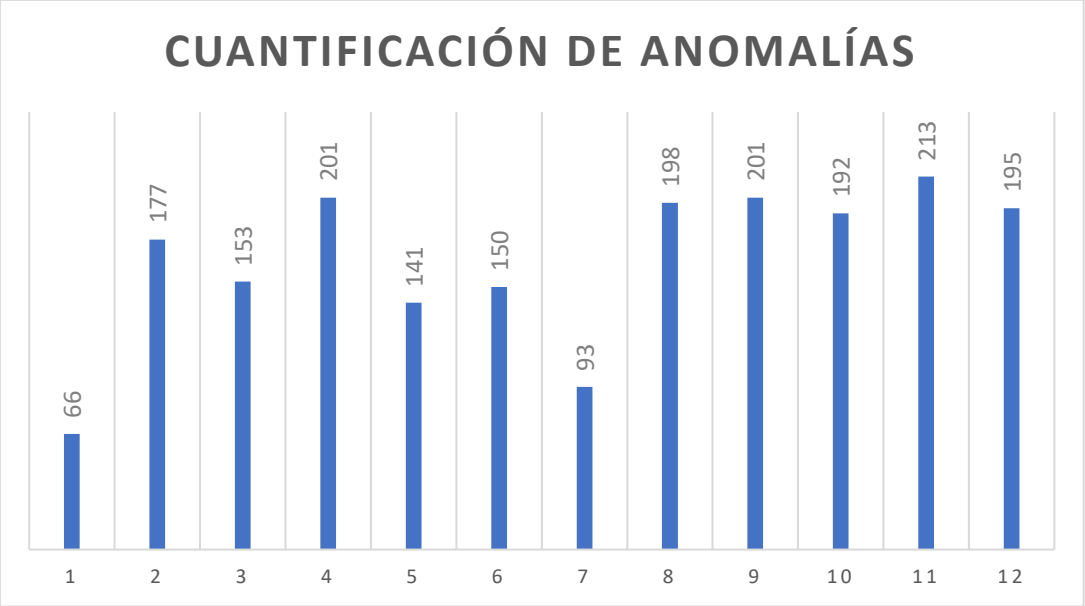


Ilustración 69. Cuantificación de anomalías en los anillos. Año 5.

Año 6

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

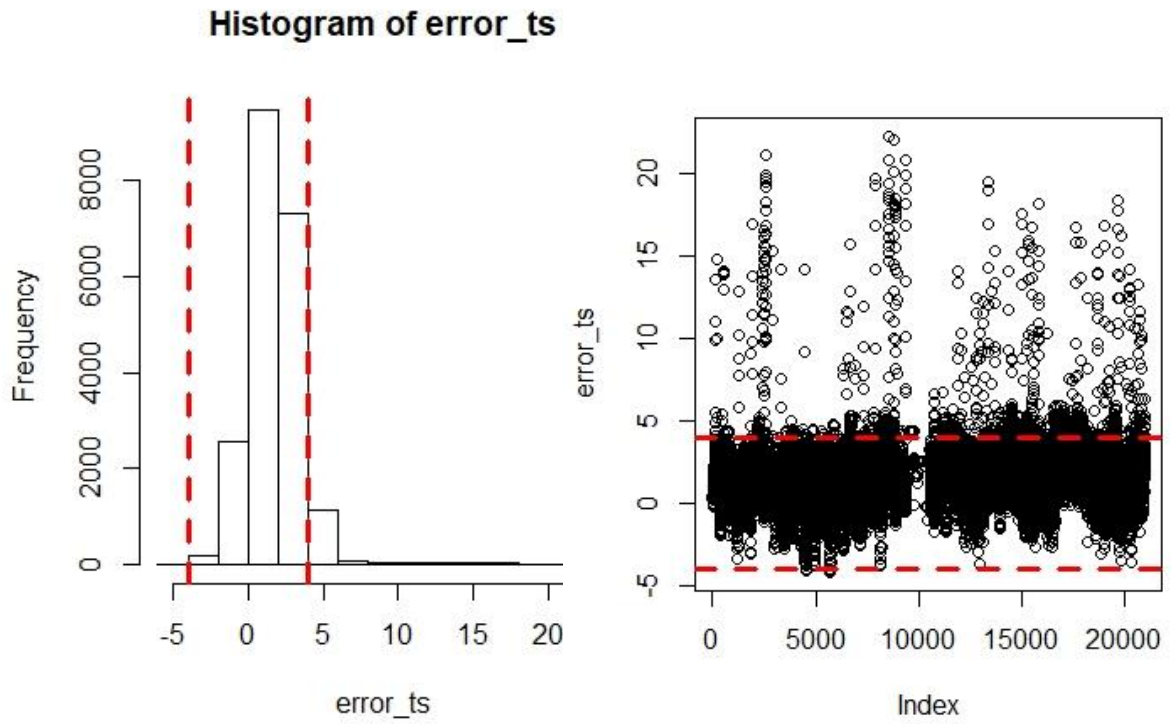


Ilustración 70. Histograma de residuos y errores de predicción en los anillos. Año 6.

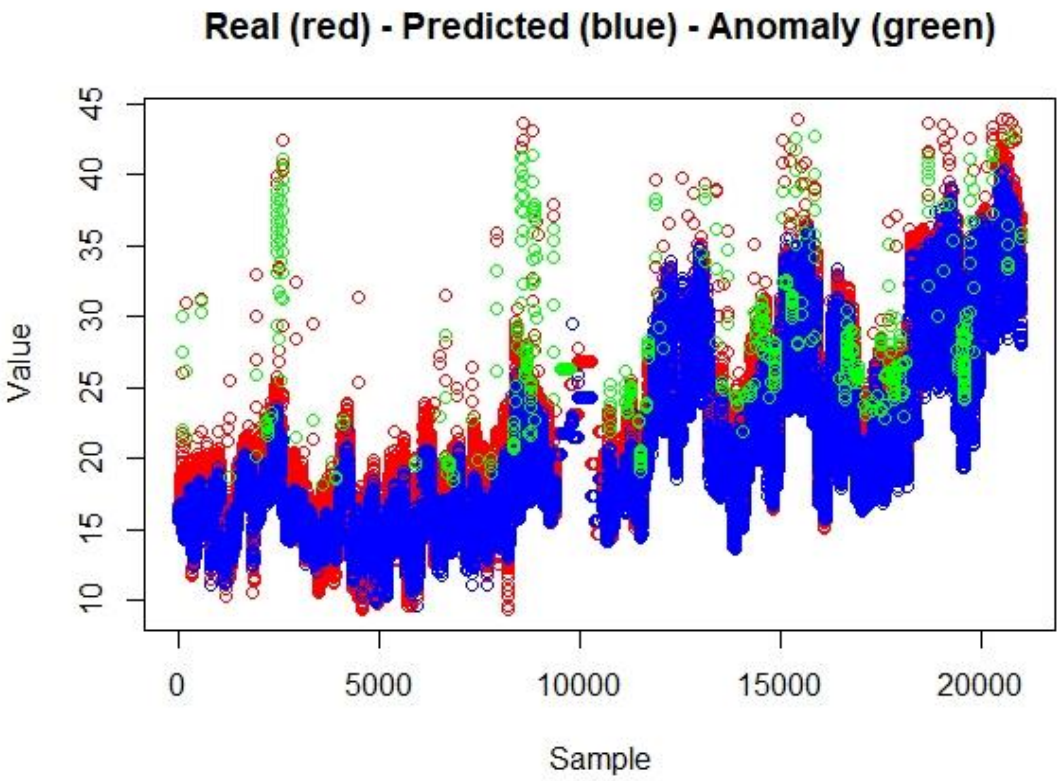


Ilustración 71. Predicciones, valores reales y anomalías en los anillos. Año 6.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-3{,}927 \leq X \leq 4{,}006]$, generado con los datos del primer año, encontramos 1466 muestras. Esto es un 6,98% de las muestras, de un total de 21006 muestras

Año	**6**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	30	86	27	48	130	303	104	78	209	154	150	147
% muestras anómalas con respecto al total mensual de muestras	8,42	1,71	4,91	1,54	2,74	7,43	17,31	5,94	4,46	11,94	8,80	8,57

Tabla 34. Cuantificación muestras anómalas en los anillos. Año 6.

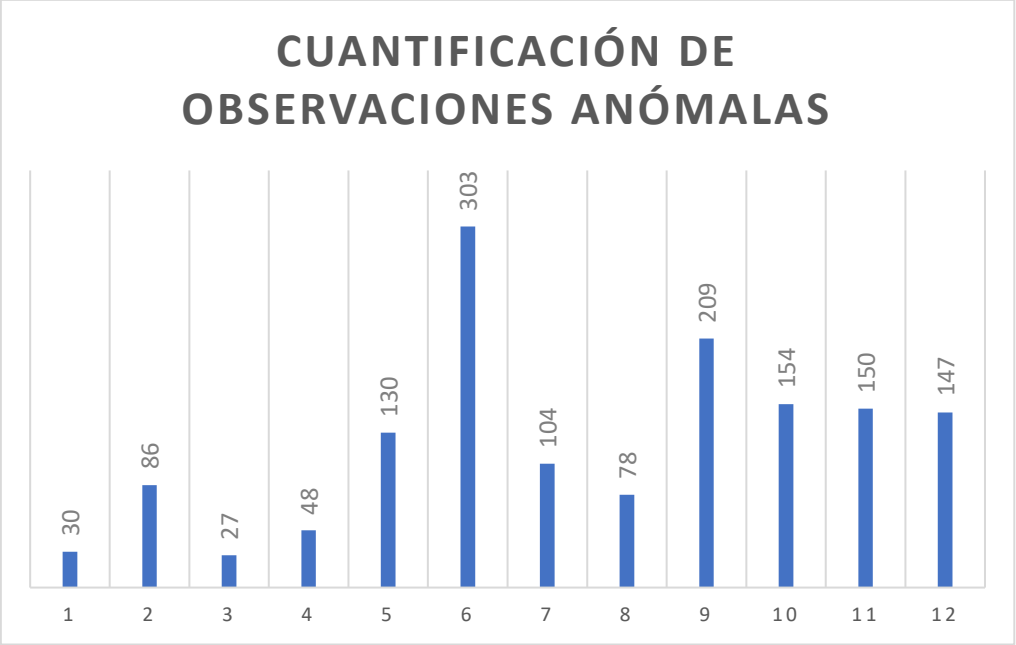


Ilustración 72. Cuantificación de observaciones anómalas en los anillos. Año 6.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas se concentran el último mes de junio y septiembre, donde hay dos picos.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o mas desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**6**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	12	39	12	21	39	21	42	30	93	60	72	54
% anomalías con respecto al total mensual de muestras	0,69	2,22	0,69	1,2	2,22	1,2	2,4	1,71	5,31	3,42	4,11	3,09

Tabla 35. Cuantificación anomalías en los anillos. Año 6.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías se concentran en los últimos meses del año. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

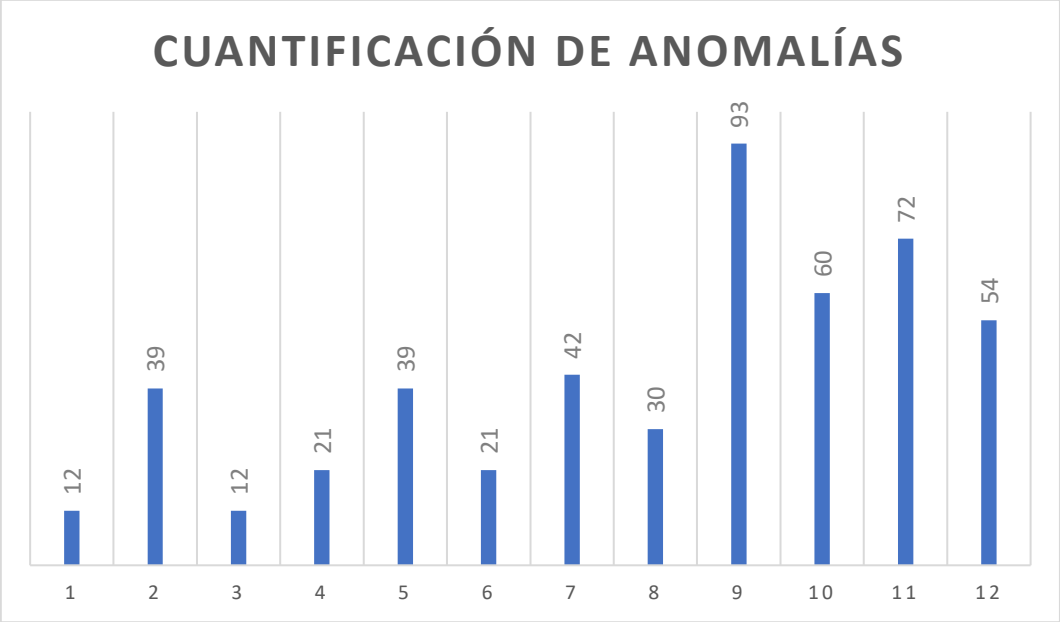


Ilustración 73. Cuantificación de anomalías en los anillos. Año 6.

6.4.- Modelo de previsión de la temperatura de la multiplicadora

A partir de los intervalos de confianza calculados para el primer año, establecemos el intervalo de confianza que se empleará para los años sucesivos.

$$[-2{,}914 \leq X \leq 2{,}974]\ U\ [-2{,}869 \leq X \leq 2{,}954] = [-2{,}914 \leq X \leq 2{,}974]$$

Recordamos que el intervalo de confianza del 95% para el modelo ha sido establecido con el intervalo de confianza del 95% del conjunto de entrenamiento unido al intervalo de intervalo de confianza del 95% del conjunto de test de los datos del primer año. Esto se ha hecho cogiendo la opción menos restringida, para evitar incluir errores aportados por el propio modelo.

Año 2

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

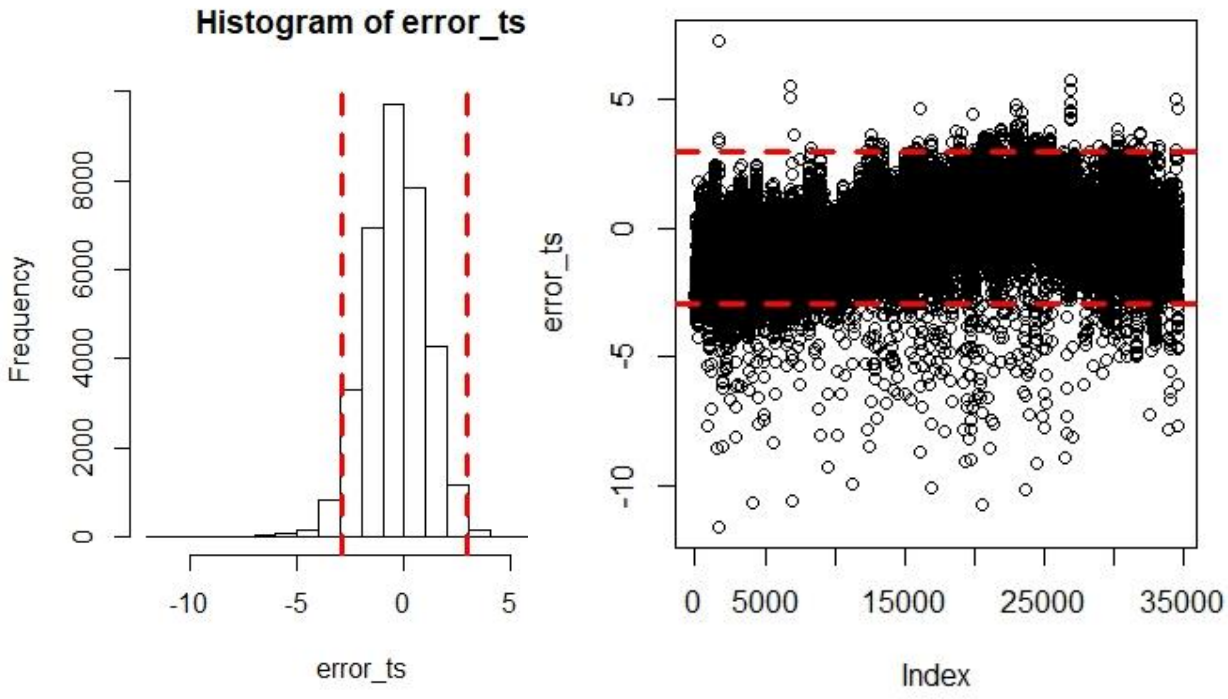


Ilustración 74. Histograma de residuos y errores de predicción en la multiplicadora. Año 2.

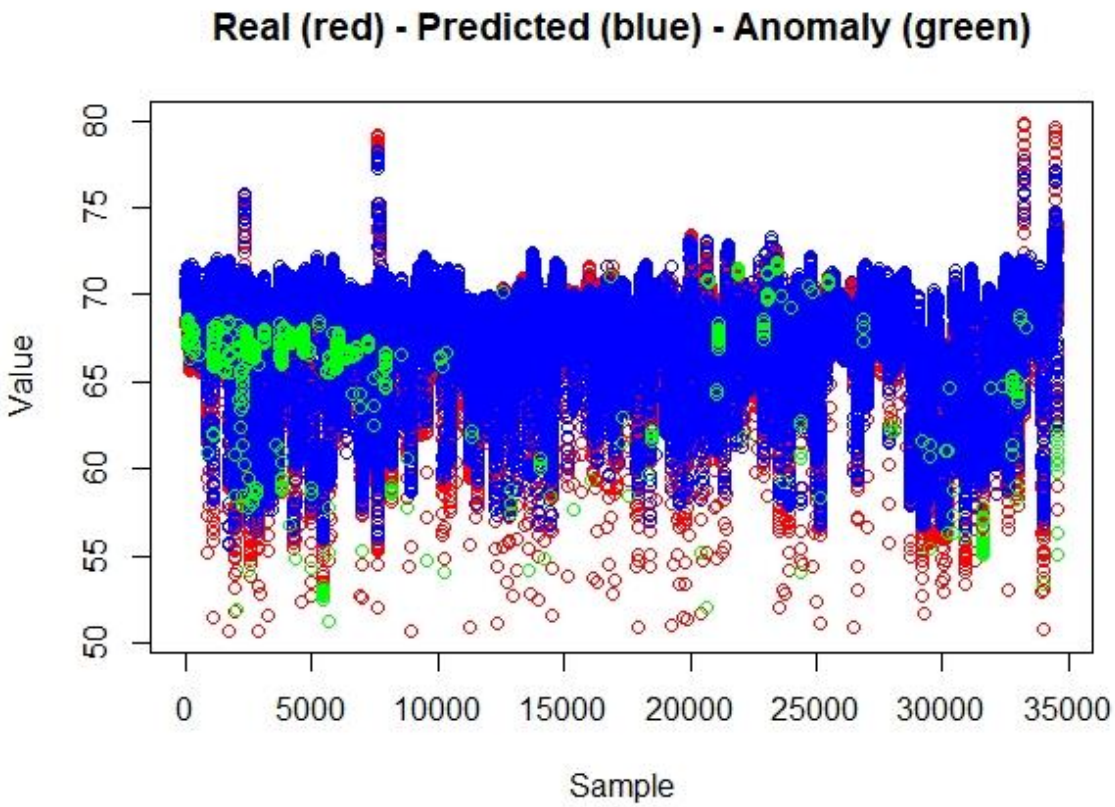


Ilustración 75. Predicciones, valores reales y anomalías en la multiplicadora. Año 2.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-2{,}914 \leq X \leq 2{,}974]$, generado con los datos del primer año, encontramos 1526 muestras. Esto es un 4,41% de las muestras, de un total de 34591 muestras.

Año	**2**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	380	262	199	62	53	61	63	107	106	35	105	93
% muestras anómalas con respecto al total mensual de muestras	13,18	9,09	6,90	2,15	1,84	2,12	2,19	3,71	3,68	1,21	3,64	3,23

Tabla 36. Cuantificación muestras anómalas en la multiplicadora. Año 2.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas se concentran en los primeros meses del año.

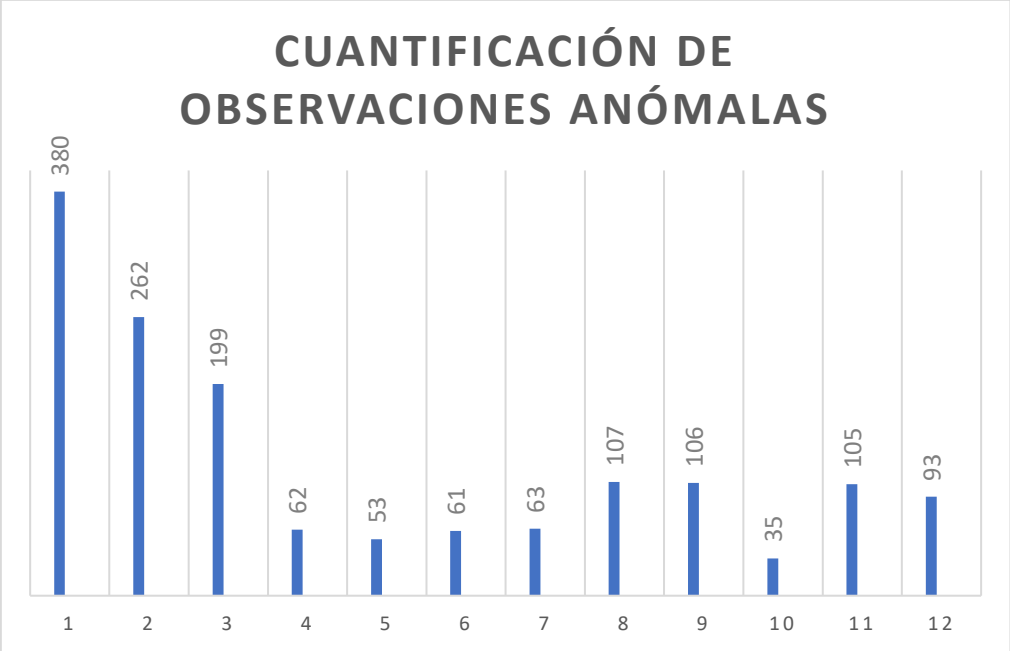


Ilustración 76. Cuantificación de observaciones anómalas en la multiplicadora. Año 2.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o mas desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**2**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	135	96	63	21	15	15	9	33	39	9	36	30
% anomalías con respecto al total mensual de muestras	4,68	3,33	2,19	0,72	0,51	0,51	0,3	1,14	1,35	0,3	1,26	1,05

Tabla 37. Cuantificación anomalías en la multiplicadora. Año 2.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías se concentran en el primer mes del año. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

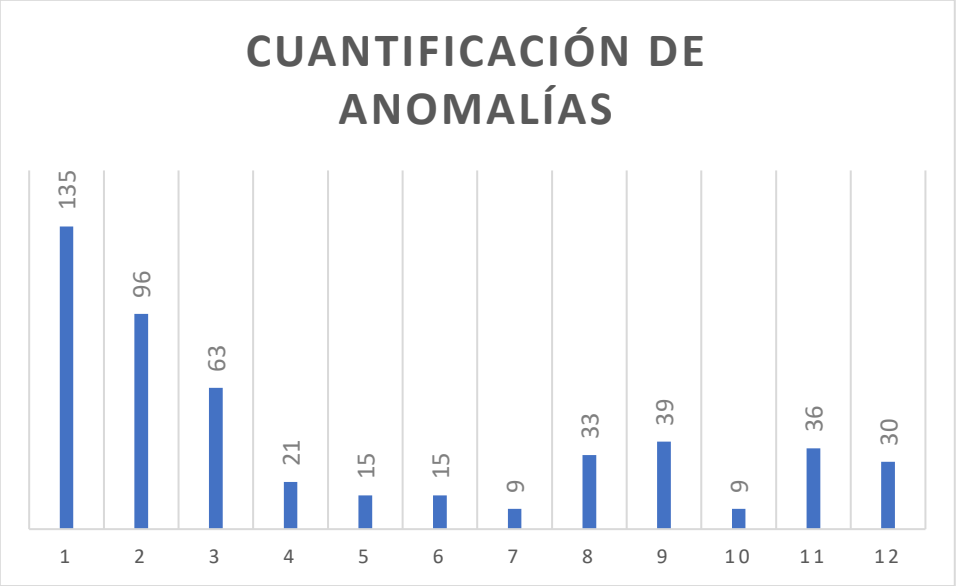


Ilustración 77. Cuantificación de anomalías en la multiplicadora. Año 2.

Año 3

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

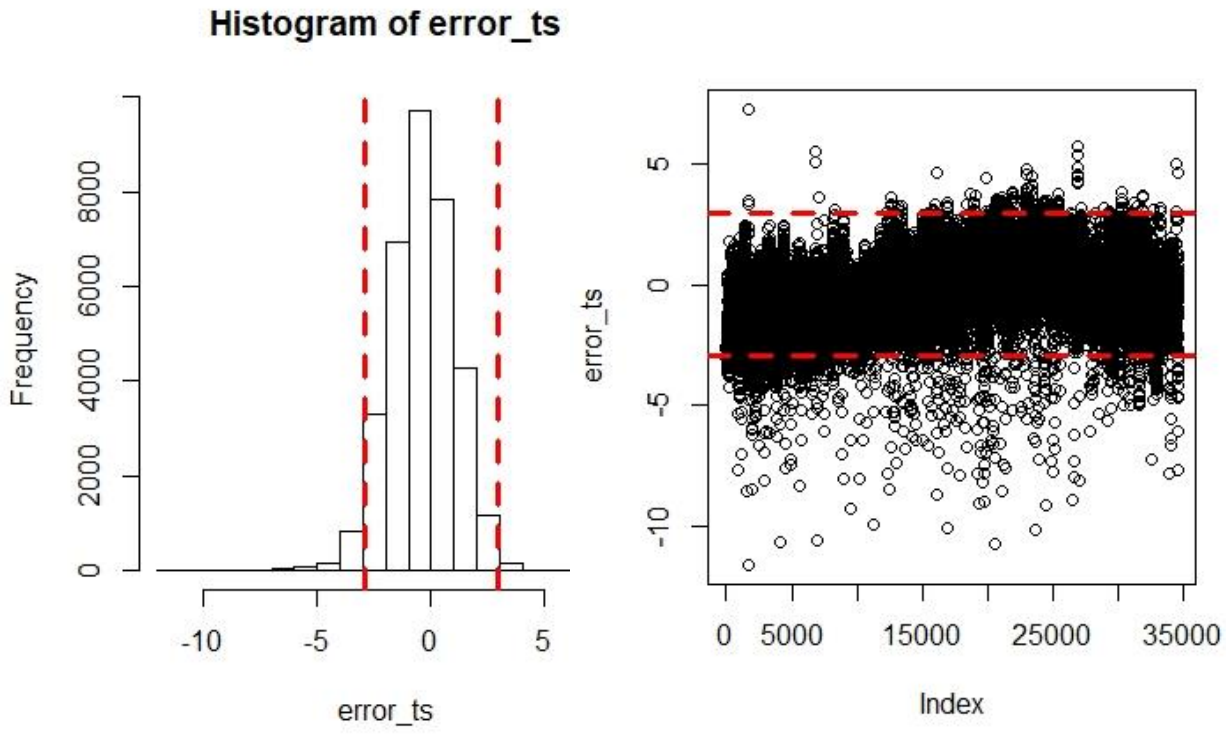


Ilustración 78. Histograma de residuos y errores de predicción en la multiplicadora. Año 3.

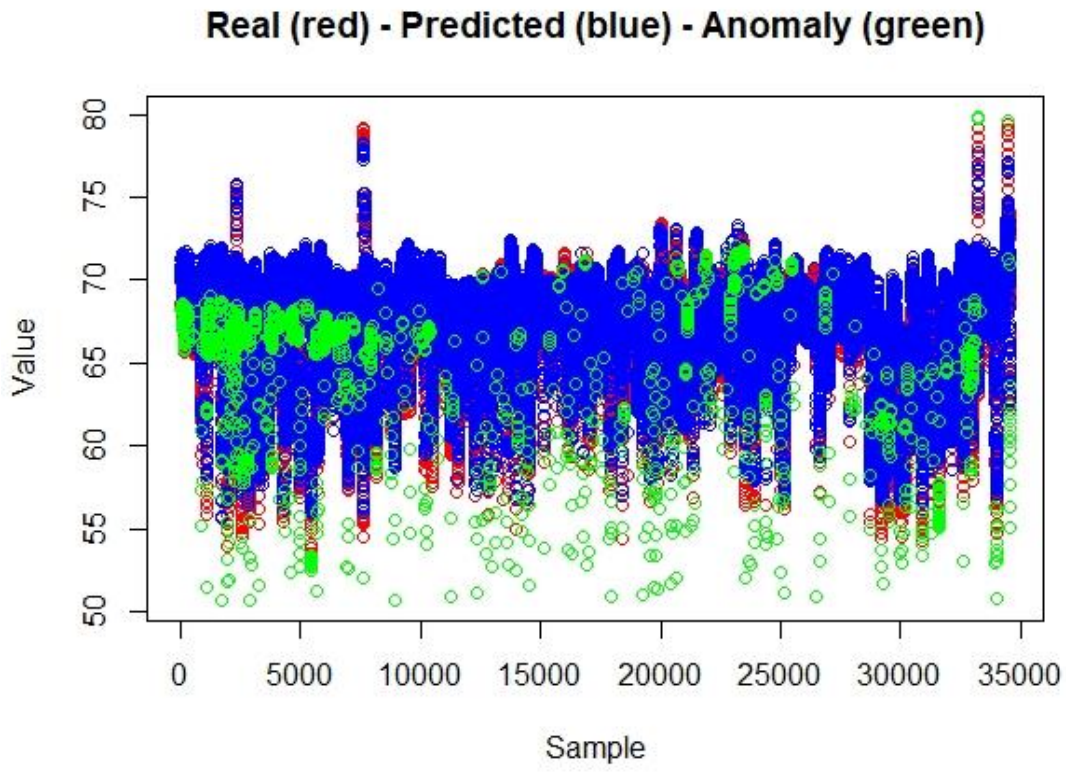


Ilustración 79. Predicciones, valores reales y anomalías en la multiplicadora. Año 3.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-2{,}914 \leq X \leq 2{,}974]$, generado con los datos del primer año, encontramos 2198 muestras. Esto es un 5,95% de las muestras, de un total de 36928 muestras.

Año	**3**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	65	105	56	63	79	156	366	266	516	287	106	133
% muestras anómalas con respecto al total mensual de muestras	2,11	3,41	1,82	2,05	2,57	5,07	11,89	8,64	16,77	9,33	3,44	4,32

Tabla 38. Cuantificación muestras anómalas en la multiplicadora. Año 3.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas se concentran en los meses de verano.

