



Universidad de Valladolid

FACULTAD DE CIENCIAS

TRABAJO FIN DE GRADO

Grado en Estadística

**DESARROLLO DE UN SENSOR SOFTWARE PARA LA ESTIMACIÓN
DE MATERIA SECA EN UN PROCESO INDUSTRIAL**

Autor:

Javier Perrote Gómez

Tutor/es/as:

Diego García Álvarez

José Belarmino Pulido Junquera

2023

Índice de Contenidos

Capítulo 1. Introducción	1
1.1 Objetivos	1
1.2 Estructura del Documento.....	2
Capítulo 2. Planificación.....	3
2.1 Metodología Usada	3
2.2 Presupuesto y Calendario del proyecto	3
2.2.1 Presupuesto en horas y calendario.....	3
2.2.2 Estimación y monitorización de los recursos a utilizar.....	3
2.2.3 Presupuesto total del proyecto.....	4
Capítulo 3. Contexto del Problema	5
3.1 Proceso de evaporación	5
3.2 Medición física del Brix	7
3.3 Medición Brix con sensores software.....	7
3.3.1 Regresión por mínimos cuadrados parciales.....	7
3.3.2 Red Neuronal Artificial	8
3.3.3 SVR	11
3.3.4 Random Forest Regression	14
3.3.5 Búsqueda y selección de hiperparámetros	16
Capítulo 4 . Metodología	18
4.1 CRISP-DM.....	18
4.2 Comprensión del proyecto	19
4.3 Comprensión de los datos	19
4.3.1 Captura Datos Iniciales	19
4.3.2 Descripción de los Datos	19
4.3.3 Exploración de los datos	22
4.3.4 Verificación de la calidad de los datos	26
4.4 Preparación de los datos	28
4.4.1 Limpieza de datos	29
4.4.2 Selección de atributos	31
4.5 Modelado y evaluación	32
Capítulo 5. Conclusiones y trabajo futuro	43
Capítulo 6 - Bibliografía.....	44
Anexo I - Funciones de Activación.....	46
Anexo II – Gráficos de control.....	47
Anexo III – Q-Q normal	51
Anexo IV – Boxplots.....	55
Anexo IV – Histogramas	59
Anexo V – Filtrado de Variables.....	63

Índice de figuras

Figura 1 Diagrama general del proceso de obtención del azúcar	6
Figura 2 Sistema evaporador de múltiple efecto de evaporación directa. Extraído de [1]	6
Figura 3 Estructura de una neurona. Adaptada de [2].....	9
Figura 4 Ejemplo de utilización SVR. Adaptado de [12].	12
Figura 5 Ejemplo de holgura en SVR. Adaptado de [12].	12
Figura 6 Ejemplo de utilización de un kernel no lineal. Adaptado de [12].	13
Figura 7 Estructura de un árbol de decisión con 4 regiones.	15
Figura 8 Ejemplo de predicción con Random Forest. Adaptado de [14]	16
Figura 9 Búsqueda de hiperparámetros Grid vs Random search. Adaptado de [16]	17
Figura 10 Fases del modelo CRISP-DM. Extraído de [17].	18
Figura 11 Ejemplo de gráfico de Shewhart.....	23
Figura 12 Gráficos de Shewhart	24
Figura 13 Q-Q normal de la variable P_vah_e1	25
Figura 14 Representación partes de un boxplot.	26
Figura 15 Boxplot de algunas variables.	27
Figura 16 Histogramas de algunas variables.....	28
Figura 17 Gráficos de control con la aplicación del filtro.	29
Figura 18 Resumen procedimiento. Adaptado de [8].	33
Figura 19 Red Neuronal: Estructura	33
Figura 20 Red Neuronal: Valor observado VS Predicción: Datos Completos	34
Figura 21 Red Neuronal: Valor observado VS Predicción: Datos Test	34
Figura 22 Red Neuronal: Valor observado VS Predicción: Datos Completos: Experimento 2	35
Figura 23 Red Neuronal: Valor observado VS Predicción: Datos Test: Experimento 2.....	35
Figura 24 SVR: Estructura	35
Figura 25 SVR: Valor observado VS Predicción: Datos Completos	36
Figura 26 SVR Valor observado VS Predicción: Datos Test	36
Figura 27 Random Forest Regression: Estructura	37

Figura 28 Random Forest Regression: Valor observado VS Predicción- Conjunto Completo	37
Figura 29 Random Forest Regression: Valor observado VS Predicción- Conjunto test	38
Figura 30 Random Forest Regression: Valor observado VS Predicción:Conjunto Completo: Experimento 2	38
Figura 31 Random Forest Regression: Valor observado VS Predicción: Conjunto Test: Experimento 2	39
Figura 32 Regresión por Mínimos Cuadrados Parciales: Estructura.....	39
Figura 33 Regresión por Mínimos Cuadrados Parciales: Valor observado VS Predicción: Datos Completos.....	40
Figura 34 Regresión por Mínimos Cuadrados Parciales: Valor observado VS Predicción: Datos Test	40
Figura 35 Regresión por Mínimos Cuadrados Parciales: Valor observado VS Predicción: Datos Completos: Experimento 2.....	41
Figura 36 Regresión por Mínimos Cuadrados Parciales: Valor observado VS Predicción: Datos Test: Experimento 2.....	41

Índice de tablas

Tabla 1 Calendario y presupuesto del proyecto	3
Tabla 2 Estimación y monitorización de los recursos	4
Tabla 3 Presupuesto final del proyecto.....	4
Tabla 4 Descripción de las variables del conjunto de datos	21
Tabla 5 Test de Kolmogórov-Smirnov en los datos	25
Tabla 6 Porcentaje de outliers en los dos análisis.....	30
Tabla 7 Porcentaje de varianza PCA sobre el conjunto completo de datos.....	31
Tabla 8 Red Neuronal: Búsqueda de Hiperparámetros.....	33
Tabla 9 Red Neuronal: Búsqueda de Hiperparámetros: Experimento 2	34
Tabla 10 SVR: Búsqueda de Hiperparámetros.....	36
Tabla 11 Random Forest Regression: Búsqueda de Hiperparámetros	37
Tabla 12 Random Forest Regression: Búsqueda de Hiperparámetros: Experimento 2	38
Tabla 13 Regresión por Mínimos Cuadrados Parciales: Búsqueda de Hiperparámetros	39
Tabla 14 Regresión por Mínimos Cuadrados Parciales: Búsqueda de Hiperparámetros: Experimento 2	40

Resumen

El proceso de la industria azucarera consta de varias etapas, siendo una de ellas la evaporación. Durante esta etapa, se utiliza un evaporador para aumentar la concentración de azúcar en el líquido, mediante la introducción del jugo obtenido de caña de azúcar o remolacha.

La concentración de azúcar se mide en grados Brix, los cuales se obtienen midiendo la concentración de materia seca en el líquido. Sin embargo, los métodos actuales que consisten en estudios de laboratorio son lentos, y los sensores en tiempo real son costosos, limitados en cantidad y pueden presentar problemas de ensuciamiento o cristalización.

Para solucionar este problema, este proyecto propone el uso de modelos de aprendizaje automático o estadístico que permiten estimar el Brix a la salida de la etapa de evaporación en base a otras medidas del proceso. Para garantizar modelos fiables, se siguen metodologías como CRISP-DM para analizar datos y modelos.

En resumen, el proyecto busca mejorar el proceso de medición en tiempo real de la concentración de azúcar en la industria azucarera. A su vez, la mejora en la estimación del Brix mediante el uso de modelos permitiría incrementar la productividad y la rentabilidad al facilitar la toma de decisiones de manera inmediata, así como detectar anomalías o desviaciones en el proceso de evaporación. De esta manera, se podrían adoptar rápidamente medidas correctivas para evitar pérdidas en la producción.

Abstract

The sugar industry process consists of several stages, one of which is evaporation. During this stage, an evaporator is used to increase the sugar concentration in the liquid by introducing the juice obtained from sugar cane or beet.

The sugar concentration is measured in Brix degrees, which are obtained by measuring the concentration of dry matter in the liquid. However, current methods consisting of laboratory studies are slow, and real-time sensors are expensive, limited in quantity and can present problems of fouling or crystallisation.

To solve this problem, this project proposes the use of automatic or statistical learning models that allow Brix to be estimated at the exit of the evaporation stage based on other process measurements. To guarantee reliable models, methodologies such as CRISP-DM are used to analyse data and models.

In summary, the project aims to improve the real-time measurement process of sugar concentration in the sugar industry. In turn, improved Brix estimation through the use of models would increase productivity and profitability by facilitating immediate decision making, as well as detecting anomalies or deviations in the evaporation process. In this way, corrective measures could be taken quickly to avoid production losses.

Capítulo 1. Introducción

La industria azucarera tiene como objetivo final la fabricación de azúcar para el consumo.

El proceso de producción llevado a cabo es complejo y se divide en varias etapas. Una de ellas es conocida como evaporación.

Durante el proceso de evaporación se introduce el jugo de caña o remolacha resultado de las anteriores etapas en evaporadores con el objetivo de aumentar la concentración de azúcar en el líquido. Para determinar esta cantidad es habitual medir la concentración de materia seca, siendo la unidad de medida más utilizada el denominado Brix, definido como el porcentaje de materia seca diluida en un líquido, de tal forma que la mezcla de 1 gramo de sacarosa en 100 gramos de solución se cuantifica como un grado Brix.

Existen numerosos métodos capaces de medir el contenido de materia seca, la mayoría de ellos consisten en estudios realizados en laboratorios que acostumbran a ser simples y poco costosos. Estos estudios necesitan extraer una muestra, transportarla hasta el propio laboratorio, realizar el análisis y esperar hasta la obtención de los resultados, lo que provoca una demora en el tiempo entre la extracción y la medición que ralentiza la toma de decisiones [1].

También hay otras técnicas, algunas novedosas como la técnica por microondas y otras que incorporan nuevas tecnologías a métodos ya existentes como la reflectometría, que a diferencia de los anteriores son capaces de proporcionar resultados en tiempo real, pero los sensores que incorporan estas suponen un elevado coste, impidiendo su propagación y situándolos únicamente en puntos críticos.

En la actualidad la industria azucarera está altamente automatizada, con el propósito de aumentar la productividad y conseguir una mayor rentabilidad. Pero a cambio, esto exige disponer de datos en tiempo real para tomar decisiones de forma automática.

Para proporcionar mediciones en tiempo real, pero que no supongan un coste muy alto, la utilización de modelos de aprendizaje automático o estadístico, supone una de las propuestas más atractivas. Un buen modelo es capaz de proporcionar estimaciones muy precisas casi de manera instantánea, con un margen de error mínimo y a la vez no supone un gran gasto en comparación con los sensores que se utilizan actualmente.

1.1 Objetivos

En definitiva, el objetivo final del proyecto consiste en el desarrollo de un sensor software que implemente un modelo de aprendizaje automático o estadístico para estimar el Brix del proceso de evaporación de la industria azucarera. Sin embargo, para llegar a este objetivo, es necesario cumplir una serie de tareas que se pueden resumir en:

1. Análisis del proceso de producción del azúcar con el fin de entender las variables involucradas en el proceso.
2. Análisis estadístico de las variables relevantes con un conjunto de datos.
3. Estudio de distintos modelos para la estimación del Brix, considerando otros trabajos realizados anteriormente.
4. Diseño de un sensor software para la estimación del Brix.

1.2 Estructura del Documento

El resto de la memoria se compone de los siguientes apartados:

- **Capítulo 2. Planificación.** Se detalla la planificación que ha seguido el proyecto, identificando las distintas fases por las que ha pasado y los riesgos encontrados.
- **Capítulo 3. Contexto del problema.** Se explican los principales conceptos necesarios para entender el problema.
- **Capítulo 4. Metodología.** Se desarrollan las distintas fases del proyecto siguiendo la estructura propuesta por la metodología CRISP-DM.
- **Capítulo 5. Conclusiones.** Se resumen los resultados y se analizan otras posibles rutas o ampliaciones del trabajo.
- **Capítulo 6. Bibliografía**
- **ANEXOS**

Capítulo 2. Planificación

Una vez identificado que se quiere hacer y explicado cómo se quiere hacer. En este capítulo se detalla que metodología se ha llevado a cabo para desarrollar el proyecto.

2.1 Metodología Usada

En una primera reunión para arrancar el proyecto entre ambos tutores y el estudiante, se estableció cuál iba a ser el trabajo por realizar, los objetivos a conseguir y el procedimiento o modo de trabajar.

Esto último, consiste en dividir el trabajo en varios capítulos, de forma que no haga falta volver a ese mismo punto, una vez revisado y corregido, para así, poder ir completando la memoria de manera incremental.

2.2 Presupuesto y Calendario del proyecto

2.2.1 Presupuesto en horas y calendario

En la Tabla 1 se puede ver el calendario y presupuesto del proyecto. Para calcular los costes se ha considerado que el salario que cobraría una persona para la realización de este proyecto es de 11,28 € / hora (Salario medio de un Analista de datos en España).

	Fecha Inicio	Fecha Final	Horas Estimadas	Horas Realizadas	Coste
Matriculación e inicio de Cuatrimestre	Jun-21	Jun-21	0	0	0 €
Trabajo inicial de estudio y realización de Introducción y Estado del Arte	Mar-22	Mar-22	40	44	496,32 €
Análisis Inicial	Mar-22	Abr-22	20	20	225,60 €
Tratamiento de los datos	Abr-22	May-22	40	40	451,20 €
Creación de modelos	May-22	May-22	50	60	676,80 €
Análisis de resultados	May-22	Jun-22	50	55	620,40 €
Elaboración de la memoria	Mar-22	Jun-22	40	40	451,20 €
Elaboración de la presentación PPT	Jun-22	Jul-22	10	10	112,80 €
			240	259	2.291,52 €

Tabla 1 Calendario y presupuesto del proyecto

2.2.2 Estimación y monitorización de los recursos a utilizar

En la Tabla 2 se ven los recursos utilizados y el precio que tendrían en el mercado si el estudiante no hubiera proporcionado estos. El precio del ordenador se establece a partir del coste inicial de 600 €, una estimación de vida útil alrededor de 6 años y del tiempo de uso en el proyecto (256 horas).

Recursos a utilizar (entorno de trabajo)	Fecha Inicio Uso	Fecha Final Uso	Descripción	Precio (si aplica)	Precio final (si aplica)
Ordenador personal	Feb-22	Jul-22	Ordenador Portátil	2,92 €	2,92 €
Microsoft Word Licencia Microsoft Office 365 UVA	Feb-22	Jul-22	Software para desarrollar la memoria	0 €	0 €
Microsoft Outlook Licencia Microsoft Office 365 UVA	Feb- 22	Jul-22	Software para la comunicación entre tutores y estudiante	0 €	0 €
Google Collab	Mar-22	Jun-22	Software de programación	0 €	0 €
				2,92 €	2,92 €

Tabla 2 Estimación y monitorización de los recursos

2.2.3 Presupuesto total del proyecto

En la Tabla 3 se visualiza el presupuesto final del proyecto. El total es la suma de los recursos, más el sueldo por las horas del trabajo realizado por el estudiante.

Totales	Unidades/ Horas	Precio Unitario	Precio Total
Trabajo del proyectante	259,00	11,28 €	2.921,52 €
Recursos	1,00	2,92 €	2,92 €
			2.924,44 €

Tabla 3 Presupuesto final del proyecto

Capítulo 3. Contexto del Problema

3.1 Proceso de evaporación

La obtención del azúcar a partir de las materias primas es un proceso complejo que consta de varias etapas. Una explicación detallada del proceso completo escapa de lo requerido para situar el contexto del problema, aunque es importante mencionar un esquema general para poder situar la etapa de evaporación.

En la Figura 1 se puede ver un diagrama general de las principales etapas del proceso de obtención del azúcar, en este caso a partir de la remolacha, aunque pudiera tratarse de la caña de azúcar. Ambas constituyen las dos principales formas de obtención del azúcar, con procesos muy similares. Estas etapas son las siguientes [2]:

- **Descarga, lavado y troceado de las remolachas.** En esta etapa se descargan, lavan y trocean las remolachas a porciones triangulares denominadas *coseñas*.
- **Difusión.** En esta etapa se extrae el jugo azucarado mediante la acción de agua caliente en contra corriente de las *coseñas*. Este jugo extraído contiene una gran cantidad de sustancias no azucaradas con una pureza aproximada del 85% y una densidad aproximada de 16° Brix.
- **Depuración.** El objetivo en esta etapa es eliminar las sustancias no azucaradas del jugo empleando cal, a través de decantadores y otros instrumentos, elevando la pureza desde un 85% a un 91%, aproximadamente. El Brix en esta etapa no varía respecto a la anterior y está en torno a los 16° Brix.
- **Evaporación.** Durante esta etapa se aumenta la concentración de azúcar en el jugo, partiendo de una densidad de 16° Brix, hasta un 60°-65° Brix, a través de la evaporación del agua contenida en el jugo, obteniendo como producto final el denominado jarabe. Es en la salida de esta fase donde se considera incluir el sensor software para controlar la concentración de azúcar en el jugo.
- **Cocimiento** [3]. En esta etapa se introduce el jarabe en unos recipientes denominadas *tachas* para favorecer la formación de cristales. El resultado es una solución azucarada sobresaturada denominada masa cocida con una concentración aproximada 91°-92° Brix.
- **Cristalización.** En esta etapa se extrae la masa cocida de las *tachas* y se sigue con la cristalización mediante el enfriamiento de esta en unos recipientes denominados cristalizadores.
- **Centrifugación y Secado.** En esta etapa se procede a separar el azúcar de la masa cocida y posteriormente transportarlo hasta el secadero para reducir la humedad resultado de la centrifugación.



Figura 1 Diagrama general del proceso de obtención del azúcar

En cuanto al proceso de evaporación únicamente se menciona el propósito, pero no los mecanismos llevados a cabo para aumentar la concentración del jugo, siendo el procedimiento más habitual, la utilización de un sistema de evaporación de múltiple efecto de evaporación directa.

En la Figura 2 se muestra sistema evaporador de múltiple efecto de evaporación directa, que como su nombre indica es un instrumento compuesto de varios efectos (conjunto de uno o varios evaporadores que trabajan a una misma presión y temperatura).

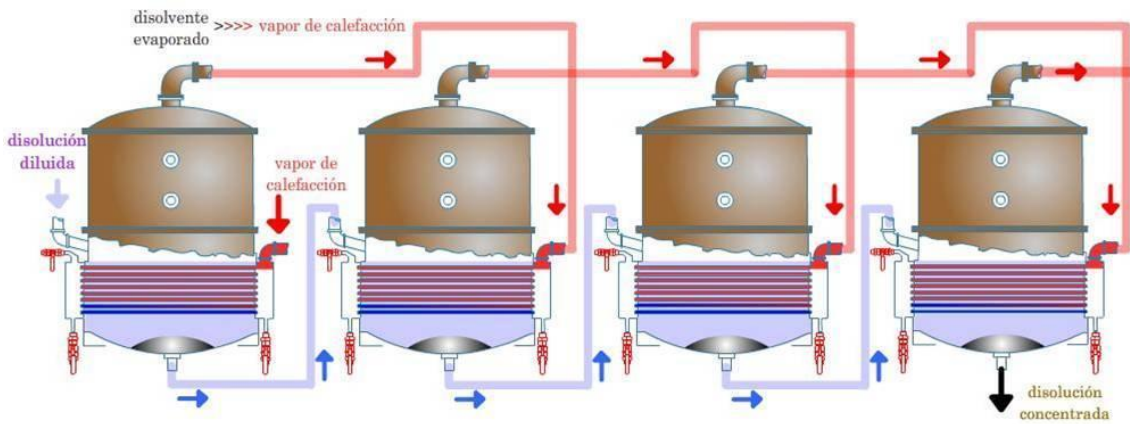


Figura 2 Sistema evaporador de múltiple efecto de evaporación directa. Extraído de [1]

En primer lugar, se suministra vapor vivo o *fresh steam* (vapor generado por una caldera o un cambiador de calor) y al jugo resultado de la etapa de decantación al primer efecto, produciendo vapor en éste mediante una reducción de la presión. A este vapor producido por una disminución drástica de la presión se denomina *vapor flash*, que sirve como forma de calentar el segundo efecto y así sucesivamente hasta el último.

La disminución de la presión se consigue mediante un incremento del vacío, lo que conlleva a su vez, que la temperatura vaya disminuyendo también en cada efecto. De esta forma se evita que el jugo resulte dañado en el proceso, al alcanzar temperaturas de ebullición inferiores en cada uno de los efectos [4].

3.2 Medición física del Brix

En el capítulo de Introducción se mencionan las dos formas de medición física del Brix y los inconvenientes de cada una de ellas.

Por un lado, estudios en laboratorios, cuya principal desventaja es el retraso en las mediciones y son inadecuados para monitorizar un proceso ya que el tamaño de la muestra que se analiza es muy pequeño. En la actualidad los que más destacan son:

- Polarimetría.
- Cromatografía.
- Biosensores.
- Hidrómetro.
- NIRS.
- Refractómetro.

Por otro lado, técnicas que permiten una medición en línea, cuya principal desventaja es su elevado coste y mantenimiento. Entre ellos destacan:

- Refractómetro digital.
- Tecnología por microondas.

3.3 Medición Brix con sensores software

Como alternativa a los sensores físicos se puede proporcionar una estimación del Brix a través de la relación con otras variables con mayor facilidad de tratamiento. En la actualidad existen diferentes técnicas que se emplean para desarrollar sensores software capaces de proporcionar buenas estimaciones.

En base a otros estudios previos se destacan algunas de las que proporcionan mejores resultados [5].

3.3.1 Regresión por mínimos cuadrados parciales

Este método se utiliza en problemas cuando los datos tienen un gran número de variables, altas correlaciones entre pares y pocas observaciones. Puede utilizarse para realizar el modelo de problemas aproximadamente lineales, sin la necesidad de que las variables sean independientes entre sí.

Mínimos Cuadrados Parciales o *Partial least Squares (PLS)* consiste en reducir el número de variables a través de un proceso iterativo en el que se calculan direcciones ortogonales para aplicar una regresión de mínimos cuadrados al subespacio elegido [6].

$$\min(e) = \min\|Y - XB\|^2 \quad (1)$$

Se basa en la relación entre la matriz de entrada X y la matriz de salida Y a través de una serie de variables denominadas latentes que son un conjunto de variables ortogonales.

$$\begin{aligned} X &= TP^T + E \\ Y &= UC^T + F \\ Y &= XB \end{aligned} \quad (2)$$

Donde T y U son matrices compuestas por vectores latentes, E y F son matrices residuales y, P y C son matrices compuestas por los vectores de peso.

En los procesos industriales es común que los datos estén altamente correlados y con mucho ruido, por lo que este método es muy eficaz. Pero existe un gran inconveniente, ya que también es habitual que existan relaciones no lineales de interés entre variables que no pueden considerarse. A consecuencia de esto existen otros métodos que trabajan bien con ellas como el caso de las redes neuronales.