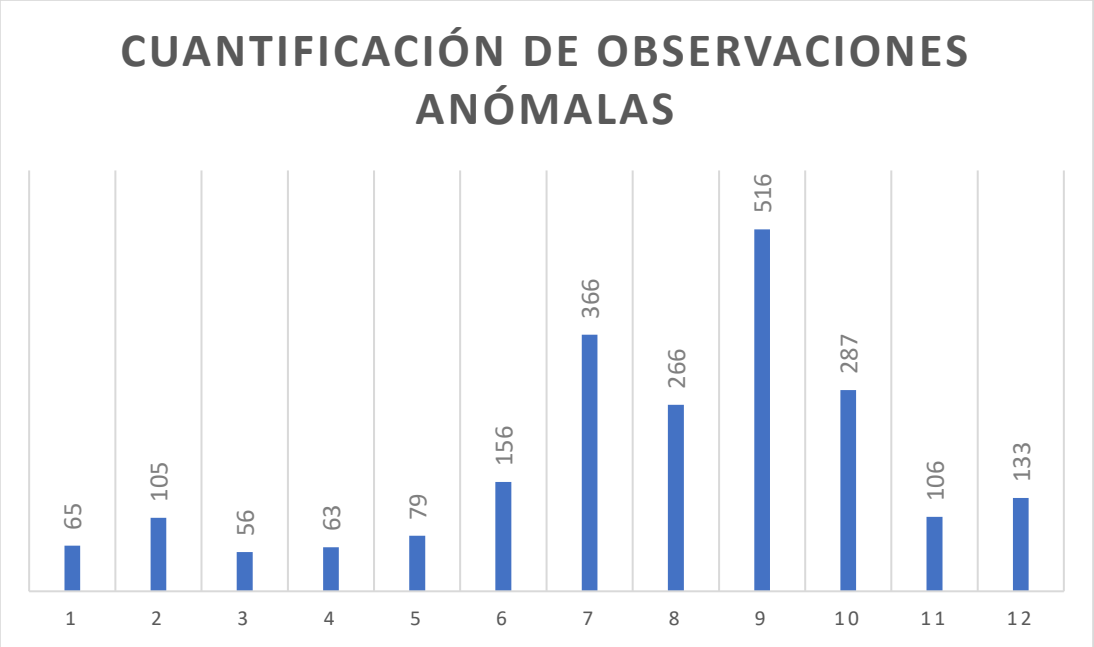


Ilustración 80. Cuantificación observaciones anómalas en la multiplicadora. Año 3.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o mas desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**3**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	21	39	24	15	18	48	120	78	132	81	33	63
% anomalías con respecto al total mensual de muestras	0,69	1,26	0,78	0,48	0,57	1,56	3,9	2,52	4,29	2,64	1,08	2,04

Tabla 39. Cuantificación anomalías en la multiplicadora. Año 3.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías se concentran en los meses de verano. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

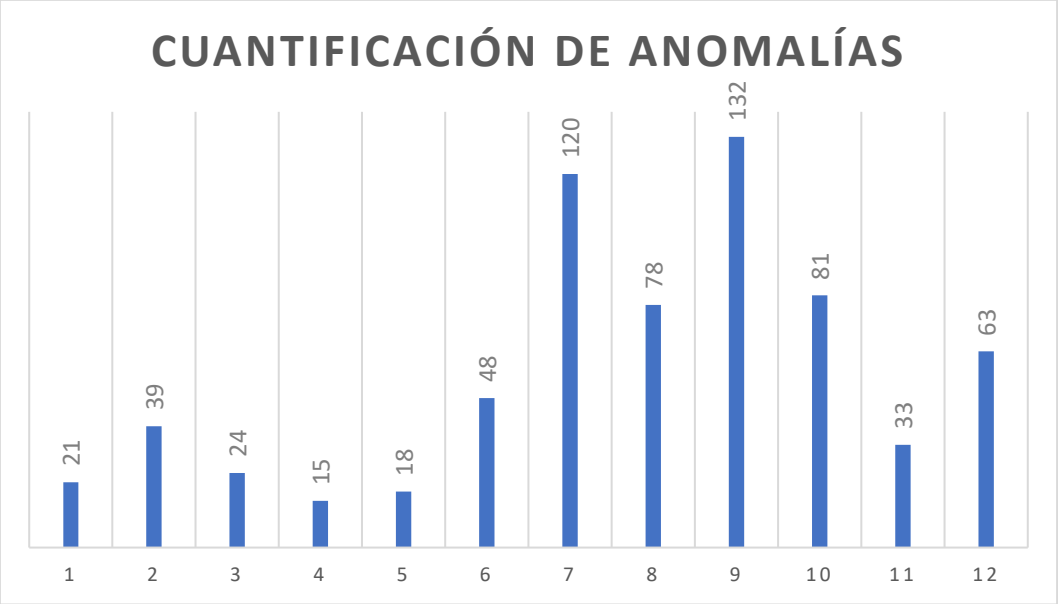


Ilustración 81. Cuantificación de anomalías en la multiplicadora. Año 3.

Año 4

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

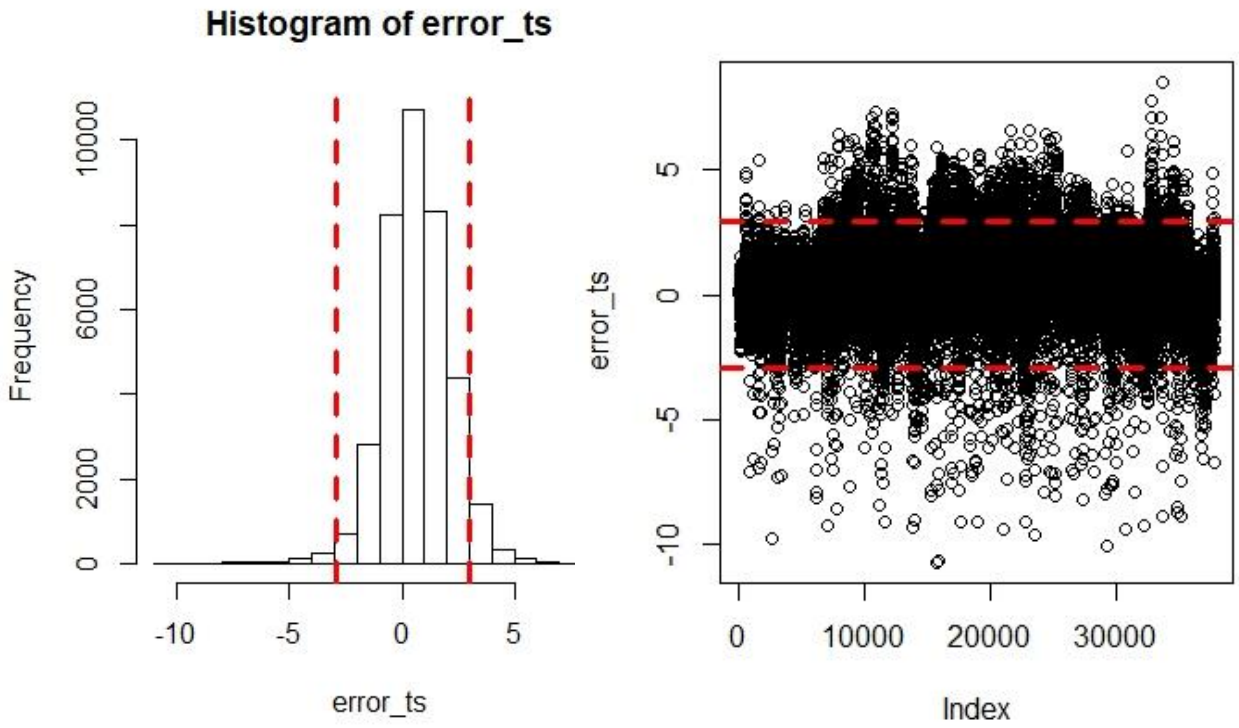


Ilustración 82. Histograma de residuos y errores de predicción en la multiplicadora. Año 4.

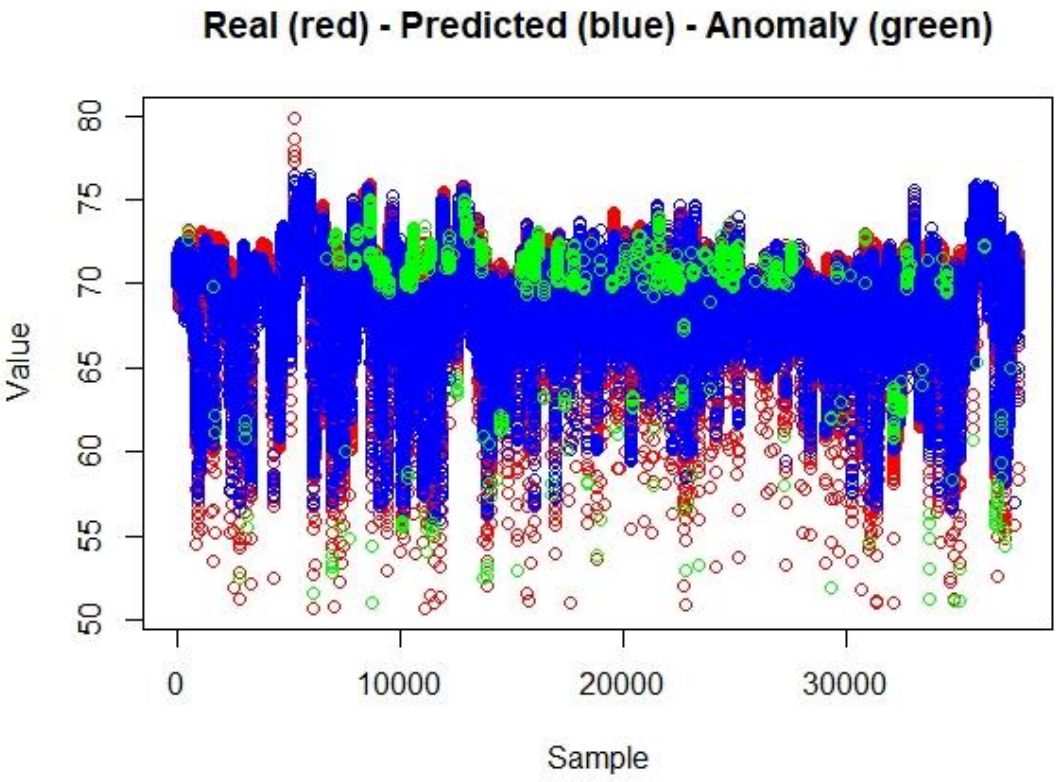


Ilustración 83. Predicciones, valores reales y anomalías en la multiplicadora. Año 4.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-2{,}914 \leq X \leq 2{,}974]$, generado con los datos del primer año, encontramos 2602 muestras. Esto es un 6,9% de las muestras, de un total de 37722 muestras.

Año	**4**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	48	25	272	281	262	361	345	465	186	92	167	97
% muestras anómalas con respecto al total mensual de muestras	1,53	0,80	8,65	8,94	8,33	11,48	10,98	14,79	5,92	2,93	5,31	3,09

Tabla 40. Cuantificación muestras anómalas en la multiplicadora. Año 4.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas se concentran en los meses centrales del año, especialmente en los de verano.

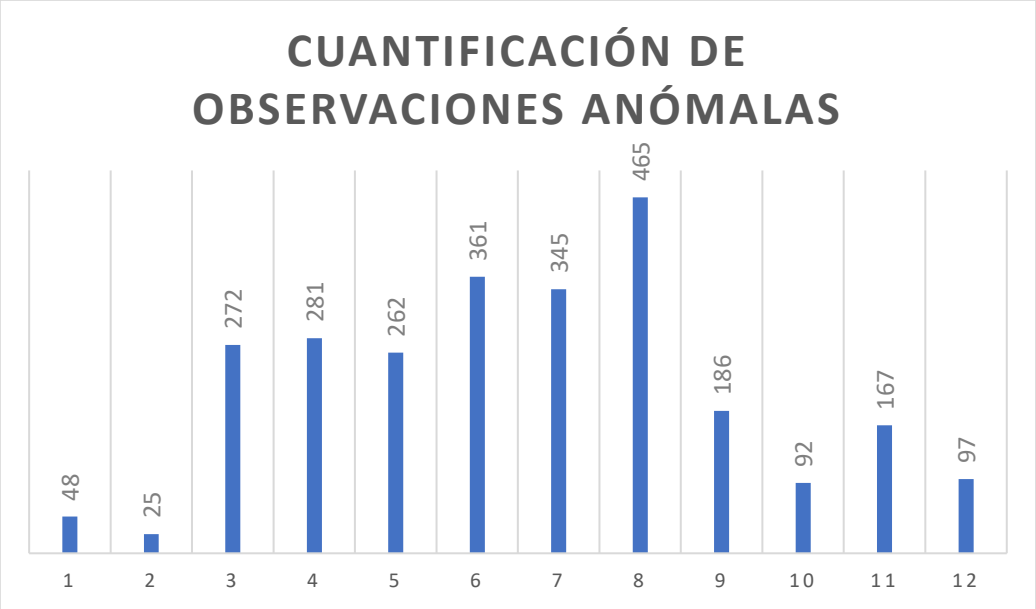


Ilustración 84. Cuantificación observaciones anómalas en la multiplicadora. Año 4.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o mas desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**4**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	21	3	90	90	84	108	123	159	51	30	57	42
% anomalías con respecto al total mensual de muestras	0,66	0,09	2,85	2,85	2,67	3,45	3,9	5,07	1,62	0,96	1,8	1,35

Tabla 41. Cuantificación anomalías en la multiplicadora. Año 4.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías se concentran en los meses centrales del año. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

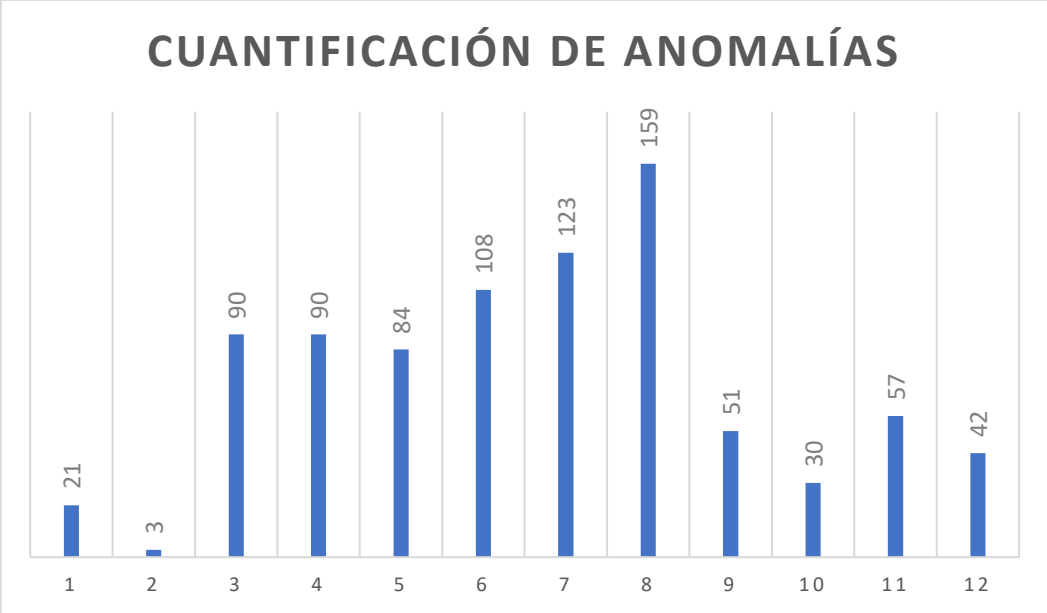


Ilustración 85. Cuantificación de anomalías en la multiplicadora. Año 4.

Año 5

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

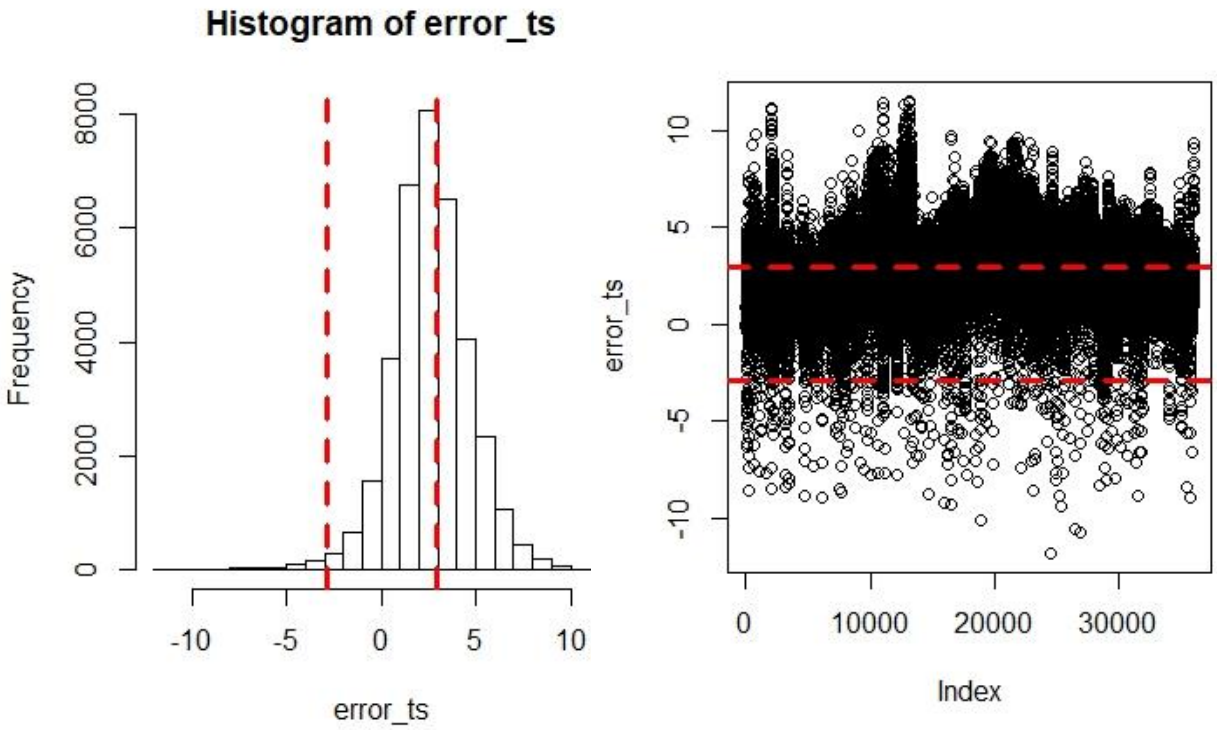


Ilustración 86. Histograma de residuos y errores de predicción en la multiplicadora. Año 5.

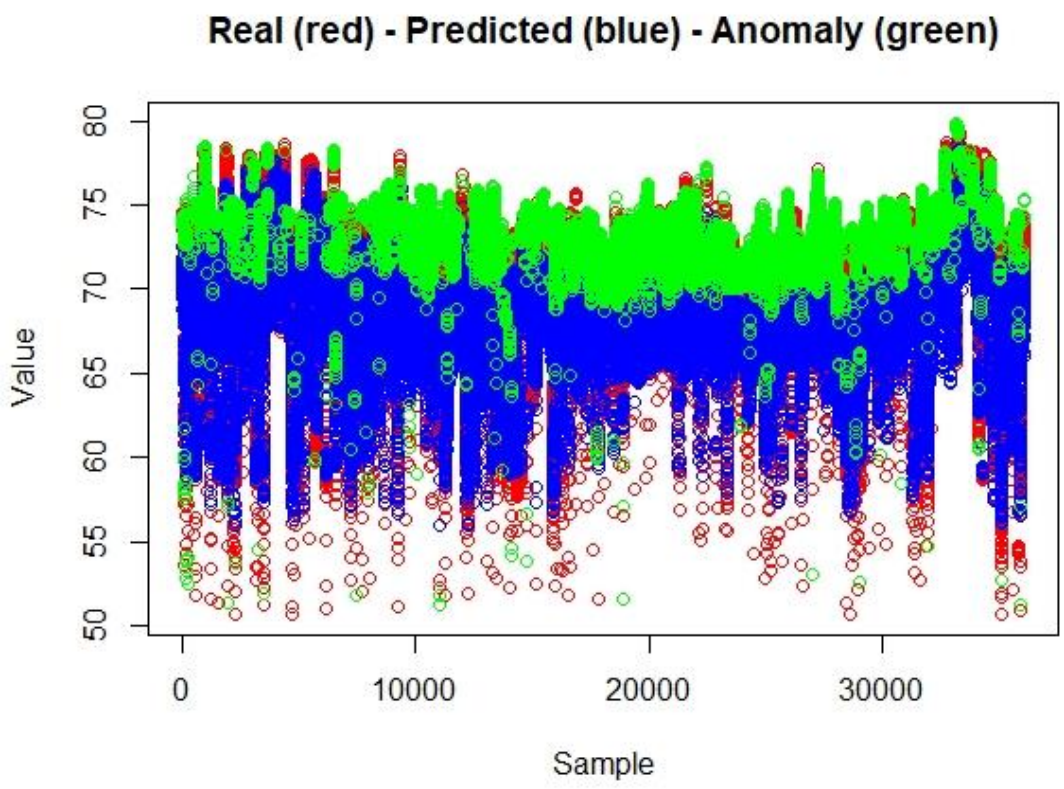


Ilustración 87. Predicciones, valores reales y anomalías en la multiplicadora. Año 5.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-2{,}914 \leq X \leq 2{,}974]$, generado con los datos del primer año, encontramos 15257 muestras. Esto es un 42,28% de las muestras, de un total de 36083 muestras.

Año	**5**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	856	407	811	1318	1328	1382	2361	1934	1647	1486	1009	713
% muestras anómalas con respecto al total mensual de muestras	28,47	13,54	26,97	43,83	44,16	45,96	78,52	64,32	54,77	49,42	33,56	23,71

Tabla 42. Cuantificación muestras anómalas en la multiplicadora. Año 5.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas son abundantes durante todo el año, pero van ascendiendo hasta verano. Luego empiezan a disminuir.

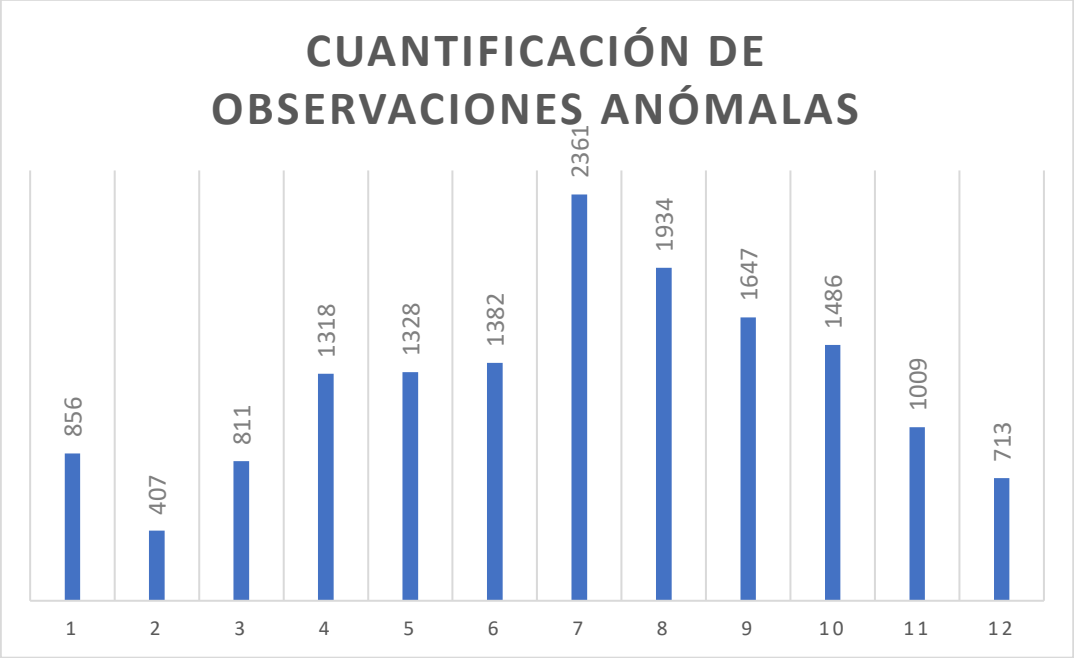


Ilustración 88. Cuantificación observaciones anómalas en la multiplicadora. Año 5.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o más desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**2**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	177	105	207	207	150	240	252	294	249	231	201	150
% anomalías con respecto al total mensual de muestras	5,88	3,48	6,87	6,87	4,98	7,98	8,37	9,78	8,28	7,68	6,69	4,98

Tabla 43. Cuantificación anomalías en la multiplicadora. Año 5.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías son abundantes durante todo el año, pero van ascendiendo hasta verano. Luego empiezan a disminuir. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

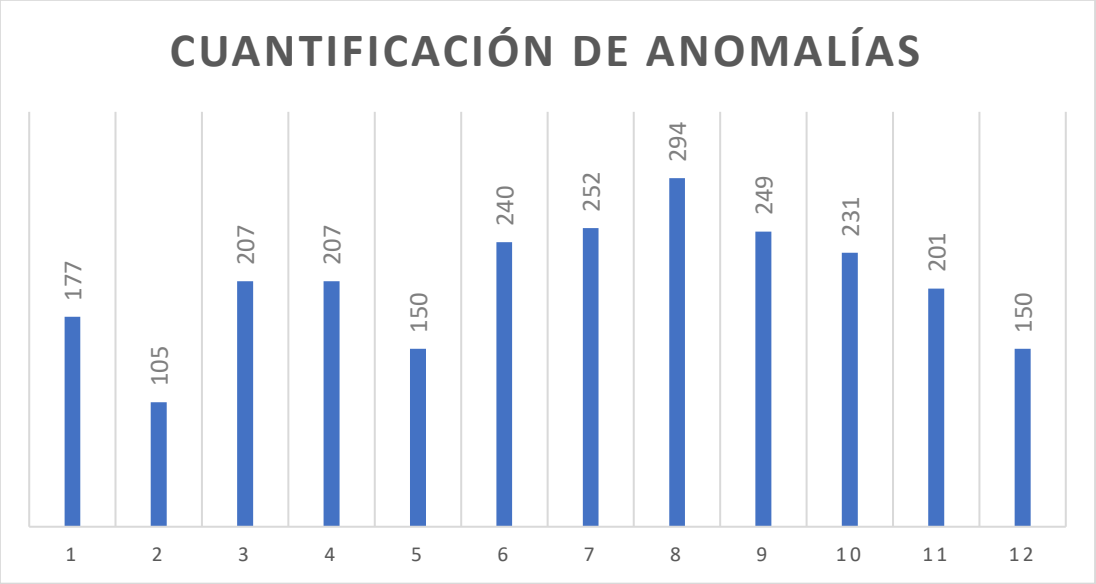


Ilustración 89. Cuantificación de anomalías en la multiplicadora. Año 5.

Año 6

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

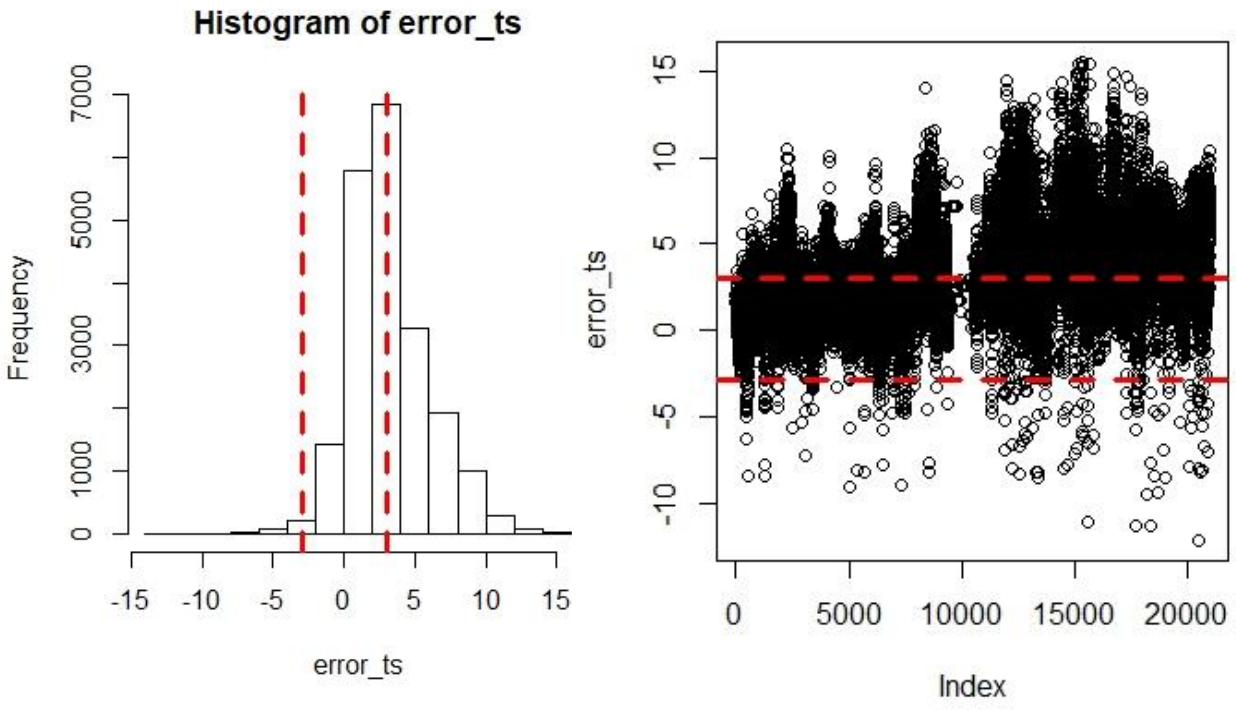


Ilustración 90. Histograma de residuos y errores de predicción en la multiplicadora. Año 6.

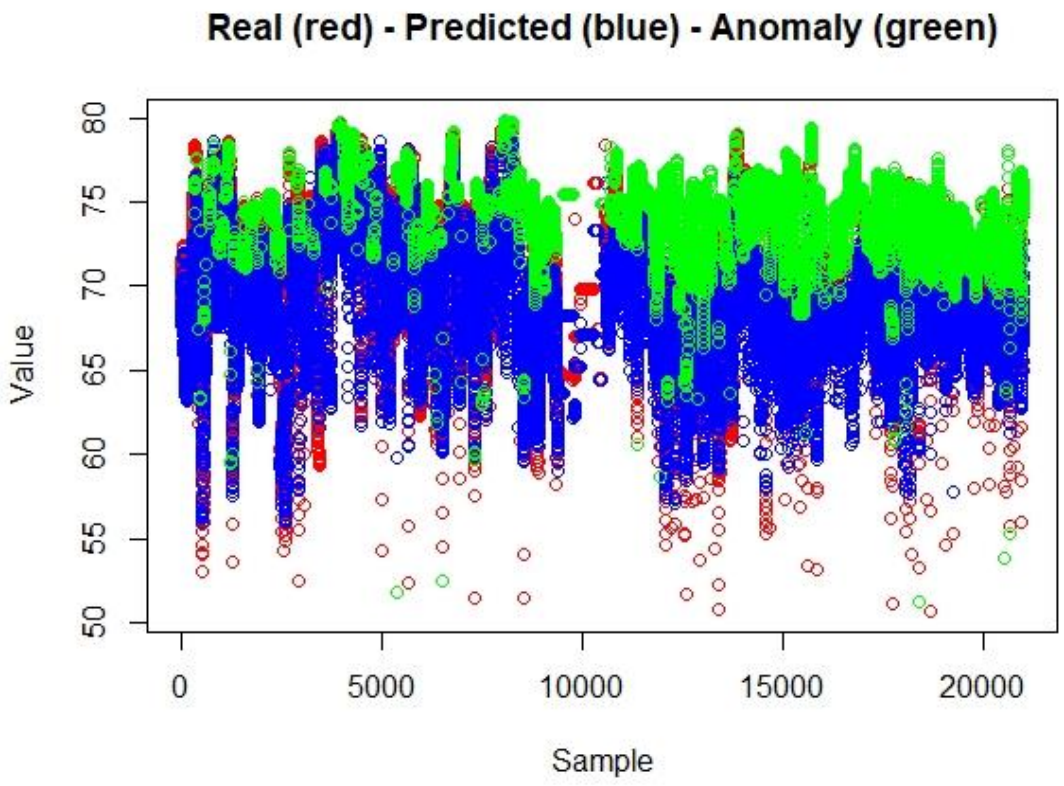


Ilustración 91. Predicciones, valores reales y anomalías en la multiplicadora. Año 6.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-2{,}914 \leq X \leq 2{,}974]$, generado con los datos del primer año, encontramos 10000 muestras. Esto es un 47,61% de las muestras, de un total de 21006 muestras.

Año	**6**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	383	456	484	457	607	536	1104	1038	1366	973	1338	1254
% muestras anómalas con respecto al total mensual de muestras	21,88	26,05	27,65	26,11	34,68	30,62	63,07	59,30	78,03	55,58	76,44	71,64

Tabla 44. Cuantificación muestras anómalas en la multiplicadora. Año 6.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas son abundantes durante todo el año, y van en ascenso a lo largo del mismo.