3.3.2 Red Neuronal Artificial

Las redes neuronales [7] son modelos de computación paralelos que constan de algoritmos de aprendizaje que sirven para extraer características o aproximar funciones no lineales.

Este aprendizaje se realiza de manera automática, donde a medida que se obtiene, almacena y organiza información.

Las redes neuronales están formadas por unidades de cómputo denominadas neuronas que se relacionan entre ellas en base al proceso de sinapsis (que emula el proceso biológico con el mismo nombre), que consiste en el intercambio de señales entre neuronas. Es decir, las neuronas están conectadas entre sí mediante señales formando una red neuronal. Estas conexiones tienen asociada un peso, cuyo valor se estima en el proceso de aprendizaje o ajuste.

Por lo tanto, cada neurona recibe una serie de señales que le dan información sobre el estado de activación de todas las neuronas con las que está conectada.

Todo esto se puede formular de manera matemática [8].

La neurona descrita tiene un número N de entradas que pueden ser excitadoras o inhibidoras, por lo tanto, son binarias.

$$x_k \in \{0,1\}, \quad k = 1 \dots N \quad (3)$$

A su vez la neurona produce una respuesta, que puede ser activarse o permanecer en reposo, que también es binaria.

$$y_k \in \{0,1\} \quad (4)$$

Dicha respuesta depende de si la suma ponderada de las entradas supera el valor del umbral θ (límite de activación) en cuyo caso la neurona se activa, por lo tanto $y=1$.

$$\sigma(x) = \begin{cases} 1 & \text{si } \sum_1^N w_k \cdot x_k > \theta \\ 0 & \text{En caso contrario} \end{cases} \quad (5)$$

La señal que llega a la función de activación tiene en cuenta esta última relación:

$$\sum_1^N w_k \cdot x_k + \theta \quad (6)$$

En la Figura 3 se puede ver la estructura general de una neurona que resume la explicación anterior.

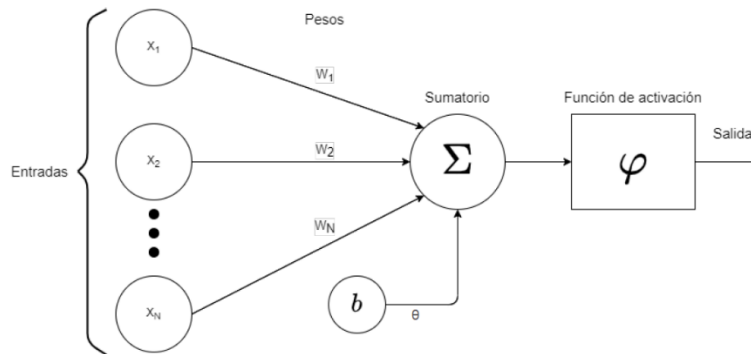


Figura 3 Estructura de una neurona. Adaptada de [2]

Función de activación

La función de activación es una función matemática aplicada a la salida de la neurona que permiten modificar la salida para poder introducir no-linealidades y de esta manera ser capaz de resolver problemas más complejos. Según la modificación que se quiera realizar existen diversas funciones destacando la identidad, escalón (*treeshold*), sigmoide y rectificadora (ReLU) que se pueden ver de manera más detallada en el Anexo I – Funciones de Activación.

Estructura de una red neuronal

Una red neuronal es un conjunto de neuronas que pueden dividirse en capas según el tipo de señal que reciben, distinguiendo tres tipos diferentes:

- **Capa de entrada:** Formada por neuronas que reciben las señales que provienen del entorno. Es decir, es la capa que conecta los datos con la propia red.
- **Capa de salida:** Está formada por neuronas que envían la señal fuera de la red. En otras palabras, son las encargadas de proporcionar la salida.
- **Capas ocultas:** Son aquellas neuronas cuyas entradas y salidas se encuentra dentro de la red sin contacto con el exterior.

Según el tipo de problema tanto el número de capas como el número de neuronas en cada una de ellas se puede modificar. Hay que tener especial cuidado a la hora de seleccionar el número de neuronas en la capa de salida, así como su función de activación.

Mecanismo de aprendizaje o entrenamiento

Durante el aprendizaje se construyen y destruyen conexiones entre neuronas en respuesta a una información de entrada lo que modifica sus pesos. De tal manera que la creación de una nueva conexión supone que su peso pase a ser distinto de 0, mientras que en la destrucción sea igual a 0, considerando el proceso terminado cuando los pesos permanecen estables. Dicho esto, es importante conocer cómo se modifican, esto es, qué criterios o reglas se siguen para cambiar los valores de dichas conexiones con el objetivo de que la red aprenda, a esto se lo conoce como regla de aprendizaje. Se suelen considerar dos tipos de reglas, reglas de aprendizaje supervisado si se conoce el valor real de la salida de la red o reglas de aprendizaje no supervisado en caso contrario.

De entre las reglas de aprendizaje supervisado se encuentra el aprendizaje por corrección de error, también denominado función de coste o función de pérdida, que consiste en ajustar los pesos de las conexiones en función de la diferencia entre el valor real y el obtenido por la red. Entre las distintas funciones de coste existentes se destaca una conocida como *MSE* (Error cuadrático medio, del inglés *Mean Squared Error*) utilizada en problemas de regresión.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (7)$$

Donde n es el número de observaciones, y_i es el valor real de la observación i e \hat{y}_i es el valor predicho de la observación i .

Descenso del gradiente

Conocida la función de coste, el problema del aprendizaje de una red pasa a ser un problema de optimización, donde el objetivo es minimizar dicha función. Dicho de otra manera, el objetivo es ajustar los pesos θ de la red para minimizar la función de coste.

Si llamamos $C_x(\theta)$ al coste de cada observación x del conjunto, θ a los parámetros de la red y n al número de observaciones en el conjunto de, entonces el problema puede expresarse como:

$$\arg_{\theta} \min C(\theta) = \arg_{\theta} \min \frac{1}{n} \sum_x C_x(\theta) \quad (8)$$

El descenso del gradiente [3] es un algoritmo que parte de la idea de que el gradiente negativo de una función que tiene múltiples variables apunta en la dirección de máximo decrecimiento. Por lo tanto, partiendo de unos parámetros iniciales θ_0 y calculando el gradiente en ese punto. Se obtiene un nuevo punto θ_1 desplazándose en la dirección contraria al gradiente, consiguiendo así reducir la función de coste.

$$\theta_1 = \theta_0 - \nabla f(\theta_0) \quad (9)$$

Como es posible desplazarse demasiado lejos del mínimo de la función de forma que a partir de un punto la función ya no decrezca e incluso pueda aumentar. Es necesario añadir un parámetro adicional denominado tasa de aprendizaje η que permita regular el desplazamiento.

$$\theta_1 = \theta_0 - \eta \nabla f(\theta_0) \quad (10)$$

Si se repite el proceso de manera iterativa se consigue un punto cercano al mínimo de la función.

$$\theta_{k+1} = \theta_k - \eta \nabla f(\theta_k) \quad (11)$$

Debido a que el conjunto de entrenamiento suele tener un elevado número de observaciones y teniendo en cuenta que la expresión del gradiente es:

$$\nabla C = \frac{1}{n} \sum_x C_x \quad (12)$$

Esto supone un elevado coste computacional y temporal al tener que calcular el gradiente de miles de observaciones. Es por ello por lo que surge una alternativa denominada descenso

por gradiente estocástico o *SGD*, que permite actualizar los parámetros con mayor rapidez al aproximar el gradiente a partir de un subconjunto más pequeño.

Una vez conocidos los pesos, solo queda calcular el coste de cada una de las observaciones, pero esto requiere el cálculo de la derivada parcial, para cada uno de los pesos, lo que resulta muy ineficiente, ya que las redes neuronales suelen tener miles de estos.

El algoritmo de *Backpropagation* o retro propagación del error permite reducir notablemente el número de cálculos, a partir de la idea de que el error en las capas anteriores depende directamente del error en las capas posteriores, por lo que se puede ir analizando y optimizando la propagación del error [9].

Esto quiere decir que si el error en una capa está controlado y se ha visto que no es influyente, en consecuencia, el error en las capas anteriores también lo está, evitando de esta manera el cálculo de todas las derivadas parciales que involucran a esas capas.

Optimizadores

Encontrar una tasa de aprendizaje correcta que permita que la red aprenda y a su vez este proceso no demore en el tiempo es una tarea difícil. A raíz de esto, surgen una serie de optimizadores que automatizan la selección de este hiperparámetro a la vez que permiten su modificación a medida que el entrenamiento avanza. Entre ellos destacan algunos como.

- Descenso del gradiente estocástico o SGD: El método básico de optimización donde la tasa de aprendizaje es fija (no cambia en el tiempo).
- Adagrad [10]: Es un optimizador con tasas de aprendizaje específicas para cada parámetro que se adaptan en función de la frecuencia con la que se actualiza un parámetro durante el entrenamiento. Cuantas más actualizaciones reciba un parámetro, menores serán las actualizaciones.
- Adam [11]: Es un método estocástico de descenso de gradiente que se basa en la estimación adaptativa de momentos de primer y segundo orden.

3.3.3 SVR

SVR (Support Vector Regression) [12] es un algoritmo de aprendizaje automático que permite reducir el error a un cierto nivel, es decir, no importa cual grande es el error, sino, si está dentro de un rango aceptable.

SVR permite definir que error es aceptable para el modelo, y teniendo en cuenta esto, encontrar una recta o hiperplano que se ajuste a los datos.

La idea consiste en minimizar la norma l_2 del vector de coeficientes (w). Por otro lado, existe una serie de restricciones que permiten establecer el error absoluto menor o igual a un margen establecido ϵ , pudiendo ajustar este último para adecuar la precisión del modelo:

Minimizar:

$$\min \frac{1}{2} \|w\|^2 \quad (13)$$

Restricciones:

$$|y_i - (w_i x_i + b)| \leq \epsilon \quad (14)$$

Donde b es la pendiente de la recta.

En la Figura 4 se puede ver de manera gráfica las expresiones y términos utilizados.

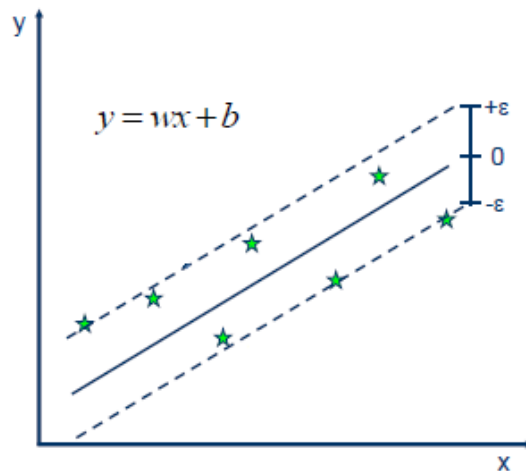


Figura 4 Ejemplo de utilización SVR. Adaptado de [12].

Pero al tratarse de un problema de regresión, la salida es un número real y por tanto puede tomar valores fuera de los márgenes. Por ello hay que tener en cuenta la posibilidad de errores mayores al margen. Esto se corrige añadiendo un nuevo hiperparámetro denominado holgura. La idea es sencilla: cualquier valor fuera del margen se va a denotar su desviación como ξ .

Minimizar:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n |\xi_i| \quad (15)$$

Restricciones:

$$|y_i - (w_i x_i + b)| \leq \varepsilon + |\xi_i| \quad (16)$$

En la Figura 5 se observa un ejemplo donde existe una observación fuera de los márgenes.

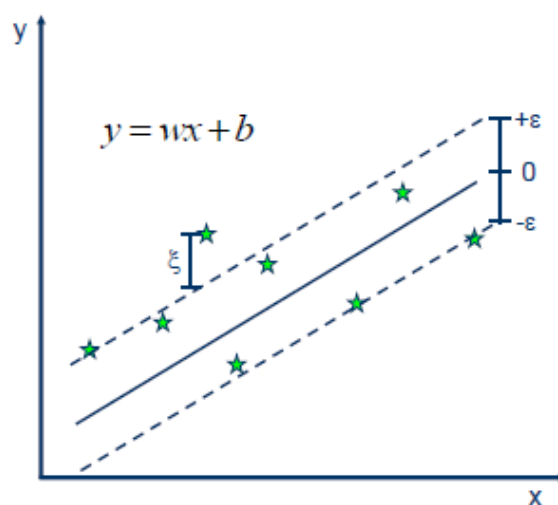


Figura 5 Ejemplo de holgura en SVR. Adaptado de [12].

También se ve como se ha añadido otro hiperparámetro C , que muestra la tolerancia que se va a tener para los puntos fuera de ε . Así pues, un valor muy grande de la constante C , en el caso límite ($C \rightarrow \infty$) se considera que el conjunto está perfectamente representado por el hiperplano predictor ($\xi \rightarrow 0$). Por contra, un número pequeño cercano a 0 permite valores de ξ elevados, es decir, se admite un número muy elevado de ejemplos mal representados [13].

Todo esto es admisible porque se supone que se trata de un problema lineal.

$$y = \sum_{i=1}^n (w_i - w_i^*) \cdot \langle x_i, x \rangle + b \quad (17)$$

Para tratar con problemas no lineales, existe otro hiperparámetro denominado kernel, que no es más que una función que transforma los datos a una dimensión mayor para poder hacer la separación lineal [4].

$$y = \sum_{i=1}^n (w_i - w_i^*) \cdot K(x_i, x) + b \quad (18)$$

Donde $K(x_i, x)$ representa la función Kernel.

En la Figura 6 se muestra un ejemplo de utilización de una función Kernel para transformar los datos para que sean linealmente separables.

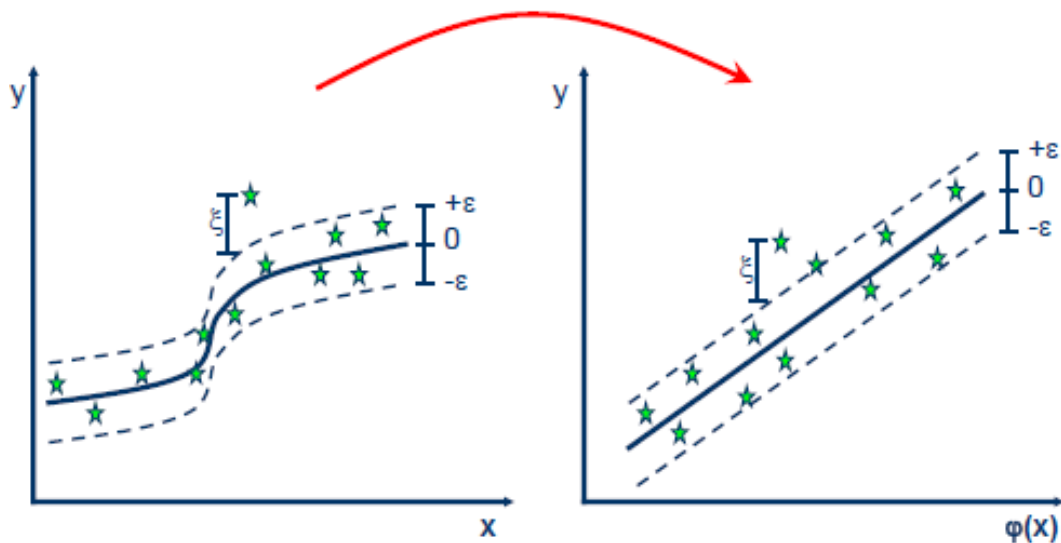


Figura 6 Ejemplo de utilización de un kernel no lineal. Adaptado de [12].

Existen distintos tipos de kernel según la transformación que realizan [13].

- Lineal:

$$K(x, x') = \langle x, x' \rangle \quad (19)$$

- Polinomial:

$$K_d(x, x') = (\langle x, x' \rangle + r)^d \quad (20)$$

- Función de base radial o RBF:

$$K(x, x') = \exp(-\gamma \|x - x'\|^2) \quad (21)$$

- Sigmoidea:

$$K(x, x') = \tanh(\gamma(x, x') + r) \quad (22)$$

Depende del tipo de kernel, existen una serie de nuevos parámetros que pueden o no aparecer.

d que indica el grado del polinomio

γ Gamma indica cuánta influencia tiene un único ejemplo de entrenamiento. A mayor valor más cerca tienen que estar otros ejemplos para verse afectados.

r se trata de un valor independiente que puede tomar valor 0.

3.3.4 Random Forest Regression

Random Forest [14] es un algoritmo de aprendizaje supervisado que combina la idea de aprendizaje por conjuntos o ensemble, con la construcción de múltiples árboles de decisión [15].

Por un lado, el aprendizaje por conjuntos consiste en utilizar para unos mismos datos, múltiples modelos, para posteriormente combinar los resultados de cada modelo y en este caso realizar una predicción más precisa.

Por otro lado, los árboles de decisión son un algoritmo de aprendizaje supervisado que segmentan el espacio de un predictor en un número de regiones, y más concretamente, los árboles de regresión utilizan la media de las observaciones de entrenamiento de cada una de esas regiones para obtener las predicciones.

Por lo tanto, para construir un árbol de regresión es necesario dividir el espacio del predictor (conjunto de valores posibles según las variables X_1, X_2, \dots, X_p en J regiones distintas y sin solapamiento R_1, R_2, \dots, R_j). Debido a que el objetivo es minimizar el error, si se toman regiones rectangulares de múltiples dimensiones, es posible encontrar la mejor división posible escogiendo la que minimice la Suma de Cuadrados Residuales (*Residual Sum of Squares*, en siglas *RSS*) [15].

$$\sum_{j=1}^J \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2 \quad (23)$$

Donde \hat{y}_{R_j} es la media de la variable respuesta en la región j .

Por tanto, en cuanto al valor de predicción de una nueva observación, por ejemplo, si esta pertenece a la región R_2 , el valor será la media de la variable respuesta de las observaciones de entrenamiento en R_2 .

Uno de los inconvenientes de este método es que, considerar todas las regiones del espacio es demasiado costoso computacionalmente. En consecuencia, se recurre a un algoritmo de división denominado: división recursiva binaria [15].

La división recursiva binaria consiste en empezar desde la raíz (donde todas las observaciones pertenecen a la misma región) y de manera sucesiva, realizar la mejor división posible generando dos ramas.

Para escoger la mejor división, es necesario dividir cada predictor X_j según un punto de corte s , de forma que, las dos regiones que se generan $\{X|X_j < s\}$ y $\{X|X_j \geq s\}$ tienen la menor RSS posible.

Una vez hecho esto, es necesario comparar todos los predictores X_1, X_2, \dots, X_p y todos los posibles puntos de corte s para cada uno, escogiendo finalmente el predictor y el punto de corte que minimicen la RSS.

$$R_1(j, s) = \{X|X_j < s\} \text{ y } R_2(j, s) = \{X|X_j \geq s\} \quad (24)$$

Donde $R_1(j, s)$ es la región del predictor X_j con el punto de corte s y $R_2(j, s)$ es la región del predictor X_j con el punto de corte s .

$$\sum_{i: x_i \in R_1(j, s)} (y_i - \hat{y}_{R_1})^2 + \sum_{i: x_i \in R_2(j, s)} (y_i - \hat{y}_{R_2})^2 \quad (25)$$

Corresponde a la suma de la RSS de $R_1(j, s)$ y $R_2(j, s)$.

Una vez escogidas el siguiente paso es dividir respectivamente $R_1(j, s)$ y $R_2(j, s)$. Este proceso continúa hasta que se cumple algún criterio de parada, como puede ser, un número mínimo de observaciones por región o un máximo número de regiones.

Como se puede ver en la Figura 7, el nombre hace referencia debido a la estructura en forma de árbol con un nodo raíz, nodos internos conocidos también como nodos de decisión, y nodos terminales, donde estos últimos constituyen cada una de las regiones.

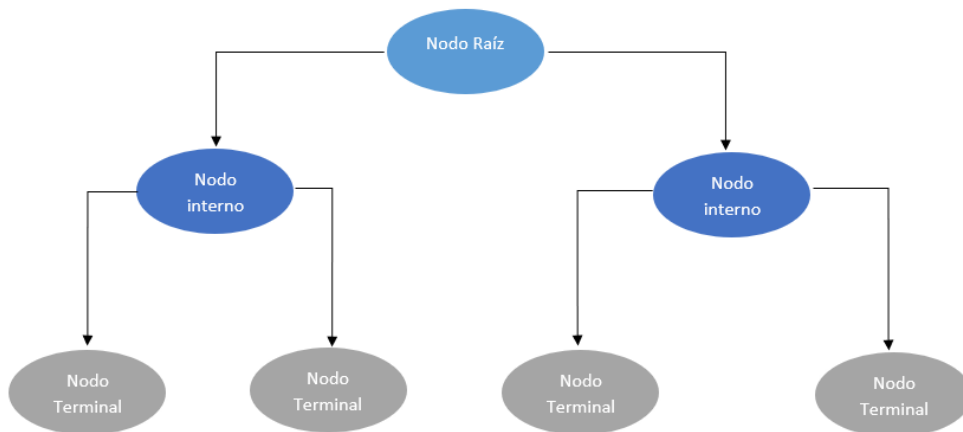


Figura 7 Estructura de un árbol de decisión con 4 regiones.

Finalmente, en cuanto al entrenamiento de un Random Forest Regression, consiste en construir varios árboles de decisión durante el entrenamiento que se ejecutan en paralelo sin interacción, esto se consigue mediante la elección de manera aleatoria de un subconjunto de los datos por bootstrapping y promediar la predicción de cada uno de ellos para obtener una predicción más acertada [14] como se muestra en la Figura 8.

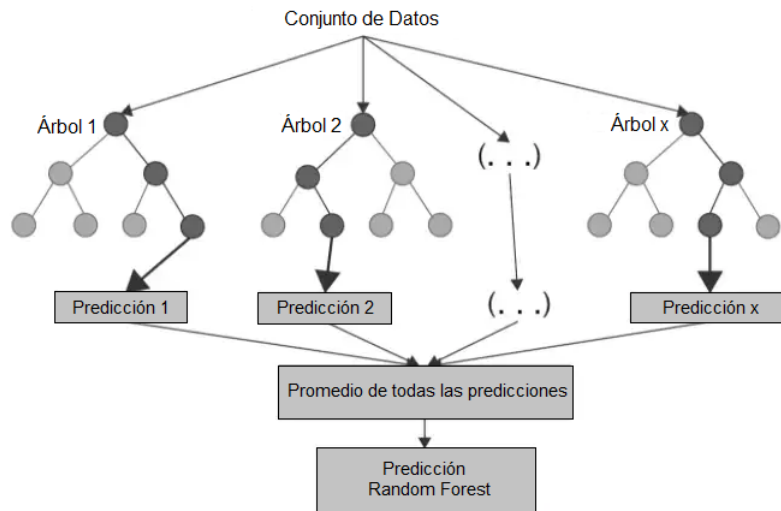


Figura 8 Ejemplo de predicción con Random Forest. Adaptado de [14]

3.3.5 Búsqueda y selección de hiperparámetros

En la mayoría de los algoritmos de aprendizaje existen una serie de hiperparámetros. Estos son parámetros que controlan el comportamiento de un modelo de aprendizaje, pero que no son aprendidos por el modelo, y que permiten controlar la forma de proceder del algoritmo.