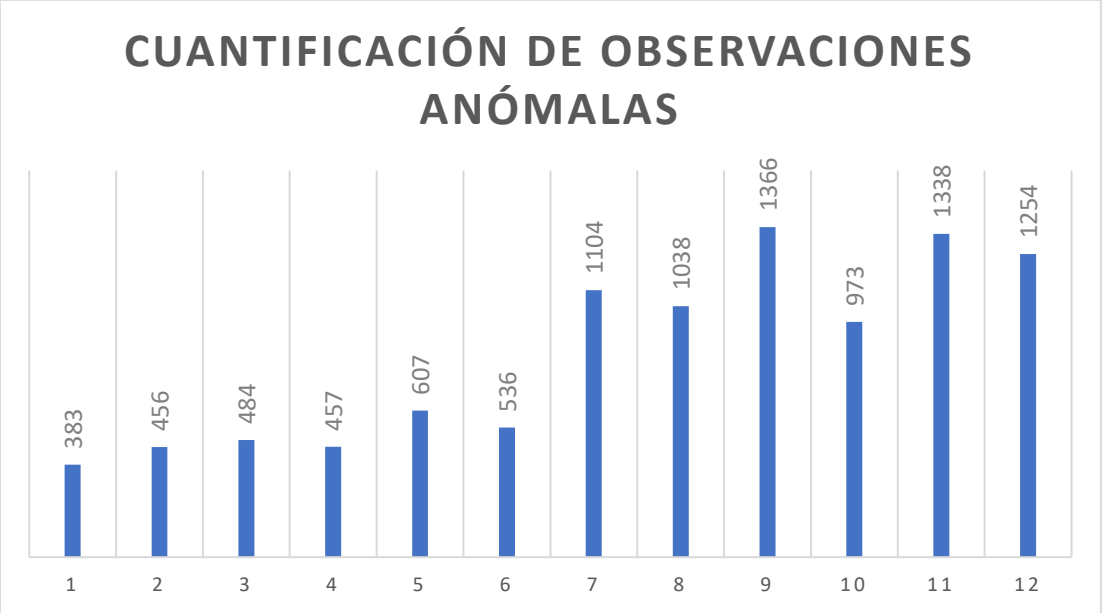


Ilustración 92. Cuantificación observaciones anómalas en la multiplicadora. Año 6.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o mas desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**6**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	132	87	102	111	99	48	123	120	72	96	168	150
% anomalías con respecto al total mensual de muestras	7,53	4,98	5,82	6,33	5,67	2,73	7,02	6,87	4,11	5,49	9,6	8,58

Tabla 45. Cuantificación anomalías en la multiplicadora. Año 6.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías son abundantes durante todo el año. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

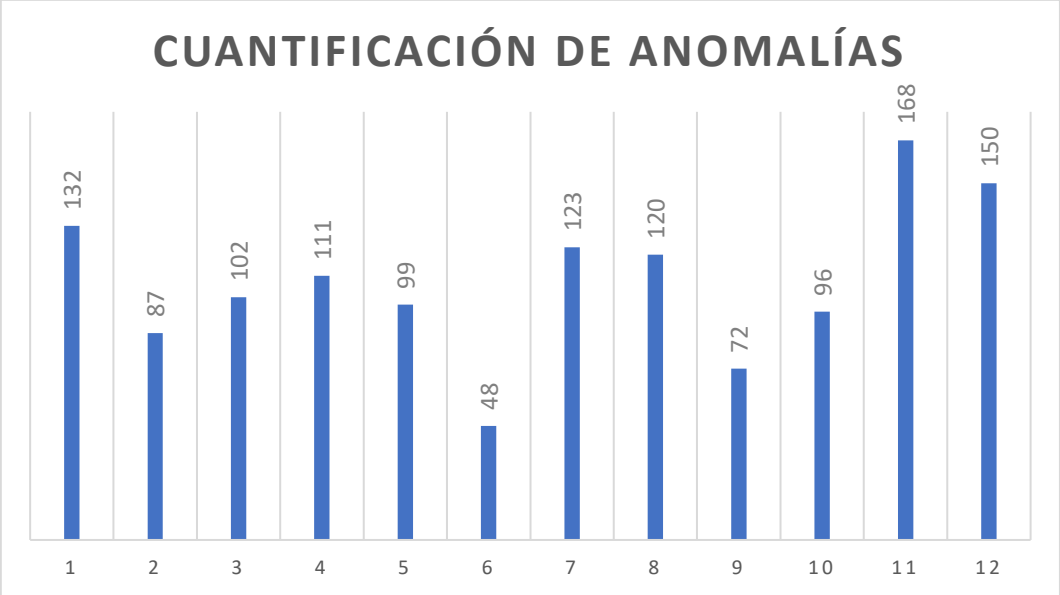


Ilustración 93. Cuantificación de anomalías en la multiplicadora. Año 6.

6.5.- Modelo de previsión de la temperatura del rodamiento DE

A partir de los intervalos de confianza calculados para el primer año, establecemos el intervalo de confianza que se empleará para los años sucesivos.

$$[-3{,}257 \leq X \leq 2{,}638]\ U\ [-3{,}359 \leq X \leq 2{,}689] = [-3{,}359 \leq X \leq 2{,}689]$$

Recordamos que el intervalo de confianza del 95% para el modelo ha sido establecido con el intervalo de confianza del 95% del conjunto de entrenamiento unido al intervalo de intervalo de confianza del 95% del conjunto de test de los datos del primer año. Esto se ha hecho cogiendo la opción menos restringida, para evitar incluir errores aportados por el propio modelo.

Año 2

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

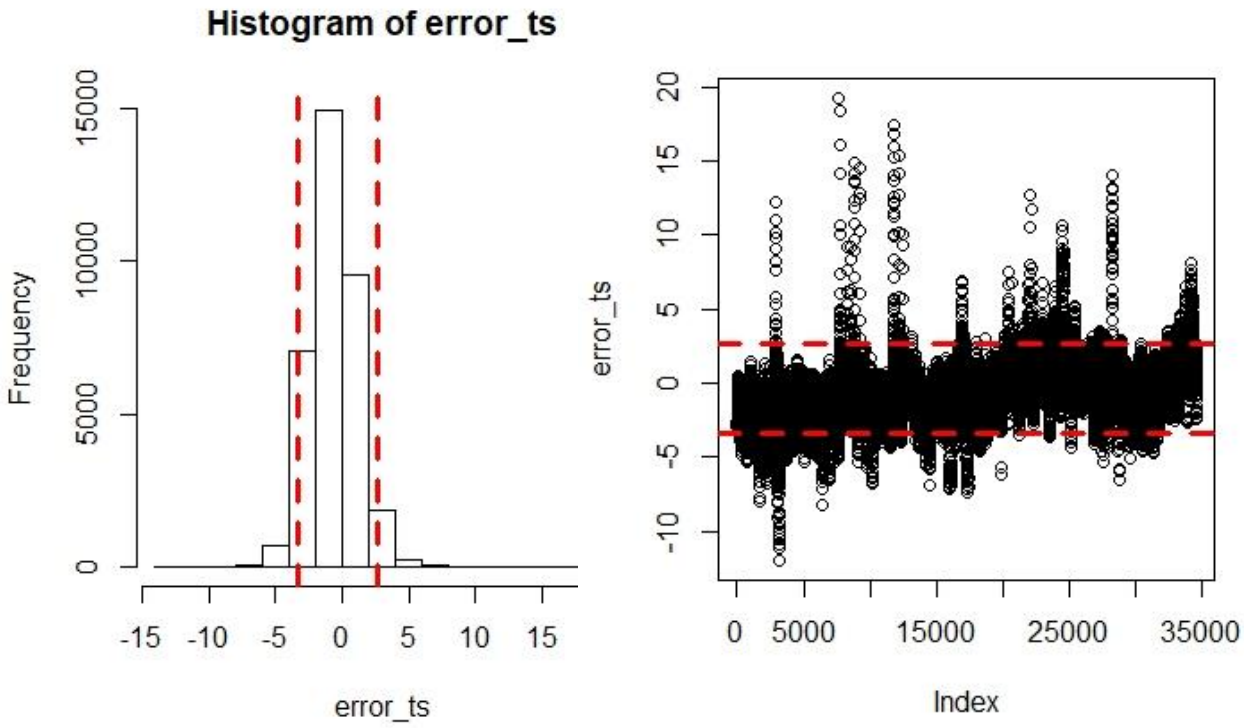


Ilustración 94. Histograma de residuos y errores de predicción en el rodamiento DE. Año 2.

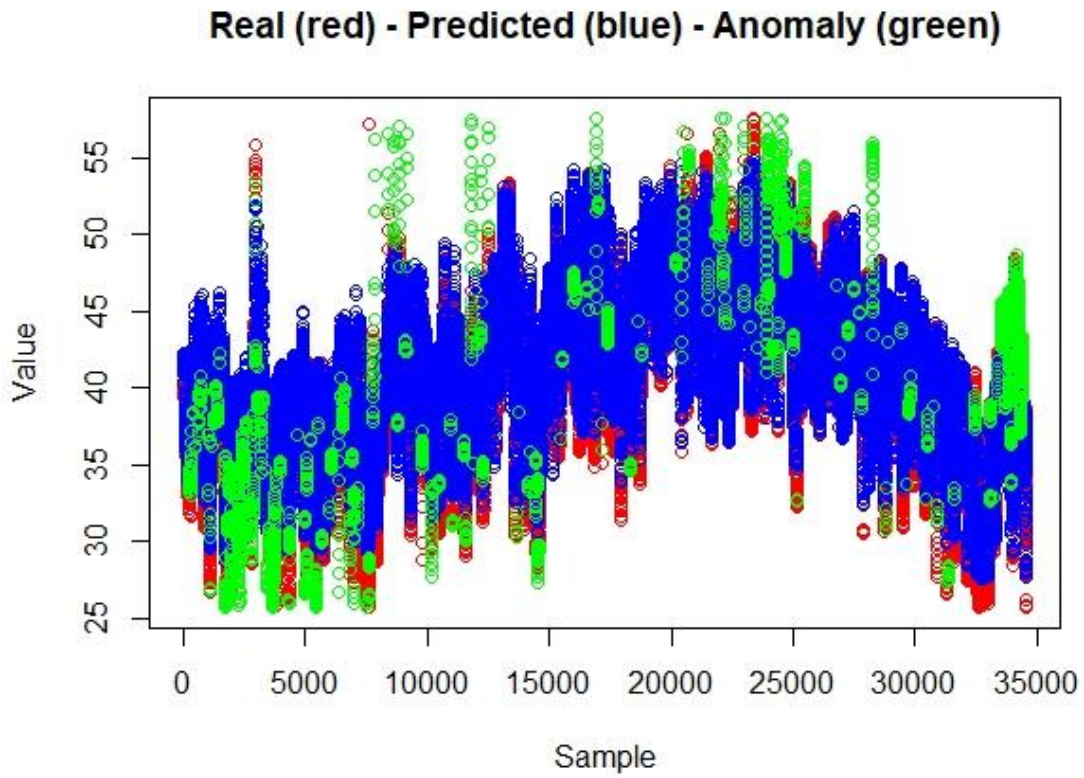


Ilustración 95. Predicciones, valores reales y anomalías en el rodamiento DE. Año 2.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-3{,}359 \leq X \leq 2{,}689]$, generado con los datos del primer año, encontramos 3040 muestras. Esto es un 8,79% de las muestras, de un total de 34592 muestras

Año	**2**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	656	437	276	124	126	161	80	175	267	101	148	489
% muestras anómalas con respecto al total mensual de muestras	22,76	15,16	9,57	4,30	4,37	5,59	2,78	6,07	9,26	3,50	5,13	16,96

Tabla 46. Cuantificación muestras anómalas en el rodamiento DE. Año 2.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas se concentran en los meses de invierno.

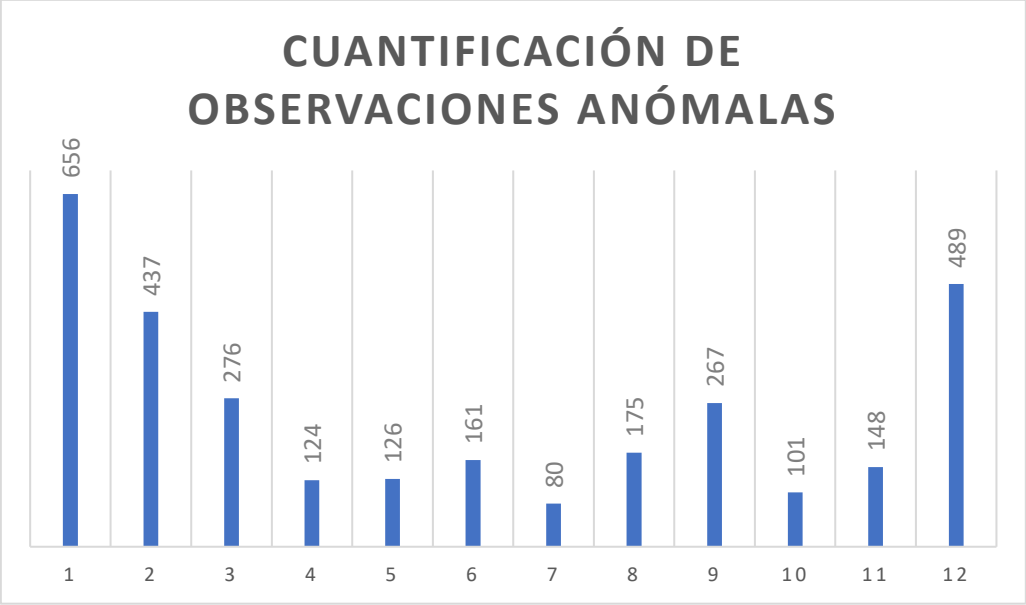


Ilustración 96. Cuantificación observaciones anómalas en el rodamiento DE. Año 2.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o mas desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**2**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	174	93	69	39	39	45	12	48	54	27	45	72
% anomalías con respecto al total mensual de muestras	6,03	3,24	2,4	1,35	1,35	1,56	0,42	1,68	1,86	0,93	1,56	2,49

Tabla 47. Cuantificación anomalías en el rodamiento DE. Año 2.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías se concentran en los primeros meses del año. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

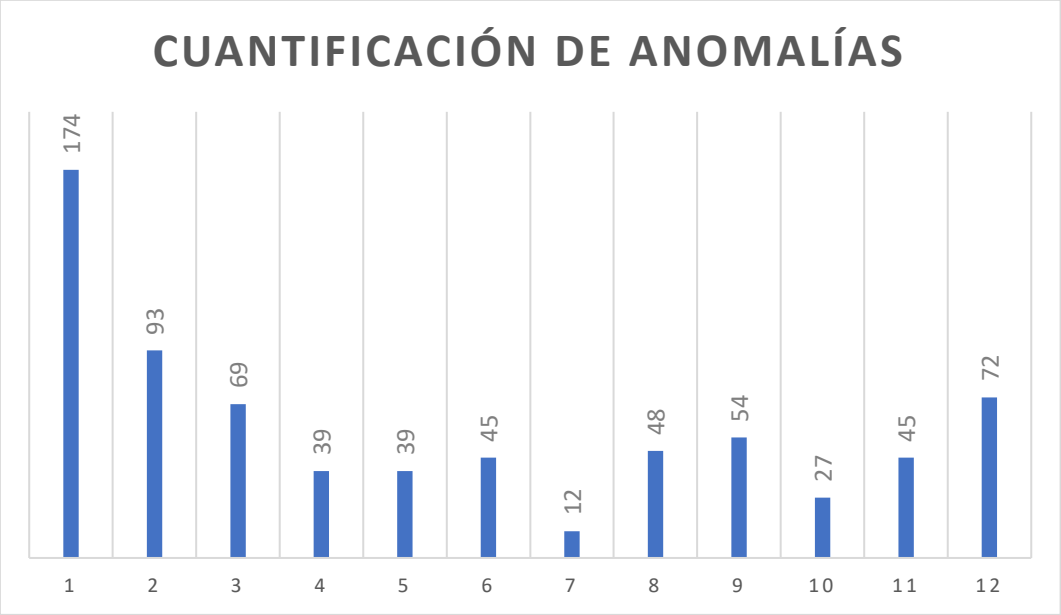


Ilustración 97. Cuantificación de anomalías en el rodamiento DE. Año 2.

Año 3

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

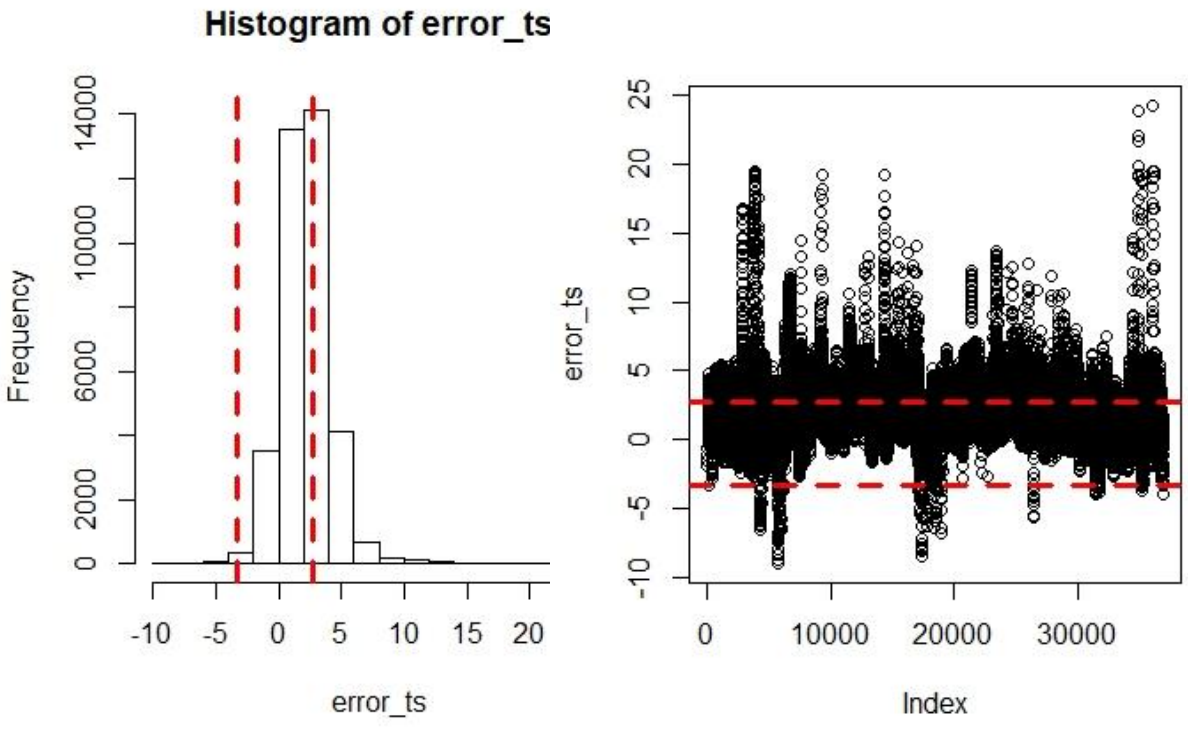


Ilustración 98. Histograma de residuos y errores de predicción en el rodamiento DE. Año 3.

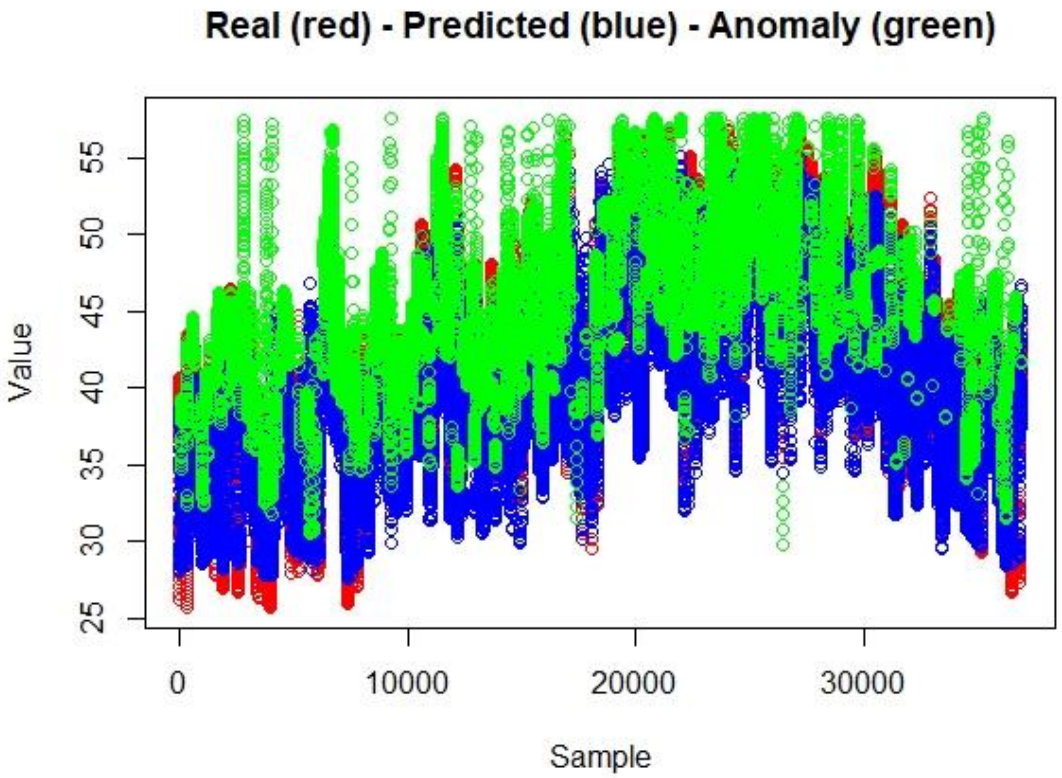


Ilustración 99. Predicciones, valores reales y anomalías en el rodamiento DE. Año 3.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-3{,}359 \leq X \leq 2{,}689]$, generado con los datos del primer año, encontramos 13645 muestras. Esto es un 36,95% de las muestras, de un total de 36929 muestras.

Año	**3**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	994	998	1643	1503	1362	1053	1217	1542	1481	772	324	755
% muestras anómalas con respecto al total mensual de muestras	32,30	32,43	53,39	48,84	44,26	34,22	39,55	50,11	48,12	25,09	10,53	24,53

Tabla 48. Cuantificación muestras anómalas en el rodamiento DE. Año 3.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas se distribuyen equitativamente a lo largo del año, aunque son abundantes durante todo el año.

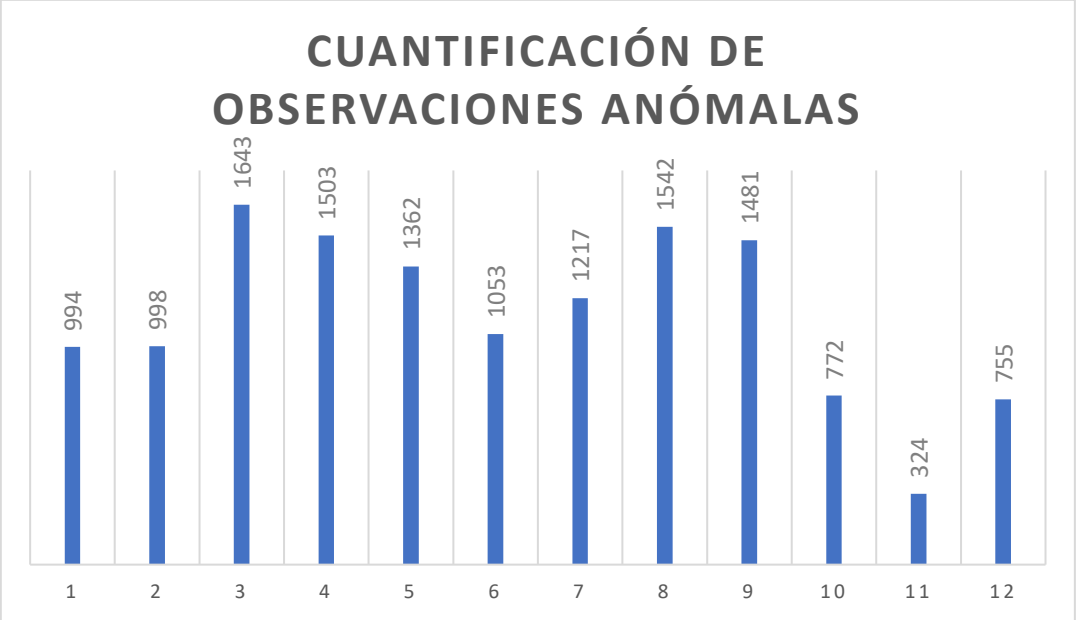


Ilustración 100. Cuantificación observaciones anómalas en el rodamiento DE. Año 3.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o mas desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**3**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	180	123	156	180	156	123	207	195	165	135	90	114
% anomalías con respecto al total mensual de muestras	5,85	3,99	5,07	5,85	5,07	3,99	6,72	6,33	5,37	4,38	2,91	3,69

Tabla 49. Cuantificación anomalías en el rodamiento DE. Año 3.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías son abundantes y se distribuyen equitativamente a lo largo del año. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

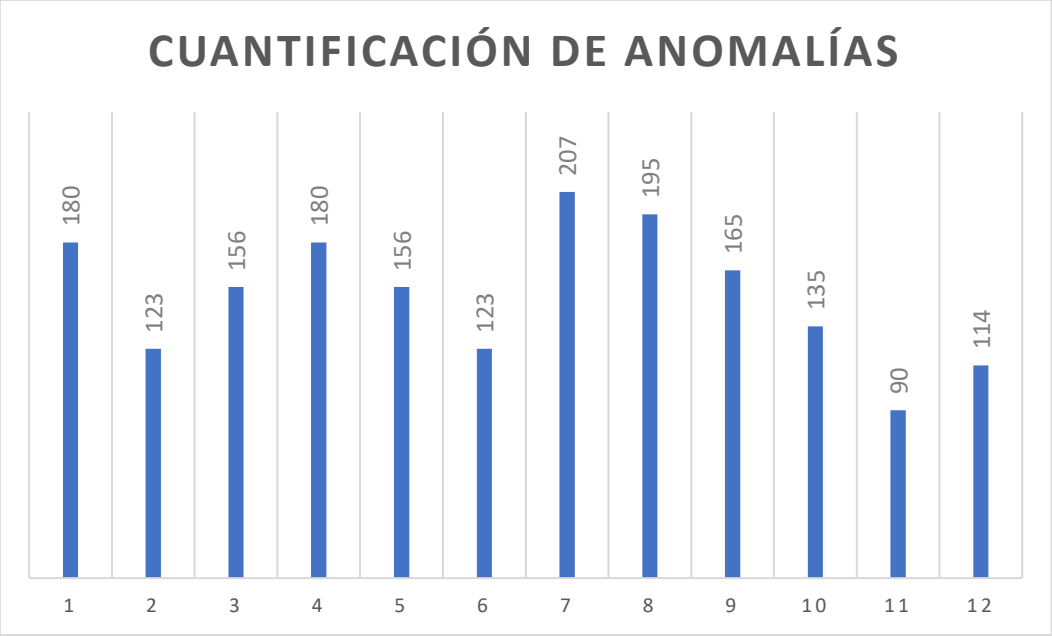


Ilustración 101. Cuantificación de anomalías en el rodamiento DE. Año 3.

Año 4

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

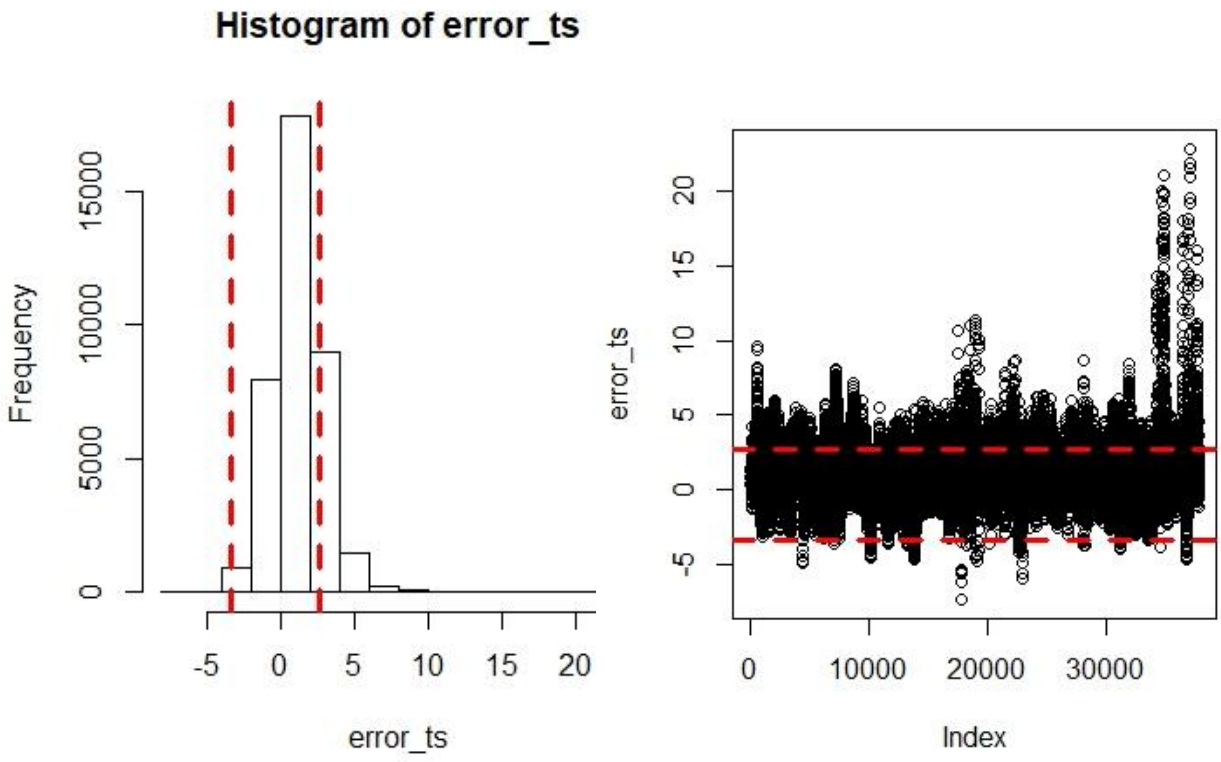


Ilustración 102. Histograma de residuos y errores de predicción en el rodamiento DE. Año 4.

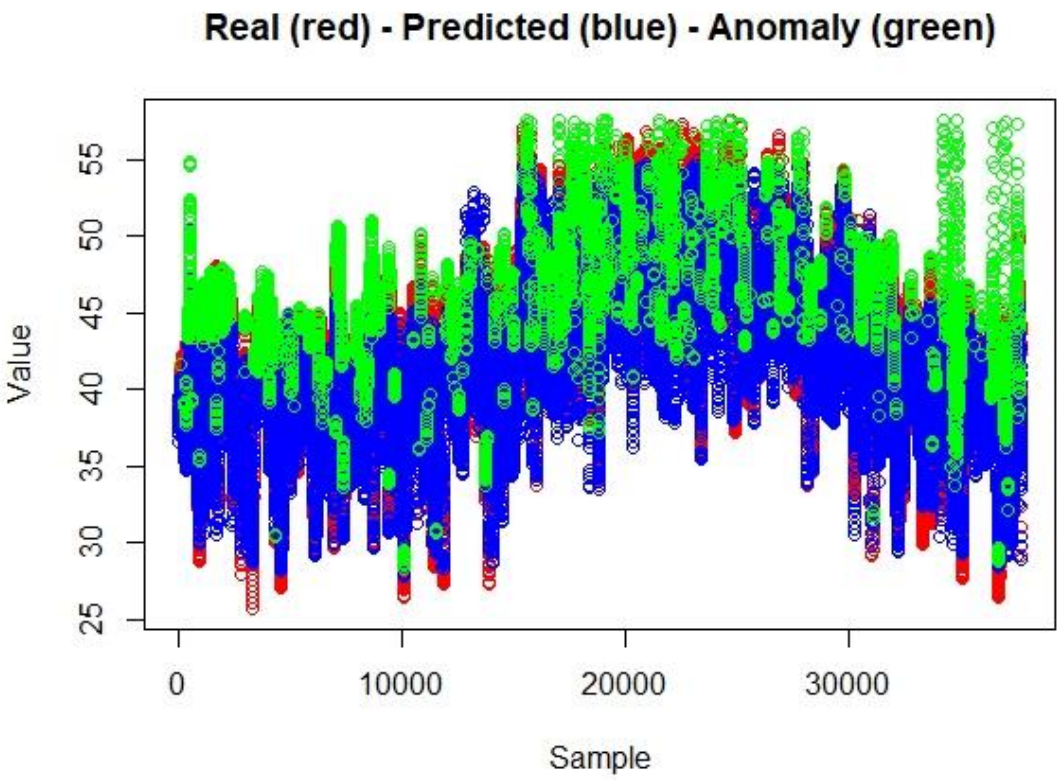


Ilustración 103. Predicciones, valores reales y anomalías en el rodamiento DE. Año 4.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-3{,}359 \leq X \leq 2{,}689]$, generado con los datos del primer año, encontramos 6584 muestras. Esto es un 17,45% de las muestras, de un total de 37724 muestras.

Año	**4**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	668	541	675	288	356	814	624	523	432	496	311	855
% muestras anómalas con respecto al total mensual de muestras	21,25	17,21	21,47	9,16	11,32	25,89	19,85	16,64	13,74	15,78	9,89	27,20

Tabla 50. Cuantificación muestras anómalas en el rodamiento DE. Año 4.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas se distribuyen equitativamente a lo largo del año.

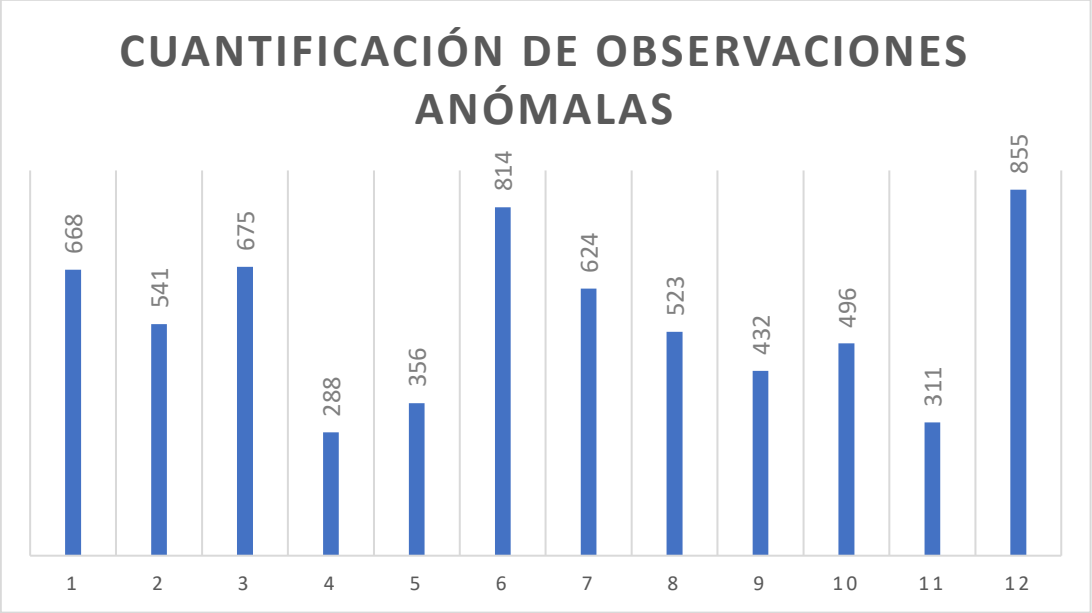


Ilustración 104. Cuantificación observaciones anómalas en el rodamiento DE. Año 4.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o mas desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**4**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	114	105	117	90	90	183	186	153	123	102	69	99
% anomalías con respecto al total mensual de muestras	3,63	3,33	3,72	2,85	2,85	5,82	5,91	4,86	3,9	3,24	2,19	3,15

Tabla 51. Cuantificación anomalías en el rodamiento DE. Año 4.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías se concentran en los meses de verano, aunque son abundantes durante todo el año. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

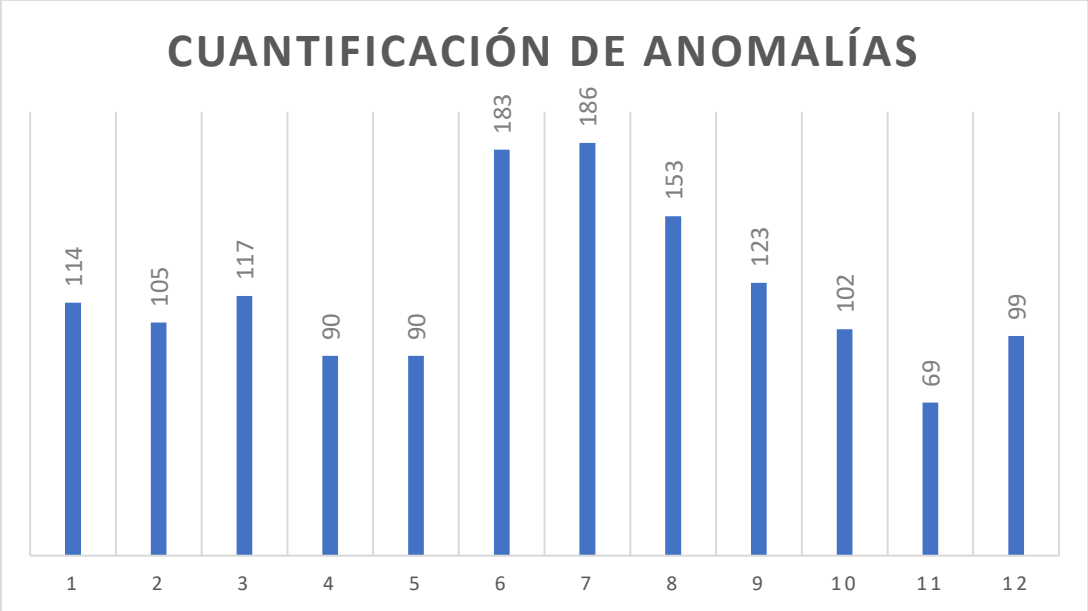


Ilustración 105. Cuantificación de anomalías en el rodamiento DE. Año 4.

Año 5

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

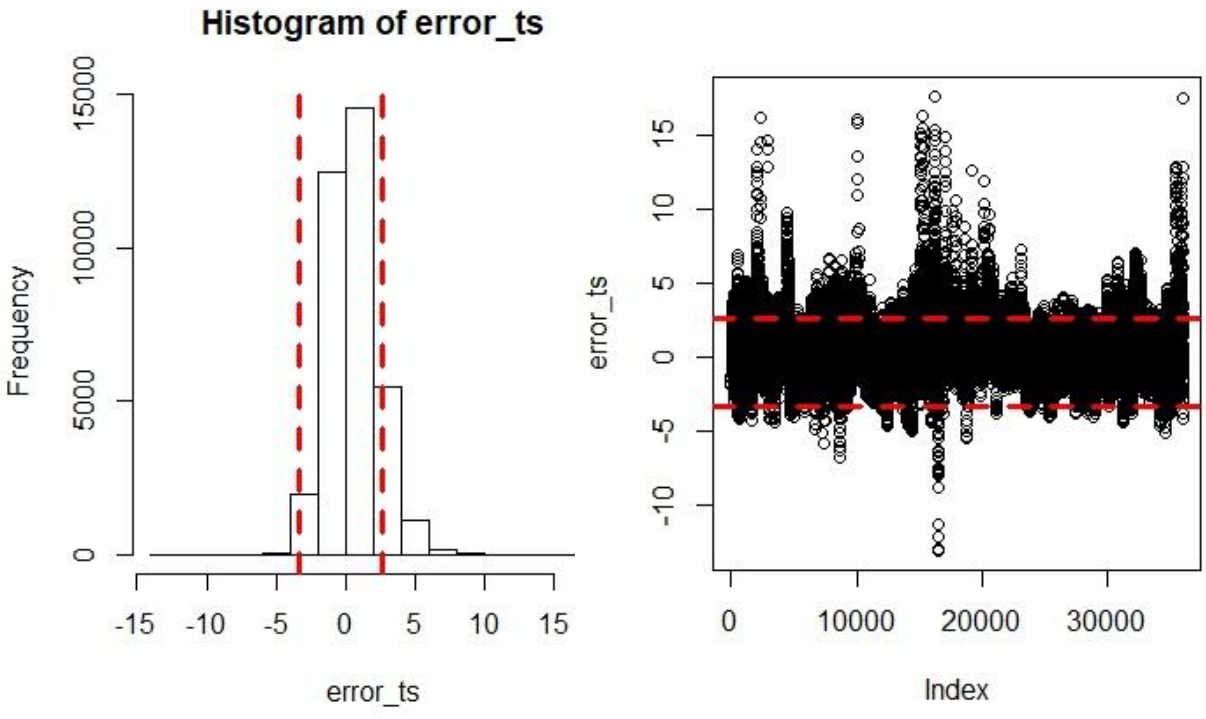


Ilustración 106. Histograma de residuos y errores de predicción en el rodamiento DE. Año 5.

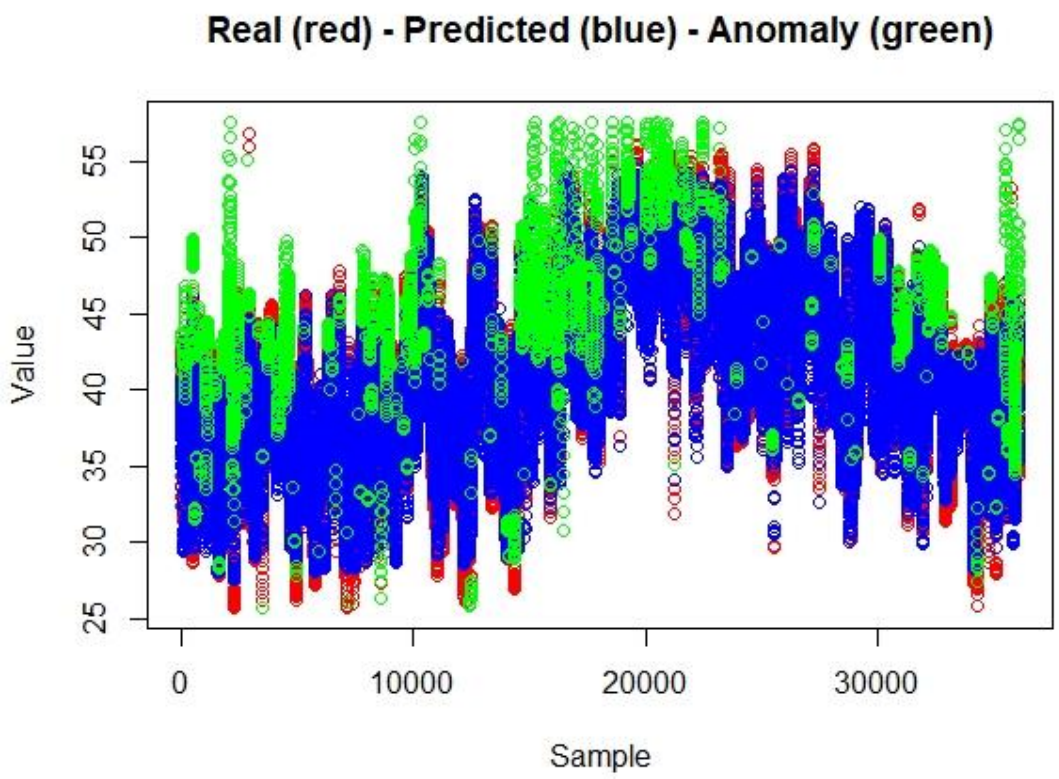


Ilustración 107. Predicciones, valores reales y anomalías en el rodamiento DE. Año 5.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-3{,}359 \leq X \leq 2{,}689]$, generado con los datos del primer año, encontramos 4502 muestras. Esto es un 12,48% de las muestras, de un total de 36083 muestras.

Año	**5**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	743	380	377	232	330	830	320	190	39	104	603	353
% muestras anómalas con respecto al total mensual de muestras	24,71	12,64	12,54	7,72	10,97	27,60	10,64	6,32	1,30	3,46	20,05	11,74

Tabla 52. Cuantificación muestras anómalas en el rodamiento DE. Año 5.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas se concentran en meses puntuales.

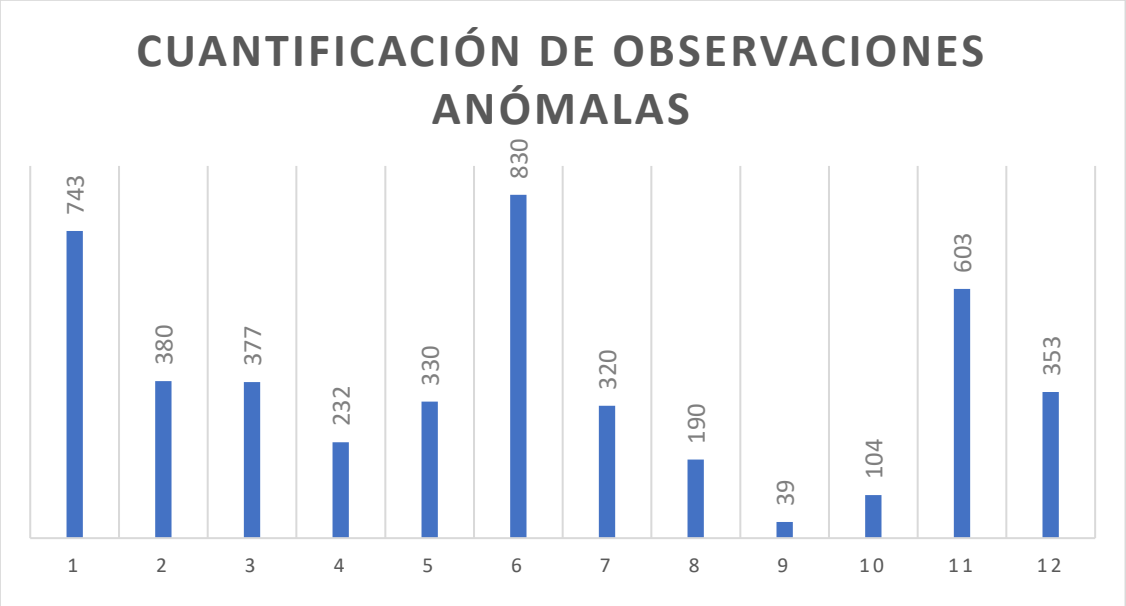


Ilustración 108. Cuantificación observaciones anómalas en el rodamiento DE. Año 5.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o mas desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**5**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	150	81	96	54	78	120	60	60	21	33	105	72
% anomalías con respecto al total mensual de muestras	4,98	2,7	3,18	1,8	2,58	3,99	2,01	2,01	0,69	1,11	3,48	2,4

Tabla 53. Cuantificación anomalías en el rodamiento DE. Año 5.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías se distribuyen equilibradamente a lo largo del año, aunque son mas abundantes en meses puntuales. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

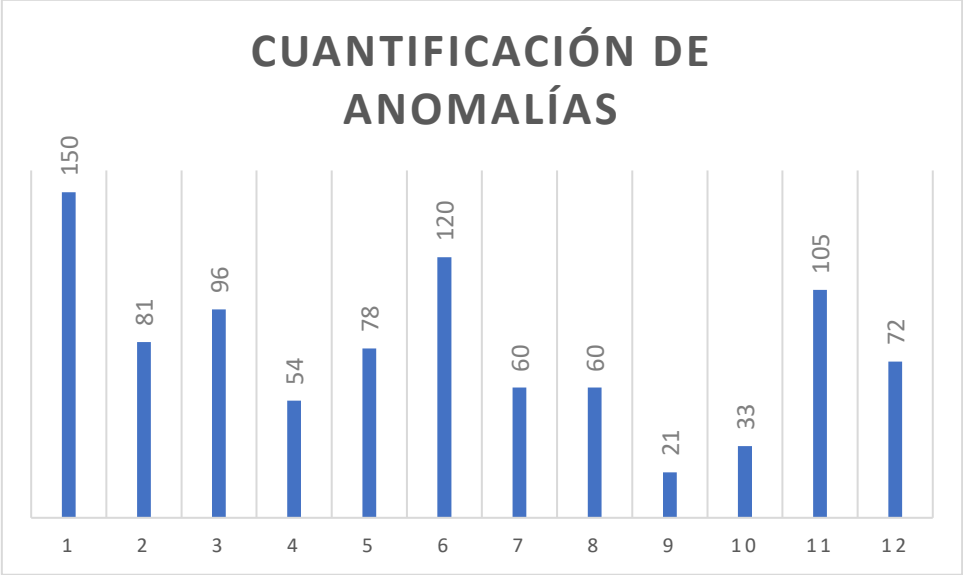


Ilustración 109. Cuantificación de anomalías en el rodamiento DE. Año 5.

Año 6

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

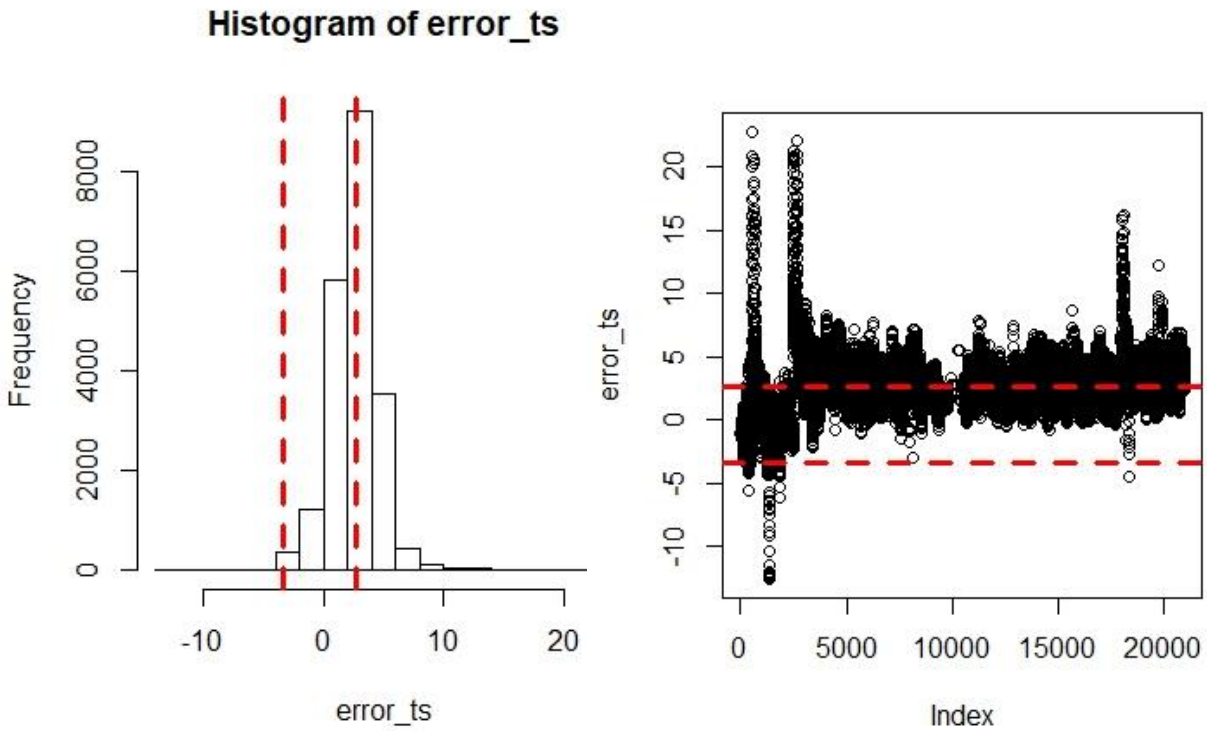


Ilustración 110. Histograma de residuos y errores de predicción en el rodamiento DE. Año 6.

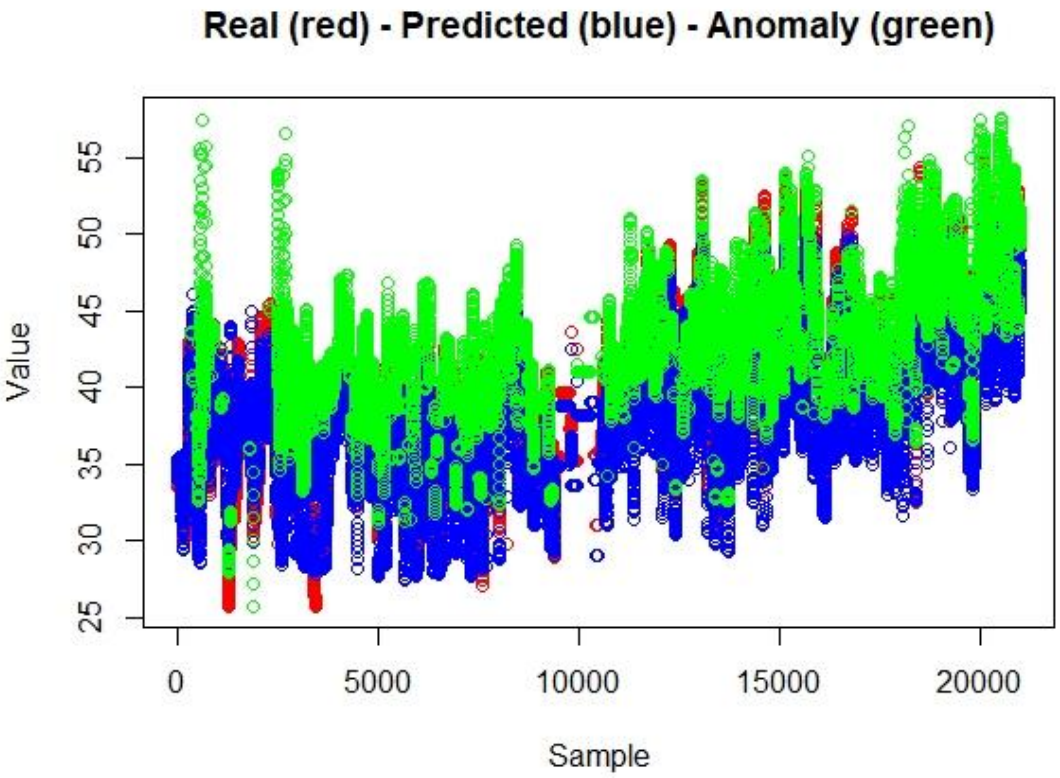


Ilustración 111. Predicciones, valores reales y anomalías en el rodamiento DE. Año 6.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-3{,}359 \leq X \leq 2{,}689]$, generado con los datos del primer año, encontramos 10286 muestras. Esto es un 48,97% de las muestras, de un total de 21006 muestras.

Año	**6**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	372	759	1216	731	942	587	910	722	1109	640	1114	1180
% muestras anómalas con respecto al total mensual de muestras	21,25	43,36	69,47	41,76	53,81	33,53	51,99	41,25	63,35	36,56	63,64	67,41

Tabla 54. Cuantificación muestras anómalas en el rodamiento DE. Año 6.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas se distribuyen equitativamente a lo largo del año.

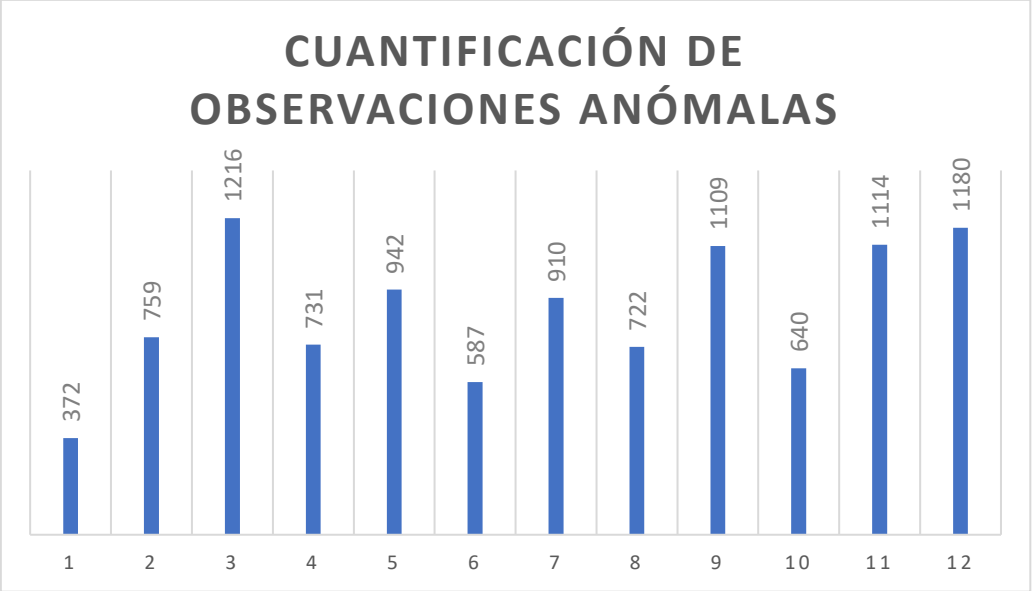


Ilustración 112. Cuantificación observaciones anómalas en el rodamiento DE. Año 6.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o mas desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**6**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	51	36	90	96	150	27	150	90	126	108	93	123
% anomalías con respecto al total mensual de muestras	2,91	2,07	5,13	5,49	8,58	1,53	8,58	5,13	7,2	6,18	5,31	7,02

Tabla 55. Cuantificación anomalías en el rodamiento DE. Año 6.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías se concentran en el segundo semestre, con picos en verano. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

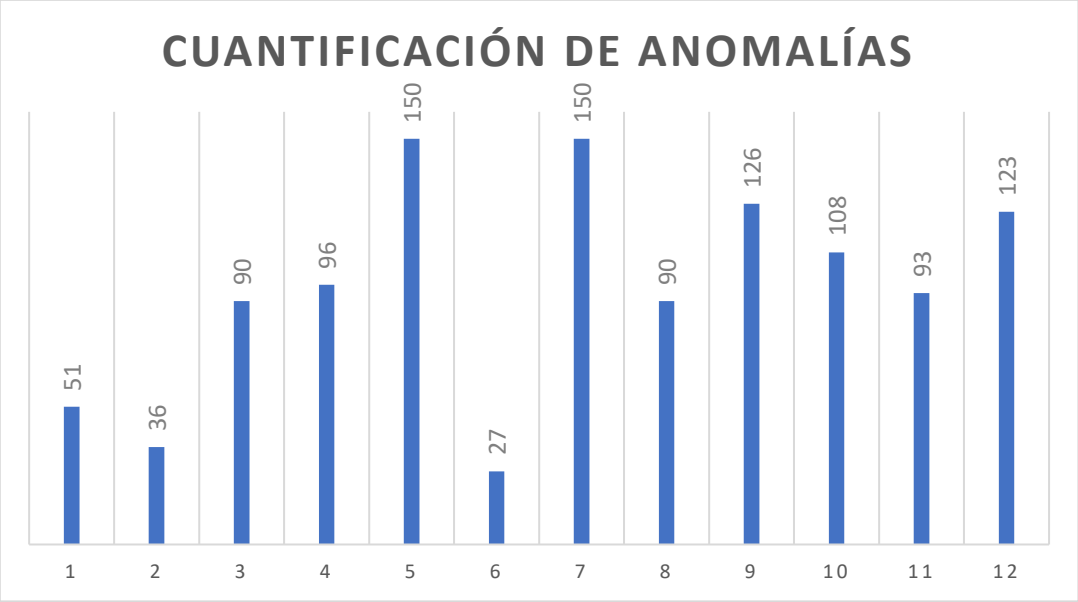


Ilustración 113. Cuantificación de anomalías en el rodamiento DE. Año 6.

6.6.- Modelo de previsión de la temperatura del rodamiento NDE

A partir de los intervalos de confianza calculados para el primer año, establecemos el intervalo de confianza que se empleará para los años sucesivos.

$$[-3{,}403 \leq X \leq 4{,}416]\ U\ [-3{,}481 \leq X \leq 4{,}608] = [-3{,}481 \leq X \leq 4{,}608]$$

Recordamos que el intervalo de confianza del 95% para el modelo ha sido establecido con el intervalo de confianza del 95% del conjunto de entrenamiento unido al intervalo de intervalo de confianza del 95% del conjunto de test de los datos del primer año. Esto se ha hecho cogiendo la opción menos restringida, para evitar incluir errores aportados por el propio modelo.

Año 2

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

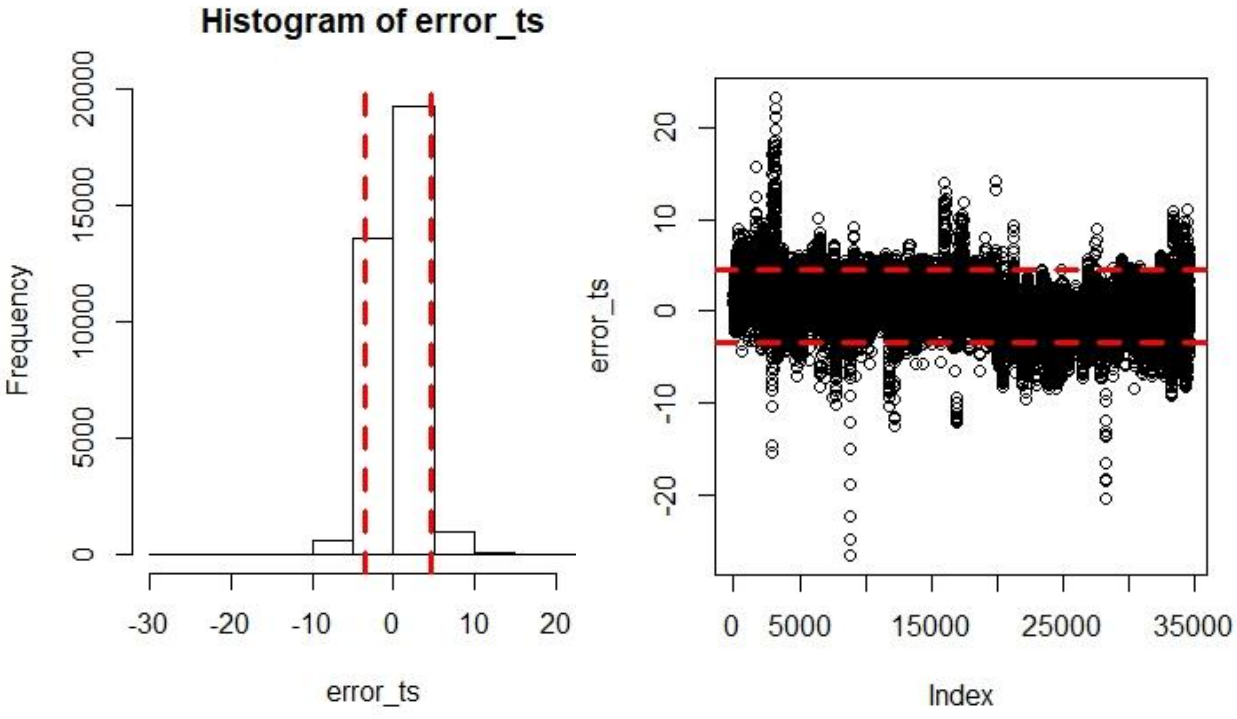


Ilustración 114. Histograma de residuos y errores de predicción en el rodamiento NDE. Año 2.

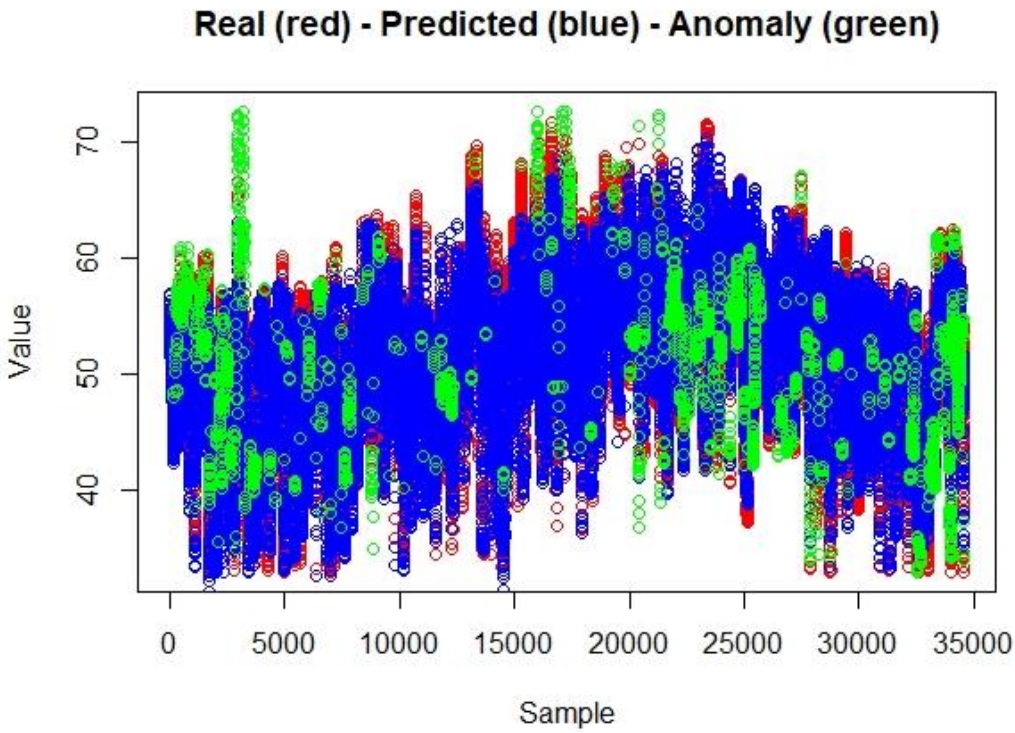


Ilustración 115. Predicciones, valores reales y anomalías en el rodamiento NDE. Año 2.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-3{,}481 \leq X \leq 4{,}608]$, generado con los datos del primer año, encontramos 3403 muestras. Esto es un 9,84% de las muestras, de un total de 34592 muestras

Año	**2**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	569	300	209	116	131	144	128	345	416	208	150	686
% muestras anómalas con respecto al total mensual de muestras	19,74	10,41	7,25	4,02	4,54	5,00	4,44	11,97	14,43	7,22	5,20	23,80

Tabla 56. Cuantificación muestras anómalas en el rodamiento NDE. Año 2.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas se concentran en los meses de invierno.