En este proyecto se describen 3 algoritmos de aprendizaje automático, en concreto, redes neuronales, SVR y Random Forest Regression, y algunos de los hiperparámetros de estos. Teniendo en cuenta que la estimación va a verse afectada según los hiperparámetros utilizados, es necesario conocer algunos de ellos:

- **Redes neuronales:** Número de capas ocultas, número de neuronas en cada capa, funciones de activación y optimizadores.
- **SVR:** Holgura, C y Kernel.
- **Random Forest Regression:** Número de árboles, criterio para definir la calidad de una división y profundidad máxima de cada árbol.

Donde cada uno de ellos tiene un conjunto posible de valores, por lo que, cuando se quiere encontrar el mejor modelo posible, es necesario estudiar un número de combinaciones extremadamente alto, lo que se traduce en un costo computacional y de tiempo extremadamente alto. Para llevar a cabo esta tarea en la práctica se suele recurrir a dos algoritmos.

Búsqueda en cuadrícula o grid search

En la búsqueda de cuadrícula [16] se prueba todas las posibles combinaciones de una lista preestablecida de valores de hiperparámetros y se evalúa el modelo para cada una de ellas.

El nombre hace referencia a que el patrón que se sigue es similar a una cuadrícula, donde todos los valores se colocan en una matriz se recorre esta misma pasando por cada uno de sus puntos que corresponde a cada una de las combinaciones y se registra su precisión. Después de evaluar todas las combinaciones (recorrer toda la matriz), se considera el mejor modelo aquel que proporcionó una mayor precisión.

Una de las principales desventajas de la búsqueda en cuadrícula es que sufre cuando el número de hiperparámetros crece exponencialmente en términos de dimensionalidad. Con solo cuatro parámetros, este problema puede volverse inviable ya que la cantidad de evaluaciones requeridas crece geoméricamente con cada parámetro adicional. Esto se le conoce como la maldición de la dimensionalidad.

Búsqueda aleatoria o random search

La búsqueda aleatoria [16] es una técnica en la que se utilizan combinaciones aleatorias de los hiperparámetros para encontrar la mejor solución para el modelo construido. Dicho de otra manera, va recorriendo la matriz de manera aleatoria. Una optimización de esta búsqueda consiste en únicamente evaluar en un cierto número de configuraciones.

Este procedimiento es más viable que la búsqueda en cuadrícula sobre todo en dimensiones altas, debido a que el patrón de búsqueda hace posible que haya una exploración más amplia del espacio de los hiperparámetros, en un menor número de iteraciones, ya que es poco probable que se repitan uno o varios valores.

En la Figura 9 se observa las diferencias que existen entre los patrones de búsqueda de ambos algoritmos.

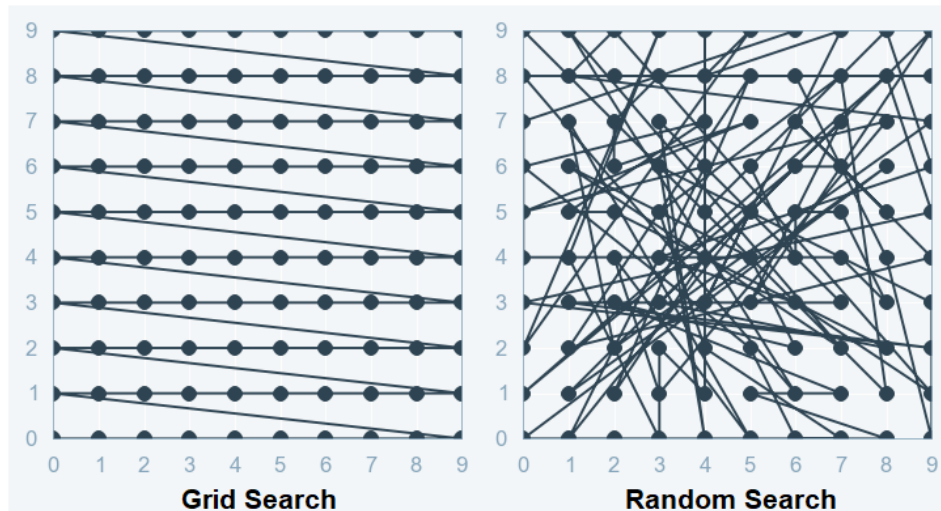


Figura 9 Búsqueda de hiperparámetros Grid vs Random search. Adaptado de [16]

Capítulo 4 . Metodología

En esta sección se detalla el proceso seguido para realizar el proyecto mediante el uso de la metodología CRISP-DM.

4.1 CRISP-DM

La metodología *CRISP-DM* (*Cross-Industry Standard Process for Data Mining*) proporciona una guía a seguir para un mejor desarrollo de un proyecto, independientemente del ámbito al que pertenece o de las herramientas tecnológicas de las que se dispone.

En la Figura 10 se puede ver las 6 fases de las que consta el modelo y resumen el ciclo de vida de un proyecto, relacionadas entre sí de manera secuencial y cíclica, de forma que en cada iteración se llega a una mejor aproximación al utilizar los resultados obtenidos en análisis anteriores [17].

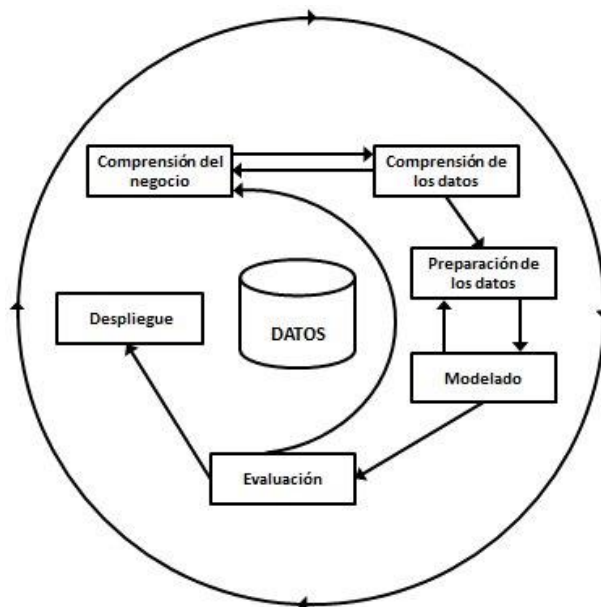


Figura 10 Fases del modelo CRISP-DM. Extraído de [17].

- **Comprensión del negocio:** El objetivo de esta fase es determinar los objetivos del negocio, es decir, cuál es el problema que se quiere resolver y los requisitos que deben cumplirse, así como especificar los recursos de los que se dispone, criterios de éxito, potenciales beneficios y utilidad desde un punto de vista empresarial, y los posibles riesgos que pueden surgir a la hora de llevarlo a cabo.

- **Comprensión de los datos:** El objetivo de esta fase consiste en conseguir los datos con los que se quiere trabajar, verificar su calidad y detectar relaciones que puedan servir para establecer unas primeras hipótesis.

- **Preparación de los datos:** Esta fase tiene como objetivo transformar los datos para que sirvan como datos de entrada para la fase de modelado. Las principales tareas que se realizan son la selección, limpieza y transformación de los datos.

- **Modelado:** En esta fase se seleccionan las técnicas y algoritmos de aprendizaje que se van a utilizar según el problema, cada una de ellas con requisitos diferentes en cuanto a la preparación de los datos. Además, es posible que se tenga que realizar varios ciclos de preparación de datos para conseguir el que mejor se adapte al algoritmo utilizado.

- **Evaluación:** Durante esta fase, una vez se ha construido el modelo, se evalúa y se revisa para verificar que cumple con los objetivos propuestos y si existen problemas que deben considerarse.

- **Despliegue:** En esta fase se organiza el conocimiento obtenido y se presenta de forma que sea útil desde el punto de vista del negocio, ya sea integrando la aplicación o utilizando la información para implementar mejoras en la organización.

4.2 Comprensión del proyecto

El objetivo del proyecto es desarrollar un sensor de software capaz de estimar el contenido de materia seca del jugo obtenido mediante el proceso de evaporación de un evaporador de múltiple efecto. Para lograr este propósito, se utilizan modelos de aprendizaje automático y/o estadísticos, con el fin de predecir el Brix en la salida del último efecto. El criterio de rendimiento se establece en base a la calidad del ajuste de los modelos, lo cual garantiza la precisión y la fiabilidad de los resultados. Es relevante mencionar que, para completar el trabajo, el proyecto dispone únicamente de un recurso humano y un conjunto de datos ya adquirido.

4.3 Comprensión de los datos

4.3.1 Captura Datos Iniciales

En cuanto a la adquisición de los datos iniciales, es importante mencionar que se dispone únicamente de una única fuente para obtenerlos. Asimismo, cabe resalta que esta fuente fue incluida dentro de los recursos proporcionados al inicio del proyecto, por lo que no ha generado ningún inconveniente en esta fase.

4.3.2 Descripción de los Datos

El conjunto proporcionado corresponde a un archivo con extensión .csv que cuenta con un total de 34.560 observaciones y 51 variables extraídas del proceso de evaporación de un evaporador de 6 efectos, donde cada observación dista un segundo. Entre estas variables, se incluyen distintos indicadores que se consideran de interés, como la presión o la temperatura en cada uno de los efectos, así como una medida del Brix a la salida del último.

En la Tabla 4 se detalla información sobre las variables, que incluye su nombre, orden en el conjunto, unidad de medida y una breve descripción. Las casillas están coloreadas para indicar que miden la misma magnitud en un punto específico del sistema y para una misma sustancia o elemento del proceso. Esto facilita su comprensión al relacionarlas visualmente entre sí.

	Nombre	Descripción	Unidad de Medida	
1	<i>Output</i>	Contenido de materia seca	g /100 g	Brix a la salida del sexto efecto
2	<i>P_vah_e1</i>	Presión del vapor (efecto 1)	bar	Presión del vapor a la salida de cada efecto
3	<i>P_vah_e2</i>	Presión del vapor (efecto 2)	bar	
4	<i>P_vah_e3</i>	Presión del vapor (efecto 3)	bar	
5	<i>P_vah_e4</i>	Presión del vapor (efecto 4)	bar	
6	<i>P_vah_e5</i>	Presión del vapor (efecto 5)	bar abs	
7	<i>P_vah_e6</i>	Presión del vapor (efecto 6)	bar abs	
8	<i>T_ent_e1</i>	Temperatura del jugo (entrada)	°C	Temperatura

		del efecto 1)		del jugo a la entrada
9	T_{ent_R10}	Temperatura del jugo (entrada cambiador de calor R10)	°C	Temperatura del jugo a la entrada de cambiadores de calor previos a la entrada del jugo al primer efecto
10	T_{ent_R13}	Temperatura del jugo (entrada cambiador de calor R13)	°C	
11	T_{ent_R14}	Temperatura del jugo (entrada cambiador de calor R14)	°C	
12	T_{ent_R3}	Temperatura del jugo (entrada cambiador de calor R3)	°C	
13	T_{ent_R3B}	Temperatura del jugo (entrada cambiador de calor R3B)	°C	
14	T_{ent_R8}	Temperatura del jugo (entrada cambiador de calor R8)	°C	
15	T_{ent_R9}	Temperatura del jugo (entrada cambiador de calor R9)	°C	
16	$T_{jugo_anteevap}$	Temperatura del jugo (salida del tanque de jugo)	°C	Temperatura salida del tanque que acumula el jugo antes de la entrada. Idéntica a T_{ent_R10}
17	T_{salida_R10}	Temperatura del jugo (salida cambiador de calor R10)	°C	Temperatura del jugo a la salida de cambiadores de calor previos a la entrada del jugo al primer efecto
18	T_{salida_R11}	Temperatura del jugo (salida cambiador de calor R11)	°C	
19	T_{salida_R12}	Temperatura del jugo (salida cambiador de calor R12)	°C	
20	T_{salida_R3}	Temperatura del jugo (salida cambiador de calor R3)	°C	
21	T_{salida_R3A}	Temperatura del jugo (salida cambiador de calor R3A)	°C	
22	T_{salida_R3B}	Temperatura del jugo (salida cambiador de calor R3B)	°C	
23	T_{salida_R4}	Temperatura del jugo (salida cambiador de calor R4)	°C	
24	T_{salida_R5}	Temperatura del jugo (salida cambiador de calor R5)	°C	
25	T_{salida_R6}	Temperatura del jugo (salida cambiador de calor R6)	°C	
26	T_{salida_R7}	Temperatura del jugo (salida cambiador de calor R7)	°C	
27	T_{salida_R8}	Temperatura del jugo (salida cambiador de calor R8)	°C	
28	T_{salida_R9}	Temperatura del jugo (salida cambiador de calor R9)	°C	
29	T_{vah_e1}	Temperatura del vapor (efecto 1)	°C	
30	T_{vah_e2}	Temperatura del vapor (efecto 2)	°C	

31	T_{vah_e3}	Temperatura del vapor (efecto 3)	°C	
32	T_{vah_e4}	Temperatura del vapor (efecto 4)	°C	
33	T_{vah_e5}	Temperatura del vapor (efecto 5)	°C	
34	T_{vah_e6}	Temperatura del vapor (efecto 6)	°C	
35	T_{vap_cald}	Temperatura del vapor de las calderas	°C	Temperatura del "Fresh steam" (vapor de calderas)
36	W_{jugo_R10}	Flujo másico (cambiador de calor R10)	t / h	Flujo másico del jugo a través de los cambiadores de calor previos a la entrada del jugo al primer efecto
37	W_{jugo_R3}	Flujo másico (cambiador de calor R3)	t / h	
38	W_{jugo_R3B}	Flujo másico (cambiador de calor R3B)	t / h	
39	W_{jugo_R4}	Flujo másico (cambiador de calor R4)	t / h	
40	W_{jugo_R8}	Flujo másico (cambiador de calor R8)	t / h	
41	W_{jugo_R9}	Flujo másico (cambiador de calor R9)	t / h	
42	$W_{jugo_anteevap}$	Flujo másico (tanque del jugo)	t / h	Flujo másico a la salida del tanque que acumula el jugo antes de la entrada
43	$W_{jugo_salida_e6}$	Flujo másico (salida efecto 6)	t / h	Flujo másico a la salida del sexto efecto
44	W_{cap_cald}	Flujo másico del vapor de las calderas	t / h	Flujo másico de "Fresh steam" (vapor de calderas)
45	$e1_T_jugo_salida$	Temperatura del jugo (efecto 1)	°C	Temperatura de jugo a la salida de cada efecto
46	$e2_T_jugo_salida$	Temperatura del jugo (efecto 2)	°C	
47	$e3_T_jugo_salida$	Temperatura del jugo (efecto 3)	°C	
48	$e4_T_jugo_salida$	Temperatura del jugo (efecto 4)	°C	
49	$e5_T_jugo_salida$	Temperatura del jugo (efecto 5)	°C	
50	$e6_T_jugo_salida$	Temperatura del jugo (efecto 6)	°C	
51	p_{vap_cald}	Presión del vapor de las calderas	bar	Presión de "Fresh steam"

Tabla 4 Descripción de las variables del conjunto de datos

En primer lugar, en amarillo se encuentra la variable respuesta de los modelos generados, que está relacionada con el contenido de materia seca del jugo a la salida del último efecto y se mide en unidades Brix.

En segundo lugar, en verde claro se encuentran las variables relacionadas con la presión del vapor en los efectos. Este grupo incluye las variables que miden la presión del vapor a la salida de cada uno de los efectos, como "P_vah_e1", "P_vah_e2", "P_vah_e3", "P_vah_e4", "P_vah_e5", "P_vah_e6" y del *Fresh steam* "p_vap_cald". Hay dos medidas distintas de presión: la presión absoluta, que se mide en relación a la presión de referencia 0 (la presión del vacío en el espacio exterior), y la presión relativa, que se mide en relación a la presión atmosférica, que tiene un valor de 1,01325 bar. La relación entre ambas se puede definir de la siguiente manera:

$$\text{Presión Absoluta} = \text{Presión relativa} + 1.01325 \text{ Bar} \quad (26)$$

Solo las variables "P_vah_e5" y "P_vah_e6" miden la presión absoluta, mientras que el resto de las variables miden la presión relativa.

En tercer lugar, en azul se encuentran variables relacionadas con la temperatura del vapor en los efectos. Este grupo incluye las variables que miden la temperatura del vapor a la salida de cada uno de los efectos, como "T_vah_e1", "T_vah_e2", "T_vah_e3", "T_vah_e4", "T_vah_e5", "T_vah_e6" y del *Fresh steam* "T_vap_cald" medidas todas en °C.

En cuarto lugar, en naranja se encuentran variables relacionadas con la temperatura del jugo en los efectos. Este grupo incluye las variables que miden la temperatura del jugo a la salida de cada uno de los efectos "e1_T_jugo_salida", "e2_T_jugo_salida", "e3_T_jugo_salida", "e4_T_jugo_salida", "e5_T_jugo_salida", "e6_T_jugo_salida" y a la entrada del primero "T_ent_e1" medidas todas en °C.

En quinto lugar, en gris se encuentran variables relacionadas con la temperatura del jugo en otros puntos del sistema. Este grupo incluye variables que miden la temperatura del jugo en la entrada de cambiadores de calor "T_ent_R10", "T_ent_R13", "T_ent_R14", "T_ent_R3", "T_ent_R3B", "T_ent_R8", "T_ent_R9", en la salida "T_salida_R10", "T_salida_R11", "T_salida_R12", "T_salida_R3", "T_salida_R3A", "T_salida_R3B", "T_salida_R4", "T_salida_R5", "T_salida_R6", "T_salida_R7", "T_salida_R8", "T_salida_R9" y en la salida del tanque que lo acumula hasta que entra al primer efecto "T_jugo_anteevap", medidas todas en °C.

Por último, en morado se encuentran variables relacionadas con el flujo másico (magnitud que expresa la variación de la masa con respecto al tiempo en un área específica). Este grupo incluye variables relacionadas con el jugo en distintos puntos como en los cambiadores de calor "W_jugo_R10", "W_jugo_R3", "W_jugo_R3B", "W_jugo_R4", "W_jugo_R8", "W_jugo_R9", en el tanque que lo almacena hasta que entra al primer efecto "W_jugo_anteevap" y a la salida del sexto efecto "W_jugo_salida_e6". Además, también tiene una variable que mide el flujo másico del *Fresh Stream*, medidas todas en Tn/h.

Después de haber descrito las variables, es esencial resaltar las zonas del sistema y las sustancias o elementos del proceso en las que se realizan mediciones y que sirven para relacionar las diferentes variables. Las mediciones se llevan a cabo en diferentes zonas del sistema como en los cambiadores de calor, el tanque de acumulación del jugo antes de ingresar al primer efecto y la salida de cada uno de los efectos. Asimismo, las sustancias y elementos del proceso en los que se hacen las mediciones son el jugo, el vapor de los efectos y el *Fresh Steam*.

4.3.3 Exploración de los datos

Para descubrir características o relaciones que resulten útiles para etapas posteriores es necesario realizar una exploración de los datos disponibles. Dentro de este análisis es esencial identificar la variabilidad existente, especialmente en este tipo de datos producidos en entornos industriales complejos. Hay que tener en cuenta que, debido a las particularidades y

variabilidad inherente de estos datos, no se pueden esperar resultados exactamente iguales a los de otro momento o contexto. En consecuencia, es esencial determinar si la variabilidad se debe a causas comunes o causas especiales para poder ajustar el proceso industrial y mejorar el rendimiento y la calidad.

Las causas comunes son aquellas que generan una variabilidad natural y constante en el proceso, lo que se traduce en un patrón de variabilidad estable y predecible que puede ser medido a través de la capacidad del proceso. Si solo existen causas comunes, se considera que el proceso se encuentra bajo control estadístico. Por otro lado, en el caso de las causas especiales, la variabilidad no es inherente al proceso y puede generar un patrón de variabilidad errático e impredecible que puede hacer que el proceso se salga del control estadístico. Por lo tanto, es importante identificar y eliminar estas causas especiales para que el proceso pueda volver a estar bajo control estadístico [18].

Una herramienta gráfica que puede utilizarse es el gráfico de Shewhart [19] de las variables medidas y monitorizadas durante el proceso. Este gráfico muestra la magnitud en el eje vertical y el tiempo en el eje horizontal. Además, incluye varias líneas horizontales que se basan en la distribución de muestreo, como la Línea Central (LC) que muestra la media (μ), y los límites de control que se utilizan para detectar posibles causas especiales en los datos: el Límite de Control Inferior (LCI) y el Límite Control Superior (LCS).

Existen diversos criterios para determinar los límites de control, siendo uno de los más empleados el siguiente:

$$\begin{aligned} LCI &= LC + 3 \sigma \\ LCS &= LC - 3 \sigma \end{aligned} \tag{27}$$

Donde σ es la desviación estándar de la distribución.

La Figura 11 representa el gráfico de Shewhart, donde las líneas horizontales en rojo representan LCI y LCS , mientras que la línea horizontal en verde representa LC . Si los datos se encuentran dentro de los límites, el proceso está bajo control estadístico, lo que sugiere que la variabilidad del proceso se debe a causas comunes. Sin embargo, si existen puntos fuera de los límites, como los resaltados con un círculo, el proceso está fuera de control debido a la presencia de una o varias causas especiales.

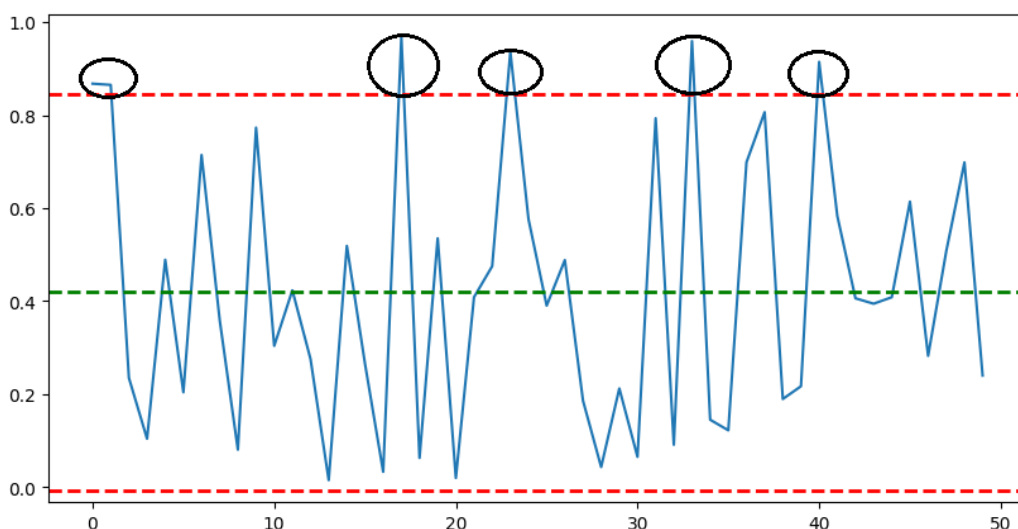


Figura 11 Ejemplo de gráfico de Shewhart

La Figura 12 muestra los gráficos de Shewhart correspondientes a las variables de más interés y que resumen el estudio de la variabilidad. Se identifican muchos puntos que sobresalen de los límites de control en variables como “P_vah_e5” y “P_vah_e6”, lo que sugiere la presencia de causas especiales de variación, como posible ruido en los sensores. Además, se observan valores medidos fuera de los límites en determinados momentos puntuales y que parecen ser recurrentes en muchos de los intercambiadores como por ejemplo “T_entrada_R10”, “T_salida_R10” y “T_entrada_R13”. Por último, se identifican variables como “W_jugo_R10” y “W_jugo_R3” que muestran un comportamiento similar al de las variables “P_vah_e5” y “P_vah_e6”, pero en este caso, las observaciones se alejan considerablemente de los límites de control establecidos, lo que indica que las causas especiales de variación presentes en estos casos son más significativas.