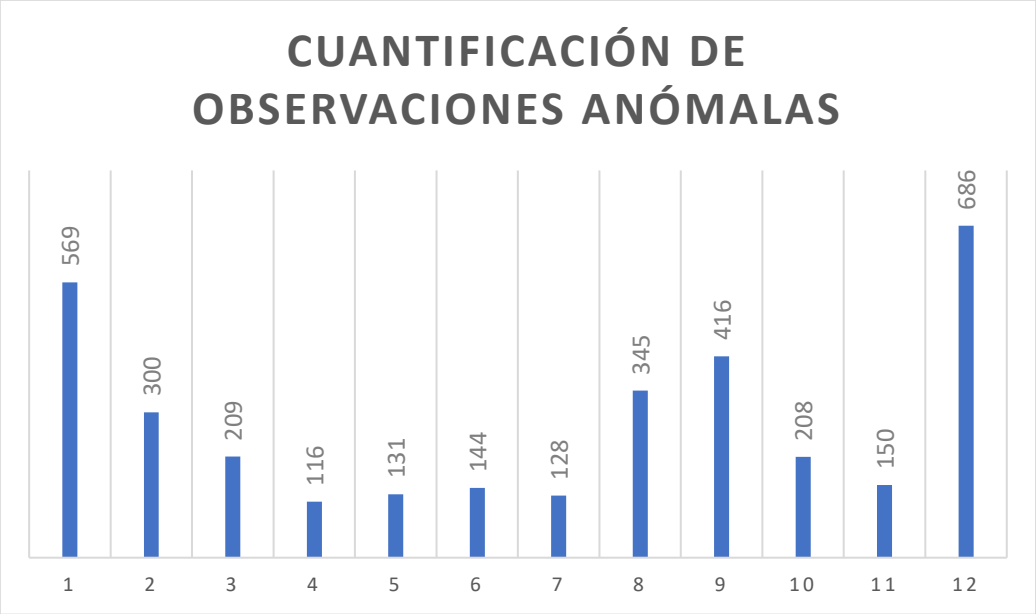


Ilustración 116. Cuantificación observaciones anómalas en el rodamiento NDE. Año 2.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o mas desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**2**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	207	75	81	39	39	51	54	105	120	75	51	162
% anomalías con respecto al total mensual de muestras	7,17	2,61	2,82	1,35	1,35	1,77	1,86	3,63	4,17	2,61	1,77	5,61

Tabla 57. Cuantificación anomalías en el rodamiento NDE. Año 2.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías se concentran los meses de invierno. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

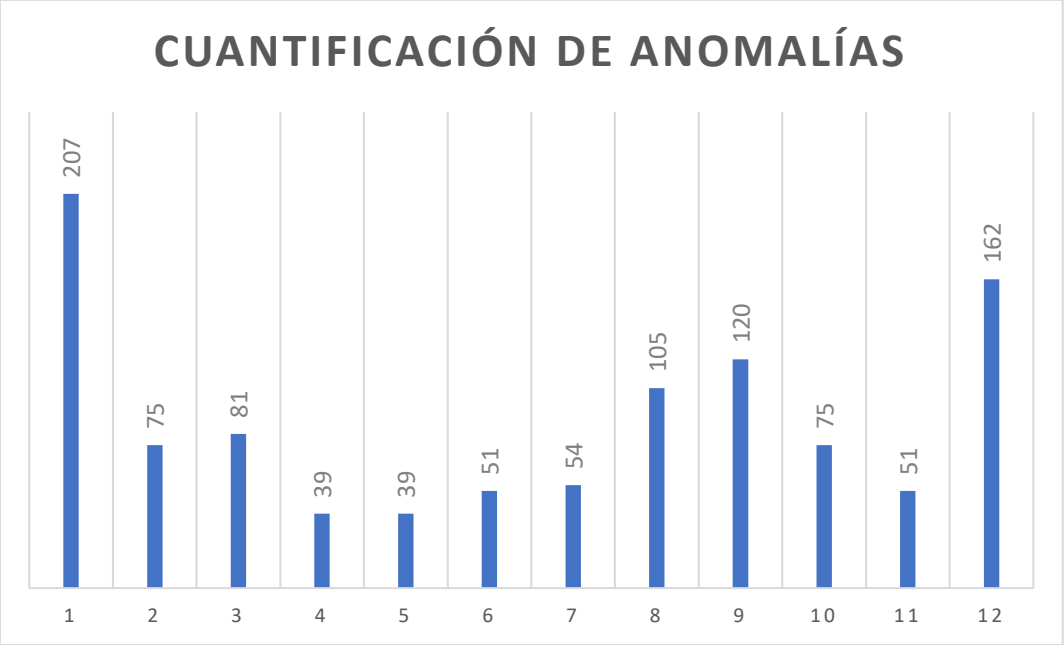


Ilustración 117. Cuantificación de anomalías en el rodamiento NDE. Año 2.

Año 3

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

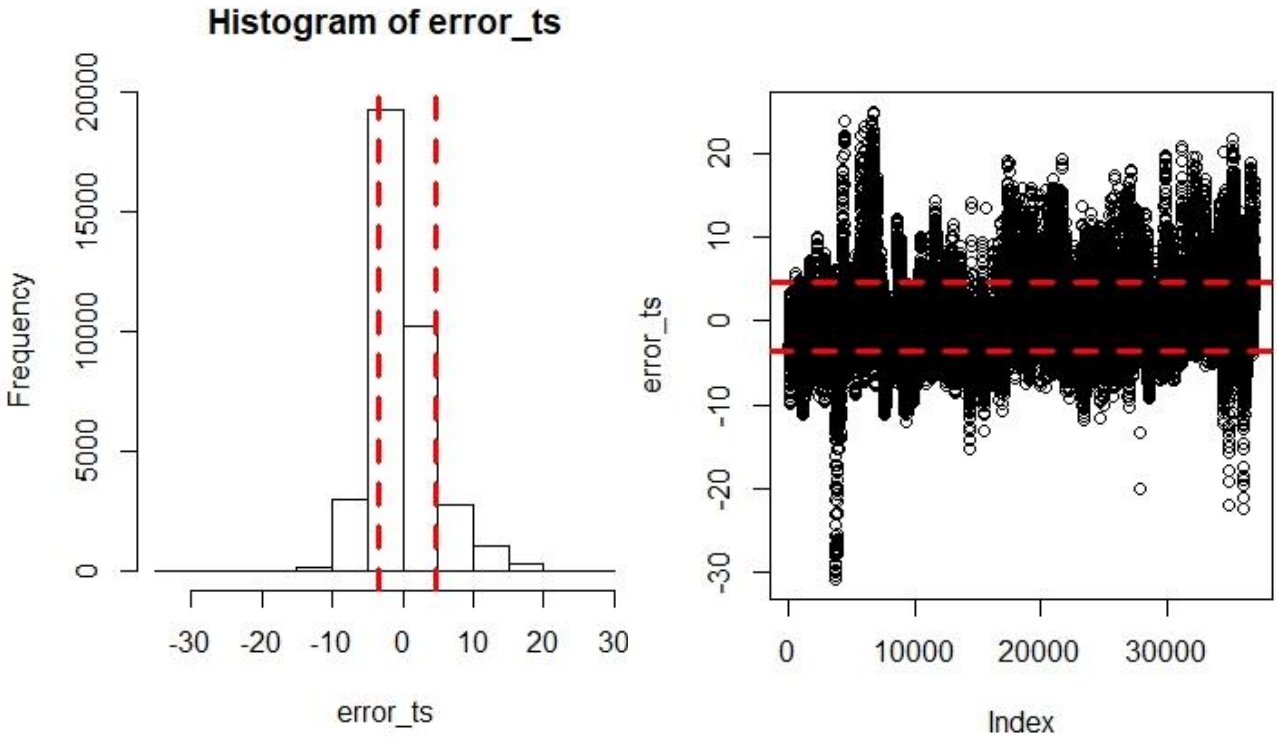


Ilustración 118. Histograma de residuos y errores de predicción en el rodamiento NDE. Año 3.

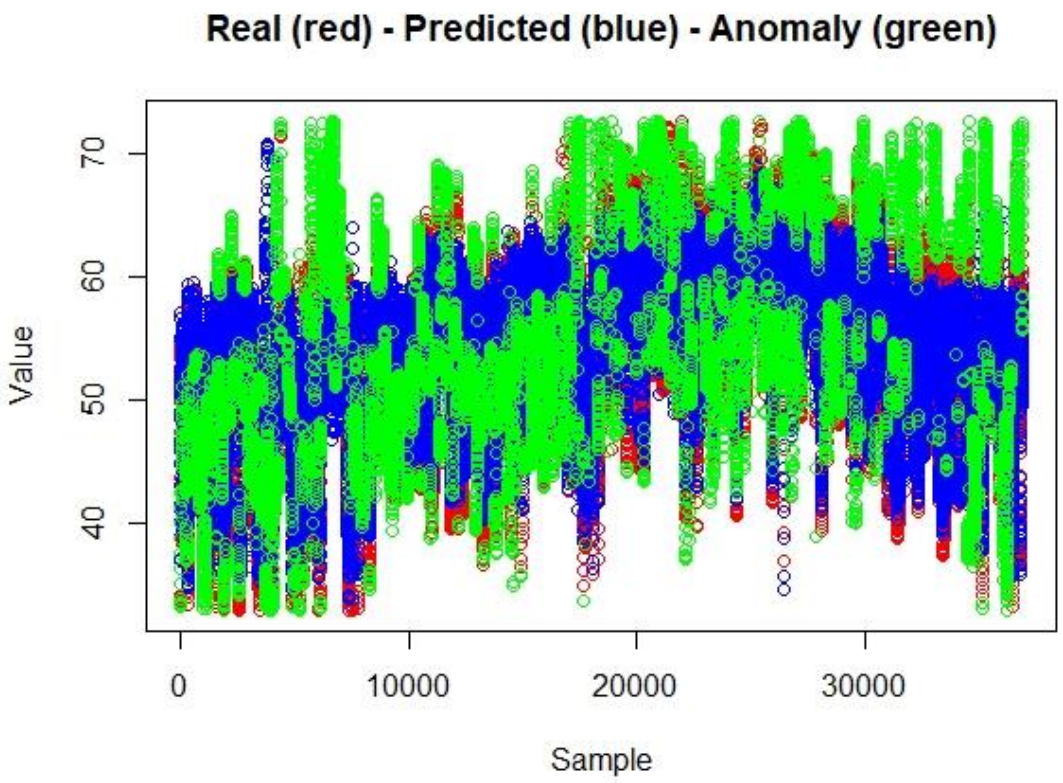


Ilustración 119. Predicciones, valores reales y anomalías en el rodamiento NDE. Año 3.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-3{,}481 \leq X \leq 4{,}608]$, generado con los datos del primer año, encontramos 11970 muestras. Esto es un 32,41% de las muestras, de un total de 36929 muestras

Año	**3**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	1147	1053	1369	1062	1076	1039	877	985	1124	686	506	1044
% muestras anómalas con respecto al total mensual de muestras	37,27	34,22	44,49	34,51	34,96	33,76	28,50	32,01	36,52	22,29	16,44	33,92

Tabla 58. Cuantificación muestras anómalas en el rodamiento NDE. Año 3.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas se distribuyen equitativamente a lo largo del año.

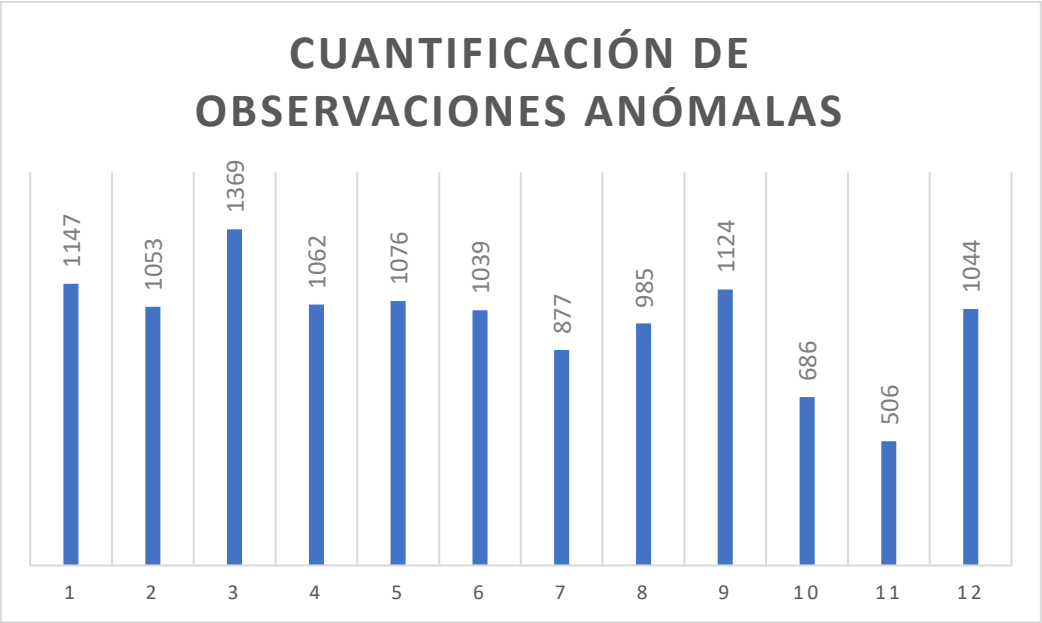


Ilustración 120. Cuantificación observaciones anómalas en el rodamiento NDE. Año 3.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o mas desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**3**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	243	219	180	237	168	192	255	240	225	168	87	168
% anomalías con respecto al total mensual de muestras	7,89	7,11	5,85	7,71	5,46	6,24	8,28	7,8	7,32	5,46	2,82	5,46

Tabla 59. Cuantificación anomalías en el rodamiento NDE. Año 3.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías se distribuyen equitativamente a lo largo del año, y son muy abundantes. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

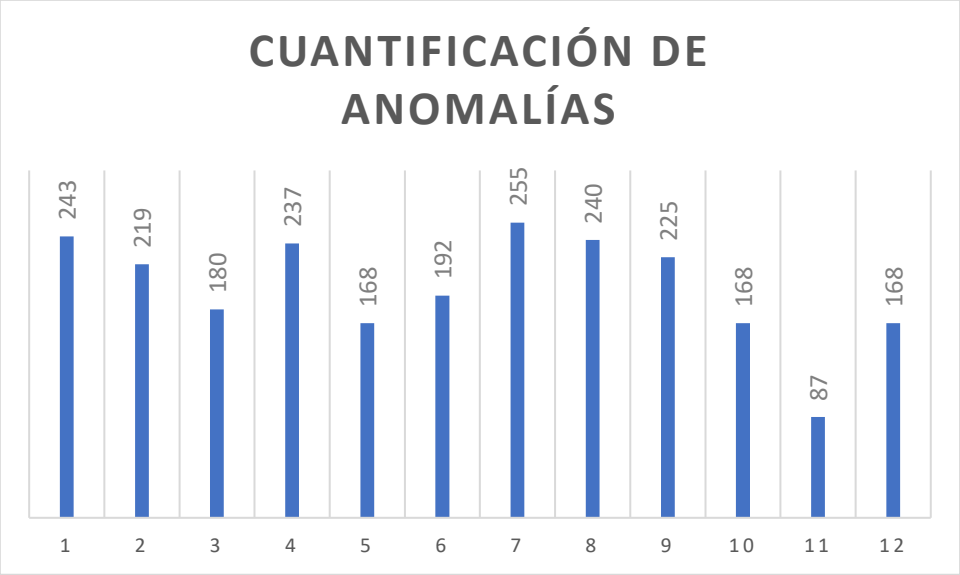


Ilustración 121. Cuantificación de anomalías en el rodamiento NDE. Año 3.

Año 4

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

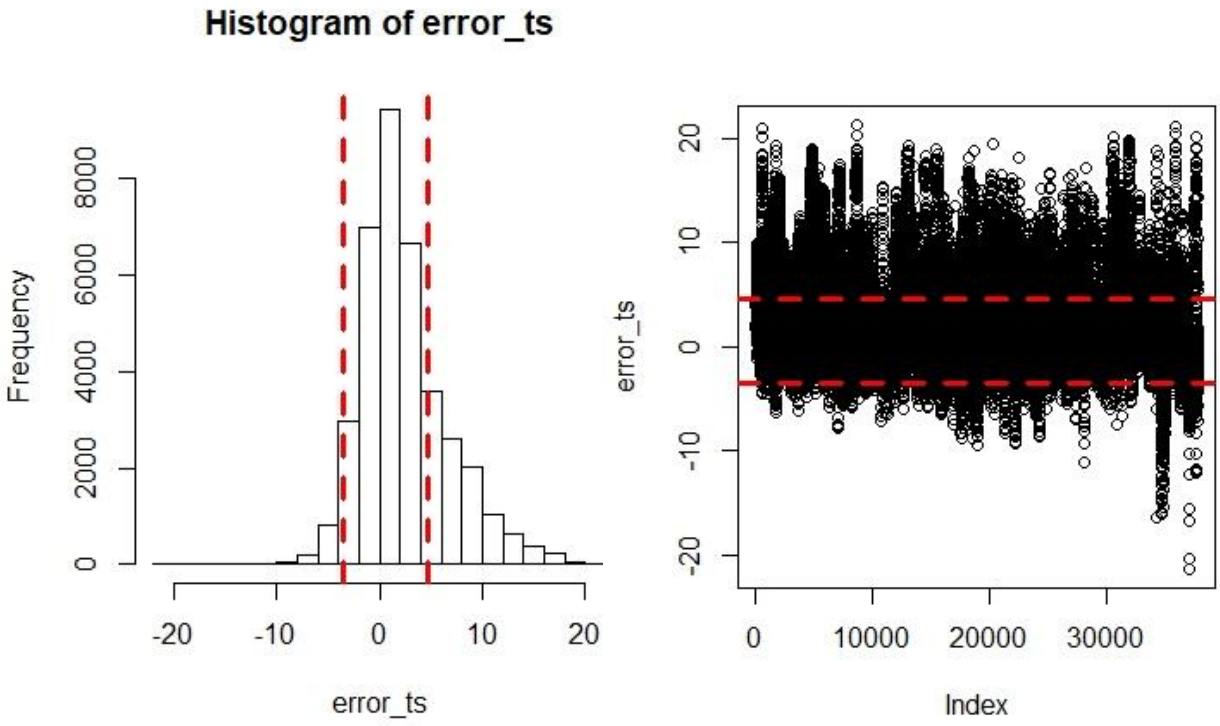


Ilustración 122. Histograma de residuos y errores de predicción en el rodamiento NDE. Año 4.

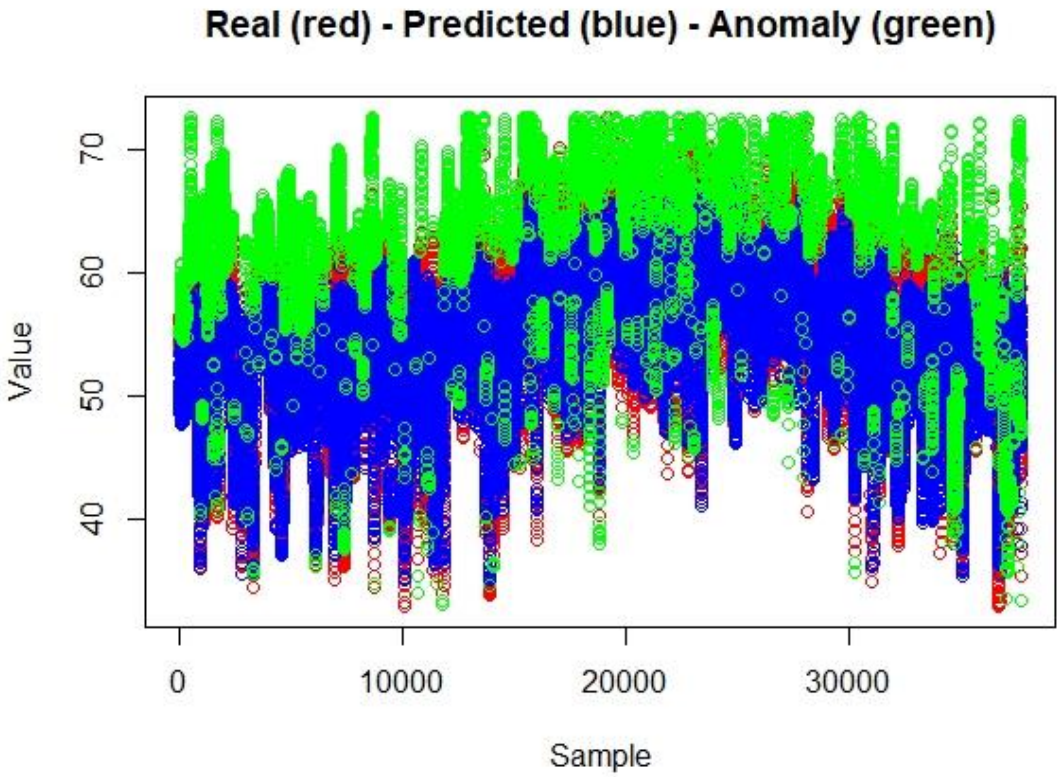


Ilustración 123. Predicciones, valores reales y anomalías en el rodamiento NDE. Año 4.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-3{,}481 \leq X \leq 4{,}608]$, generado con los datos del primer año, encontramos 10863 muestras. Esto es un 28,8% de las muestras, de un total de 37724 muestras.

Año	**4**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	1340	1442	822	563	863	886	917	864	663	605	699	1198
% muestras anómalas con respecto al total mensual de muestras	42,63	45,87	26,15	17,91	27,45	28,18	29,17	27,48	21,09	19,25	22,24	38,11

Tabla 60. Cuantificación muestras anómalas en el rodamiento NDE. Año 4.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas se distribuyen equitativamente a lo largo del año, aunque son mas cuantiosas en los meses de invierno.

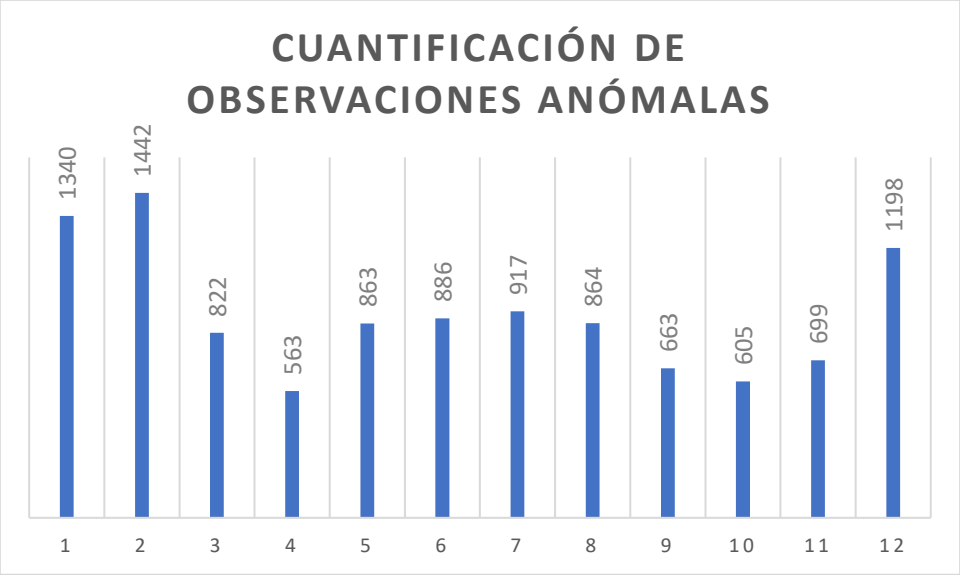


Ilustración 124. Cuantificación observaciones anómalas en el rodamiento NDE. Año 4.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o mas desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**4**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	177	174	105	117	138	201	207	252	162	87	153	216
% anomalías con respecto al total mensual de muestras	5,64	5,55	3,33	3,72	4,38	6,39	6,57	8,01	5,16	2,76	4,86	6,87

Tabla 61. Cuantificación anomalías en el rodamiento NDE. Año 4.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías se distribuyen equitativamente a lo largo del año, con picos en verano. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

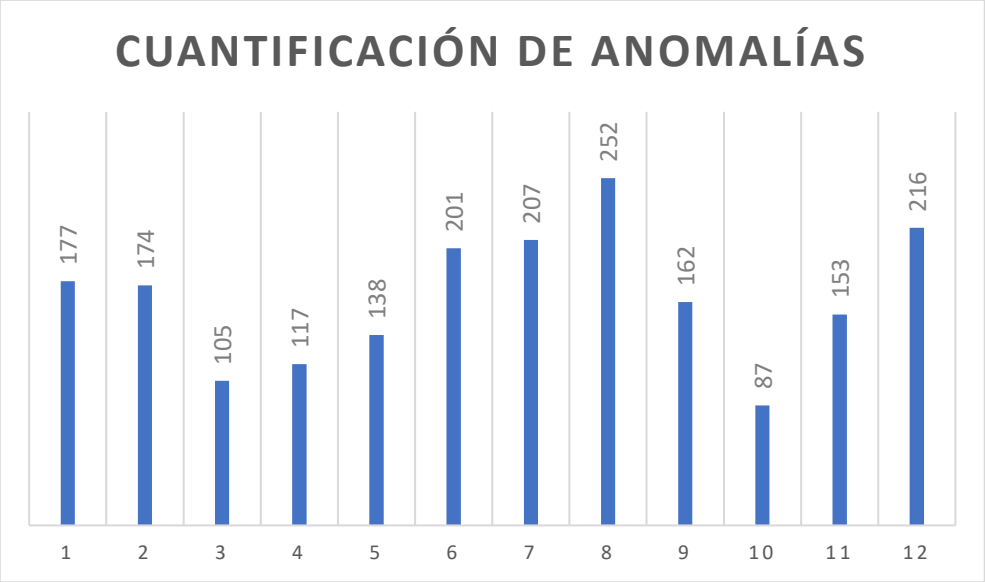


Ilustración 125. Cuantificación de anomalías en el rodamiento NDE. Año 4.

Año 5

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

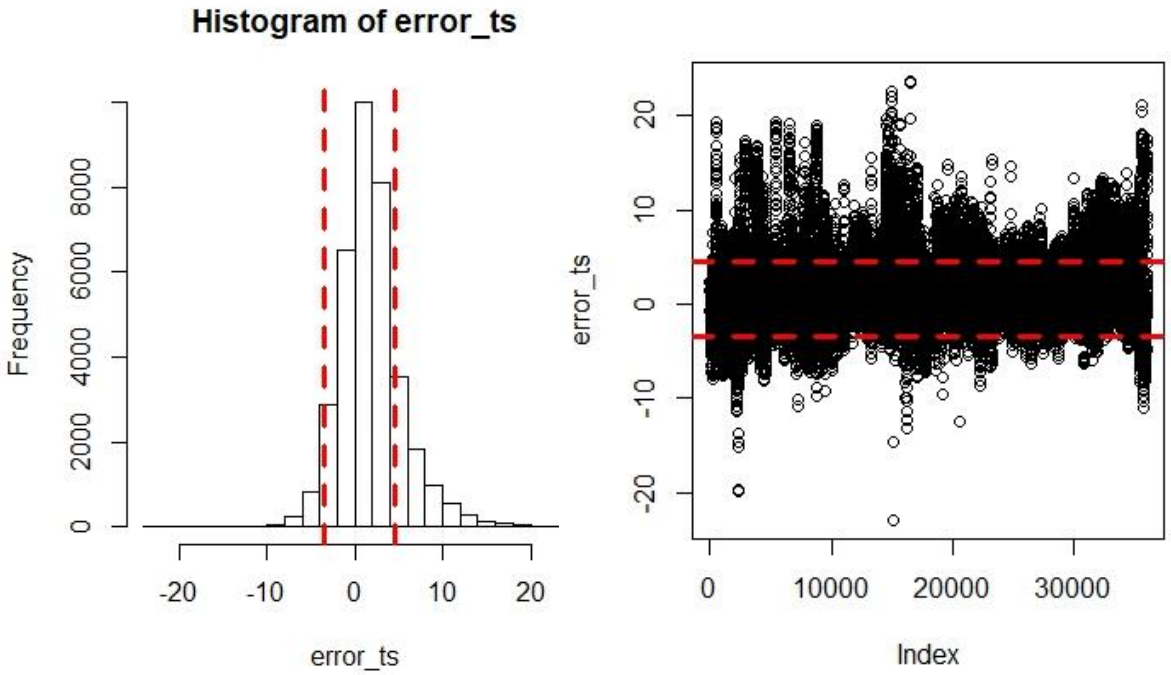


Ilustración 126. Histograma de residuos y errores de predicción en el rodamiento NDE. Año 5.

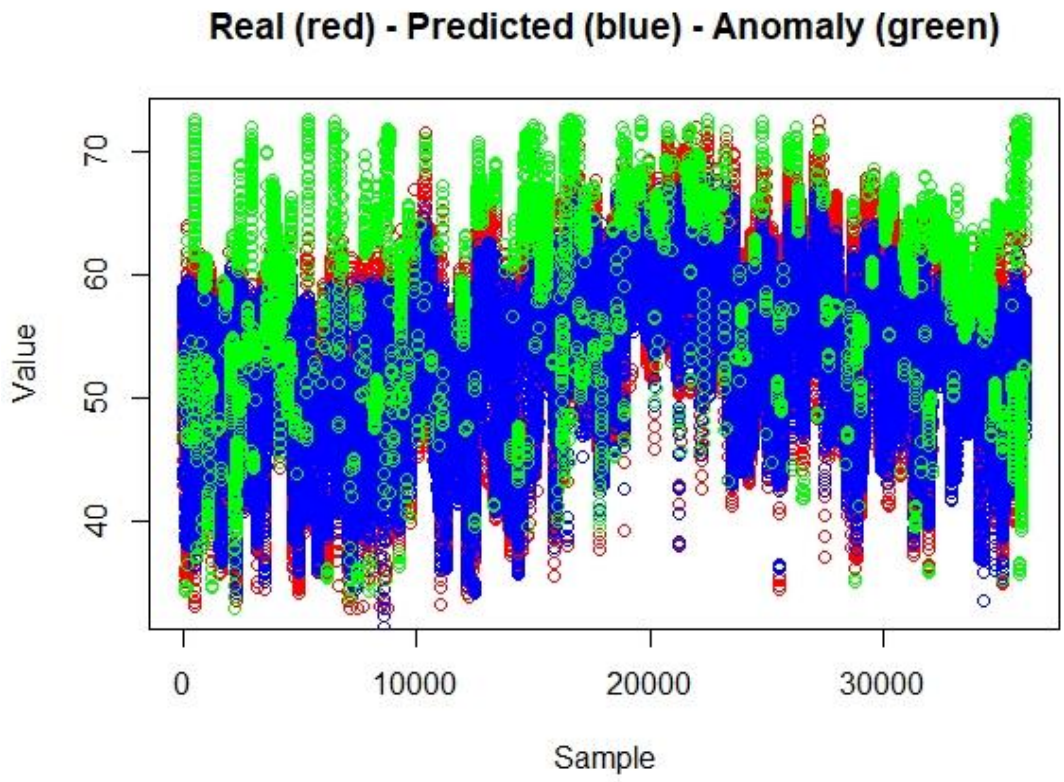


Ilustración 127. Predicciones, valores reales y anomalías en el rodamiento NDE. Año 5.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-3{,}481 \leq X \leq 4{,}608]$, generado con los datos del primer año, encontramos 7653 muestras. Esto es un 21,21% de las muestras, de un total de 36083 muestras.

Año	**5**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	927	717	583	369	524	730	440	435	245	259	970	1451
% muestras anómalas con respecto al total mensual de muestras	30,83	23,85	19,39	12,27	17,43	24,28	14,63	14,47	8,15	8,61	32,26	48,26

Tabla 62. Cuantificación muestras anómalas en el rodamiento NDE. Año 5.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas se concentran en los meses de invierno.