Después de discutir los resultados con un experto se llegó a la conclusión de que los puntos fuera de los límites de control se deben a condiciones normales del proceso, tales como caídas de presión debido a la demanda en otras zonas de la planta o disminuciones de temperatura del jugo al aplicarse agua fría para evitar la cristalización. En el *Anexo II – Gráficos de Shewhart* se muestran los gráficos de Shewhart para todas las variables.

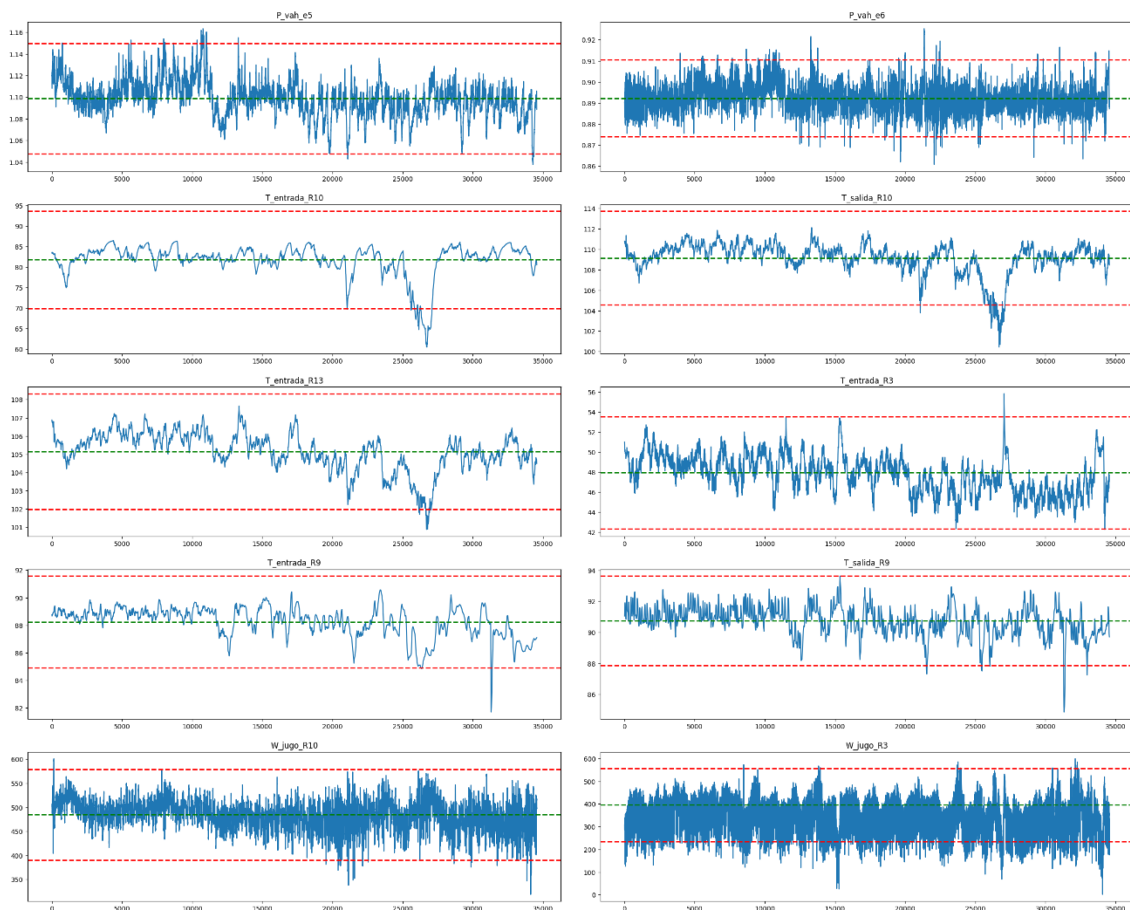


Figura 12 Gráficos de Shewhart

Otro análisis importante es el estudio de la normalidad de los datos, ya que esto puede mejorar significativamente la precisión del modelo y facilitar la interpretación de los resultados. Además, es un complemento útil para comprender los resultados de los gráficos de Shewhart, ya que estos se basan en la suposición de que los datos siguen una distribución normal. Para un estudio formal de la normalidad se recomienda utilizar pruebas de bondad de ajuste específicas, tales como el test de Kolmogórov-Smirnov [20], para verificar la normalidad de los datos de manera rigurosa:

$$D = \text{máx} | F_n(x) - F_0(x) | \quad (28)$$

Donde $F_n(x)$ es la función de distribución muestral y $F_0(x)$ es la función teórica de la población normal esperada en la hipótesis nula. Este estadístico puede servir para contrastar que los datos siguen otras distribuciones cambiando $F_0(x)$.

La Tabla 5 muestra los resultados del test en algunas de las variables en donde se puede observar que no siguen una distribución normal, debido al valor significativo del p-valor ($p < 0.05$) en todas ellas. Al rechazar la hipótesis nula de que las variables se distribuyen normalmente, se sugiere que las distribuciones no son normales en el conjunto de datos completo.

	Nombre	D	p-valor
1	Output	1.000000	0.0
2	P_vah_e1	0.886979	0.0
3	P_vah_e2	0.794119	0.0
...
51	p_vap_cald	0.927968	0.0

Tabla 5 Test de Kolmogórov-Smirnov en los datos

Para tomar decisiones más sólidas acerca de la normalidad de los datos, se recomienda utilizar varios métodos de análisis. Además de la prueba de normalidad de Kolmogórov-Smirnov o la prueba de Shapiro-Wilk [20], es útil complementar el análisis con gráficos de probabilidad como el gráfico Q-Q. Este gráfico permite comparar la distribución de los datos con una distribución normal, mostrando en el eje vertical los valores observados y en el eje horizontal los valores teóricos. Si los datos siguen una distribución normal, se espera que los puntos se aproximen a una línea recta. En la Figura 13 se muestra el gráfico Q-Q normal de la variable "P_vah_e1", donde se puede observar una desviación de los datos de la línea de distribución normal. Este resultado corrobora la prueba de Kolmogórov-Smirnov, la cual también indica la no normalidad de los datos. En el Anexo III – Q-Q normal se incluyen gráficos Q-Q normal similares para el resto de las variables con resultados análogos que sugieren que los datos de las demás variables podrían también no ser normales.

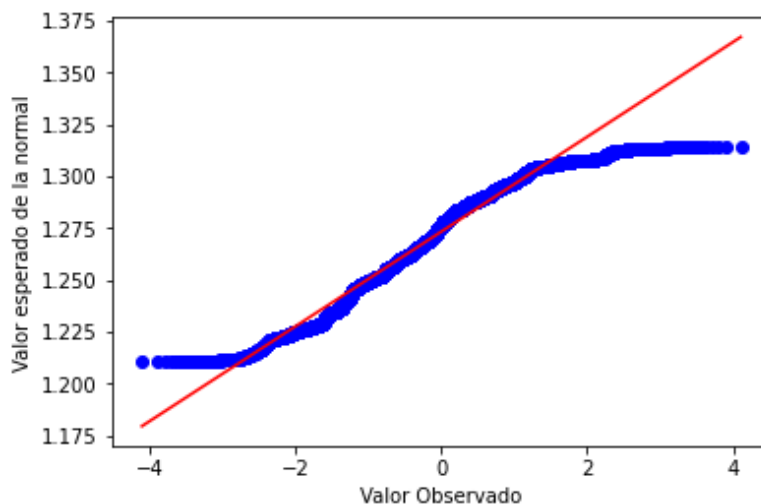


Figura 13 Q-Q normal de la variable P_vah_e1

En resumen, se puede concluir que los datos no siguen una distribución normal, lo cual indica que la falta de cumplimiento del supuesto de normalidad puede ser una de las razones por las cuales algunos puntos se encuentran fuera de los límites de control en los gráficos de Shewhart.

Finalmente, en el análisis previo, se encontraron varias variables repetidas. Por un lado, están las variables “*W_jugo_R3B*”, “*W_jugo_R4*”, “*W_jugo_R8*” y “*W_jugo_R9*”, que representan flujos másicos a través del mismo intercambiador de calor. Por otro lado, están las variables “*W_jugo_R10*” y “*W_jugo_anteevap*”, que miden el flujo másico del jugo al entrar al primer efecto y al acumularse en el tanque anterior, respectivamente. De manera similar, las variables “*T_entrada_R10*” y “*T_jugo_anteevap*” miden la temperatura del jugo en los mismos puntos mencionados anteriormente.

4.3.4 Verificación de la calidad de los datos

Una vez realizado el análisis de los datos es importante comprobar su calidad a través de diferentes comprobaciones entre las que se incluyen:

- Valores ausentes.
- Observaciones duplicadas.
- Errores ortográficos.
- Outliers.
- Ruido en los datos.

En relación con los tres primeros puntos, cabe destacar que no se han encontrado valores ausentes ni observaciones duplicadas. Además, la totalidad del conjunto procede de una única base de datos, lo que elimina cualquier riesgo de duplicación entre ellas. Por último, no se han detectado errores ortográficos en las variables numéricas, las cuales utilizan el punto “.” como separador decimal y no presentan ninguna otra representación que pudiera afectar el análisis.

Por otro lado, los gráficos de control son útiles para detectar outliers y obtener una visión general de las observaciones que se encuentran fuera de los límites de control y que podrían ser valores atípicos. Sin embargo, no todas las observaciones que sobrepasan estos límites son necesariamente valores atípicos. Algunas pueden ser causadas por la variabilidad aleatoria e impredecible del proceso lo que se conoce como “*ruido blanco*”, un fenómeno que puede deberse a errores de medición o de preprocesamiento de datos.

Por esta razón, se utilizan los boxplots, que proporcionan una imagen más clara de la presencia de outliers. La Figura 14 muestra las partes que conforman un boxplot. Es importante destacar que, aunque los extremos del boxplot no corresponden necesariamente a los límites de control establecidos en la sección de exploración, se utiliza el mismo factor de multiplicación de 3 para determinar los valores que se alejan más de los cuartiles y que, por lo tanto, podrían ser considerados valores atípicos.

En el contexto de los procesos industriales, donde la variabilidad de los datos es alta, es importante ser menos restrictivo en la detección de valores atípicos para evitar descartar información relevante. En consecuencia, se ha elegido un factor de multiplicación de 3 tanto para los límites de control como para la construcción del boxplot.

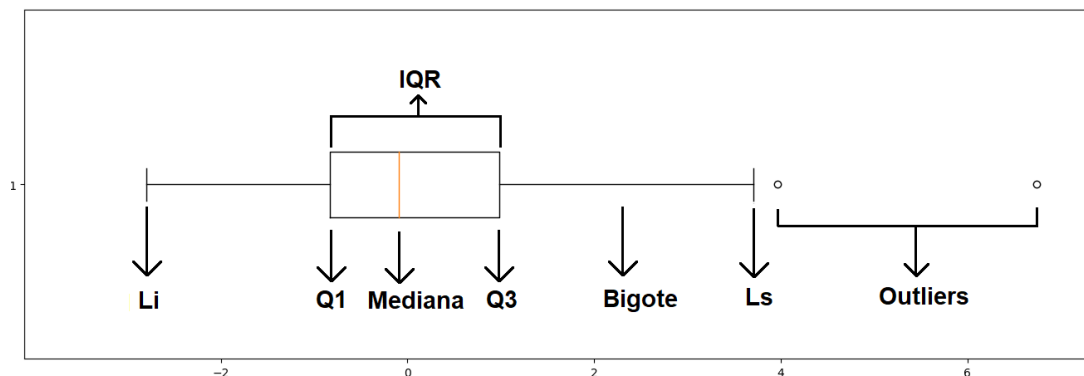


Figura 14 Representación partes de un boxplot.

Donde Q1 es el primer cuartil y Q3 es el tercer cuartil que corresponden a los percentiles 25 y 75 respectivamente, IQR es el rango intercuartílico que se calcula como la diferencia entre el tercer cuartil y el primer cuartil (Q3 - Q1). En cuanto a los límites superior (L_s) e inferior (L_i) se construyen utilizando la siguiente información:

$$\begin{aligned} L_s &= Q3 + 3 IQR \\ L_i &= Q1 - 3 IQR \end{aligned} \quad (29)$$

En la Figura 15 se muestran los boxplots de las mismas variables que aparecen en la Figura 12. Al comparar ambas figuras, se puede apreciar una clara relación entre los outliers y los puntos que se encuentran fuera de los límites de control en la gran mayoría de las variables. Por ejemplo, en la variable " $T_{entrada_R10}$ ", los outliers se encuentran únicamente a la izquierda del L_i , y en el gráfico se observa una zona donde las observaciones se escapan del LCI. Esta relación sugiere que muchos de los outliers presentes en algunas variables no son realmente atípicos, sino que son el resultado de condiciones normales del propio proceso, tal como indicó el experto. Sin embargo, es importante tener en cuenta que la falta de normalidad en los datos puede también estar influyendo en este resultado, dado que los boxplots se construyen a partir del test de Tukey [21], el cual asume una distribución normal de los datos.

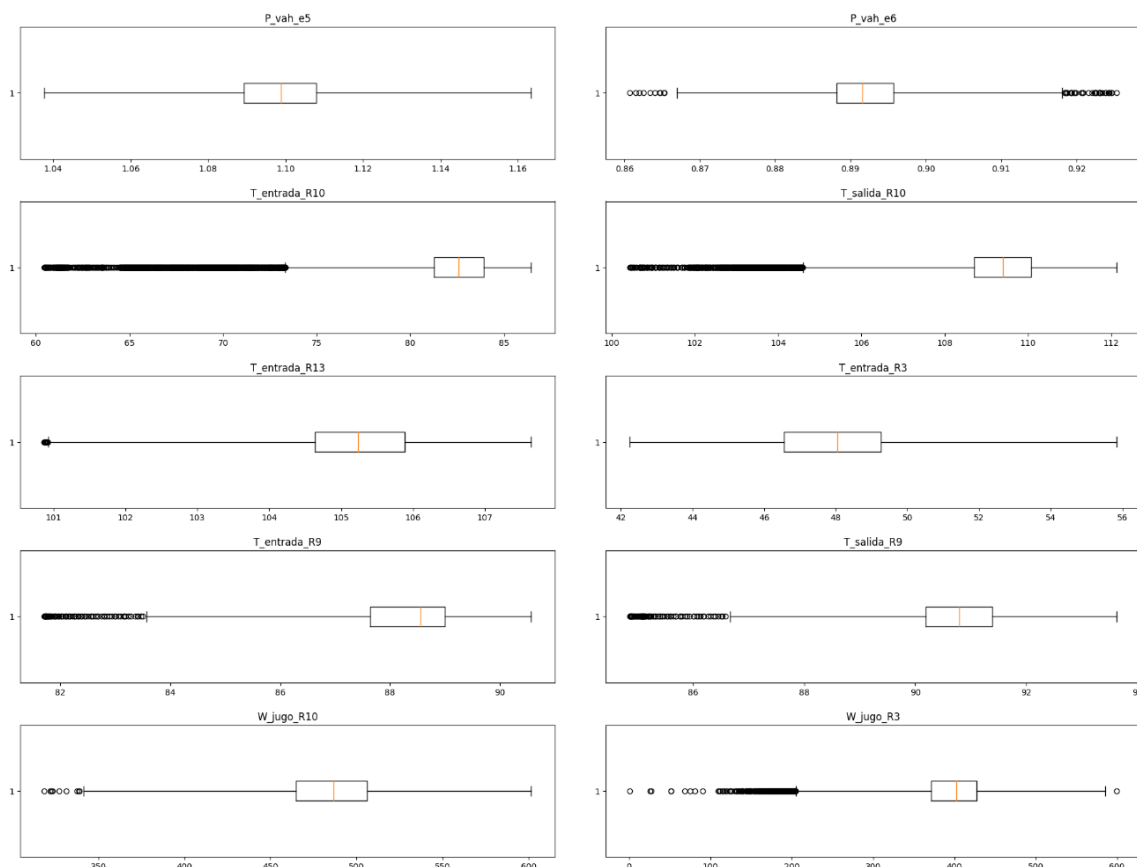


Figura 15 Boxplot de algunas variables.

En conclusión, a pesar de que el porcentaje elevado de outliers en variables como " $T_{entrada_R10}$ " y " T_{salida_R10} ", que representan el 5.11% y el 2.61% respectivamente, pueda atribuirse a condiciones normales, el incumplimiento del supuesto de normalidad en el test plantea dudas sobre la validez de los resultados. Por lo tanto, se decide repetir el experimento una vez realizadas todas las transformaciones de los datos. En el Anexo IV – *Boxplots* se muestran los boxplots para todas las variables.

Por último, para detectar la presencia de ruido en los datos, tanto los gráficos de Shewhart como los boxplot pueden proporcionar una idea sobre la variabilidad aleatoria en los datos. Sin

embargo, la adición del análisis de un histograma puede ayudar a identificar la distribución de frecuencia de los datos y detectar patrones de variabilidad y ruido. En la Figura 16 se muestran algunas de las variables que se consideran de más interés a la hora de analizar el ruido junto con una línea vertical que indica la media y una curva que simula una distribución normal de los datos. En el Anexo V – *Histogramas* se presenta el histograma del resto de variables.

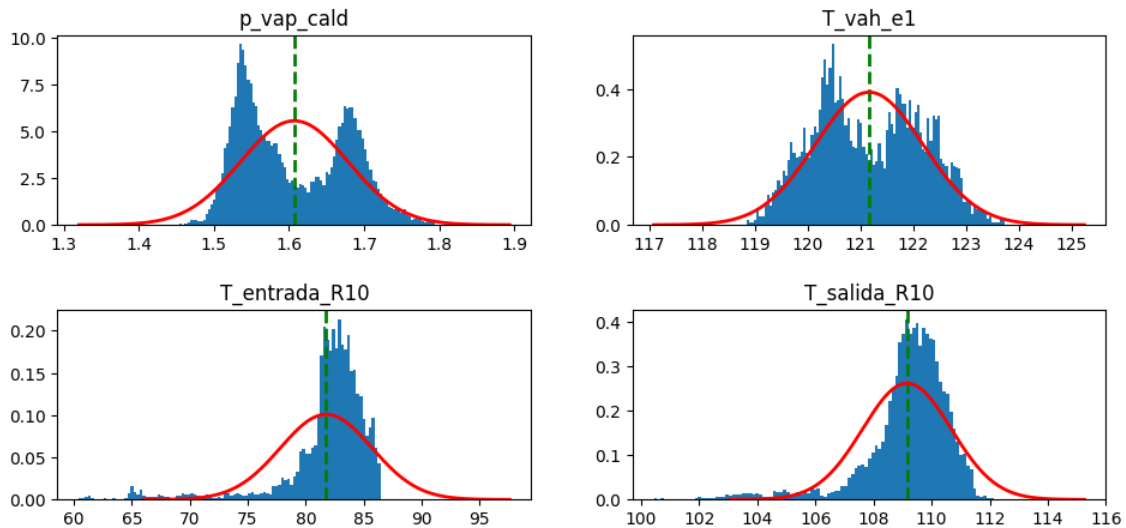


Figura 16 Histogramas de algunas variables

Al analizar la distribución de los datos, es importante considerar que los histogramas no siempre se ajustan a una distribución normal. Esto respalda los resultados del estudio de normalidad realizado en la sección de exploración, ya que indica que los datos no se distribuyen simétricamente alrededor de la media, lo que sugiere una alta variabilidad. En algunos casos, se pueden observar distribuciones sesgadas o bimodales, lo que sugiere la presencia de subgrupos de datos.

En particular, la presencia de una distribución bimodal en variables como "*p_vap_cald*" y "*T_vah_e1*" puede indicar la existencia de dos subprocesos distintos, posiblemente como resultado de operaciones que tuvieron lugar durante el proceso, como la demanda de vapor por parte de otras zonas de la planta o la posibilidad de que existan dos puntos de funcionamiento distintos. Esto refuerza las conclusiones observadas en los gráficos de Shewhart, donde se identifica una disminución puntual en los valores.

Por otro lado, en variables como "*T_entrada_R10*" y "*T_salida_R10*", se puede observar una distribución sesgada que también puede atribuirse a operaciones específicas del proceso. Esto, al igual que en el caso anterior, se refleja en los gráficos de Shewhart con una disminución puntual de los valores, aunque de manera más abrupta. Además, se corresponde con la presencia de outliers extremos que se acumulan exclusivamente en el extremo izquierdo L_i de los boxplots.

En resumen, los resultados obtenidos de los histogramas respaldan la idea de que las operaciones realizadas durante el proceso contribuyen a un aumento en la variabilidad, lo cual se traduce en la presencia de ruido en los datos. Además, es importante tener en cuenta que el ruido también puede originarse a partir de las mediciones realizadas por los propios sensores.

4.4 Preparación de los datos

En esta fase se construye el conjunto de datos a partir de los datos iniciales que se utiliza para la elaboración de los modelos.

4.4.1 Limpieza de datos

En el apartado de verificación de la calidad de los datos, se identifica que el proceso tiene una alta variabilidad, por lo tanto, es necesario limpiar los datos para reducir la variabilidad de los datos.

Entre las medidas implementadas, la primera consiste en aplicar un filtro de Butterworth [22] para conseguir suavizar la señal y reducir la variabilidad del conjunto de variables. El filtro de Butterworth es ampliamente utilizado debido a su capacidad para permitir que las frecuencias de interés pasen sin una atenuación significativa (banda de paso), al mismo tiempo que atenúa o suprime las frecuencias no deseadas (bandas de atenuación).

La aplicación de este filtro permite una visualización más clara de las relaciones y patrones entre las variables. En la Figura 17 se muestran las gráficas de control para las variables analizadas en el apartado de exploración, comparando los datos con y sin la aplicación del filtro. Se puede observar cómo el filtro atenúa la señal y en la mayoría de las variables se reduce la variabilidad, lo que se traduce en una mayor estabilidad de los datos respecto a los límites de control. Sin embargo, en las variables “*T_entrada_R10*” y “*T_salida_R10*”, los valores extremos son tan altos que el filtro no es suficiente para reducir la variabilidad y los datos todavía se escapan de los límites de control. En el *Anexo VI – Filtrado de Variables* se presenta el gráfico para el resto de las variables.