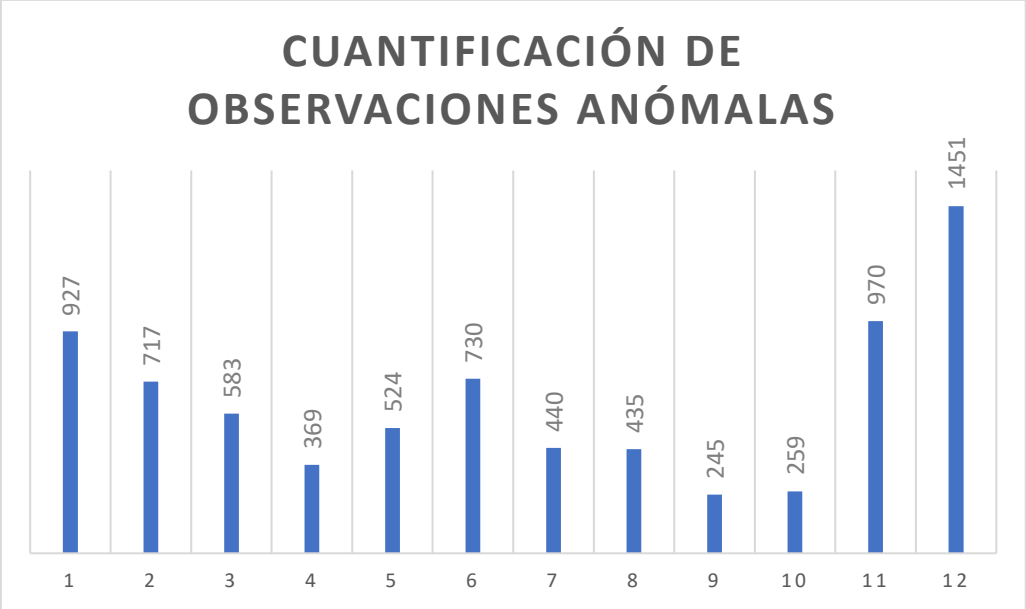


Ilustración 128. Cuantificación observaciones anómalas en el rodamiento NDE. Año 5.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o mas desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**5**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	228	156	144	114	96	186	159	138	78	90	144	177
% anomalías con respecto al total mensual de muestras	7,59	5,19	4,8	3,78	3,18	6,18	5,28	4,59	2,58	3	4,8	5,88

Tabla 63. Cuantificación anomalías en el rodamiento NDE. Año 5.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías se distribuyen equilibradamente a lo largo del año, con algunos picos en verano e invierno. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

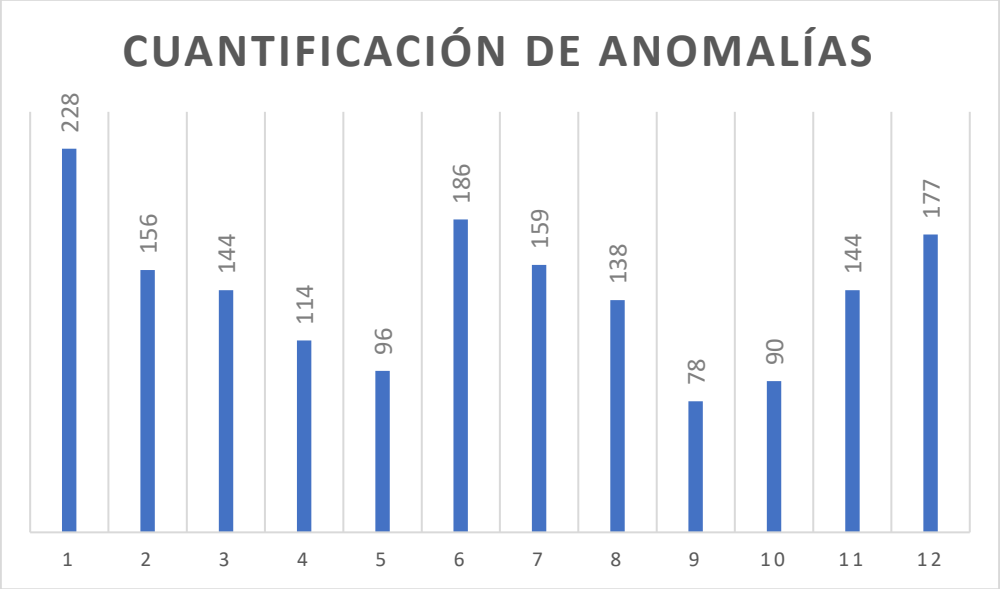


Ilustración 129. Cuantificación de anomalías en el rodamiento NDE. Año 5.

Año 6

Tras el haber definido la estrategia de detección de anomalías, se puede empezar a analiza en profundidad el conjunto de test utilizando el script , cuya explicación se puede encontrar en un apartado anterior. Los distintos indicadores de anomalías se pueden encontrar a continuación, y así como las gráficas obtenidas del intervalo de confianza del 95% aplicado al histograma de resíduos:

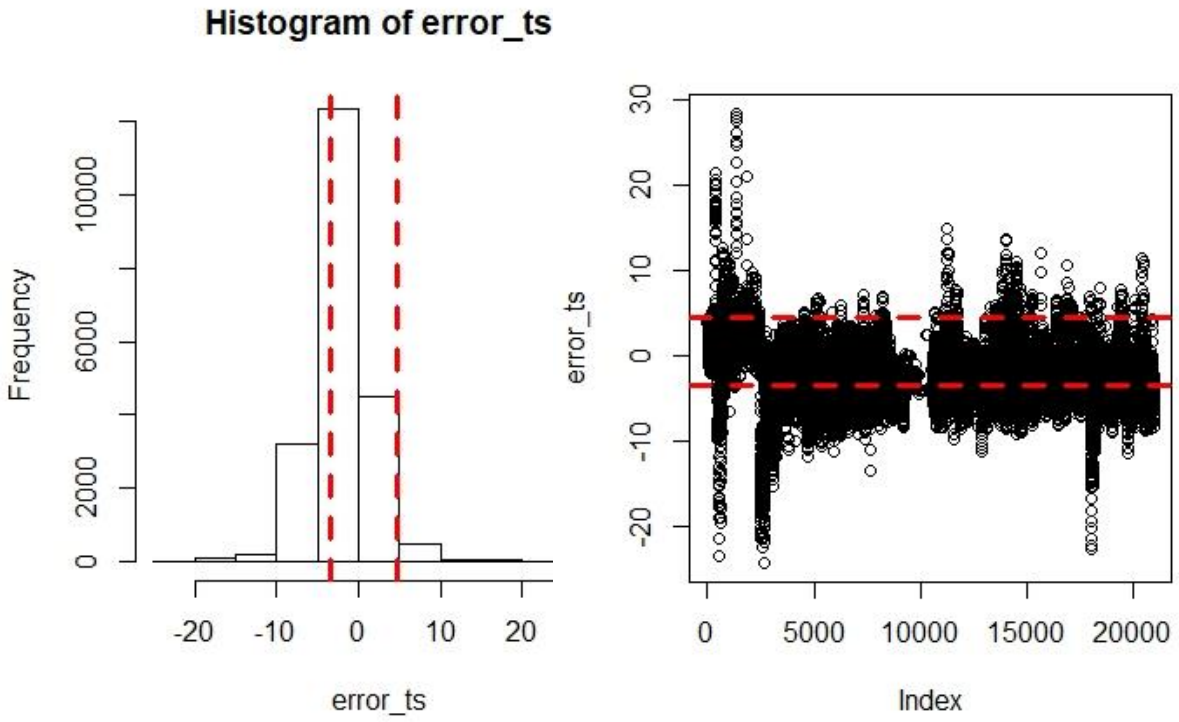


Ilustración 130. Histograma de residuos y errores de predicción en el rodamiento NDE. Año 6.

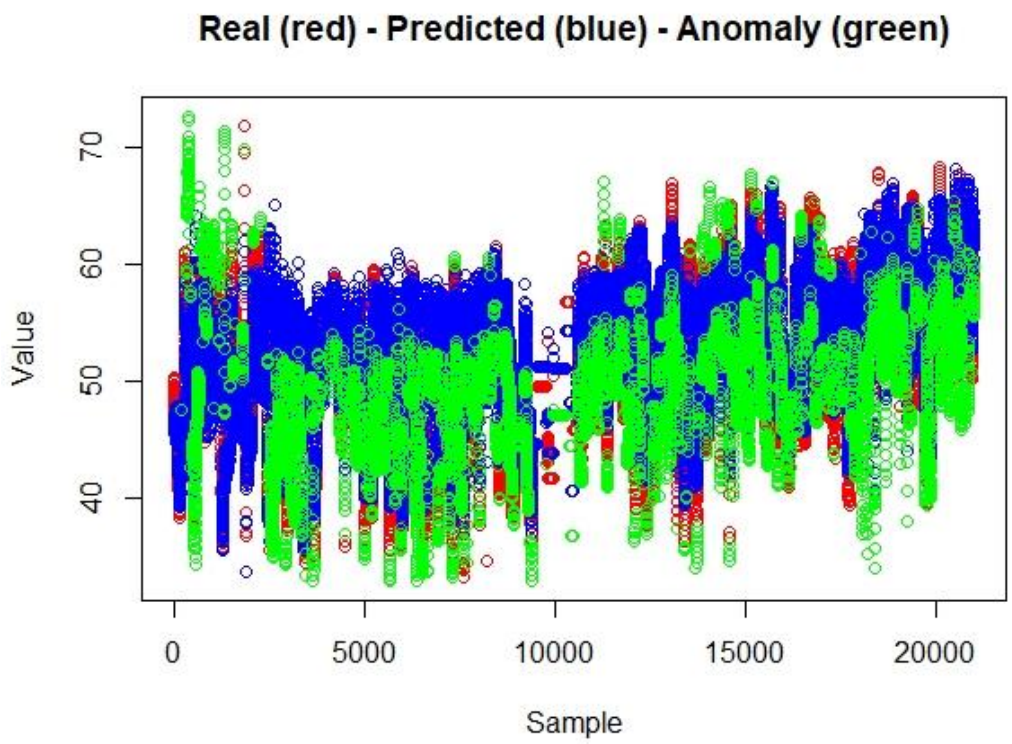


Ilustración 131. Predicciones, valores reales y anomalías en el rodamiento NDE. Año 6.

En los gráficos de arriba se puede dar ver el intervalo de confianza situado sobre el histograma de residuos y sobre el ploteado de errores en el eje temporal. Adicionalmente, es posibles observar sobre la gráfica de valores reales vs predicciones en que puntos se sitúan las anomalías.

Fuera del intervalo de confianza $[-3{,}481 \leq X \leq 4{,}608]$, generado con los datos del primer año, encontramos 8265 muestras. Esto es un 39,35% de las muestras, de un total de 21006 muestras.

Año	**6**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº muestras anómalas	483	716	531	825	750	678	710	532	797	519	858	864
% muestras anómalas con respecto al total mensual de muestras	27,59	40,90	30,33	47,13	42,84	38,73	40,56	30,39	45,53	29,65	49,01	49,36

Tabla 64. Cuantificación muestras anómalas en el rodamiento NDE. Año 6.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las observaciones anómalas se distribuyen equitativamente a lo largo del año.

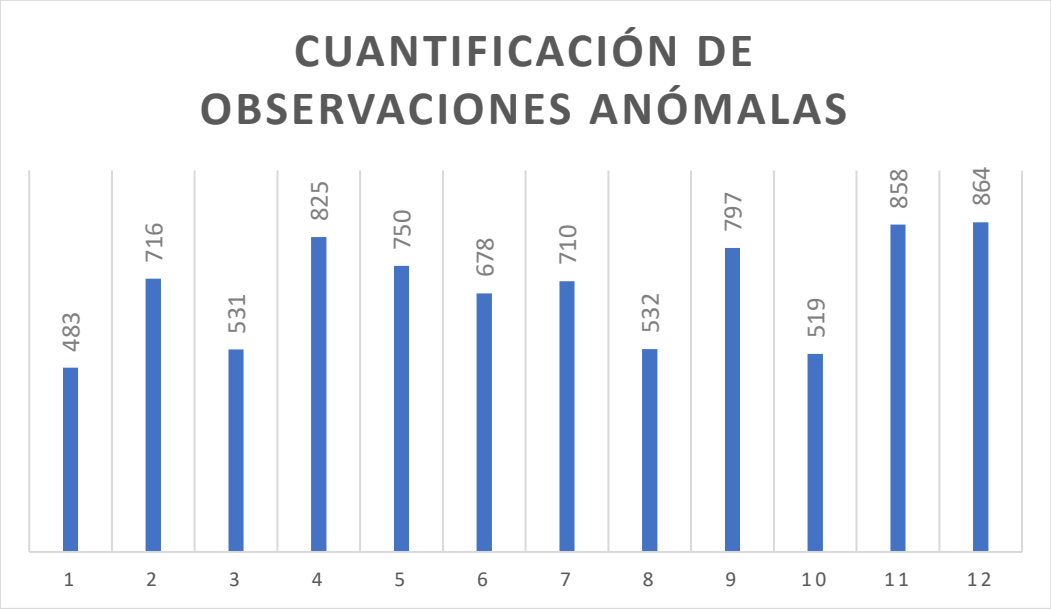


Ilustración 132. Cuantificación observaciones anómalas en el rodamiento NDE. Año 6.

Como ya hemos explicado, se considera una observación anómala que se trata de una anomalía cuando se sitúan tres o mas desviaciones seguidas las cuales se encuentran fuera del intervalo de confianza. Todo un conjunto de estas desviaciones anómalas (de fuera del intervalo) seguidas se consideran como una sola anomalía; un conjunto está formado por tantas observaciones anómalas como se sucedan.

A continuación, podemos observar esta cuantificación de las anomalías para el año presente.

Año	**6**											
Mes	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	81	75	108	180	192	33	156	108	138	141	123	150
% anomalías con respecto al total mensual de muestras	4,62	4,29	6,18	10,29	10,98	1,89	8,91	6,18	7,89	8,04	7,02	8,58

Tabla 65. Cuantificación anomalías en el rodamiento NDE. Año 6.

A continuación, se puede observar un gráfico en el que se puede apreciar visualmente que las anomalías se distribuyen equitativamente a lo largo del año, mas abundantes en los meses de primavera. Esto es algo que también se puede ver en el gráfico de valores reales frente a predicciones, donde se han superpuesto las anomalías.

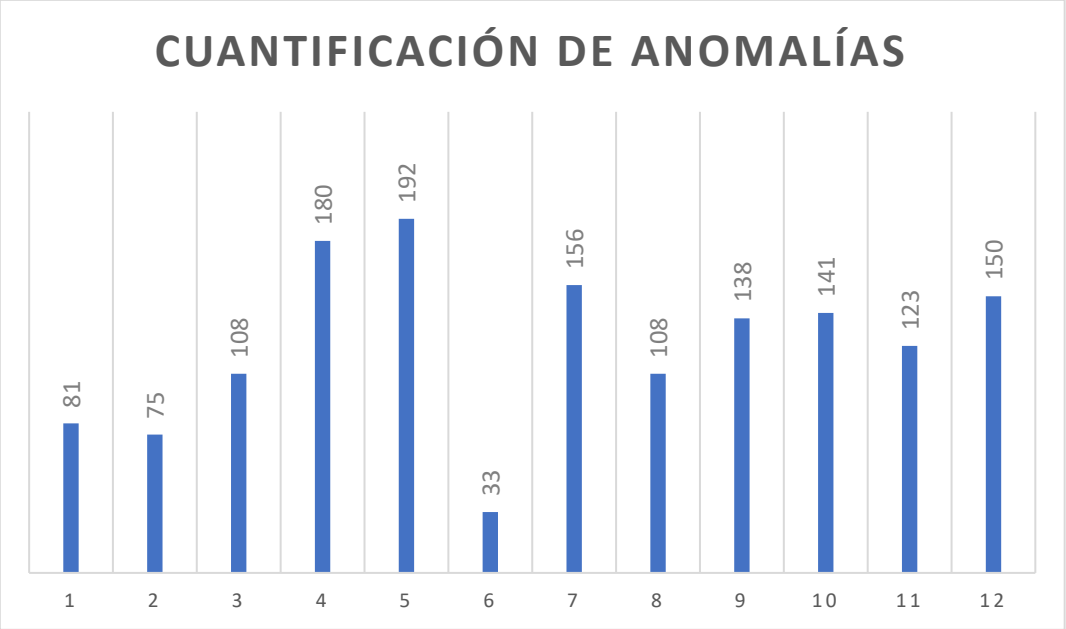


Ilustración 133. Cuantificación de anomalías en el rodamiento NDE. Año 6.

Capítulo 7.- Resultados

Tras haber detectado las anomalías, y haber extraído la presencia mensual de anomalías en cada uno de los elementos del aerogenerador, podemos proceder a interpretar los resultados. Esta sección interpreta resultados de los modelos con objeto de identificar el tipo de fallo presente, en el caso de existencia de uno. Aquí reside la utilidad última del proyecto ó finalidad del trabajo realizado.

7.1.- Estrategia para la interpretación de los resultados

Viendo la evolución de las anomalías detectadas en cada modelo, se puede extraer información del tipo de fallo que puede aparecer o que ya está presente.

Recordemos que los modelos predecían los valores que debería tomar un indicador de estado, indicador de estado del elemento para el que el modelo están diseñados y del que conocemos su ubicación. Por lo tanto, con este modelo conocemos:

1. Localización del fallo
2. Valor del indicador de estado

Por lo tanto, cada modelo está desarrollado para detectar un tipo particular de fallo:

F. Modelo de previsión de la potencia
    a. Modelo base para verificar el estado del aerogenerador, es decir, que se está operando con regularidad
G. Modelo de previsión de la temperatura de los anillos
    a. Fallo de los anillos con alta temperatura
H. Modelo de previsión de la temperatura de la multiplicadora
    a. Fallo de la multiplicadora con alta temperatura
I. Modelo de previsión de la temperatura de del rodamiento DE
    a. Fallo del rodamiento DE con alta temperatura
J. Modelo de previsión de la temperatura de del rodamiento NDE
    a. Fallo del rodamiento NDE con alta temperatura

Aunque todavía no se trate de un fallo catastrófico, el modelo puede facilitar el conocimiento prematuro de del mismo. Esto puede ser útil, por ejemplo, para la planificación de las tareas de mantenimiento.

La aplicación de este proyecto reside en poder poner en funcionamiento los modelos en tiempo real en un aerogenerador. De esta forma, en cuanto se distinga una línea de tendencia que eleve el porcentaje de anomalía, se podrá traducir que nos encontramos ante el modo de fallo asociado al modelo que esté señalando la incidencia.

Una vez se conoce el fallo, se monitoriza en tiempo real. Se establecerá un techo para el porcentaje de anomalía, cuantificación del volumen del fallo, dado por el modelo. Antes de que se llegue al mismo se debería actuar sobre el elemento afectado.

Año	-	2	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3
Mes		1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	Potencia	45	45	54	33	12	15	15	9	12	39	39	117	99	111	69	48	36	21	45	51	51	15	51	87
% anomalías con respectoal total mensual de muestras		1,56	1,56	1,86	1,14	0,42	0,51	0,51	0,3	0,42	1,35	1,35	4,05	3,21	3,6	2,25	1,56	1,17	0,69	1,47	1,65	1,65	0,48	1,65	2,82
Nº anomalías	Temperatura de los anillos	84	12	6	9	9	12	21	45	66	84	48	135	216	168	165	207	297	270	63	171	90	180	114	174
% anomalías con respectoal total mensual de muestras		2,91	0,42	0,21	0,3	0,3	0,42	0,72	1,56	2,28	2,91	1,68	4,68	7,02	5,46	5,37	6,72	9,66	8,76	2,04	5,55	2,91	5,85	3,69	5,64
Nº anomalías	Temperatura de la multiplicadora	135	96	63	21	15	15	9	33	39	9	36	30	21	39	24	15	18	48	120	78	132	81	33	63
% anomalías con respectoal total mensual de muestras		4,68	3,33	2,19	0,72	0,51	0,51	0,3	1,14	1,35	0,3	1,26	1,05	0,69	1,26	0,78	0,48	0,57	1,56	3,9	2,52	4,29	2,64	1,08	2,04
Nº anomalías	Rodamiento DE	174	93	69	39	39	45	12	48	54	27	45	72	180	123	156	180	156	123	207	195	165	135	90	114
% anomalías con respectoal total mensual de muestras		6,03	3,24	2,4	1,35	1,35	1,56	0,42	1,68	1,86	0,93	1,56	2,49	5,85	3,99	5,07	5,85	5,07	3,99	6,72	6,33	5,37	4,38	2,91	3,69
Nº anomalías	Rodamiento NDE	207	75	81	39	39	51	54	105	120	75	51	162	243	219	180	237	168	192	255	240	225	168	87	168
% anomalías con respectoal total mensual de muestras		7,17	2,61	2,82	1,35	1,35	1,77	1,86	3,63	4,17	2,61	1,77	5,61	7,89	7,11	5,85	7,71	5,46	6,24	8,28	7,8	7,32	5,46	2,82	5,46

Año	-	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5
Mes		1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	Potencia	45	111	66	87	282	123	66	39	51	24	48	84	36	84	114	84	93	78	6	48	42	60	66	75
% anomalías con respectoal total mensual de muestras		1,44	3,54	2,1	2,76	8,97	3,9	2,1	1,23	1,62	0,75	1,53	2,67	1,2	2,79	3,78	2,79	3,09	2,58	0,21	1,59	1,41	2,01	2,19	2,49
Nº anomalías	Temperatura de los anillos	144	144	219	195	177	180	93	51	129	210	138	36	66	177	153	201	141	150	93	198	201	192	213	195
% anomalías con respectoal total mensual de muestras		4,59	4,59	6,96	6,21	5,64	5,73	2,97	1,62	4,11	6,69	4,38	1,14	2,19	5,88	5,1	6,69	4,68	4,98	3,09	6,57	6,69	6,39	7,08	6,48
Nº anomalías	Temperatura de la multiplicadora	21	3	90	90	84	108	123	159	51	30	57	42	177	105	207	207	150	240	252	294	249	231	201	150
% anomalías con respectoal total mensual de muestras		0,66	0,09	2,85	2,85	2,67	3,45	3,9	5,07	1,62	0,96	1,8	1,35	5,88	3,48	6,87	6,87	4,98	7,98	8,37	9,78	8,28	7,68	6,69	4,98
Nº anomalías	Rodamiento DE	114	105	117	90	90	183	186	153	123	102	69	99	150	81	96	54	78	120	60	60	21	33	105	72
% anomalías con respectoal total mensual de muestras		3,63	3,33	3,72	2,85	2,85	5,82	5,91	4,86	3,9	3,24	2,19	3,15	4,98	2,7	3,18	1,8	2,58	3,99	2,01	2,01	0,69	1,11	3,48	2,4
Nº anomalías	Rodamiento NDE	177	174	105	117	138	201	207	252	162	87	153	216	228	156	144	114	96	186	159	138	78	90	144	177
% anomalías con respectoal total mensual de muestras		5,64	5,55	3,33	3,72	4,38	6,39	6,57	8,01	5,16	2,76	4,86	6,87	7,59	5,19	4,8	3,78	3,18	6,18	5,28	4,59	2,58	3	4,8	5,88

Año	-	6	6	6	6	6	6	6	6	6	6	6	6
Mes		1	2	3	4	5	6	7	8	9	10	11	12
Nº anomalías	Potencia	57	54	36	66	48	6	42	30	39	66	30	39
% anomalías con respectoal total mensual de muestras		3,26	3,08	2,06	3,77	2,74	0,34	2,4	1,71	2,23	3,77	1,71	2,23
Nº anomalías	Temperatura de los anillos	12	39	12	21	39	21	42	30	93	60	72	54
% anomalías con respectoal total mensual de muestras		0,69	2,22	0,69	1,2	2,22	1,2	2,4	1,71	5,31	3,42	4,11	3,09
Nº anomalías	Temperatura de la multiplicadora	132	87	102	111	99	48	123	120	72	96	168	150
% anomalías con respectoal total mensual de muestras		7,53	4,98	5,82	6,33	5,67	2,73	7,02	6,87	4,11	5,49	9,6	8,58
Nº anomalías	Rodamiento DE	51	36	90	96	150	27	150	90	126	108	93	123
% anomalías con respectoal total mensual de muestras		2,91	2,07	5,13	5,49	8,58	1,53	8,58	5,13	7,2	6,18	5,31	7,02
Nº anomalías	Rodamiento NDE	81	75	108	180	192	33	156	108	138	141	123	150
% anomalías con respectoal total mensual de muestras		4,62	4,29	6,18	10,29	10,98	1,89	8,91	6,18	7,89	8,04	7,02	8,58

Tabla 66. Resultados globales de cuantificación de las anomalías

Para trabajar con los resultados se ha recopilado esta tabla resumen que computa las anomalías detectadas es cada uno de los elementos mensualmente, así como el porcentaje de anomalías con respecto al total de las muestras tomadas ese mes.

7.2.- Modelo de previsión de la potencia

Se calculará la media de las proporciones de anomalías mensuales de los años estudiados. Cuando se sobrepase la misma se considerará que la máquina está siendo estresada por el modo de fallo de potencia fuera de rango.

$$\mu = 2{,}1$$

Por lo tanto, los meses que se ha marcado como existencia de un fallo son:

2				3				4				5				6			
1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12

Tabla 67. Incidencia por fallo en el aerogenerador

Se calculará el tercer cuartil de las proporciones de anomalías mensuales de los años estudiados. Cuando se sobrepase la misma se considerará que la máquina corre riesgo de fallo que precisará un mantenimiento correctivo (reparación).

$$Q_3 = 2{,}76$$

Se puede identificar como techo razonable en función de las proporciones de anomalía observadas.

En el siguiente gráfico podemos encontrar los 12 meses del año, durante el transcurso de 6 años. Encontramos, superpuestos, un diagrama de barras y un gráfico lineal. Se puede observar en el diagrama de barras las anomalías acumuladas en un mes en concreto. El gráfico lineal sigue los valores de la proporción de anomalías detectadas en la temperatura de los anillos en función del total de muestras tomadas en ese mes.

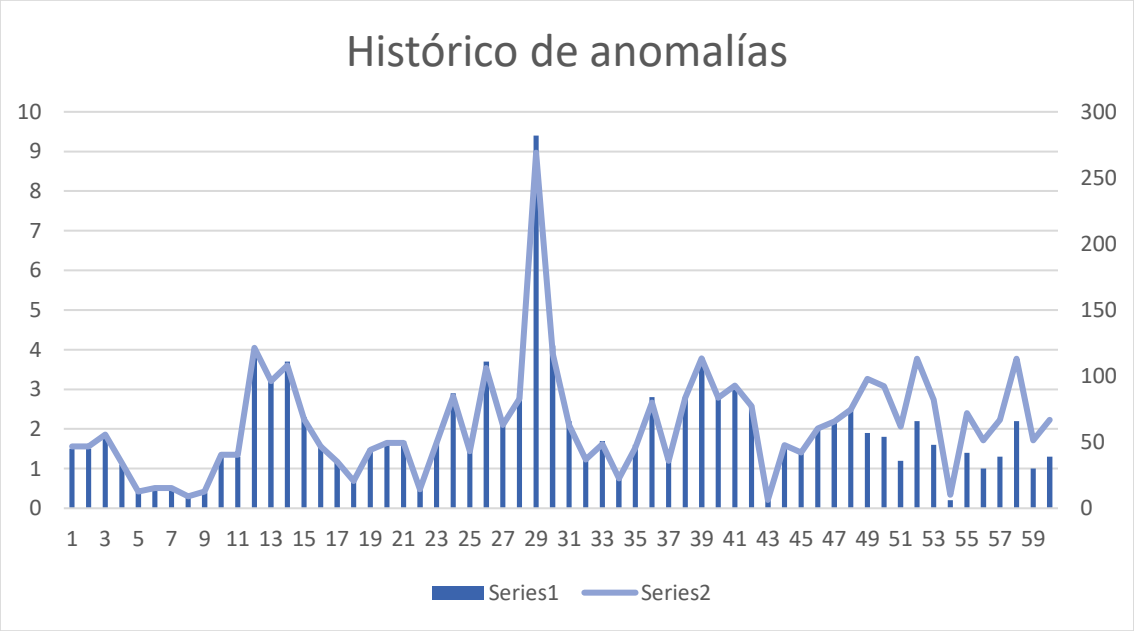


Ilustración 134. Histórico de anomalías detectadas en el aerogenerador

Se pueden identificar claras líneas de tendencia sobre las que se podía haber actuado antes de alcanzar valores altos del indicador. Estas se encuentran marcadas en el diagrama inferior. Estas líneas de tendencia conducen a fallo, el cual se puede identificar claramente.

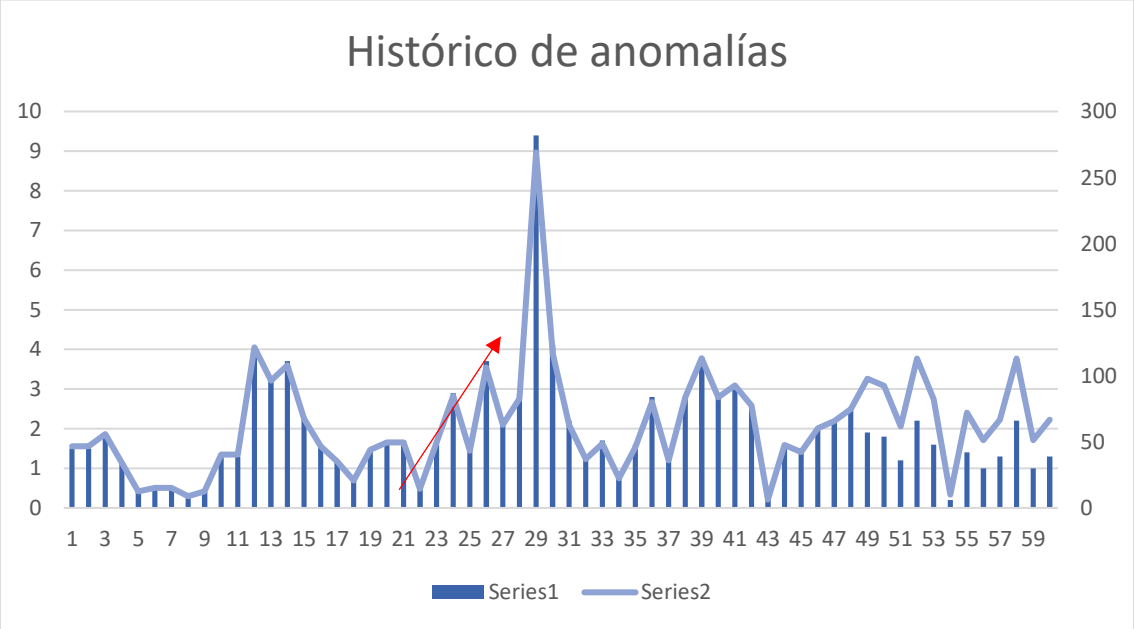


Ilustración 135. Líneas de tendencia detectadas en el aerogenerador

Una de ellas da lugar a un pico muy alto de anomalías durante el mes de mayo del cuarto año.

Se puede identificar que la tendencia estacional de aparición de valores fuera de rango es en invierno.

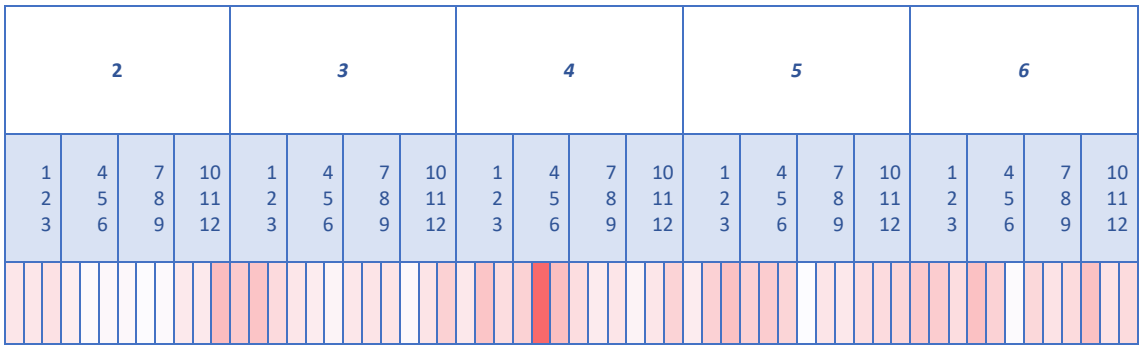


Tabla 68. Incidencia estacional de fallo en el aerogenerador

7.3.- Modelo de previsión de la temperatura de los anillos

Se calculará la media de las proporciones de anomalías mensuales de los años estudiados. Cuando se sobrepase la misma se considerará que la máquina está siendo estresada por el modo de fallo de temperatura alta en los anillos.

$$\mu = 3{,}92$$

Por lo tanto, los meses que se ha marcado como existencia de un fallo son:

2				3				4				5				6			
1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12

Tabla 69. Incidencia por fallo en los anillos

Se calculará el tercer cuartil de las proporciones de anomalías mensuales de los años estudiados. Cuando se sobrepase la misma se considerará que la máquina corre riesgo de fallo que precisará un mantenimiento correctivo (reparación).

$$Q_3 = 5{,}76$$

Se puede identificar como techo razonable en función de las proporciones de anomalía observadas.

En el siguiente gráfico podemos encontrar los 12 meses del año, durante el transcurso de 6 años. Encontramos, superpuestos, un diagrama de barras y un gráfico lineal. Se puede observar en el diagrama de barras las anomalías acumuladas en un mes en concreto. El gráfico lineal sigue los valores de la proporción de anomalías detectadas en la temperatura de los anillos en función del total de muestras tomadas en ese mes.

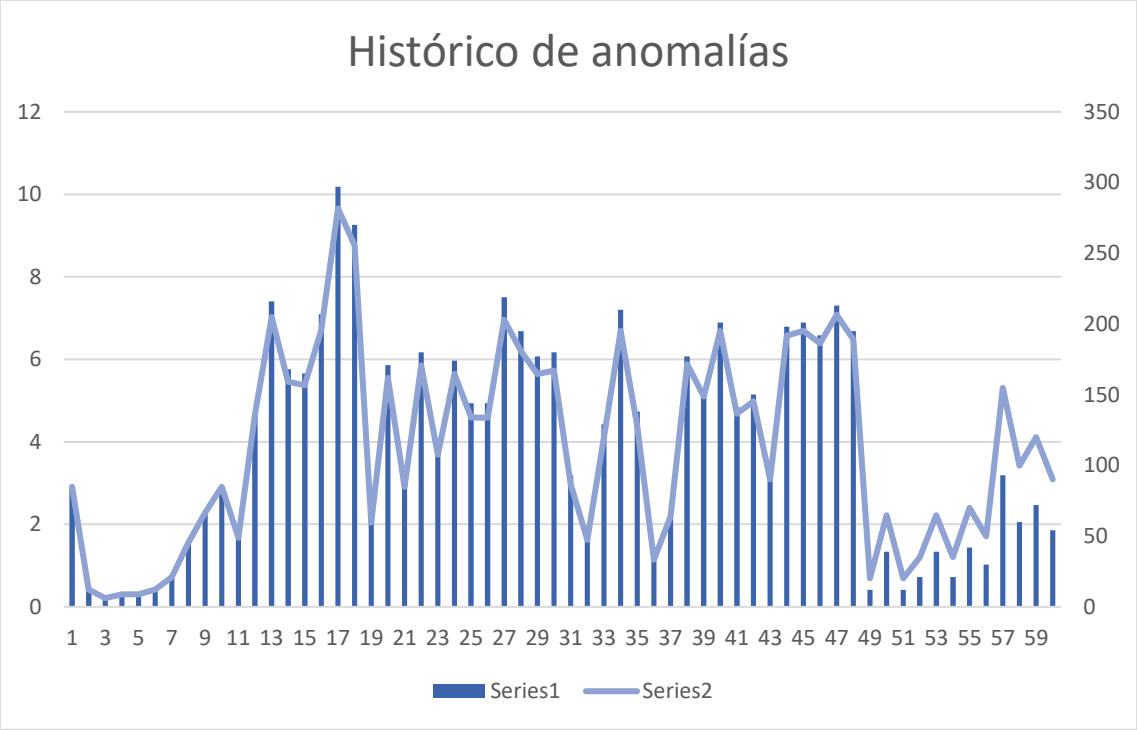


Ilustración 136. Histórico de anomalías detectadas en los anillos

Se pueden identificar claras líneas de tendencia sobre las que se podía haber actuado antes de alcanzar valores altos del indicador. Estas se encuentran marcadas en el diagrama inferior. Estas líneas de tendencia conducen a fallo, el cual se puede identificar claramente.

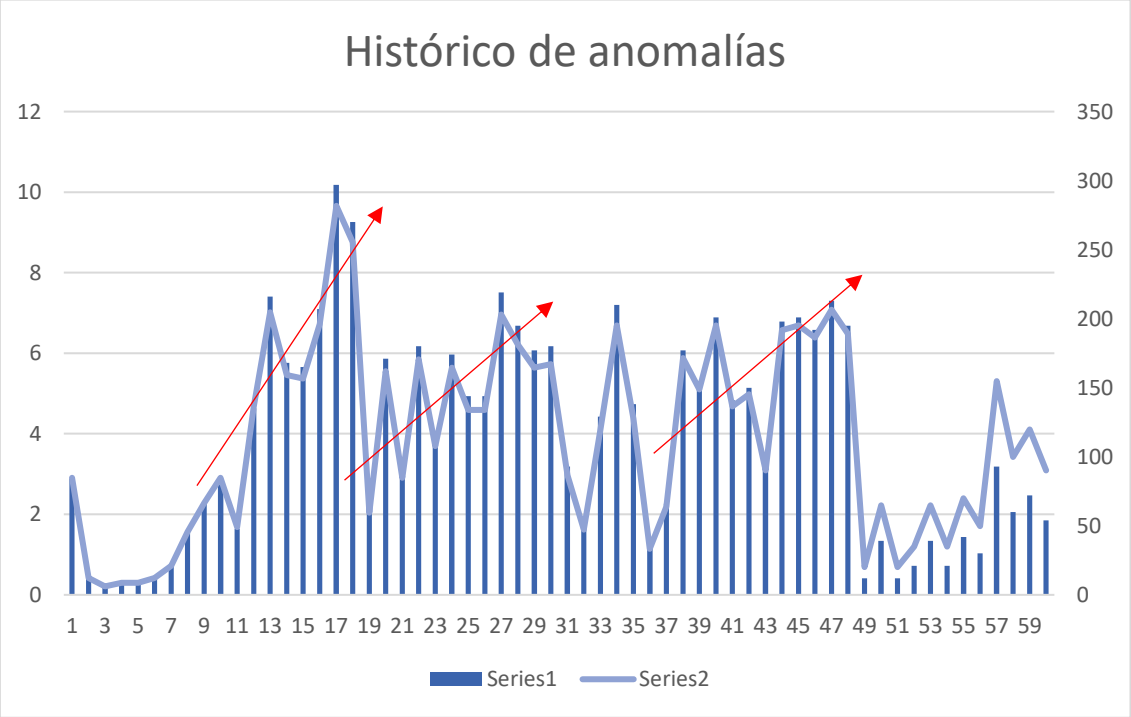


Ilustración 137. Líneas de tendencia detectadas en los anillos

Se puede identificar que la tendencia estacional de aparición de valores fuera de rango es en primavera.

2				3				4				5				6			
1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12

Tabla 70. Incidencia estacional de fallo en los anillos

7.4.- Modelo de previsión de la temperatura de la multiplicadora

Se calculará la media de las proporciones de anomalías mensuales de los años estudiados. Cuando se sobrepase la misma se considerará que la máquina está siendo estresada por el modo de fallo de temperatura alta en **la multiplicadora**.

$$\mu = 3{,}71$$

Por lo tanto, los meses que se ha marcado como existencia de un fallo son:

2				3				4				5				6			
1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12

Tabla 71. Incidencia por fallo en la multiplicadora

Se calculará el tercer cuartil de las proporciones de anomalías mensuales de los años estudiados. Cuando se sobrepase la misma se considerará que la máquina corre riesgo de fallo que precisará un mantenimiento correctivo (reparación).

$$Q_3 = 5{,}83$$

Se puede identificar como techo razonable en función de las proporciones de anomalía observadas.

En el siguiente gráfico podemos encontrar los 12 meses del año, durante el transcurso de 6 años. Encontramos, superpuestos, un diagrama de barras y un gráfico lineal. Se puede observar en el diagrama de barras las anomalías acumuladas en un mes en concreto. El gráfico lineal sigue los valores de la proporción de anomalías detectadas en la temperatura de los anillos en función del total de muestras tomadas en ese mes.

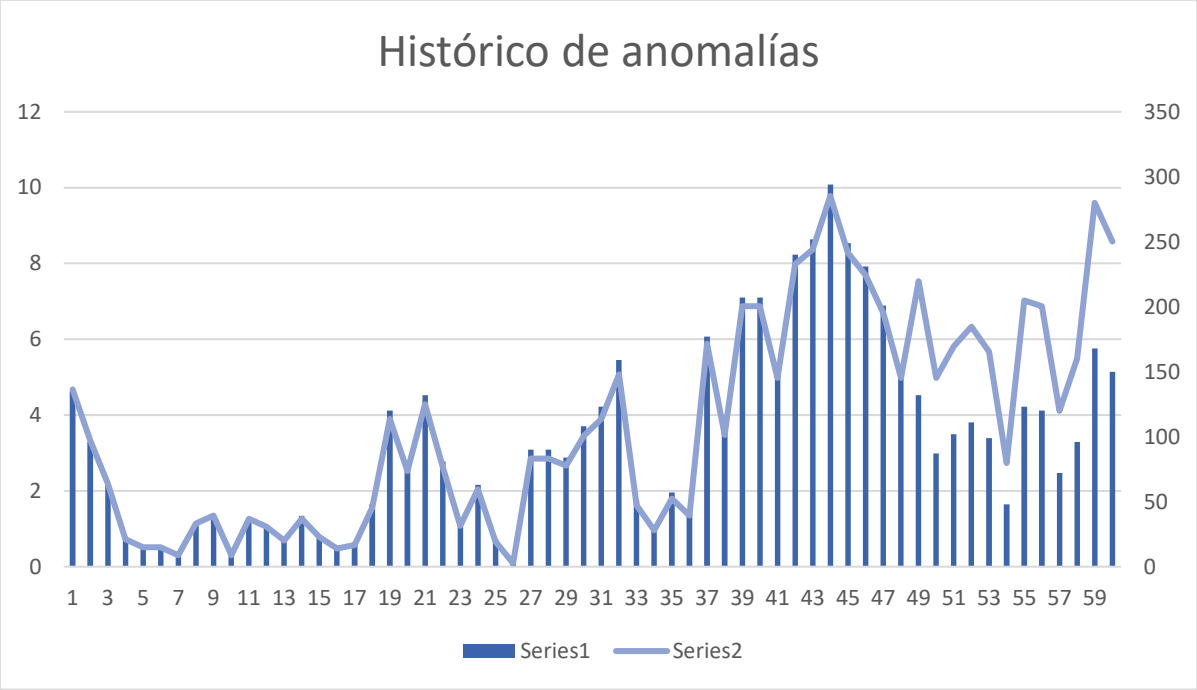


Ilustración 138. Histórico de anomalías detectadas en la multiplicadora

Se pueden identificar claras líneas de tendencia sobre las que se podía haber actuado antes de alcanzar valores altos del indicador. Estas se encuentran marcadas en el diagrama inferior. Estas líneas de tendencia conducen a fallo, el cual se puede identificar claramente.

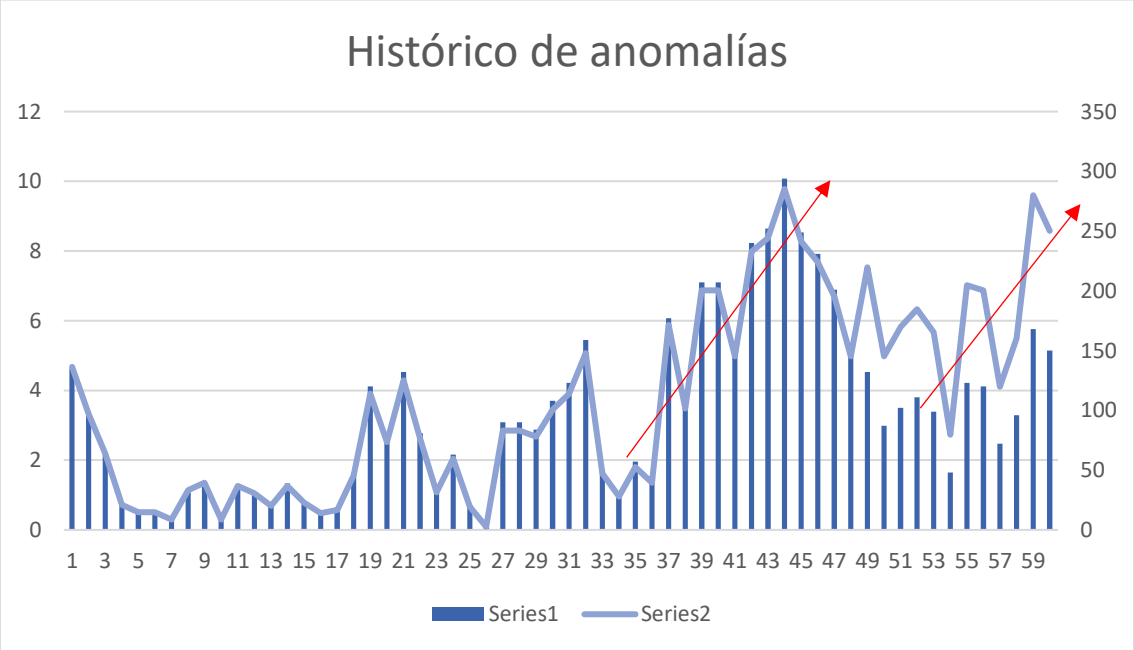


Ilustración 139. Líneas de tendencia detectadas en la multiplicadora

Se puede identificar que la tendencia estacional de aparición de valores fuera de rango es en verano e invierno de los últimos años.

2				3				4				5				6			
1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12

Tabla 72. Incidencia estacional de fallo en la multiplicadora

7.5.- Modelo de previsión de la temperatura de del rodamiento DE

Se calculará la media de las proporciones de anomalías mensuales de los años estudiados. Cuando se sobrepase la misma se considerará que la máquina está siendo estresada por el modo de fallo de temperatura alta en **el rodamiento DE**.

$$\mu = 3{,}76$$

Por lo tanto, los meses que se ha marcado como existencia de un fallo son:

2				3				4				5				6			
1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12

Tabla 73. Incidencia por fallo en el rodamiento DE

Se calculará el tercer cuartil de las proporciones de anomalías mensuales de los años estudiados. Cuando se sobrepase la misma se considerará que la máquina corre riesgo de fallo que precisará un mantenimiento correctivo (reparación).

$$Q_3 = 5{,}17$$

Se puede identificar como techo razonable en función de las proporciones de anomalía observadas.

En el siguiente gráfico podemos encontrar los 12 meses del año, durante el transcurso de 6 años. Encontramos, superpuestos, un diagrama de barras y un gráfico lineal. Se puede observar en el diagrama de barras las anomalías acumuladas en un mes en concreto. El gráfico lineal sigue los valores de la proporción de anomalías detectadas en la temperatura de los anillos en función del total de muestras tomadas en ese mes.

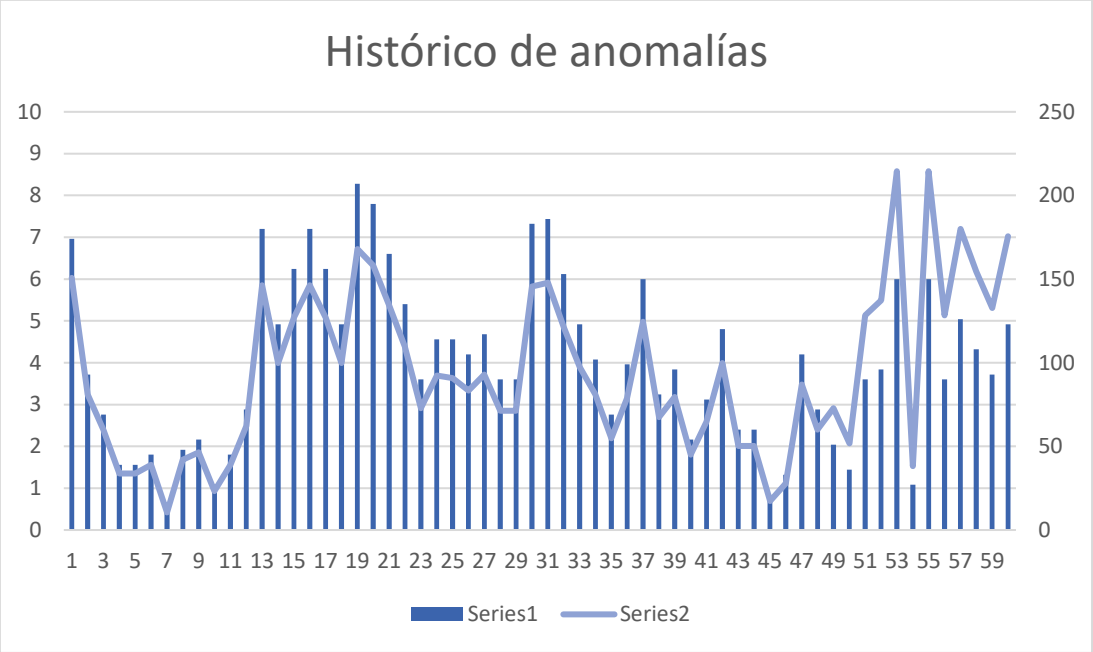


Ilustración 140. Histórico de anomalías detectadas en el rodamiento DE

Se pueden identificar claras líneas de tendencia sobre las que se podía haber actuado antes de alcanzar valores altos del indicador. Estas se encuentran marcadas en el diagrama inferior. Estas líneas de tendencia conducen a fallo, el cual se puede identificar claramente.

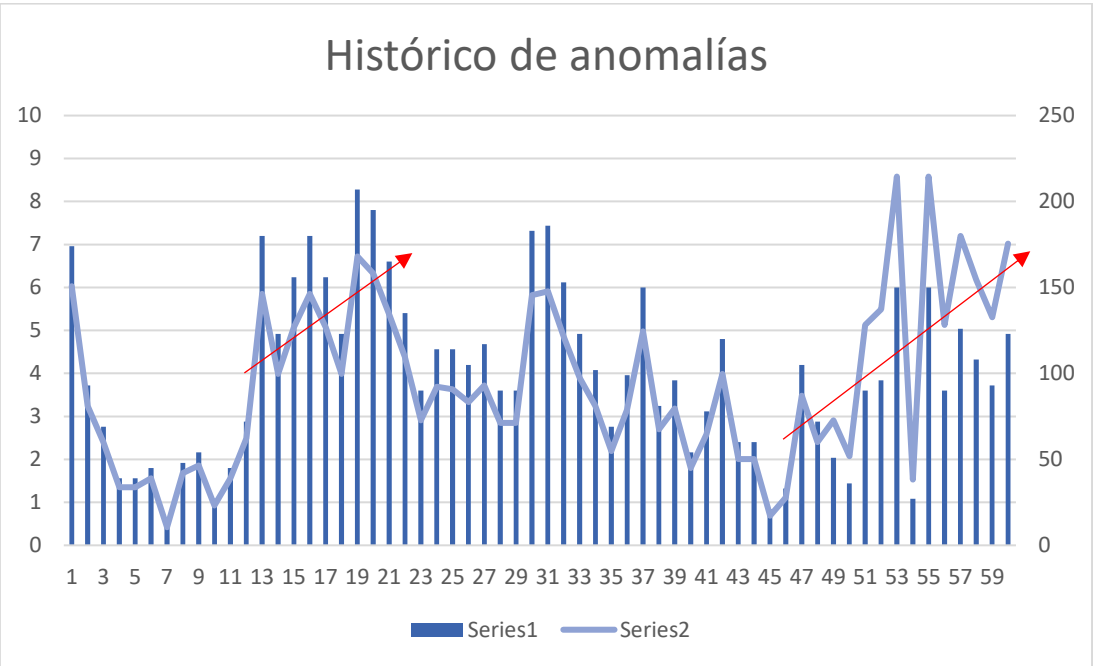


Ilustración 141. Líneas de tendencia detectadas en el rodamiento DE

Se puede identificar que la tendencia estacional de aparición de valores fuera de rango es en primavera y verano.

2				3				4				5				6			
1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12

Tabla 74. Incidencia estacional de fallo en el rodamiento DE

7.6.- Modelo de previsión de la temperatura de del rodamiento NDE

Se calculará la media de las proporciones de anomalías mensuales de los años estudiados. Cuando se sobrepase la misma se considerará que la máquina está siendo estresada por el modo de fallo de temperatura alta en **el rodamiento NDE**.

$$\mu = 5{,}32$$

Por lo tanto, los meses que se ha marcado como existencia de un fallo son:

2				3				4				5				6			
1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12	1 2 3	4 5 6	7 8 9	10 11 12

Tabla 75. Incidencia en el rodamiento NDE

Se calculará el tercer cuartil de las proporciones de anomalías mensuales de los años estudiados. Cuando se sobrepase la misma se considerará que la máquina corre riesgo de fallo que precisará un mantenimiento correctivo (reparación).

$$Q_3 = 7{,}04$$

Se puede identificar como techo razonable en función de las proporciones de anomalía observadas.

En el siguiente gráfico podemos encontrar los 12 meses del año, durante el transcurso de 6 años. Encontramos, superpuestos, un diagrama de barras y un gráfico lineal. Se puede observar en el diagrama de barras las anomalías acumuladas en un mes en concreto. El gráfico lineal sigue los valores de la proporción de anomalías detectadas en la temperatura de los anillos en función del total de muestras tomadas en ese mes.

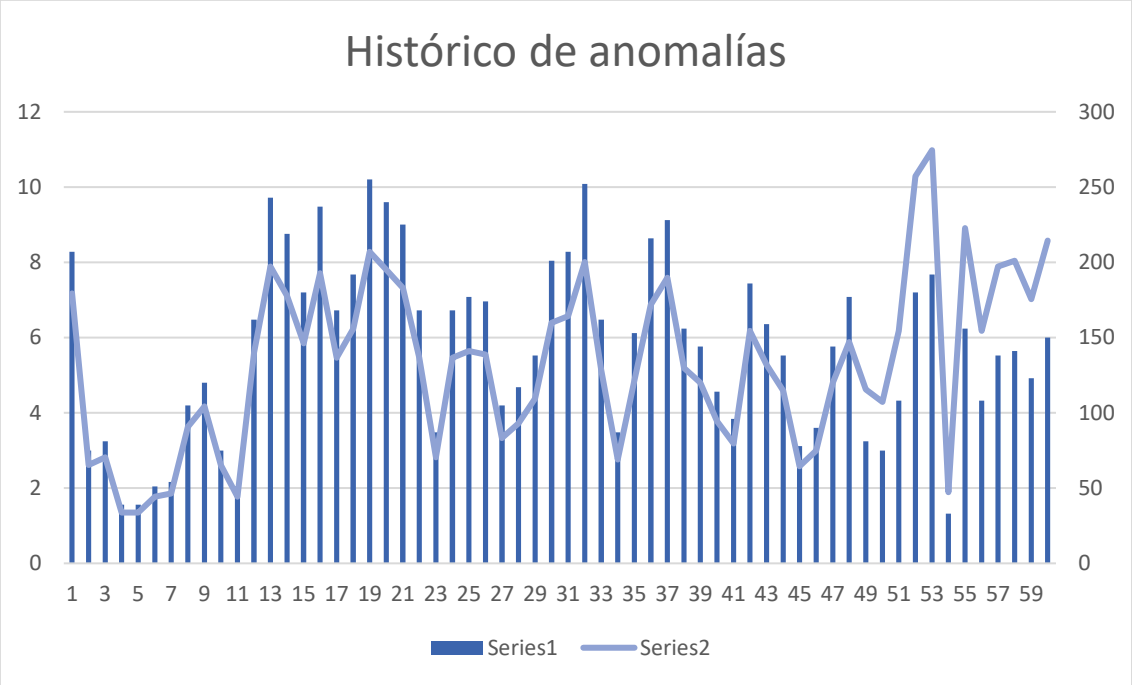


Ilustración 142. Histórico de anomalías detectadas en el rodamiento NDE

Se pueden identificar claras líneas de tendencia sobre las que se podía haber actuado antes de alcanzar valores altos del indicador. Estas se encuentran marcadas en el diagrama inferior. Estas líneas de tendencia conducen a fallo, el cual se puede identificar claramente.

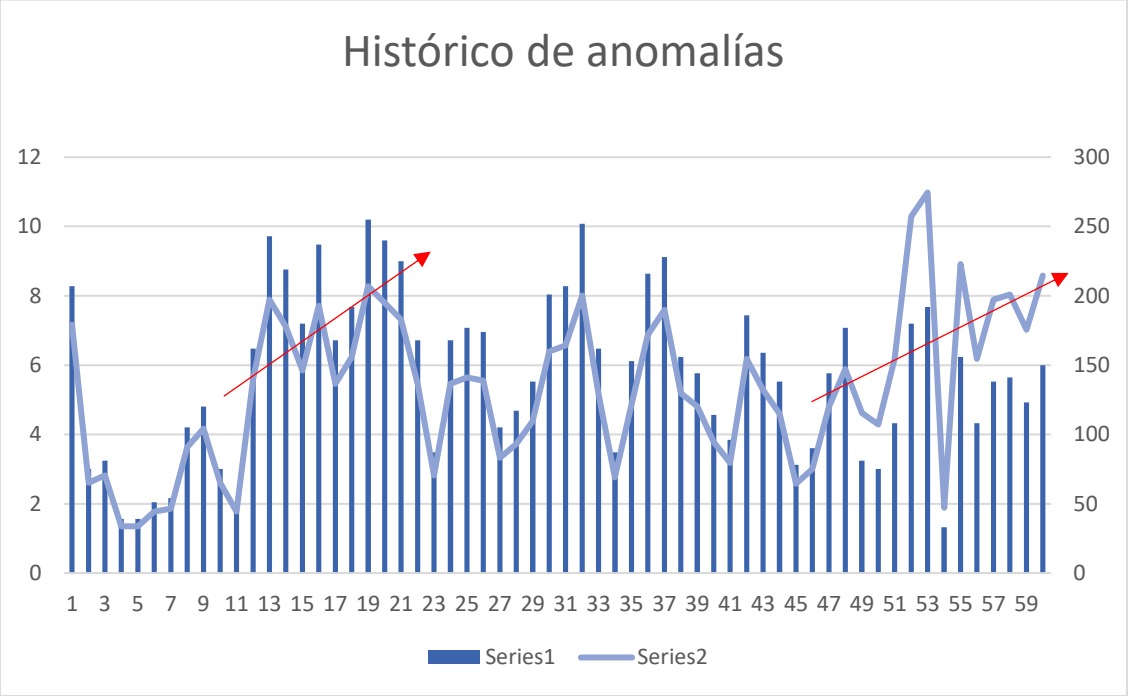


Ilustración 143. Líneas de tendencia detectadas en el rodamiento NDE

Se puede identificar que la tendencia estacional de aparición de valores fuera de rango es en primavera y verano.

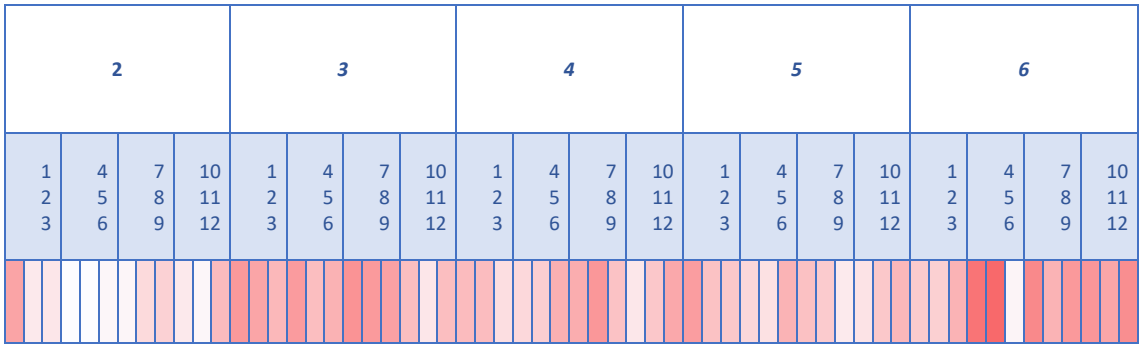


Tabla 76. Incidencia estacional de fallo en el rodamiento NDE

Capítulo 8.- Conclusiones y futuros desarrollos

8.1.- Conclusiones

Tras trabajar con el modelo analizando todos los casos, hemos trabajado tanto para obtener los resultados como para obtener varios modelos de comportamiento de un aerogenerador, capaces de dar una previsión de valores para la variable que monitorizan. Por lo tanto, se va a concluir no solo sobre los resultados, sino también sobre la técnica utilizada y sobre la metodología desarrollada.

1. Conclusiones sobre la técnica:

   La técnica empleada para obtener la previsión de los valores de la variable objetivo han sido las redes neuronales de convolución. El resultado ha sido bueno, los modelos obtenidos han resultado robustos y los errores de precisión en las predicciones no eran relevantes.

   Es interesante observar que, al tratarse de entradas numéricas y ser un número de variables de entrada limitadas, y la profundidad de la red neuronal no afectaba razonablemente a la robustez del modelo. Esto es, no por incorporar más capas o más filtros por capa, se obtenían mejores resultados de previsión. Gracias a la poca profundidad de la red neuronal, los tiempos de ejecución de los modelos han sido mucho más reducidos, a pesar del gran volumen de datos que se ha utilizado para el entrenamiento.

   Con respecto al lenguaje de programación R y al entorno gráfico de Rstudio, ofrece muy buenas prestaciones para el cálculo estadístico. Esto es algo que puede resultar muy beneficioso en tareas de análisis de datos en general, y en tareas de mantenimiento predictivo en particular.

   Ha resultado interesante trabajar con un lenguaje de programación abierto, en el que los paquetes pueden ser generados por cualquiera y se pueden intercambiar, cargar y ejecutar. Esto ha sido de utilidad puesto que, para realizar una tarea, era posible encontrar varias formas alternativas/funciones de código distintas en la red. Tan solo era necesario cargar el paquete que albergaba la lógica de las mismas.

   Sin embargo, el entorno de Rstudio aun deja mucho que desear con respecto a las prestaciones gráficas que ofrece. Los gráficos no son de buena calidad, no se puede operar sobre los mismos y no permite realizar simulaciones, tan solo generar imágenes. Matlab, con el que se ha trabajado al principio del proyecto, si reúne estas características.

1. Conclusiones sobre la metodología:

   La metodología empleada ha consistido en generar unos modelos de comportamiento. A partir de los mismos, comparando las previsiones de las variables con los valores medidos, identificar la presencia de una anomalía. A través de un análisis de las anomalías contabilizadas, definir el tipo de fallo que aparece y la criticidad del mismo.

   Para la generación de los modelos de comportamiento, ha sido necesario realizar un tratamiento previo de los datos. Este tratamiento previo se ha realizado con Matlab, lo cual ha sido una decisión acertada. Puesto que las capacidades gráficas de Matlab son mejores que las de R, la visualización de los resultados ha sido más amigable. Además, como los datos originales eran dados en formato Matlab, había que exportarlos desde este programa igualmente.

   Una vez importados los datos en R, se definen los objetivos, estrategia y utilidad de cada modelo y las variables que se emplearán en los mismos. Se genera un primer modelo, se entrena y se optimiza. La optimización de los modelos de red neuronal en R no es especialmente intuitiva, se lleva a cabo a través de un análisis empírico, cuantificando la función de coste/error medio absoluto del modelo.

   Una vez generado el modelo de comportamiento base, se adapta y se generan los demás a partir de él, empleando la misma estrategia de optimización. Se genera un script capaz de comprobar mediante un análisis estadístico la robustez de un modelo.

   Este paso intermedio de análisis de la robustez de un modelo es un paso especialmente importante, sienta los cimientos de la exactitud de las previsiones venideras. Merece la pena realizar un estudio estadístico detallado para justificar que cada uno de los modelos generados es robusto. Además, puede permitir identificar fallos en el código o variables de entrada a los modelos que, aunque inicialmente parecían significativas, en realidad no lo son. En nuestro caso, por ejemplo, para el modelo de previsión de la temperatura de la multiplicadora, la temperatura en la góndola resultó más significativa que la temperatura ambiente, por lo que esta variable de entrada al modelo tuvo que modificarse.

   Una vez se ha demostrado que se está trabajando con modelos robustos, se puede decir que los modelos están 'construidos'. En este momento se ha definido la estrategia de detección de anomalías. Se ha generado un script capaz de detectarlas y cuantificarlas de manera anual y mensual.

   En este punto, ha sido importante identificar, de las muestras anómalas cuantificadas, cuales se deben considerar anomalías y cuales se deben considerar errores de previsión o medida. Esto aporta precisión a los resultados y permite reducir el tamaño del vector de anomalías, y por tanto, el espacio de memoria permanente que ocupara su registro.

   Por último, conociendo las cuantificaciones mensuales y anuales de anomalías aportadas por cada uno de los modelos, ha sido posible identificar el tipo de fallo, el momento de aparición fallo, y el ciclo de vida del fallo a lo largo de los meses.

1. Conclusiones sobre los resultados:

   Tras trabajar con el modelo analizando todos los casos, hemos podido obtener diferentes resultados para cada uno de los modelos y para cada uno de los meses del año.

   La obtención de resultados se ha basado en la cuantificación de las anomalías que identificaba un modelo en un mes concreto. Para tener eliminar el sesgo generado por medidas erróneas, reducción de outliers, medidas vacías… se ha calculado el porcentaje mensual de medidas anómalas en función de las medidas totales tomadas en ese mes en concreto. Si ese porcentaje sobrepasaba un cierto valor, se notificaba la existencia de fallo; y si alcanzaba un cierto techo, el fallo total era inminente y se debía actuar sobre el elemento de fallo lo antes posible.

   El análisis mensual ha resultado de utilidad para conocer los picos de fallo, el carácter cíclico de algunos fallos y las líneas de tendencia que aparecían a lo largo del año.

   Es curioso ver como los fallos no se presentan aislados, sino que suelen ser producto de una línea de tendencia ascendente de porcentaje de anomalías sobre la que no se ha actuado.

   Adicionalmente, los fallos tienen una tendencia estacional: debido a que el fallo se conoce mediante una indicación de temperatura fuera de rango, y los fallos se producen con más asiduidad bajo altas temperaturas, no es de sorprender que los fallos generalmente aparezcan en las estaciones de primavera y verano.

   En términos generales, los resultados de los modelos son concluyentes y soportan la utilidad de la aplicación de estos modelos a un aerogenerador del que se toman medidas en tiempo real. El objetivo último es demostrar que incorporar estos modelos en las técnicas de mantenimiento predictivo de un parque eólico es eficaz y práctico, tanto en ahorro de costes de reparación como en ahorro de tiempos de planificación de tareas de mantenimiento, disminución de pérdidas por bajo rendimiento de un aerogenerador, identificación de las limitaciones de la maquinaria…

8.2.- Recomendaciones para futuros estudios

En este apartado queremos abrir puntos de interés y que serían atractivos para continuar con su desarrollo en futuros estudios.