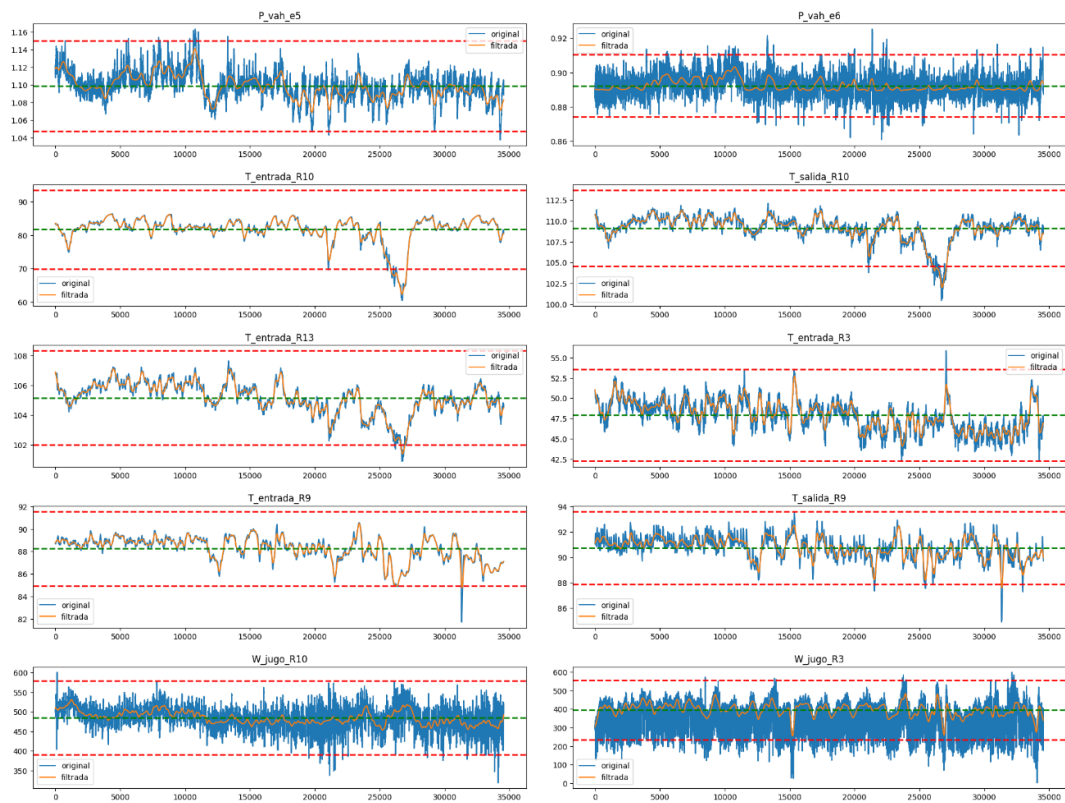


Figura 17 Gráficos de control con la aplicación del filtro.

La aplicación del filtro de suavizado puede afectar la identificación de outliers. Al reducir la variabilidad y suavizar los datos, algunas observaciones que antes se consideraban outliers pueden acercarse más a la media y dejar de ser consideradas como tales. Por otro lado, otras observaciones pueden desviarse de manera más significativa y considerarse outliers después de aplicar el filtro. En definitiva, el número de outliers puede aumentar o disminuir después de la aplicación de un filtro, lo que proporciona información sobre la distribución y naturaleza de las variables de estudio. Sin embargo, antes de proceder a su estudio es necesario realizar otras transformaciones.

En primer lugar, se convierten las medidas de presión absoluta "*P_vah_e5*" y "*P_vah_e6*" a presión relativa, para homogeneizar las unidades en todas las presiones de los efectos. También se ha aplicado un filtro Butterworth para reducir el ruido en las variables.

Por último, se considera la necesidad de normalizar los datos, ya que las variables se miden en diferentes unidades y escalas, lo que puede resultar en un mayor peso en el aprendizaje de variables con valores más altos y mayor varianza. Además, al normalizar los datos, se puede realizar el análisis sin violar el supuesto de normalidad que algunos de los gráficos utilizados asumen. Para lograrlo, existen varias técnicas de normalización, como la normalización Min-Max o la Z-score [23].

La normalización *Min-Max* consiste en cambiar la escala de los datos para que todos los valores estén dentro del intervalo [0,1]. Esto se consigue restando el valor mínimo a todos los valores y dividiendo el resultado entre la diferencia del valor máximo y el valor mínimo:

$$x_{nuevo} = \frac{x - \min}{\max - \min} \quad (30)$$

La normalización *Z-score* en escalar la distribución de los valores de forma que la media (μ) sea 0 y la desviación estándar (σ) 1. Esto se consigue al restar la media y dividir entre la desviación estándar.

$$x_{nuevo} = \frac{x - \mu}{\sigma} \quad (31)$$

Una vez aplicadas todas las transformaciones, se repite el experimento de detección de outliers teniendo en cuenta que ahora se cumple con el supuesto de normalidad del test de Tukey. En la Tabla 6 se muestran los porcentajes de outliers para las variables que presentan alguno antes y/o después de aplicar las transformaciones (se excluyen algunas variables idénticas, como se explica en la siguiente sección). Al realizar las transformaciones se detecta que la mayoría de las variables ya no muestran outliers, a excepción de "*T_entrada_R10*" y "*T_salida_R10*", donde incluso se observa un incremento. Estos resultados refuerzan el análisis anterior de que los datos no son atípicos, sino más bien anómalos a causa de alguna operación del proceso. Por lo tanto, se concluye que no existen outliers en los datos y no es necesario realizar ningún tratamiento.

	Porcentaje outliers Primer Análisis	Porcentaje outliers Segundo Análisis
<i>P_vah_e6</i>	0.13	0.00
<i>T_entrada_R10</i>	5.11	5.19
<i>T_entrada_R13</i>	0.06	0.00
<i>T_entrada_R9</i>	0.21	0.00
<i>T_salida_R10</i>	2.61	3.38
<i>T_salida_R11</i>	0.12	0.00
<i>T_salida_R9</i>	0.25	0.00
<i>W_jugo_R10</i>	0.03	0.00
<i>W_jugo_R3</i>	1.05	0.00
<i>W_jugo_R3B</i>	0.38	0.00

Tabla 6 Porcentaje de outliers en los dos análisis.

4.4.2 Selección de atributos

Después del tratamiento de datos, es necesario seleccionar solo algunos atributos del conjunto, ya que el conjunto inicial contiene una gran cantidad de variables. Utilizar toda la información disponible genera modelos con un gran número de atributos, muchos de los cuales son irrelevantes o redundantes, lo que dificulta la interpretación del modelo. Además, cuantos más atributos se precisen, más aumenta tanto el costo computacional como el riesgo de sobreajuste.

Existen varios mecanismos de selección de variables, siendo el más sencillo la selección manual de variables. Por ejemplo, en este caso, se seleccionan "W_jugoR3B" del grupo de variables "W_jugoR3B", "W_jugo_R4", "W_jugo_R8", "W_jugo_R9"; "W_jugo_R10" del grupo de variables "W_jugo_R10" y "W_jugo_anteevap"; y "T_entrada_R10" del grupo de variables "T_entrada_R10" y "T_jugo_anteevap" debido a que corresponden a variables repetidas. Además, se decide no eliminar las variables "T_entrada_R10" y "T_salida_R10" que se encuentran fuera de los límites de control, ya que se trata de condiciones normales del proceso que no van a afectar a la precisión de los modelos. Después de eliminar las variables mencionadas, el nuevo conjunto de variables se compone de un total de 46 variables.

En algunos casos, seleccionar manualmente las variables es una buena opción, pero puede resultar insuficiente cuando hay un gran número de atributos. Por lo tanto, existen otras formas de reducir la dimensionalidad del conjunto de datos utilizado en la minería de datos, como el Análisis de Componentes Principales (*PCA*, por sus siglas en inglés) [24].

PCA genera nuevas variables que son combinaciones lineales de las variables originales e independientes entre sí. Luego, estas nuevas variables se ordenan por el porcentaje de variabilidad total que recogen. Como resultado, se obtienen componentes altamente correlacionados con la variable respuesta, pero no correlacionados entre sí. Es importante destacar la normalización de las variables, ya que, si no están normalizadas, aquellas con valores de varianza más altos pueden tener un mayor impacto en los resultados del *PCA* y esto puede llevar a interpretaciones erróneas de los resultados obtenidos.

En la Tabla 7 se muestra el porcentaje de varianza explicada por cada componente y la acumulada. En este caso, se puede observar que 10 componentes son suficientes para explicar más del 95 % de la variabilidad total. Además, el componente 11 ya explica menos del 1 %, lo que sugiere que no es necesario incluirlo. En resumen, el *PCA* es una técnica útil para reducir la dimensionalidad del conjunto de datos sin perder demasiada información.

Componente	Porcentaje varianza explicada	Porcentaje varianza acumulada
1	59.86504	59.86504
2	1.296828	72.980505
3	6.167457	79.304171
4	4.765766	83.916503
5	2.723575	86.676103
6	2.337238	88.894778
7	2.021882	90.919431
8	1.871161	92.832369
9	1.429039	94.297211
10	1.297040	95.576799
11	0.9982183	96.471615
...
46	1.274198e-12	100.000000

Tabla 7 Porcentaje de varianza *PCA* sobre el conjunto completo de datos

4.5 Modelado y evaluación

Durante la fase de modelado del proyecto, se lleva a cabo la creación de modelos que permiten estimar el contenido de materia seca. Para ello, se utilizan tres tipos de modelos de aprendizaje automático: Red Neuronal, SVR y Random Forest Regression, junto con un modelo estadístico de regresión por mínimos cuadrados parciales.

El primer paso consiste en dividir el conjunto de datos en dos subconjuntos: uno de entrenamiento y otro de prueba. El conjunto de entrenamiento se compone del 80% de las primeras observaciones, mientras que el conjunto de prueba se compone del 20% de las observaciones finales. Es importante tener en cuenta que los datos provienen de medidas de un proceso y están ordenados en el tiempo, lo que implica que variables cercanas están altamente correlacionadas entre sí. Si se realiza la división de forma aleatoria sin considerar estas correlaciones, el modelo no es capaz de estimar correctamente nuevos datos.

El siguiente paso implica la normalización de las 47 variables explicativas del conjunto de entrenamiento que quedan tras selección manual de variables. Para ello, se utiliza la técnica de normalización Min-Max, que escala los valores al rango $[0,1]$ y se aplica los mismos parámetros para realizar la normalización por separado al conjunto de test. Realizar por separado la normalización es fundamental para evitar la sobreestimación del modelo, ya que utilizar información del conjunto de test durante el proceso de entrenamiento podría generar una evaluación optimista de su rendimiento. De igual manera, se aplica la normalización a la variable respuesta.

Una vez que se ha normalizado los conjuntos, se realiza el Análisis de Componentes Principales (PCA) mediante el cálculo y selección de los 10 primeros componentes principales basados solo en los datos de entrenamiento. A continuación, se proyectan los datos del conjunto de prueba en el mismo espacio de componentes principales, asegurando la coherencia y consistencia en la reducción de dimensionalidad entre ambos conjuntos.

Por otro lado, al entrenar un modelo de aprendizaje automático, es crucial ajustar los hiperparámetros correctamente para evitar el sobreajuste. Una técnica común utilizada para abordar este problema es la validación cruzada k-fold. En este caso, se utiliza la validación cruzada 5-fold, lo que significa que se divide el conjunto de entrenamiento en 5 partes iguales. En cada iteración del proceso, se entrena el modelo con 4 partes y se valida con la restante. De esta manera, se puede estimar el error de generalización del modelo y buscar la mejor combinación de hiperparámetros explorando todas las combinaciones posibles mediante la técnica de búsqueda en cuadrícula.

En Regresión por Mínimos Cuadrados Parciales también se puede utilizar la técnica de validación cruzada para seleccionar el número óptimo de componentes que equilibre tanto la complejidad del modelo como su capacidad de generalización. Además, esta técnica permite reducir la dimensionalidad del conjunto de datos sin necesidad de realizar el PCA. De hecho, Mínimos Cuadrados Parciales (PLS) se basa en PCA para calcular los componentes.

Una vez seleccionado el mejor modelo con validación cruzada se entrena este modelo con todos los datos de entrenamiento y por último se utilizan los datos del test para estimar el error cuadrático medio (MSE) teniendo en cuenta que los valores de la variable respuesta están en el rango $[0,1]$ consecuencia de la normalización. Todo el procedimiento anterior se puede resumir en la Figura 18.

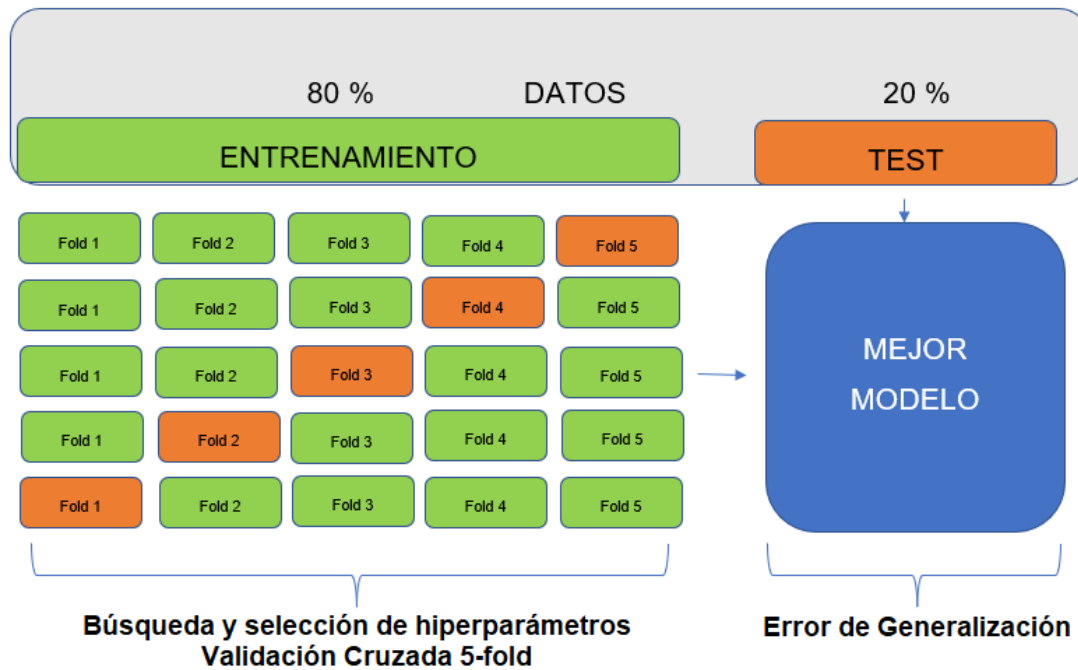


Figura 18 Resumen procedimiento. Adaptado de [8].

A continuación, se describen los modelos utilizados y los resultados obtenidos

Red Neuronal

La estructura propuesta para la red neuronal se muestra en la Figura 19. Se utilizan los 10 componentes extraídos de PCA como entradas. A continuación, se agregan una o más capas ocultas, cada una con un número "n" de neuronas. En la capa de salida de la red hay una única neurona, ya que se estima una única variable.

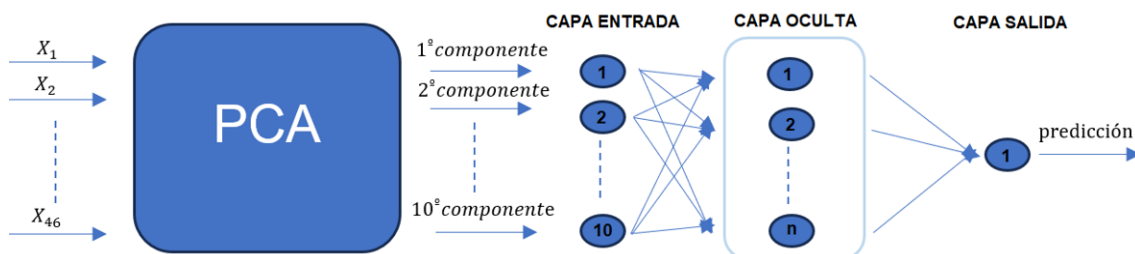


Figura 19 Red Neuronal: Estructura

Por lo tanto, el espacio de búsqueda de hiperparámetros incluye el número de capas ocultas, el número de neuronas por capa y se añade uno más, la tasa de aprendizaje. Los valores correspondientes se encuentran en la Tabla 8. Otros hiperparámetros se establecen previamente para evitar un espacio de búsqueda excesivamente grande. En este caso, se utiliza la función de activación ReLU en las capas ocultas y la función lineal en la capa de salida, junto con el optimizador Adam. Además, con el objetivo de reducir el tiempo de ejecución del algoritmo, las redes se entrenan durante 10 épocas.

Número de capas ocultas	{1,2,3}
Número de neuronas por capa	{5,10,15,20}
Tasa de aprendizaje	{0.001,0.01,0.1}

Tabla 8 Red Neuronal: Búsqueda de Hiperparámetros

Los resultados obtenidos muestran que el mejor modelo se obtiene con 3 capas ocultas, 20 neuronas por capa y una tasa de aprendizaje de 0.01.

- *MSE* en datos de entrenamiento: $5.38 \cdot 10^{-5}$ o $8.67 \cdot 10^{-3}$ con datos desnormalizados.
- *MSE* en datos de prueba: 0.01 o 1.64 con datos desnormalizados.

La diferencia entre los datos test y los datos de entrenamiento es muy significativa, siendo el error en los datos de prueba muchas veces mayor que en los datos de entrenamiento. Estos resultados indican que existe sobreajuste.

En las Figuras 20 y 21 se comparan los valores observados con los valores predichos por la red neuronal. En la Figura 20, se representan tanto los datos de entrenamiento como los del conjunto de test, separados por una línea roja, mientras que en la Figura 21 solo se muestran los datos del conjunto de test. Estos gráficos evidencian que el modelo sufre de sobreajuste, ya que las predicciones con los datos de entrenamiento son altamente precisas, pero estas empeoran considerablemente al utilizar los datos de test.

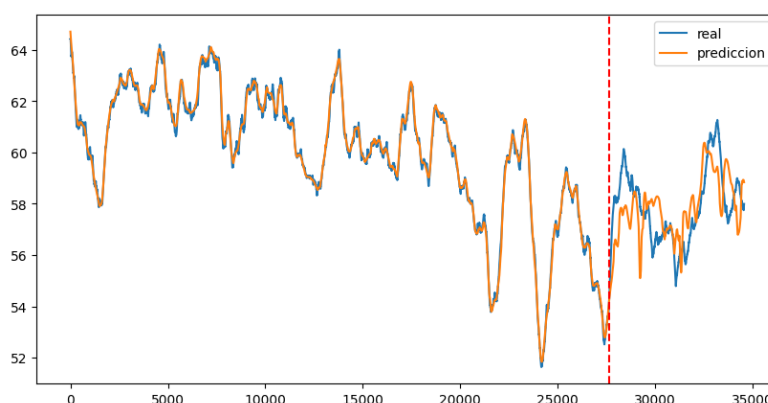


Figura 20 Red Neuronal: Valor observado VS Predicción: Datos Completos

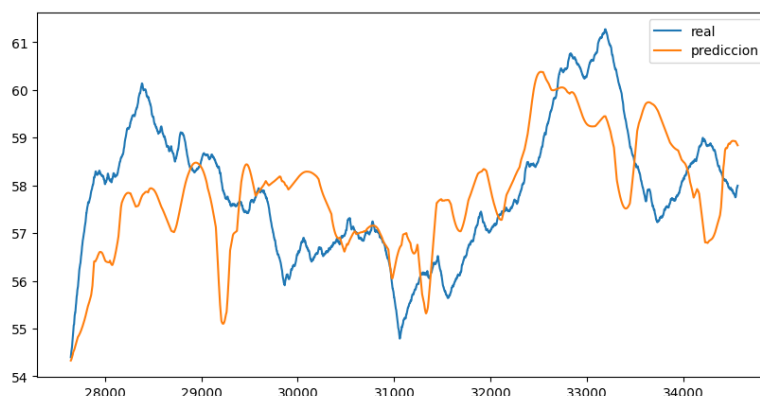


Figura 21 Red Neuronal: Valor observado VS Predicción: Datos Test

Al examinar los hiperparámetros, se observa que el sobreajuste es causado por un exceso tanto en el número de capas ocultas como en el número de neuronas en esas capas. Como solución, se decide repetir el experimento reduciendo el modelo a una o dos capas ocultas y limitando el número máximo de neuronas en dichas capas a 15. En la Tabla 9 se presentan el nuevo espacio de búsqueda de hiperparámetros.

Número de capas ocultas	{1,2}
Número de neuronas por capa	{5,10,15}
Tasa de aprendizaje	{0.001,0.01,0.1}

Tabla 9 Red Neuronal: Búsqueda de Hiperparámetros: Experimento 2

Ahora el mejor modelo que se obtiene con 2 capas ocultas, 5 neuronas en cada capa y una tasa de aprendizaje de 0.001.

- MSE en datos de entrenamiento: $9.97 \cdot 10^{-3}$ o 0.16 con datos desnormalizados.
- MSE en datos de prueba: $5.78 \cdot 10^{-2}$ o 0.93 con datos desnormalizados.

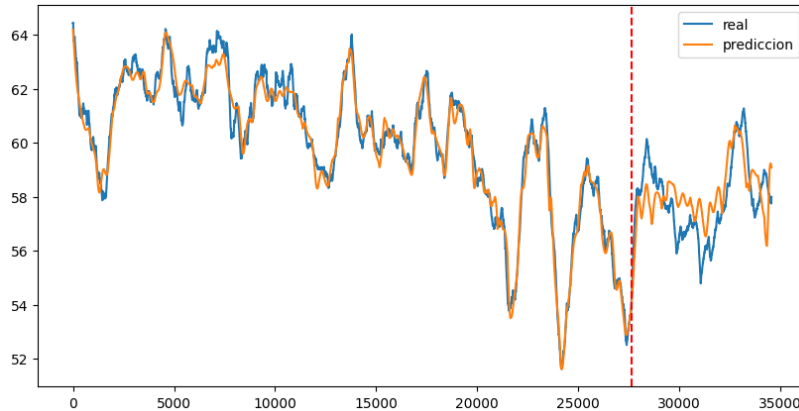


Figura 22 Red Neuronal: Valor observado VS Predicción: Datos Completos: Experimento 2

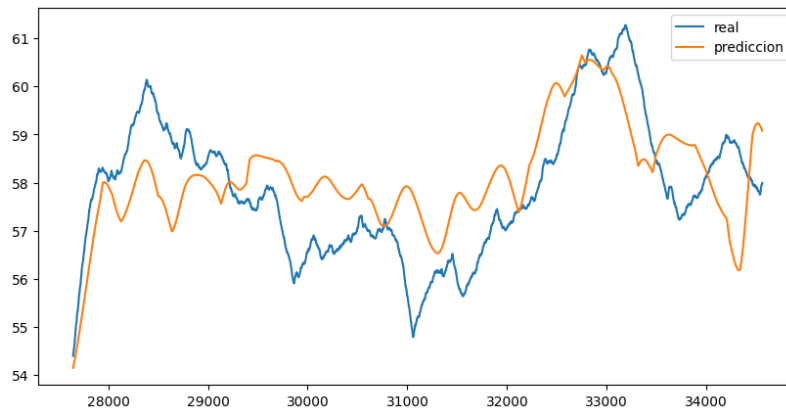


Figura 23 Red Neuronal: Valor observado VS Predicción: Datos Test: Experimento 2

En este segundo experimento, se observa que la diferencia entre los MSE se reduce considerablemente, además de obtener mejores predicciones en el conjunto de test. Esto significa que el modelo generaliza mejor y se reduce el sobreajuste.

SVR

La estructura propuesta para SVR se muestra en la Figura 24. Al igual que en el caso anterior se utilizan los 10 componentes extraídos de PCA como entradas.

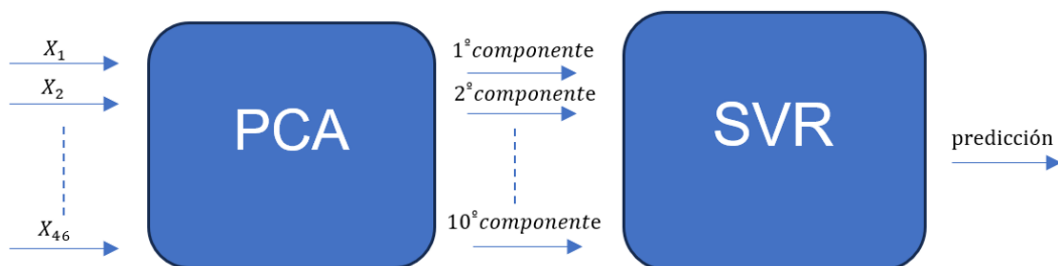


Figura 24 SVR: Estructura

La Tabla 10 muestra el espacio de búsqueda que explora el algoritmo de búsqueda aleatoria para SVR, donde se incluyen los hiperparámetros de kernel, C y ϵ con sus correspondientes valores. Cabe destacar que existen otros hiperparámetros como γ , que utiliza funciones específicas con relación al número de características de los datos para determinar los valores correspondientes.

Kernel	{ <i>lineal, polinomial, RBF, sigmoidea</i> }
C	{ <i>0.1, 1, 5, 10</i> }
ϵ	{ <i>0.1, 1, 5, 10</i> }

Tabla 10 SVR: Búsqueda de Hiperparámetros

Los resultados obtenidos muestran que el mejor modelo se obtiene con un kernel **lineal**, un C de **0.01** y un ϵ de **1**.

- MSE en datos de entrenamiento: $3.8 \cdot 10^{-3}$ o 0.62 con datos desnormalizados.
- MSE en datos de prueba: $4.9 \cdot 10^{-3}$ o 0.78 con datos desnormalizados.

En la Figura 23 y 24 se muestran unos gráficos comparativos entre los valores observados y los valores predichos por SVR.

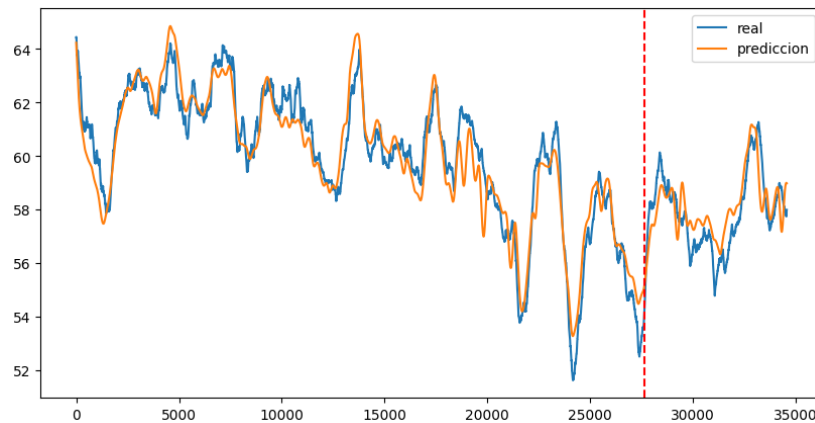


Figura 25 SVR: Valor observado VS Predicción: Datos Completos

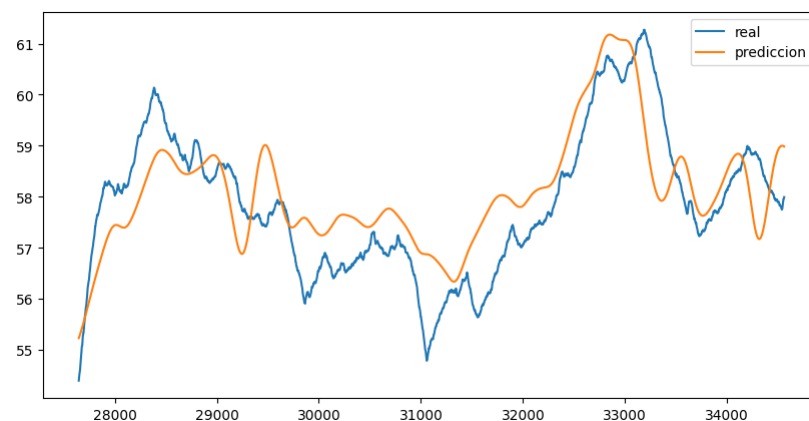


Figura 26 SVR Valor observado VS Predicción: Datos Test

En este caso, al analizar el MSE tanto en el conjunto de entrenamiento como en el conjunto de prueba, se puede concluir que no se está produciendo sobreajuste. Además, al observar los gráficos, se aprecia que el modelo se ajusta muy bien a los datos, incluso las predicciones en el conjunto de test también son de buena calidad.

Random Forest Regression

La estructura propuesta para Random Forest Regression se visualiza en la Figura 23. Al igual que con el resto, se emplean los 10 componentes extraídos mediante *PCA* como características de entrada para cada uno de los “n” árboles del modelo, donde cada uno como máximo puede tener una profundidad “m”.

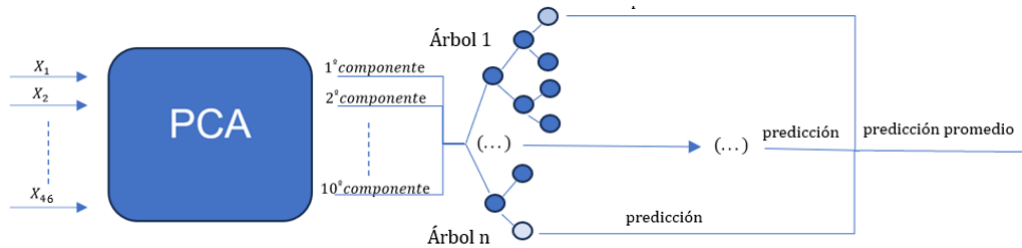


Figura 27 Random Forest Regression: Estructura

La Tabla 11 muestra el espacio de búsqueda que explora el algoritmo de búsqueda aleatoria para Random Forest Regression, donde se incluyen los hiperparámetros del número de árboles (n) y la máxima profundidad de cada árbol (m) descritos en la estructura, y además el número máximo de características a considerar para la división. Cabe destacar que para el hiperparámetro del número máximo de características se utilizan funciones específicas con relación al número de características de los datos para determinar los valores correspondientes.

Número de arboles	{5, 10, 15, 20}
Máxima profundidad	{3, 5, 7, 9}
Número de características para considerar la división	{sqrt, log2}

Tabla 11 Random Forest Regression: Búsqueda de Hiperparámetros

Los resultados obtenidos muestran que el mejor modelo se obtiene con 20 árboles, una profundidad máxima de 7 y “log2” como número de características para considerar la división.

- *MSE* en datos de entrenamiento: $1.09 \cdot 10^{-3}$ o 0.17 con datos desnormalizados.
- *MSE* en datos de prueba: $1.14 \cdot 10^{-2}$ o 1.83 con datos desnormalizados.

En la Figura 28 y 29 se muestran unos gráficos comparativos entre los valores observados y los valores predichos por Random Forest Regression.

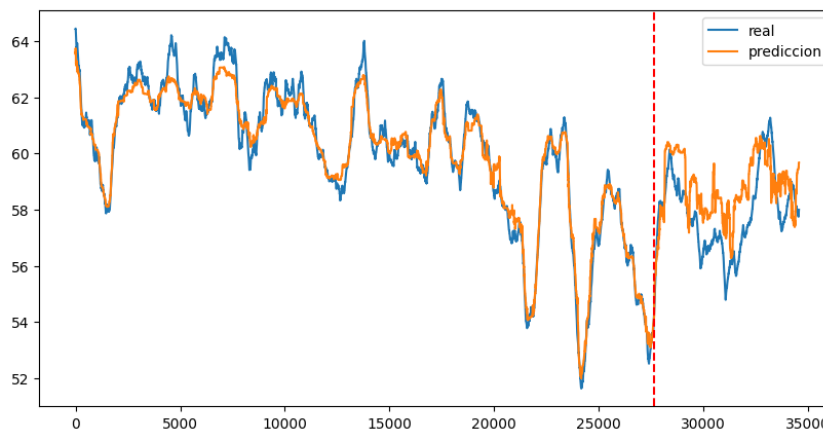


Figura 28 Random Forest Regression: Valor observado VS Predicción- Conjunto Completo

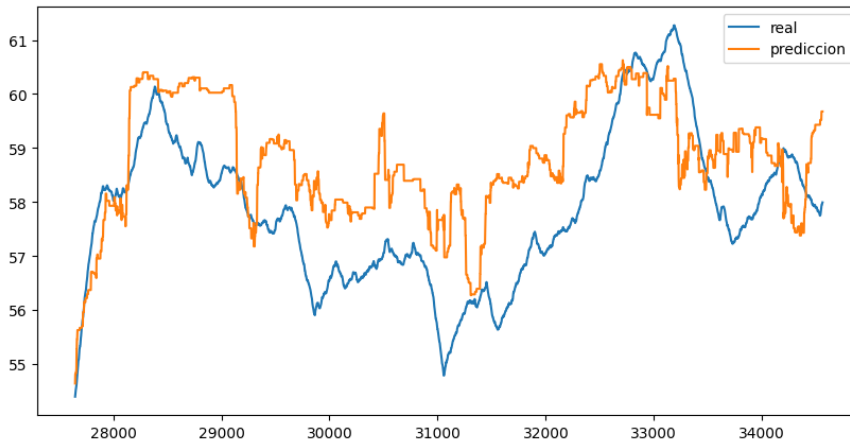


Figura 29 Random Forest Regression: Valor observado VS Predicción- Conjunto test

Nuevamente, al igual que ocurre con la Red Neuronal, se observa una diferencia significativa entre los *MSE* y un ajuste preciso con los datos de entrenamiento, pero bastante peor con los de test en los gráficos. Por lo tanto, se toma la decisión de repetir el experimento reduciendo tanto el número de árboles como la profundidad máxima. En la Tabla 12 se muestra el espacio de hiperparámetros correspondiente a este nuevo experimento.

Número de arboles	{5, 10, 15}
Máxima profundidad	{3, 5}
Número de características para considerar la división	{sqrt, log2}

Tabla 12 Random Forest Regression: Búsqueda de Hiperparámetros: Experimento 2

Para este segundo experimento los resultados obtenidos muestran que el mejor modelo se obtiene con 15 árboles, una profundidad máxima de 5 y “log2” como número de características para considerar la división.

- *MSE* en datos de entrenamiento: $2.94 \cdot 10^{-3}$ o 0.47 con datos desnormalizados.
- *MSE* en datos de prueba: $9.68 \cdot 10^{-3}$ o 1.56 con datos desnormalizados.

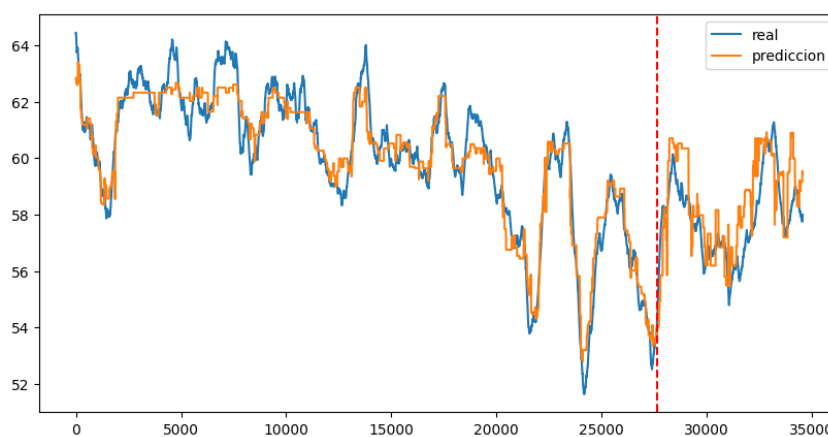


Figura 30 Random Forest Regression: Valor observado VS Predicción:Conjunto Completo: Experimento 2

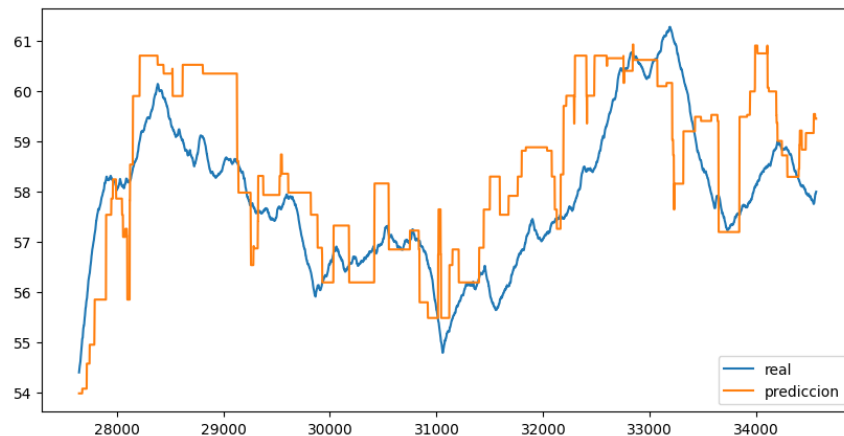


Figura 31 Random Forest Regression: Valor observado VS Predicción: Conjunto Test: Experimento 2

En este segundo experimento, se logra reducir la diferencia entre los (MSE) y conseguir mejores predicciones en el conjunto de test. Sin embargo, algo notable que se observa en los gráficos es que, al principio del proceso, el modelo parece tener un peor ajuste en comparación con la Red Neuronal y SVR. Esto también se puede apreciar en el ajuste de los datos de test.

Regresión por mínimos cuadrados parciales

La estructura propuesta para Regresión por Mínimos Cuadrados Parciales se muestra en la Figura 32. En este caso ya que el propio método tiene la capacidad de reducir la dimensionalidad no se recurre a utilizar PCA, si no que directamente se introducen las 46 variables resultado del proceso de selección manual. A partir de estas se calculan los “n” componentes que mejor resultados proporcionan.

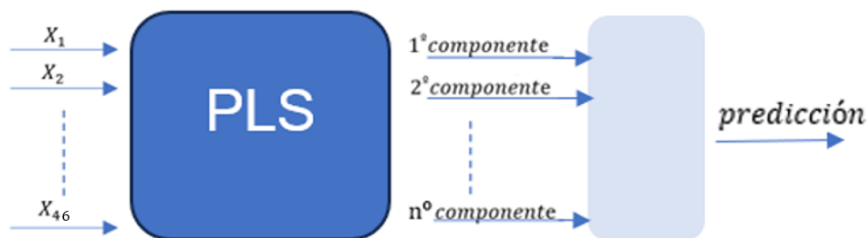


Figura 32 Regresión por Mínimos Cuadrados Parciales: Estructura

La regresión por mínimos cuadrados parciales tiene un único hiperparámetro configurable, que es el número de componentes. La Tabla 11 detalla el experimento que se ha llevado a cabo.

Número de componentes	$\{1, 2, \dots, \mathbf{22}, \dots, 46\}$
------------------------------	-------------------------------------------

Tabla 13 Regresión por Mínimos Cuadrados Parciales: Búsqueda de Hiperparámetros

Los resultados obtenidos muestran que el mejor modelo se obtiene con 22 componentes.

- MSE en datos de entrenamiento: $6.94 \cdot 10^{-4}$ o 0.11 con datos desnormalizados.
- MSE en datos de prueba: $4.46 \cdot 10^{-3}$ o 0.72 con datos desnormalizados.

En la Figura 33 y 34 se muestra un gráfico comparativo entre los valores observados y los valores predichos por regresión por Mínimos Cuadrados Parciales.

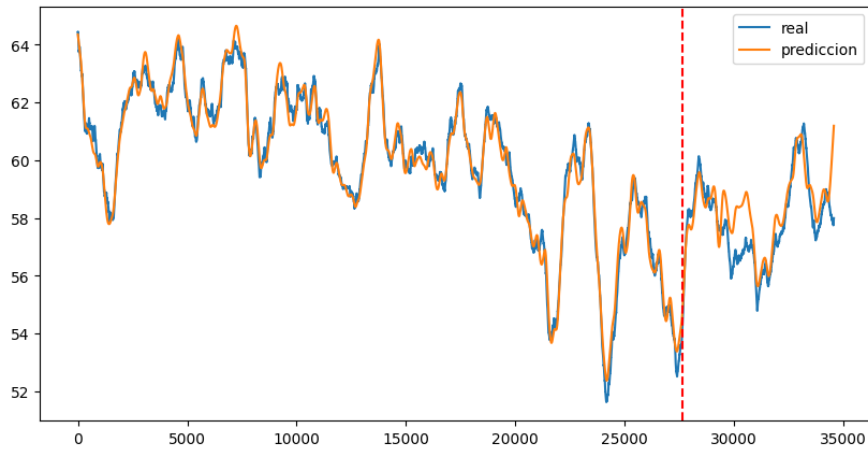


Figura 33 Regresión por Mínimos Cuadrados Parciales: Valor observado VS Predicción: Datos Completos

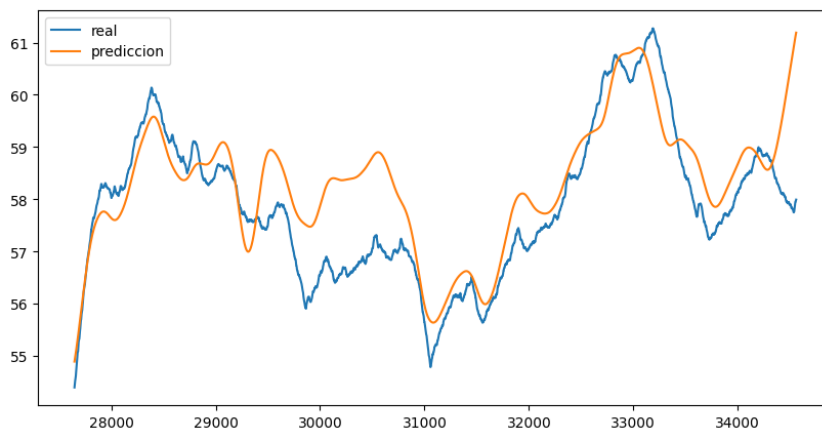


Figura 34 Regresión por Mínimos Cuadrados Parciales: Valor observado VS Predicción: Datos Test

Al analizar gráficos y la diferencias de *MSE* entre el entrenamiento y test se concluye que se está produciendo un sobreajuste, al igual que sucede con la Red Neuronal y Random Forest Regresión, y por tanto es necesario reducir el número de componentes que se seleccionan. En la Tabla 14 se muestra el nuevo experimento donde ahora se limita el número de componentes a 10, que son los que se seleccionaban con *PCA*.

Número de componentes	{1, 2, ..., 10}
------------------------------	------------------------

Tabla 14 Regresión por Mínimos Cuadrados Parciales: Búsqueda de Hiperparámetros: Experimento 2

Los resultados obtenidos muestran que ahora el mejor modelo se obtiene con 10 componentes.

- *MSE* en datos de entrenamiento: $1.43 \cdot 10^{-3}$ o 0.23 con datos desnormalizados.
- *MSE* en datos de prueba: $2.28 \cdot 10^{-3}$ o 0.45 con datos desnormalizados.

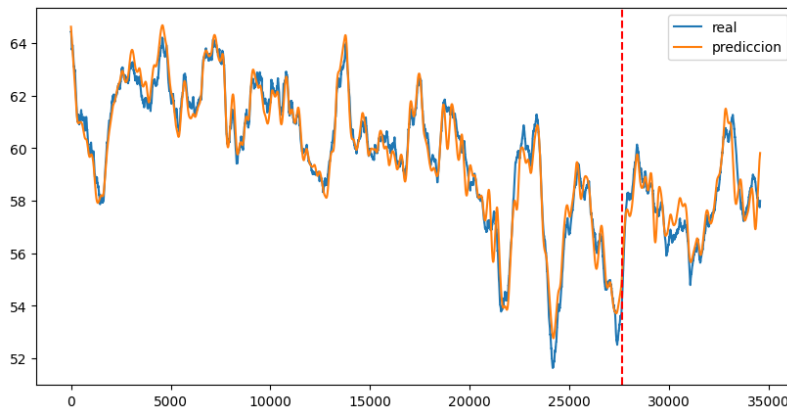


Figura 35 Regresión por Mínimos Cuadrados Parciales: Valor observado VS Predicción: Datos Completos: Experimento 2

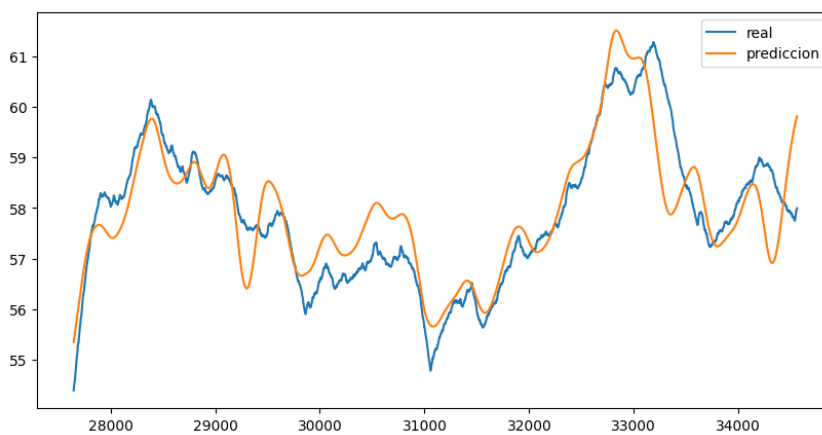


Figura 36 Regresión por Mínimos Cuadrados Parciales: Valor observado VS Predicción: Datos Test: Experimento 2

En este segundo experimento, se observa que la diferencia entre el MSE en el conjunto de entrenamiento y el conjunto de prueba es muy pequeña. Además, el modelo tiene un menor MSE en el conjunto de prueba, lo cual indica una mejor generalización del modelo y una corrección del sobreajuste.