1. Realizar una comparativa de los resultados que se obtendrían con un perceptrón multicapa clásico (ANN), en contraposición a los resultados obtenidos con nuestra red neuronal convolucional. El objetivo sería determinar qué tipo de modelo es más sensible, que tipo de modelo es más robusto, los tiempos de ejecución de cada uno, la precisión de las predicciones…

2. Definir directrices y proceso de optimización de una red neuronal de convolución utilizada para predicción de valores numéricos. De qué depende obtener un mejor resultado en función de las capas, los parámetros de una capa y el tipo de capas.

3. Profundizar en el proceso de detección de anomalías en tiempo real, aplicando el modelo a un caso de monitorización propio de la técnica del mantenimiento predictivo. Modificar el modelo convenientemente para que reciba medidas de los sensores en tiempo real y detecte líneas de tendencia de las anomalías que puedan conducir a uno de los modos de fallo vistos.

Capítulo 9.- Bibliografía y referencias

Amruthnath, N. (10 de 6 de 2020). *Anomaly Detection for Predictive Maintenance using Keras*. Obtenido de R bloggers: https://www.r-bloggers.com/anomaly-detection-for-predictive-maintenance-using-keras/

Cerliani, M. (2019 de 4 de 23). *Remaining Life Estimation with Keras*. Obtenido de Towards Data Science: https://towardsdatascience.com/remaining-life-estimation-with-keras-2334514f9c61

elEconomista Energía. (30 de 6 de 2015). *Bajan los costes de mantenimiento de los parques eólicos*. Obtenido de elEconomista: https://www.eleconomista.es/energia/noticias/6832985/06/15/Bajan-los-costes-de-mantenimiento-de-los-parques-eolicos.html

Keijzer, B. d. (11 de 1 de 2019). *GitHub*. Obtenido de CNN: https://github.com/deKeijzer/Multivariate-time-series-models-in-Keras/blob/master/notebooks/5.2%20CNN.ipynb

Keijzer, B. d. (s.f.). *GitHub*. Obtenido de Multivariate-time-series-models-in-Keras: https://github.com/deKeijzer/Multivariate-time-series-models-in-Keras

Keijzer, B. d. (s.f.). *Multivariate-time-series-models-in-Keras*. Obtenido de GitHub: https://github.com/deKeijzer/Multivariate-time-series-models-in-Keras

Lee, J.-H. (7 de 2019). *Fault Diagnosis of Induction Motor Using Convolutional Neural Network*. Obtenido de ResearchGate: https://www.researchgate.net/publication/334664585_Fault_Diagnosis_of_Induction_Motor_Using_Convolutional_Neural_Network

letYourMoneyGrow.com. (27 de 5 de 2018). *Serving Retail Investors*. Obtenido de letYourMoneyGrow.com: https://letyourmoneygrow.com/2018/05/27/classifying-time-series-with-keras-in-r-a-step-by-step-example/

Libro de nociones de estadística. (2013). En R. C. Carretero, *ESTADÍSTICA* (pág. 336). Pamplona : S.L. CIVITAS EDICIONES.

Rodrigues, V. B. (19 de 5 de 2019). *Concrete-Crack-Classification-Model*. Obtenido de GitHub: https://github.com/vbrodrigues/Concrete-Crack-Classification-Model

Rodrigues, V. B. (19 de 5 de 2019). *GitHub*. Obtenido de Concrete-Crack-Classification-Model: https://github.com/vbrodrigues/Concrete-Crack-Classification-Model

The TensorFlow Authors and RStudio, PBC. (2015). *Image Classification*. Obtenido de TensorFlow for R: https://tensorflow.rstudio.com/tutorials/beginners/basic-ml/tutorial_basic_classification/

The TensorFlow Authors and RStudio, PBC. (2015). *TensorFlow for R*. Obtenido de Image Classification: https://tensorflow.rstudio.com/tutorials/beginners/basic-ml/tutorial_basic_classification/

The TensorFlow Authors and RStudio, PBC. (2015). *Basic Regression*. Obtenido de The TensorFlow: https://tensorflow.rstudio.com/tutorials/beginners/basic-ml/tutorial_basic_regression/

Tribunal de Cuentas Europeo. (28 de 8 de 2019). *Digitising European industry*. Obtenido de Tribunal de Cuentas Europeo: https://www.eca.europa.eu/es/Pages/NewsItem.aspx?nid=12509#:~:text=En%202016%2C%20la%20Comisi%C3%B3n%20puso%20en%20marcha%20la,reciente%20que%20necesitan%20para%20acometer%20su%20transformaci%C3%B3n%20digital.

Capítulo 10.- Anexos

10.1.- Objetivos de Desarrollo Sostenible

El objetivo último del proyecto es desarrollar un modelo de diagnóstico de fallos en un aerogenerador a través de la detección de anomalías. Por lo tanto, el campo de diagnósticos de fallo en procesos industriales es el campo en el que se debe situar este proyecto.

De cara a los objetivos de desarrollo sostenible, este proyecto aporta:

1. Innovación de la técnica: soporte digital en la monitorización de la maquinaria para planificación de las tareas de mantenimiento.
2. Reducción el consumo mediante la reducción de los elementos de reparación.
3. Disminución de costes en el sector energético renovable a través de una reducción del presupuesto de mantenimiento y la parada programada de los aerogeneradores por incidencia de fallo.

Por lo tanto, a partir de estas premisas, se va a abarcar los siguientes objetivos de desarrollo sostenible:

Dimensión ODS	Identificador ODS	Rol	Aportaciones
Económica	ODS 9: Industria, innovación e infraestructura	Primario	Innovación de la técnica: soporte digital en la monitorización de la maquinaria para planificación de las tareas de mantenimiento.
Económica	ODS 12: Consumo y producción responsable	Secundario	Reducción el consumo mediante la reducción de los elementos de reparación.
Social	ODS 9: Energía limpia y asequible	Secundario	Disminución de costes en el sector energético renovable a través de una reducción del presupuesto de mantenimiento y la parada programada de los aerogeneradores por incidencia de fallo

Por lo tanto, con una visión mas ambiciosa, podemos trasladar estas aportaciones del proyecto de un de un nivel particular a un nivel mas alto de objetivos a largo plazo.

Dimensión ODS	Identificador ODS	Rol	Objetivo
Económica	ODS 9: Industria, innovación e infraestructura	Primario	Para 2030, haber incorporado una transformación digital en todo entorno industrial para la monitorización, control y planificación de los procesos

Profundizando más en el objetivo propuesto, hemos encontrado información que puede resultar de interés, pues justifica la importancia de la consecución del objetivo fijado y el apoyo al mismo por parte de la comunidad europea.

"En los últimos años, la Comisión y los Estados miembros han introducido una serie de iniciativas para crear un mercado único digital y digitalizar la industria europea, mediante acciones en el ámbito de la economía de los datos, la internet de las cosas y la computación en la nube. En 2016, la Comisión puso en marcha la iniciativa «Digitalización de la industria europea», que estimuló la creación de centros de innovación digital."

(Tribunal de Cuentas Europeo, 2019)

Para cuantificar el impacto sobre la contribución a los Objetivos de Desarrollo sostenible, se han realizado las siguientes estimaciones:

1. Del 2009 al 2014, mediante la inclusión de las técnicas de mantenimiento predictivo, los costes de mantenimiento en parques eólicos han bajado de 30.000€ el MW a 20.000€ el MW. (elEconomista Energía, 2015)

2. Para una potencia instalada en España de 25.000 MW en energía eólica, como el ahorro sería de 10.000€ por MW, se calcula que solo incorporando esta tecnología a 2 aerogeneradores por parque (suponiendo que un parque tiene 20 aerogeneradores), supondría una reducción de 25M de euros del gasto español en energía eólica.

3. Teniendo en cuenta que estos 25M se podrían reinvertir en energía eólica renovable, las emisiones de carbono (315 tons registradas en 2019) se verían progresivamente reducidas.

"Los avances obtenidos en el mantenimiento predictivo nos permiten una mejor planificación de las tareas de mantenimiento, incrementando la disponibilidad y detectando los fallos en un estado incipiente, lo que permite acometer una reparación más sencilla y barata, apuntan desde Acciona Energía. La utilización de las nuevas tecnologías ha ocasionado mejoras importantes en la gestión de la información, necesaria para optimizar el mantenimiento.

Los mayores avances vienen por la parte de los diagnósticos, condition monitoring, es decir, la instrumentación para adelantarse al propio fallo, lo que facilita el trabajo y abarata los costes."

(elEconomista Energía, 2015)

10.2.- Script Matlab

```
%% Carga de los datos recogidos
load('_ANG_alarms.mat')
load('ANG_analog_short_2.mat')
load('repara_ANG.mat')

%% CREACIÓN VECTORES
xlswrite('Alarmas_ANG.xlsx',Alarmas_ANG,'Hoja1','A1');
xlswrite('Repara_ANG.xlsx',repara_ANG,'Hoja1','A1');

ANG_fecha=ANG.fecha;
% xlswrite('ANG.xlsx',ANG_fecha,'Hoja1','A1');
% xlswrite('ANG.xlsx',ANG.fecha,'Hoja1','B2');

% ANG.aero.tem_aceite_multi.val
ANG_aero_tem_aceite_multi_val=ANG.aero.tem_aceite_multi.val;
% xlswrite('ANG.xlsx',ANG.aero.tem_aceite_multi.val,'Hoja1','C2');

%ANG.aero.tem_rod_alta_multi.val
ANG_aero_tem_rod_alta_multi_val=ANG.aero.tem_rod_alta_multi.val;
% xlswrite('ANG.xlsx',ANG.aero.tem_rod_alta_multi.val,'Hoja1','D2');

%ANG.aero.tem_rod_DE_gen.val
ANG_aero_tem_rod_DE_gen_val=ANG.aero.tem_rod_DE_gen.val;
% xlswrite('ANG.xlsx',ANG.aero.tem_rod_DE_gen.val,'Hoja1','E2');

%ANG.aero.tem_rod_NDE_gen.val
ANG_aero_tem_rod_NDE_gen_val=ANG.aero.tem_rod_NDE_gen.val;
% xlswrite('ANG.xlsx',ANG.aero.tem_rod_NDE_gen.val,'Hoja1','F2');

%ANG.aero.tem_anillos_sld_gen.val
ANG_aero_tem_anillos_sld_gen_val=ANG.aero.tem_anillos_sld_gen.val;
% xlswrite('ANG.xlsx',ANG.aero.tem_anillos_sld_gen.val,'Hoja1','G2');

%ANG.aero.tem_max_dev_gen.val
ANG_aero_tem_max_dev_gen_val=ANG.aero.tem_max_dev_gen.val;
% xlswrite('ANG.xlsx',ANG.aero.tem_max_dev_gen.val,'Hoja1','H2');

%ANG.aero.tem_dev_A_gen.val
ANG_aero_tem_dev_A_gen_val=ANG.aero.tem_dev_A_gen.val;
% xlswrite('ANG.xlsx',ANG.aero.tem_dev_A_gen.val,'Hoja1','I2');

%ANG.aero.tem_dev_B_gen.val
ANG_aero_tem_dev_B_gen_val=ANG.aero.tem_dev_B_gen.val;
% xlswrite('ANG.xlsx',ANG.aero.tem_dev_B_gen.val,'Hoja1','J2');

%ANG.aero.tem_dev_C_gen.val
ANG_aero_tem_dev_C_gen_val=ANG.aero.tem_dev_C_gen.val;
% xlswrite('ANG.xlsx',ANG.aero.tem_dev_C_gen.val,'Hoja1','K2');

%ANG.aero.tem_grupo_h.val
ANG_aero_tem_grupo_h_val=ANG.aero.tem_grupo_h.val;
% xlswrite('ANG.xlsx',ANG.aero.tem_grupo_h.val,'Hoja1','L2');

%ANG.aero.pres_grupo_h.val
ANG_aero_pres_grupo_h_val=ANG.aero.pres_grupo_h.val;
% xlswrite('ANG.xlsx',ANG.aero.pres_grupo_h.val,'Hoja1','M2');

%ANG.aero.pres_yaw.val
ANG_aero_pres_yaw_val=ANG.aero.pres_yaw.val;
% xlswrite('ANG.xlsx',ANG.aero.pres_yaw.val,'Hoja1','N2');

%ANG.aero.tem_interior_nac.val
ANG_aero_tem_interior_nac_val=ANG.aero.tem_interior_nac.val;
% xlswrite('ANG.xlsx',ANG.aero.tem_interior_nac.val,'Hoja1','O2');

%ANG.aero.tem_ambiente.val
ANG_aero_tem_ambiente_val=ANG.aero.tem_ambiente.val;
% xlswrite('ANG.xlsx',ANG.aero.tem_ambiente.val,'Hoja1','P2');

%ANG.aero.tem_max_trafo.val
ANG_aero_tem_max_trafo_val=ANG.aero.tem_max_trafo.val;
% xlswrite('ANG.xlsx',ANG.aero.tem_max_trafo.val,'Hoja1','Q2');
```

```
%ANG.aero.tem_dev_A_trafo.val
ANG_aero_tem_dev_A_trafo_val=ANG.aero.tem_dev_A_trafo.val;
% xlswrite('ANG.xlsx',ANG.aero.tem_dev_A_trafo.val,'Hoja1','R2');

%ANG.aero.tem_dev_B_trafo.val
ANG_aero_tem_dev_B_trafo_val=ANG.aero.tem_dev_B_trafo.val;
% xlswrite('ANG.xlsx',ANG.aero.tem_dev_B_trafo.val,'Hoja1','S2');

%ANG.aero.tem_dev_C_trafo.val
ANG_aero_tem_dev_C_trafo_val=ANG.aero.tem_dev_C_trafo.val;
% xlswrite('ANG.xlsx',ANG.aero.tem_dev_C_trafo.val,'Hoja1','T2');

%ANG.aero.W_eje_principal.val
ANG_aero_W_eje_principal_val=ANG.aero.W_eje_principal.val;
% xlswrite('ANG.xlsx',ANG.aero.W_eje_principal.val,'Hoja1','U2');

%ANG.aero.w_gen.val
ANG_aero_w_gen_val=ANG.aero.w_gen.val;
% xlswrite('ANG.xlsx',ANG.aero.w_gen.val,'Hoja1','V2');

%ANG.aero.angulo_palas.val
ANG_aero_angulo_palas_val=ANG.aero.angulo_palas.val;
% xlswrite('ANG.xlsx',ANG.aero.angulo_palas.val,'Hoja1','W2');

%ANG.aero.vel_viento.val
ANG_aero_vel_viento_val=ANG.aero.vel_viento.val;
% xlswrite('ANG.xlsx',ANG.aero.vel_viento.val,'Hoja1','X2');

%ANG.aero.angulo_viento.val
ANG_aero_angulo_viento_val=ANG.aero.angulo_viento.val;
% xlswrite('ANG.xlsx',ANG.aero.angulo_viento.val,'Hoja1','Y2');

%ANG.aero.angulo_yaw.val
ANG_aero_angulo_yaw_val=ANG.aero.angulo_yaw.val;
% xlswrite('ANG.xlsx',ANG.aero.angulo_yaw.val,'Hoja1','Z2');

%ANG.aero.V_red.val
ANG_aero_V_red_val=ANG.aero.V_red.val;
% xlswrite('ANG.xlsx',ANG.aero.V_red.val,'Hoja1','AA2');

%ANG.aero.pot_reactiva.val
ANG_aero_pot_reactiva_val=ANG.aero.pot_reactiva.val;
% xlswrite('ANG.xlsx',ANG.aero.pot_reactiva.val,'Hoja1','AB2');

%ANG.aero.factor_potencia.val
ANG_aero_factor_potencia_val=ANG.aero.factor_potencia.val;
% xlswrite('ANG.xlsx',ANG.aero.factor_potencia.val,'Hoja1','AC2');

%ANG.aero.pot_act_est.val
ANG_aero_pot_act_est_val=ANG.aero.pot_act_est.val;
% xlswrite('ANG.xlsx',ANG.aero.pot_act_est.val,'Hoja1','AD2');

%ANG.aero.pot_act_rot.val
ANG_aero_pot_act_rot_val=ANG.aero.pot_act_rot.val;
% xlswrite('ANG.xlsx',ANG.aero.pot_act_rot.val,'Hoja1','AE2');

%ANG.aero.pot_total.val
ANG_aero_pot_total_val=ANG.aero.pot_total.val;
% xlswrite('ANG.xlsx',ANG.aero.pot_total.val,'Hoja1','AF2');

%ANG.aero.Energia.val
ANG_aero_Energia_val=ANG.aero.Energia.val;
% xlswrite('ANG.xlsx',ANG.aero.Energia.val,'Hoja1','AG2');

%ANG.aero.state.avg
ANG_aero_state_avg=ANG.aero.state.avg;
% xlswrite('ANG.xlsx',ANG.aero.state.avg,'Hoja1','AH2');



%% CREACIÓN MATRIZ FILTRADA

%Matriz reducida, solo datos transcendentes
ANG=[ANG_aero_tem_aceite_multi_val,ANG_aero_tem_rod_alta_multi_val,ANG_aero_tem_rod_DE_g
en_val,ANG_aero_tem_rod_NDE_gen_val,ANG_aero_tem_anillos_sld_gen_val,ANG_aero_tem_dev_A_
gen_val,ANG_aero_tem_dev_B_gen_val,ANG_aero_tem_dev_C_gen_val,ANG_aero_tem_grupo_h_val,A
```

```
NG_aero_pres_grupo_h_val,ANG_aero_pres_yaw_val,ANG_aero_tem_interior_nac_val,ANG_aero_te
m_ambiente_val,ANG_aero_W_eje_principal_val,ANG_aero_w_gen_val,ANG_aero_angulo_palas_val
,ANG_aero_vel_viento_val,ANG_aero_angulo_viento_val,ANG_aero_angulo_yaw_val,ANG_aero_pot
_total_val];
ANG_titulos={'ANG_aero_tem_aceite_multi_val','ANG_aero_tem_rod_alta_multi_val','ANG_aero
_tem_rod_DE_gen_val','ANG_aero_tem_rod_NDE_gen_val','ANG_aero_tem_anillos_sld_gen_val','
ANG_aero_tem_dev_A_gen_val','ANG_aero_tem_dev_B_gen_val','ANG_aero_tem_dev_C_gen_val','A
NG_aero_tem_grupo_h_val','ANG_aero_pres_grupo_h_val','ANG_aero_pres_yaw_val','ANG_aero_t
em_interior_nac_val','ANG_aero_tem_ambiente_val','ANG_aero_W_eje_principal_val','ANG_aer
o_w_gen_val','ANG_aero_angulo_palas_val','ANG_aero_vel_viento_val','ANG_aero_angulo_vien
to_val','ANG_aero_angulo_yaw_val','ANG_aero_pot_total_val'};

M = mean(ANG,'omitnan'); %Calcula la media pero sin tener en cuenta NaN
S = std(ANG,'omitnan'); %Calcula la desviación estandar pero sin tener en cuenta NaN

%1. Filtrado de las filas con outliers
ANG_filtro_1 = [];
a=1;
res=0;

[m,n] = size(ANG);
for i= 1:m
    for j= 1:n
         if (ANG(i,j)<(M(j)-2*S(j)) || ANG(i,j)>(M(j)+2*S(j)))
            res=1;
         end
    end
    if res==0
       for j= 1:20
        ANG_filtro_1(a,j) = ANG(i,j);
       end
       ANG_fecha_filtro_1(a)=ANG_fecha(i); %Quitar tambien la fila en la columna de
fecha
        a=a+1;
    else
      res=0;
    end

end



%2. Filtrado de columnas NaN (>60%)
ANG_filtro_2 = [];
a=1;
count=0;
str = ['Se han eliminado las columnas:']; %Muestra este mensaje en la consola
disp (str) %Muestra este mensaje en la consola

[m,n] = size(ANG_filtro_1);
for j= 1:n
    for i= 1:m
         if isnan(ANG_filtro_1(i,j))  %True(se mete)cuando encuentra un NaN
            count=count+1;
         end
    end
    if (count/m)<0.6
        for i= 1:m
        ANG_filtro_2(i,a) = ANG_filtro_1(i,j);
        end
        ANG_titulos_filtro(a)=ANG_titulos(j); %Quitar tambien la columna en la fila de
títulos
        a=a+1;
        count=0;
    else
        disp (j) %Muestra este mensaje en la consola
        disp (ANG_titulos(j)) %Muestra este mensaje en la consola
        count=0;
    end

end


%Filtrado de filas NaN (>3)
ANG_filtro_3 = [];
a=1;
```

```
count=0;

[m,n] = size(ANG_filtro_2);
for i= 1:m
    for j= 1:n
         if isnan(ANG_filtro_2(i,j))
            count=count+1;
         end
    end
    if count<3
       for j= 1:n
        ANG_filtro_3(a,j) = ANG_filtro_2(i,j);
       end
       ANG_fecha_filtro_3(a)=ANG_fecha_filtro_1(i); %Quitar tambien la fila en la
columna de fecha
        a=a+1;
        count=0;
    else
      count=0;
    end

end

%% GUARDAR/CARGAR RESULTADOS
%save Results; %Guarda el workspace
%load ('Results'); %Carga el workspace guardado anteriormente

%% ESCRITURA MATRIZ
ANG_fecha_formato= datetime(ANG_fecha_filtro_3,'ConvertFrom','datenum');
xlswrite('ANG_filtrada.xlsx',{'ANG_fecha'},'Hoja1','A1');
xlswrite('ANG_filtrada.xlsx',ANG_titulos_filtro,'Hoja1','B1');
xlswrite('ANG_filtrada.xlsx',ANG_fecha_filtro_3','Hoja1','A2'); %Transpuesta para que
sean datos en una columna
%xlswrite('ANG_filtrada.xlsx',ANG_fecha_formato,'Hoja1','A2'); %No funciona
xlswrite('ANG_filtrada.xlsx',ANG_filtro_3,'Hoja1','B2');

%% GRAFICOS
%Diagrama sectores % NaN
B = categorical(ANG_filtro_1(:,1),[NaN],{'NaN'});
True = isundefined(B); %Pone a 1 los valores indefinidos y a 0 los valores definidos
B = categorical(True,[0,1],{'NaN','Real Value'});
figure()
pie(B); title(ANG_titulos(1))

B = categorical(ANG_filtro_1(:,2),[NaN],{'NaN'});
True = isundefined(B);
B = categorical(True,[0,1],{'NaN','Real Value'});
figure()
pie(B); title(ANG_titulos(2))

B = categorical(ANG_filtro_1(:,3),[NaN],{'NaN'});
True = isundefined(B);
B = categorical(True,[0,1],{'NaN','Real Value'});
figure()
pie(B); title(ANG_titulos(3))

B = categorical(ANG_filtro_1(:,4),[NaN],{'NaN'});
True = isundefined(B);
B = categorical(True,[0,1],{'NaN','Real Value'});
figure()
pie(B); title(ANG_titulos(4))

B = categorical(ANG_filtro_1(:,5),[NaN],{'NaN'});
True = isundefined(B);
B = categorical(True,[0,1],{'NaN','Real Value'});
figure()
pie(B); title(ANG_titulos(5))

B = categorical(ANG_filtro_1(:,6),[NaN],{'NaN'});
True = isundefined(B);
B = categorical(True,[0,1],{'NaN','Real Value'});
figure()
pie(B); title(ANG_titulos(6))

B = categorical(ANG_filtro_1(:,7),[NaN],{'NaN'});
True = isundefined(B);
```

```
B = categorical(True,[0,1],{'NaN','Real Value'});
figure()
pie(B); title(ANG_titulos(7))

B = categorical(ANG_filtro_1(:,8),[NaN],{'NaN'});
True = isundefined(B);
B = categorical(True,[0,1],{'NaN','Real Value'});
figure()
pie(B); title(ANG_titulos(8))

B = categorical(ANG_filtro_1(:,9),[NaN],{'NaN'});
True = isundefined(B);
B = categorical(True,[0,1],{'NaN','Real Value'});
figure()
pie(B); title(ANG_titulos(9))

B = categorical(ANG_filtro_1(:,10),[NaN],{'NaN'});
True = isundefined(B);
B = categorical(True,[0,1],{'NaN','Real Value'});
figure()
pie(B); title(ANG_titulos(10))

B = categorical(ANG_filtro_1(:,11),[NaN],{'NaN'});
True = isundefined(B);
B = categorical(True,[0,1],{'NaN','Real Value'});
figure()
pie(B); title(ANG_titulos(11))

B = categorical(ANG_filtro_1(:,12),[NaN],{'NaN'});
True = isundefined(B);
B = categorical(True,[0,1],{'NaN','Real Value'});
figure()
pie(B); title(ANG_titulos(12))

B = categorical(ANG_filtro_1(:,13),[NaN],{'NaN'});
True = isundefined(B);
B = categorical(True,[0,1],{'NaN','Real Value'});
figure()
pie(B); title(ANG_titulos(13))

B = categorical(ANG_filtro_1(:,14),[NaN],{'NaN'});
True = isundefined(B);
B = categorical(True,[0,1],{'NaN','Real Value'});
figure()
pie(B); title(ANG_titulos(14))

B = categorical(ANG_filtro_1(:,15),[NaN],{'NaN'});
True = isundefined(B);
B = categorical(True,[0,1],{'NaN','Real Value'});
figure()
pie(B); title(ANG_titulos(15))

B = categorical(ANG_filtro_1(:,16),[NaN],{'NaN'});
True = isundefined(B);
B = categorical(True,[0,1],{'NaN','Real Value'});
figure()
pie(B); title(ANG_titulos(16))

B = categorical(ANG_filtro_1(:,17),[NaN],{'NaN'});
True = isundefined(B);
B = categorical(True,[0,1],{'NaN','Real Value'});
figure()
pie(B); title(ANG_titulos(17))

B = categorical(ANG_filtro_1(:,18),[NaN],{'NaN'});
True = isundefined(B);
B = categorical(True,[0,1],{'NaN','Real Value'});
figure()
pie(B); title(ANG_titulos(18))

B = categorical(ANG_filtro_1(:,19),[NaN],{'NaN'});
True = isundefined(B);
B = categorical(True,[0,1],{'NaN','Real Value'});
figure()
pie(B); title(ANG_titulos(19))
```

```
B = categorical(ANG_filtro_1(:,20),[NaN],{'NaN'});
True = isundefined(B);
B = categorical(True,[0,1],{'NaN','Real Value'});
figure()
pie(B); title(ANG_titulos(20))

% Se eliminan las columnas 6, 7, 8 y 19

%% Limpieza de outliers (+-2*desv.tip) y filas NaN (>3 en una fila)
figure(1)
subplot(2,2,1);
plot(ANG(:,1))
hold on
plot(ANG_filtro_1(:,1))
hold on
plot(ANG_filtro_3(:,1))
title(ANG_titulos_filtro(1))

subplot(2,2,2);
plot(ANG(:,2))
hold on
plot(ANG_filtro_1(:,2))
hold on
plot(ANG_filtro_3(:,2))
title(ANG_titulos_filtro(2))

subplot(2,2,3);
plot(ANG(:,3))
hold on
plot(ANG_filtro_1(:,3))
hold on
plot(ANG_filtro_3(:,3))
title(ANG_titulos_filtro(3))

subplot(2,2,4);
plot(ANG(:,4))
hold on
plot(ANG_filtro_1(:,4))
hold on
plot(ANG_filtro_3(:,4))
title(ANG_titulos_filtro(4))

figure(2) %Pasamos a la siguiente figura para que no este muy apretado, y continuamos
por la columna 5
subplot(2,2,1);
plot(ANG(:,5))
hold on
plot(ANG_filtro_1(:,5))
hold on
plot(ANG_filtro_3(:,5))
title(ANG_titulos_filtro(5))

subplot(2,2,2);
plot(ANG(:,9)) % Se eliminan las columnas 6, 7, 8 ->6+3=9
hold on
plot(ANG_filtro_1(:,9)) % Se eliminan las columnas 6, 7, 8 ->6+3=9
hold on
plot(ANG_filtro_3(:,6))
title(ANG_titulos_filtro(6))

subplot(2,2,3);
plot(ANG(:,10)) % Se eliminan las columnas 6, 7, 8
hold on
plot(ANG_filtro_1(:,10)) % Se eliminan las columnas 6, 7, 8
hold on
plot(ANG_filtro_3(:,7))
title(ANG_titulos_filtro(7))

subplot(2,2,4);
plot(ANG(:,11)) % Se eliminan las columnas 6, 7, 8
hold on
plot(ANG_filtro_1(:,11)) % Se eliminan las columnas 6, 7, 8
hold on
plot(ANG_filtro_3(:,8))
title(ANG_titulos_filtro(8))
```

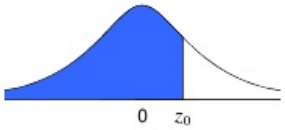
```
figure(3)
subplot(2,2,1);
plot(ANG(:,12)) % Se eliminan las columnas 6, 7, 8
hold on
plot(ANG_filtro_1(:,12)) % Se eliminan las columnas 6, 7, 8
hold on
plot(ANG_filtro_3(:,9))
title(ANG_titulos_filtro(9))

subplot(2,2,2);
plot(ANG(:,13)) % Se eliminan las columnas 6, 7, 8
hold on
plot(ANG_filtro_1(:,13)) % Se eliminan las columnas 6, 7, 8
hold on
plot(ANG_filtro_3(:,10))
title(ANG_titulos_filtro(10))

subplot(2,2,3);
plot(ANG(:,14)) % Se eliminan las columnas 6, 7, 8
hold on
plot(ANG_filtro_1(:,14)) % Se eliminan las columnas 6, 7, 8
hold on
plot(ANG_filtro_3(:,11))
title(ANG_titulos_filtro(11))

subplot(2,2,4);
plot(ANG(:,15)) % Se eliminan las columnas 6, 7, 8
hold on
plot(ANG_filtro_1(:,15)) % Se eliminan las columnas 6, 7, 8
hold on
plot(ANG_filtro_3(:,12))
title(ANG_titulos_filtro(12))

figure(4)
subplot(2,2,1);
plot(ANG(:,16)) % Se eliminan las columnas 6, 7, 8
hold on
plot(ANG_filtro_1(:,16)) % Se eliminan las columnas 6, 7, 8
hold on
plot(ANG_filtro_3(:,13))
title(ANG_titulos_filtro(13))

subplot(2,2,2);
plot(ANG(:,17)) % Se eliminan las columnas 6, 7, 8
hold on
plot(ANG_filtro_1(:,17)) % Se eliminan las columnas 6, 7, 8
hold on
plot(ANG_filtro_3(:,14))
title(ANG_titulos_filtro(14))

subplot(2,2,3);
plot(ANG(:,18)) % Se eliminan las columnas 6, 7, 8
hold on
plot(ANG_filtro_1(:,18)) % Se eliminan las columnas 6, 7, 8
hold on
plot(ANG_filtro_3(:,15))
title(ANG_titulos_filtro(15))

subplot(2,2,4);
plot(ANG(:,20)) % Se eliminan las columnas 6, 7, 8 y 19 ->16+4=20
hold on
plot(ANG_filtro_1(:,20)) % Se eliminan las columnas 6, 7, 8 y 19 ->16+4=20
hold on
plot(ANG_filtro_3(:,16))
title(ANG_titulos_filtro(16))
```

Tabla de la distribución normal N(0,1) para probabilidad acumulada inferior

μ = Media

σ = Desviación típica

$$P(z \le z_0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_0} e^{\frac{-z^2}{2}} dz$$



Tipificación: $z_0 = \frac{x - \mu}{\sigma}$

z_0	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	z_0
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359	**0,0**
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753	**0,1**
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141	**0,2**
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517	**0,3**
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879	**0,4**
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224	**0,5**
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549	**0,6**
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852	**0,7**
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133	**0,8**
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389	**0,9**
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621	**1,0**
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830	**1,1**
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015	**1,2**
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177	**1,3**
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319	**1,4**
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441	**1,5**
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545	**1,6**
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633	**1,7**
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706	**1,8**
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767	**1,9**
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817	**2,0**
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857	**2,1**
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890	**2,2**
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916	**2,3**
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936	**2,4**
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952	**2,5**
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964	**2,6**
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974	**2,7**
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981	**2,8**
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986	**2,9**
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900	**3,0**
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929	**3,1**
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950	**3,2**
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965	**3,3**
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976	**3,4**
3,5	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983	**3,5**
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989	**3,6**
3,7	0,99989	0,99990	0,99990	0,99990	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992	**3,7**
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995	**3,8**
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997	**3,9**

$1-\alpha$	90%	92%	94%	95%	96%	97%	98%	99%
α	10%	8%	6%	5%	4%	3%	2%	1%
$z_{\alpha/2}$	1,645	1,751	1,881	1,960	2,054	2,170	2,326	2,576
z_α	1,282	1,405	1,555	1,645	1,751	1,881	2,054	2,326

Siendo:

$1-\alpha$ = Nivel de confianza

α = Nivel de significación

www.vaxasoftware.com/indexes.html