Discusión de resultados

Después de analizar los resultados de todos los modelos, se puede afirmar que la Regresión por Mínimos Cuadrados Parciales obtiene mejores resultados que los modelos de aprendizaje automático. En definitiva, este modelo demuestra una buena capacidad para generalizar, siendo capaz de predecir de manera precisa nuevos datos.

En cuanto a los modelos de aprendizaje automático (Red Neuronal, SVR y Random Forest Regression), se observa que SVR también logra buenos resultados. Sin embargo, los otros dos modelos, a pesar de haber reducido el sobreajuste en los segundos experimentos, todavía muestran cierto grado de sobreajuste que afecta los resultados y resulta en predicciones menos precisas en los conjuntos de test.

En relación con estos resultados se concluye que la selección de los hiperparámetros en los modelos de aprendizaje automático parece no haber sido óptima. Por lo tanto, sería recomendable realizar un estudio más exhaustivo y explorar un espacio mayor de hiperparámetros. Esto implica considerar algunos hiperparámetros que se establecieron previamente, para limitar el espacio de búsqueda y reducir el tiempo de ejecución. Además, es

importante tener en cuenta que los resultados obtenidos pueden variar ligeramente en cada ejecución del algoritmo debido a la aleatoriedad propia de estos mismos.

En resumen, se puede concluir que la Regresión por Mínimos Cuadrados Parciales es la opción preferida para este proyecto, mientras que los modelos de aprendizaje automático requieren una optimización más cuidadosa de los hiperparámetros para obtener resultados que puedan compararse e incluso superar a este primero.

Otro aspecto a tener en cuenta es que los modelos generados en este estudio no consideraron la propia dinámica del proceso industrial. Es decir, no se investigó cómo el proceso evoluciona y cambia a lo largo del tiempo, ni se identificaron los factores que pueden influir directa o indirectamente en su rendimiento, calidad y estabilidad. Únicamente se menciona que parte de la variabilidad observada puede deberse a estos factores. Por lo tanto, al no tomar en cuenta esta dinámica, que es crítica en los procesos industriales, los modelos resultantes tienen una capacidad de generalización limitada.

Por último, es interesante comparar los resultados obtenidos en este estudio con otros trabajos que aborden el mismo proceso. Por ejemplo, en el estudio mencionado en [5], que utiliza un conjunto de entrenamiento distinto, pero la factoría y Brix estimado es igual.

En ese estudio, el *MSE* obtenido utilizando Regresión por Mínimos Cuadrados Parciales fue de 2.19, mientras que en este estudio se logra un *MSE* de 0.45. Además, el mejor resultado se obtiene con un modelo de Red Neuronal que toma como entradas los componentes de un PCA, con un valor de 0.63, que a pesar del problema en la búsqueda de hiperparámetros no difiere significativamente del obtenido en este estudio (0.93), y que a su vez no supera al logrado con Regresión por Mínimos Cuadrados Parciales.

En base a esto, se puede decir que los resultados son prometedores y demuestran ser competitivos en comparación con otros estudios.

Capítulo 5. Conclusiones y trabajo futuro

Una vez finalizado el proyecto se concluye que se ha logrado el objetivo propuesto, que consistía en el desarrollo de un sensor de software que implementara un modelo de aprendizaje automático o estadístico para estimar el Brix en el proceso de evaporación de la industria azucarera. Para lograrlo, se ha aplicado la metodología CRISP-DM a un conjunto de datos obtenidos de mediciones realizadas durante un proceso de evaporación en una fábrica.

El estudio realizado ha constatado que el proceso de evaporación en la industria azucarera es complejo y que depende de múltiples factores externos, lo que resulta en una alta variabilidad que dificulta el estudio y la creación de modelos precisos.

A pesar de las dificultades inherentes a la propia naturaleza del proceso, los resultados obtenidos indican que modelos como los creados a partir de Regresión por Mínimos Cuadrados Parciales o SVR pueden predecir con bastante precisión.

Como líneas futuras de trabajo, lo principal sería mejorar los resultados obtenidos con los modelos de aprendizaje automático, intentando minimizar el sobreajuste. Para ello, sería necesario realizar una búsqueda más amplia de hiperparámetros y estudiar con mayor detalle otros proyectos que utilicen modelos similares. Además, sería interesante probar otras alternativas, como las redes neuronales recurrentes, con el objetivo final de conseguir modelos de aprendizaje automático con buena capacidad de generalización. De este modo, se podrán aprovechar las ventajas que tienen estos modelos en comparación a métodos clásicos, como la Regresión por Mínimos Cuadrados Parciales, especialmente en la capacidad de realizar estimaciones a varios pasos en el futuro.

Otra línea de trabajo importante consistiría en considerar la dinámica del proceso. Para ello, se llevaría a cabo un estudio detallado que permita incorporar esta información en los modelos de aprendizaje automático. De esta manera, podrían realizarse ajustes en tiempo real al identificar patrones o tendencias que indiquen la influencia de ciertos factores en el proceso o una mejor selección de atributos entre otras mejoras.

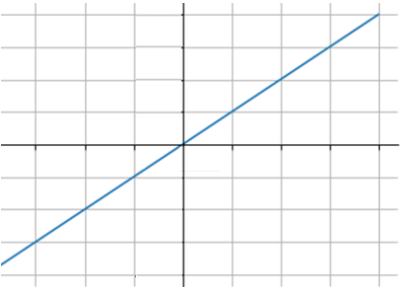
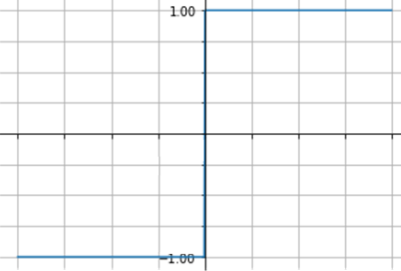
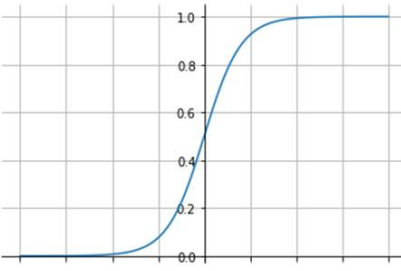
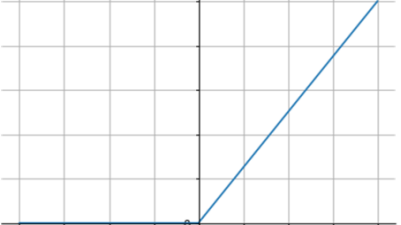
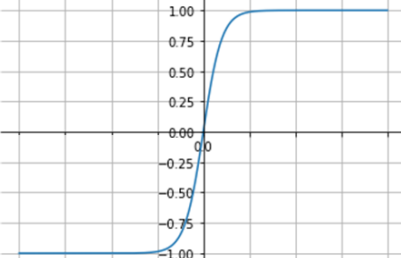
Por último, se podría aplicar esta metodología para entrenar modelos que sirvan para el resto de las fases de obtención del azúcar, como la difusión o la cristalización entre otras de manera que se pueda implementar un sensor para la estimación del Brix en cada una de ellas.

Capítulo 6 - Bibliografía

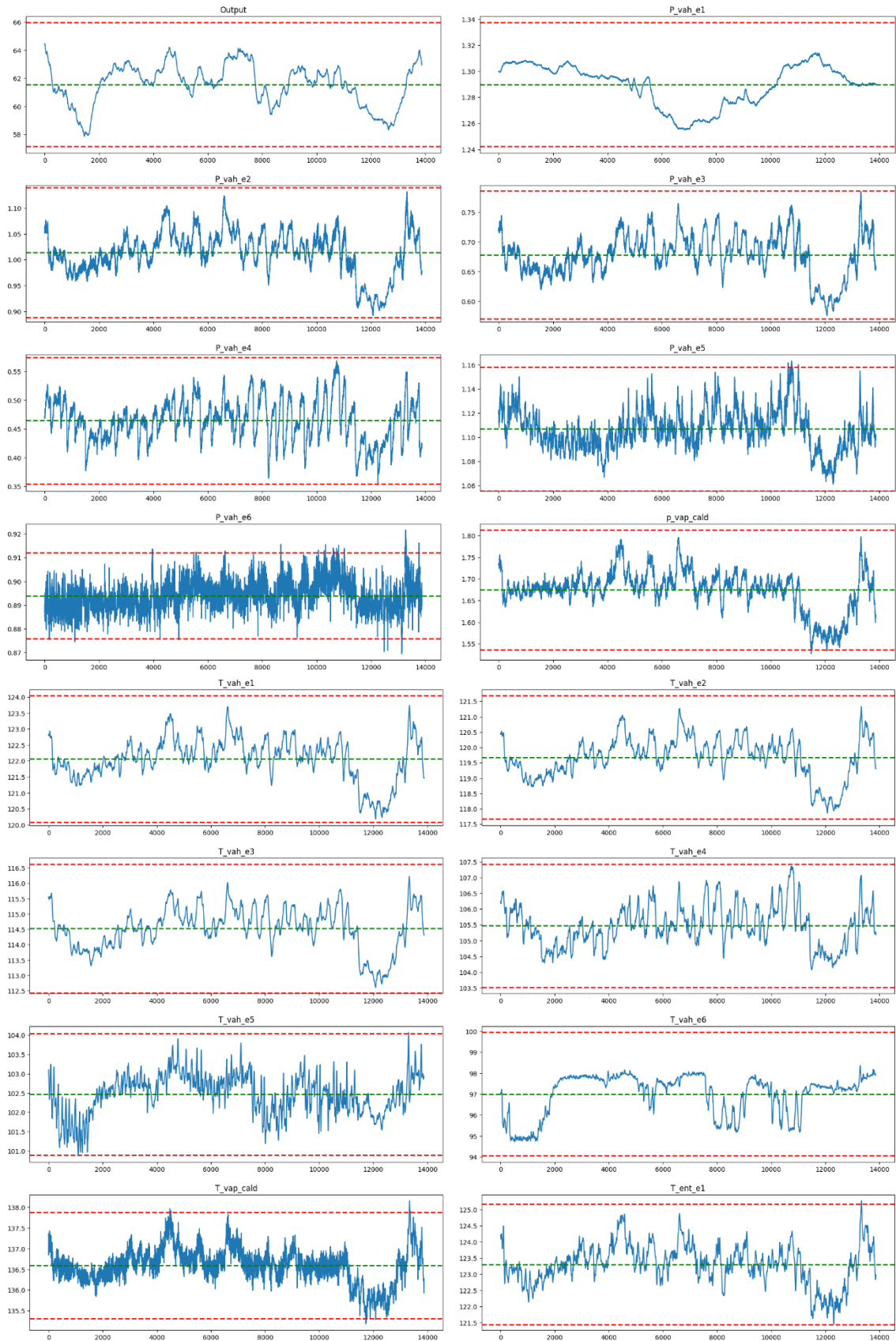
- [1] M. da Costa, C. Fontes, G. Carvalho y E. de Moraes Júnior, "UltraBrix: A Device for Measuring the Soluble Solids Content", *Sustainability*, vol. 13, nº 1227. 2021.
- [2] J. Baquero Franco, "La industria del azúcar de remolacha", *Ministerio de Agricultura, Pesca y Alimentación*. Madrid, España, 1987.
- [3] M. Golato, "Sistema de control en cocimientos de azúcar", *Cátedra de sistemas de control, Ingeniería Mecánica, Universidad Nacional de Tucumán*. 2014.
- [4] A. Muñoz, "Aplicación informática para el diseño y operación de evaporadores de múltiple efecto en la industria", *Trabajo Fin de Grado, Departamento de Ingeniería Energética, Universidad de Sevilla*. Sevilla, España, 2016.
- [5] A. Merino, D. Garcia-Alvarez, G. I. Sainz-Palmero, L. F. Acebes y M. J. Fuente, "Knowledge based recursive non-linear partial least squares (RNPLS)", *ISA Transactions*, vol. 100, pp. 481-494. 2020.
- [6] E. Rodríguez, "Modelo de regresión lineal: mínimos cuadrados parciales (PLS) con método Cuasi-Newton", *Revecitec*, vol. 6, nº 1, pp. 21-37. 2015.
- [7] J. R. Hilera y V. J. Martínez, "Redes neuronales artificiales: fundamentos, modelos y aplicaciones", *RA-MA Editorial*. Madrid, España, 1995.
- [8] M. T. Barón, "Clasificación de fallos en motores en estado transitorio mediante redes neuronales", *Trabajo Fin de Grado, Facultad de Ciencias, Universidad de Valladolid*. Valladolid, España, 2021.
- [9] F. Alonso, "Redes Neuronales y Deep Learning. Capítulo 4: Backpropagation", *FutureSpace*, 17 Septiembre 2021. [En línea]: <https://www.futurespace.es/redes-neuronales-y-deep-learning-brackpropagation/>.
- [10] Keras Team, "Keras documentation: Adagrad". [En línea]: <https://keras.io/api/optimizers/adagrad/>. [Último acceso: 24 Enero 2023].
- [11] Keras Team, "Keras documentation: Adam". [En línea]: <https://keras.io/api/optimizers/adam/>. [Último acceso: 24 Enero 2023].
- [12] J. A. Camacho, "Support Vector Regression (SVR)", *JacobSoft*, 27 Marzo 2020. [En línea]: <https://www.jacobsoft.com.mx/en/support-vector-regression/>.
- [13] J. M. Guareño, "Support vector regression: propiedades y aplicaciones", *Trabajo fin de grado, Departamento de Estadística e Investigación Operativa, Universidad de Sevilla*. Sevilla, España, 2016.

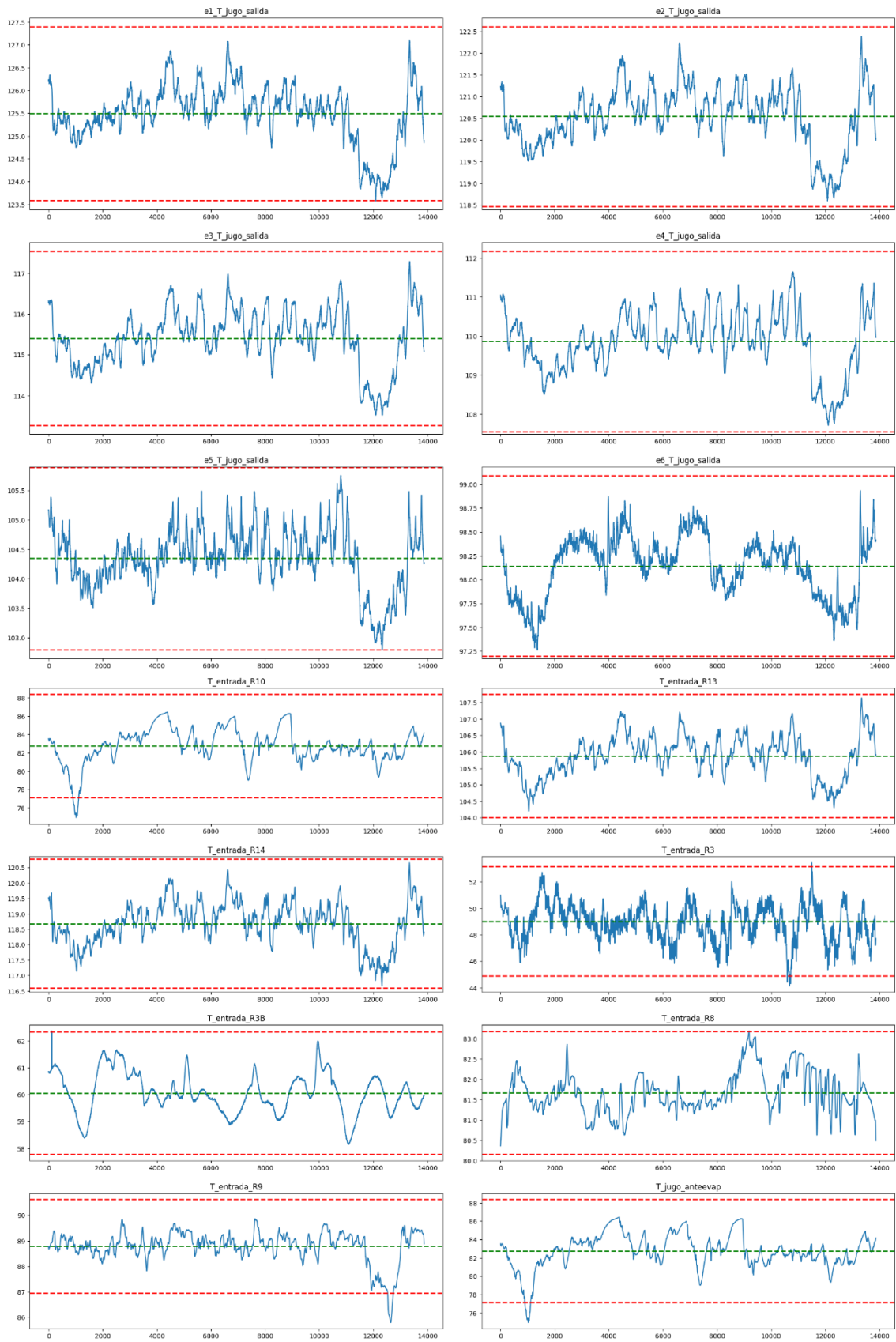
-
- [14] N. Chaya, "Random Forest Regression", *Level Up Coding*, 8 Junio 2020. [En línea]: <https://levelup.gitconnected.com/random-forest-regression-209c0f354c84>.
- [15] C. G. Martínez, "Árboles de decisión y métodos de ensemble", *RPubs by RStudio*, Junio 2018. [En línea]: https://rpubs.com/Cristina_Gil/arboles_ensemble.
- [16] D. Senapati, "Grid Search vs Random Search", *medium.com*, 29 Agosto 2018. [En línea]: <https://medium.com/@senapati.dipak97/grid-search-vs-random-search-d34c92946318>.
- [17] E. Reralta, "Proceso de Conceptualización del entendimiento del Negocio para Proyectos de Explotación de Información", *Revista Latinoamericana de Ingeniería de Software*, vol. 2, nº 5, p. 273, 2014.
- [18] E. Güemes García, "Diseño de sensores software para el control", *Trabajo Fin de Grado, Escuela de Ingenierías Industriales, Universidad de Valladolid*. Valladolid, España, 2018.
- [19] J. A. V. Emerson A. Chaparro, "Gráficos de control para la media de un proceso en poblaciones con distribución asimétrica", *Revista Colombiana de Estadística*, vol. 23, nº 2, pp. 29-44. 2000.
- [20] C. E. F. Tapia y K. L. F. Cevallos, "Pruebas para comprobar la normalidad de datos en procesos productivos: Anderson-Darling, Ryan-Joiner, Shapiro-Wilk y Kolmogórov-Smirnov", *Societas. Revista de Ciencias Sociales y Humanísticas*, vol. 23, nº 2, p. 83–106, 2021.
- [21] S. Seo, "A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets", *Tesis de maestría, Universidad de Pittsburgh*. 2006.
- [22] E. M. Singla y M. H. Singh, "Frequency based audio Noise Reduction using Butter Worth, Chebyshev & Elliptical Filters", *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 3, nº 10, p. 5989–5995. 2015.
- [23] M. Hernández-Ramírez, S. Tovar-Arriaga, J. C. Pedraza-Ortega y E. Gorrostieta-Hurtado, "Clasificación de A/V en imágenes de retina usando arquitectura inspirada en U-Net con ConvLSTM bidireccionales", *Research in Computing Science*, vol. 149, nº 8, p. 853–865. 2020.
- [24] S. de la Fuente Fernández, "Análisis de Componentes Principales", *Apuntes de Instrumentos Estadísticos Avanzados, Facultad de Ciencias Económicas y Empresariales, Universidad Autónoma de Madrid*: [En línea]: https://www.estadistica.net/Master-Econometria/Componentes_Principales.pdf

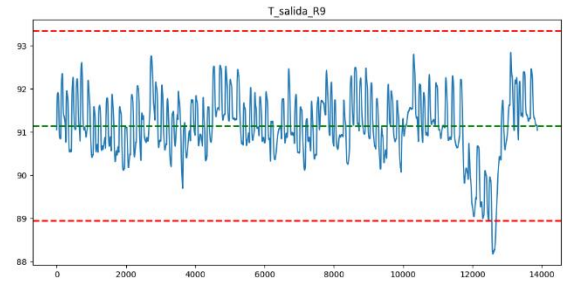
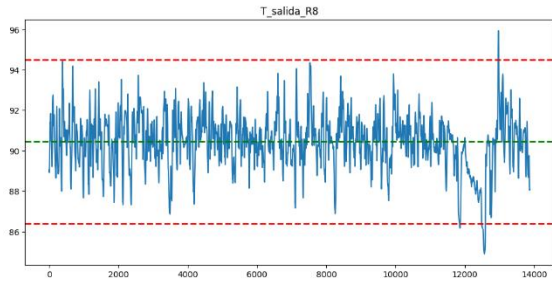
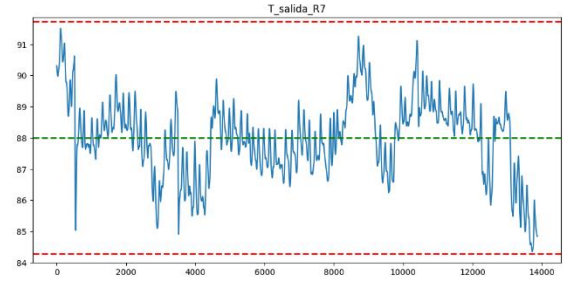
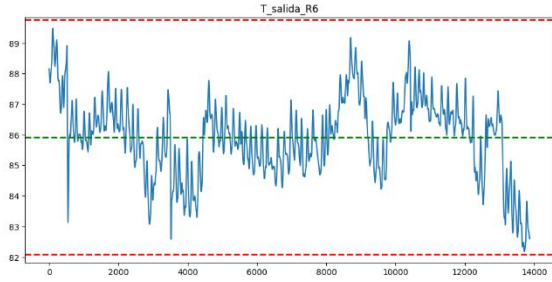
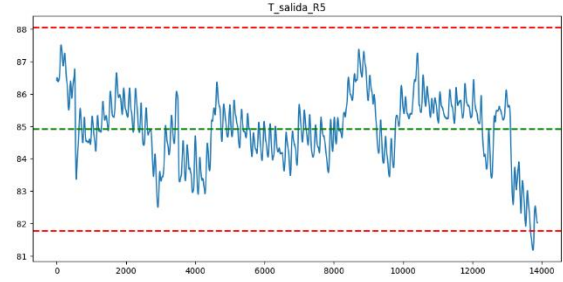
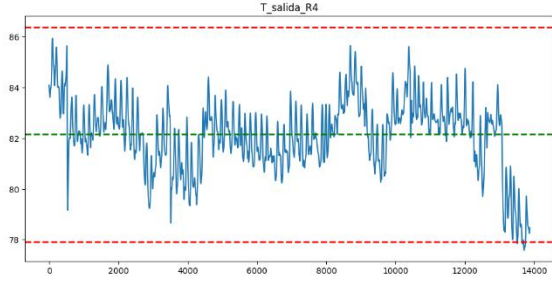
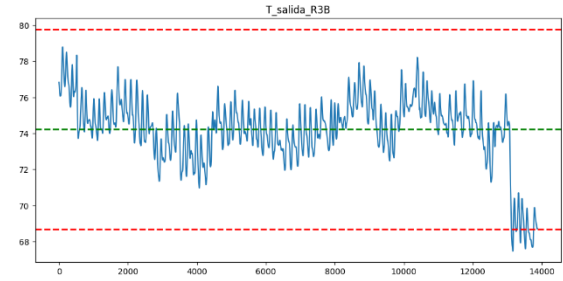
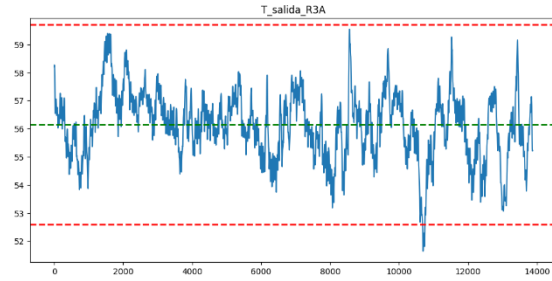
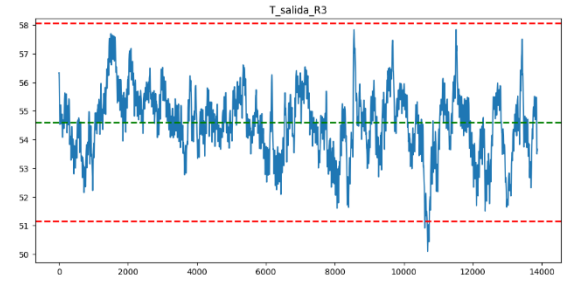
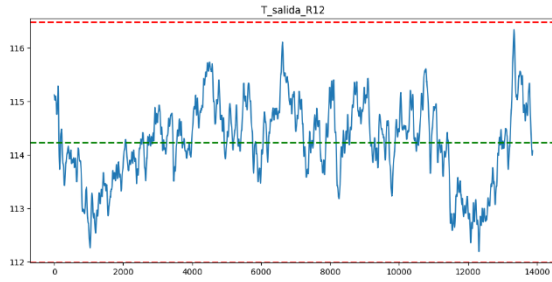
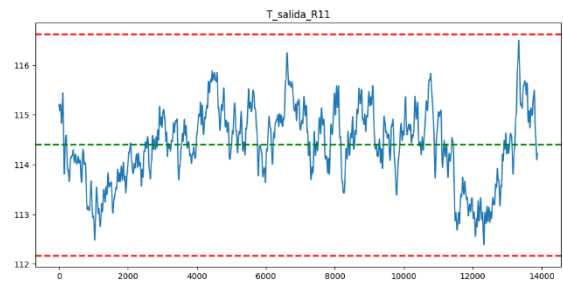
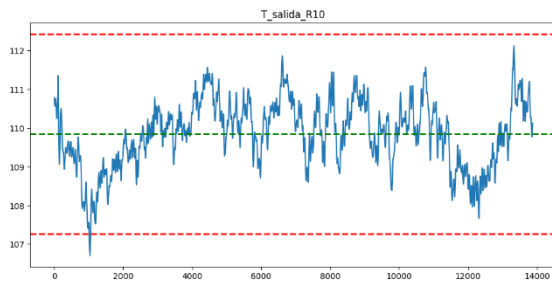
Anexo I - Funciones de Activación

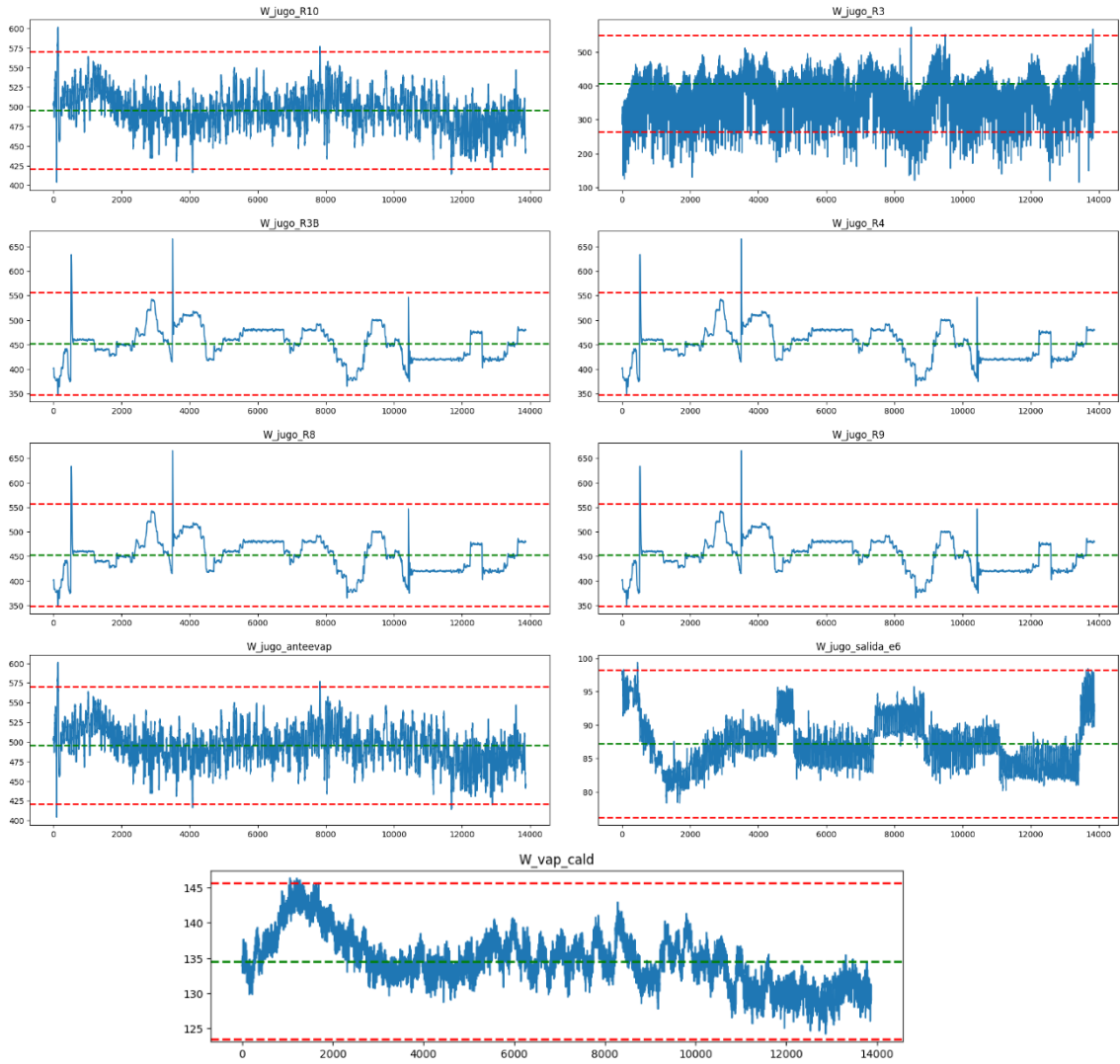
	Función	Rango	Gráfico
Identidad	$f(x) = x$	$[-\infty, +\infty]$	
Escalón (Treeshold)	$f(x) = \begin{cases} +1, Si x \geq 0 \\ -1, Si x < 0 \end{cases}$	$[+1, +1]$ $[-1, -1]$	
Sigmoide	$f(x) = \frac{1}{1 + e^{-x}}$	$(0, +1)$	
Rectificadora (ReLU)	$f(x) = \begin{cases} x, Si x \geq 0 \\ 0, Si x < 0 \end{cases}$	$[0, +\infty]$	
Tangente Hiperbólica	$f(x) = \tanh(x)$ $= \frac{e^x - e^{-x}}{e^x + e^{-x}}$	$(-1, +1)$	

Anexo II – Gráficos de control

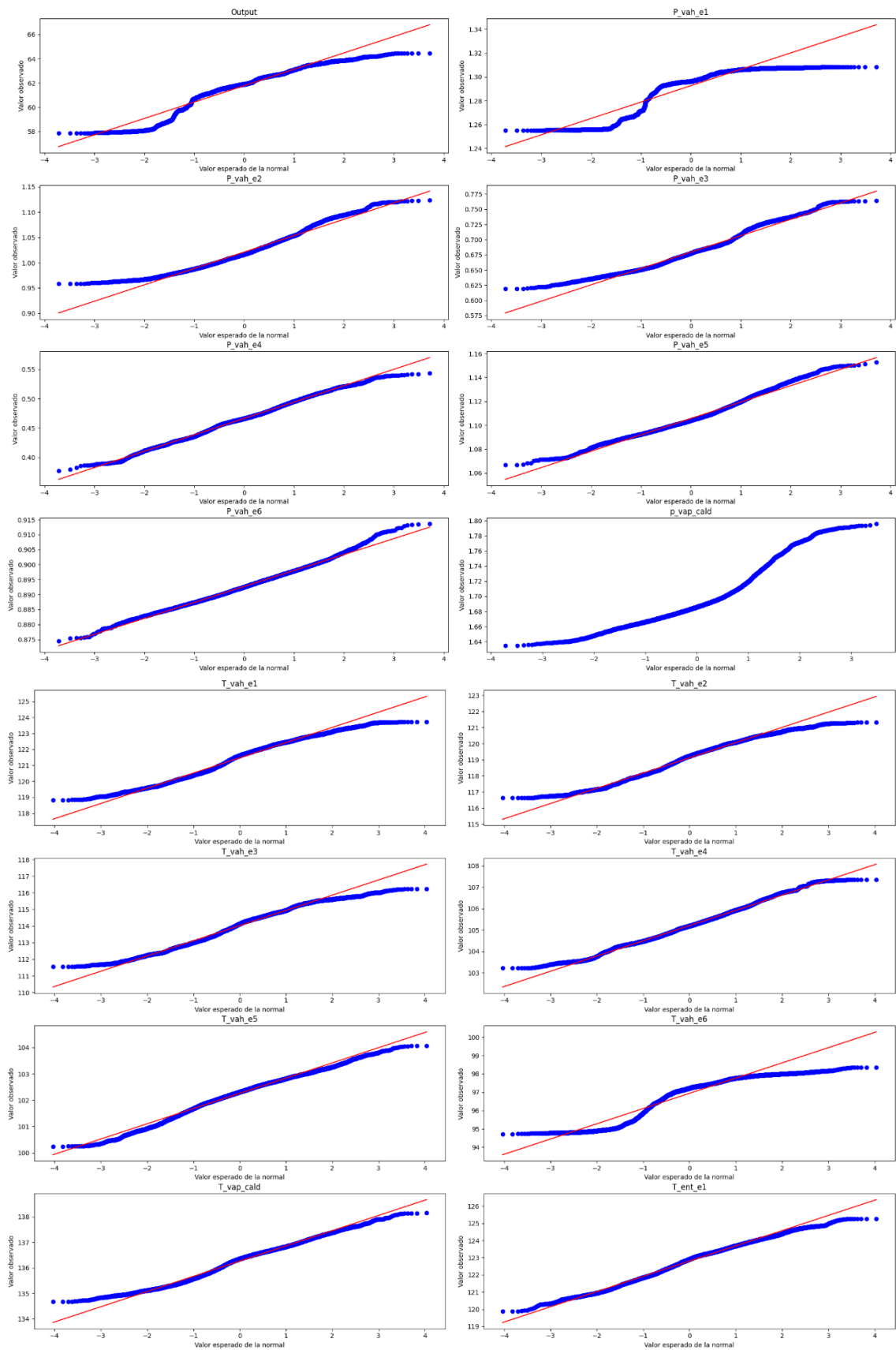


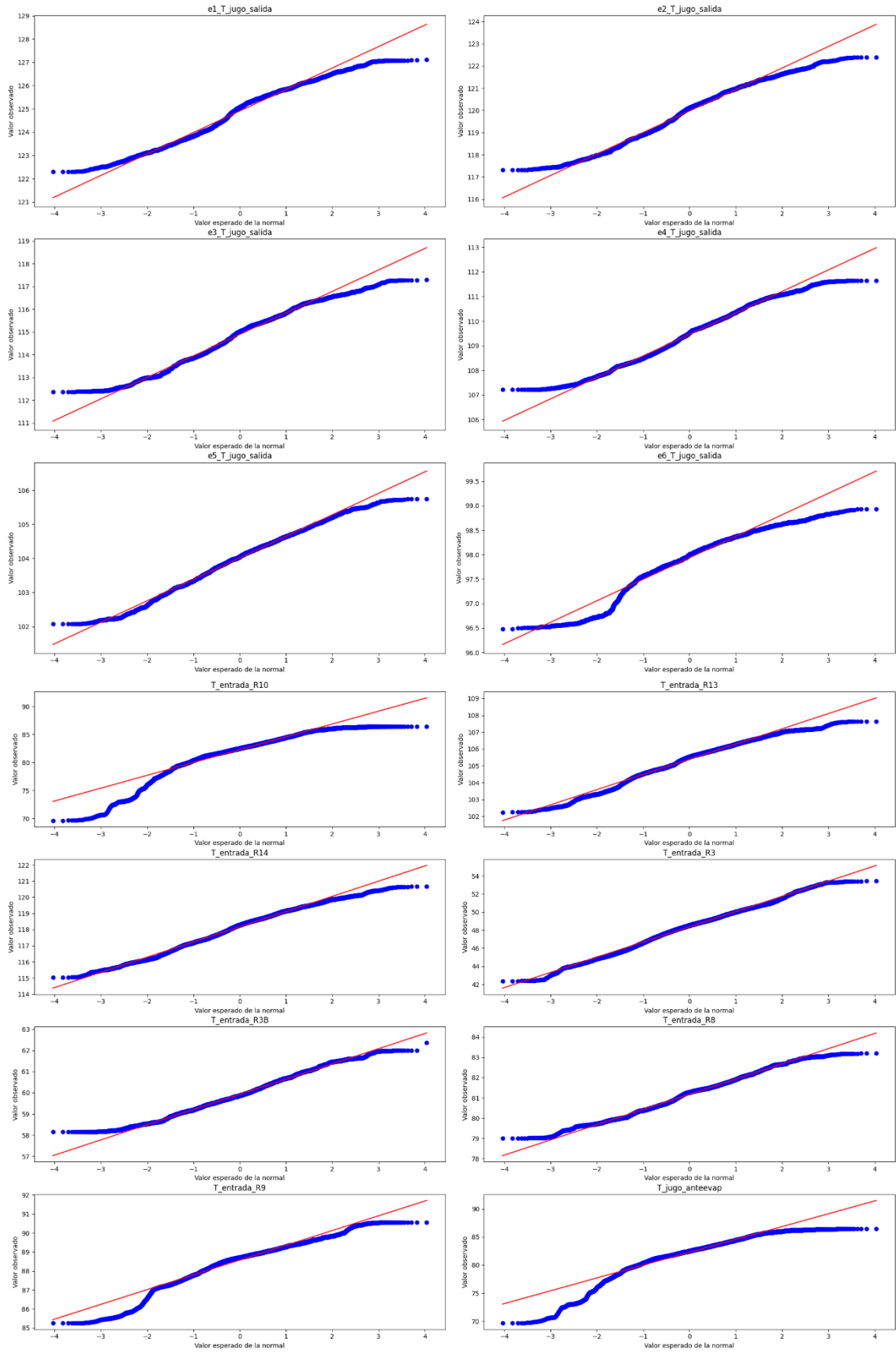


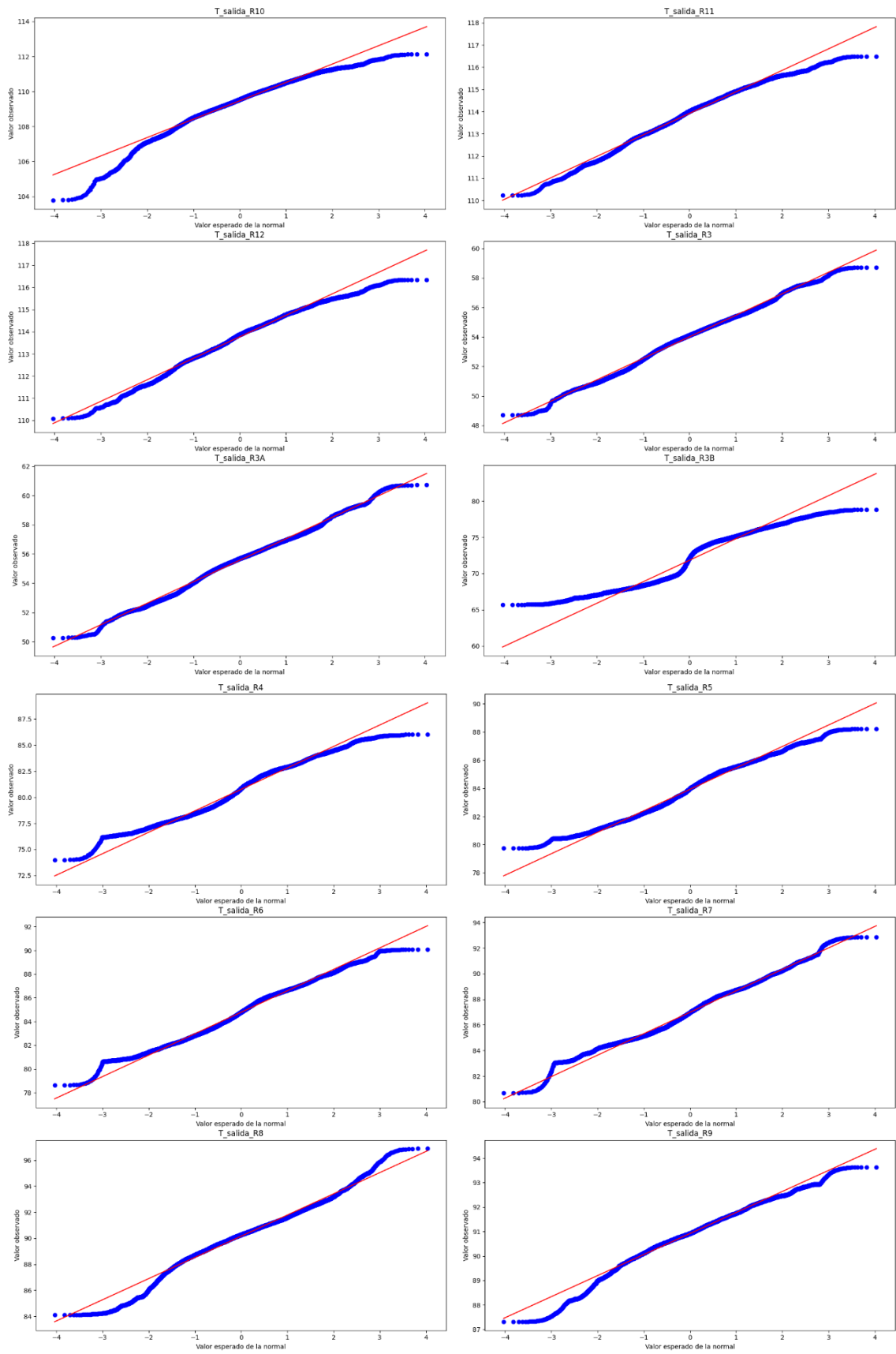


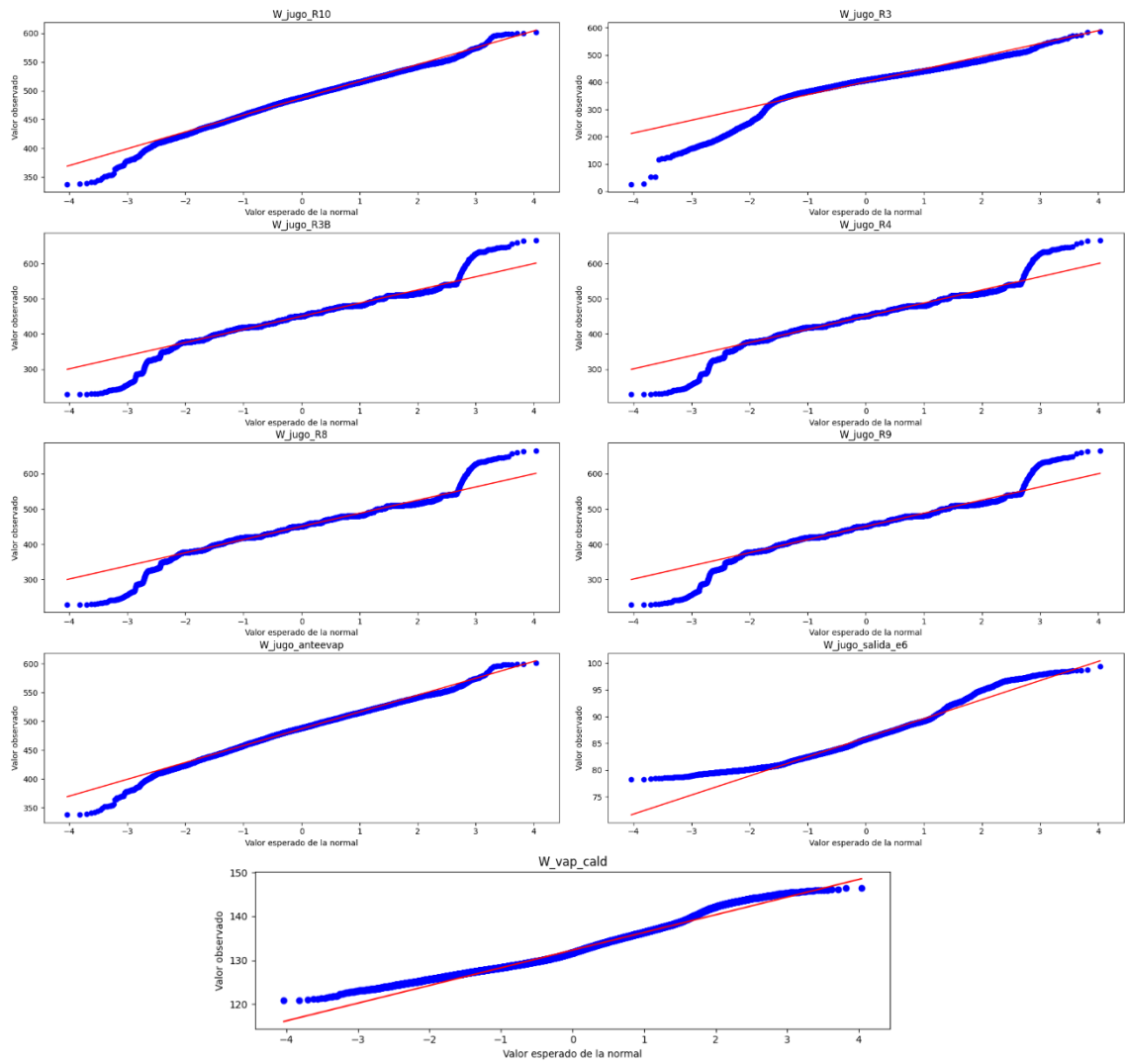


Anexo III – Q-Q normal

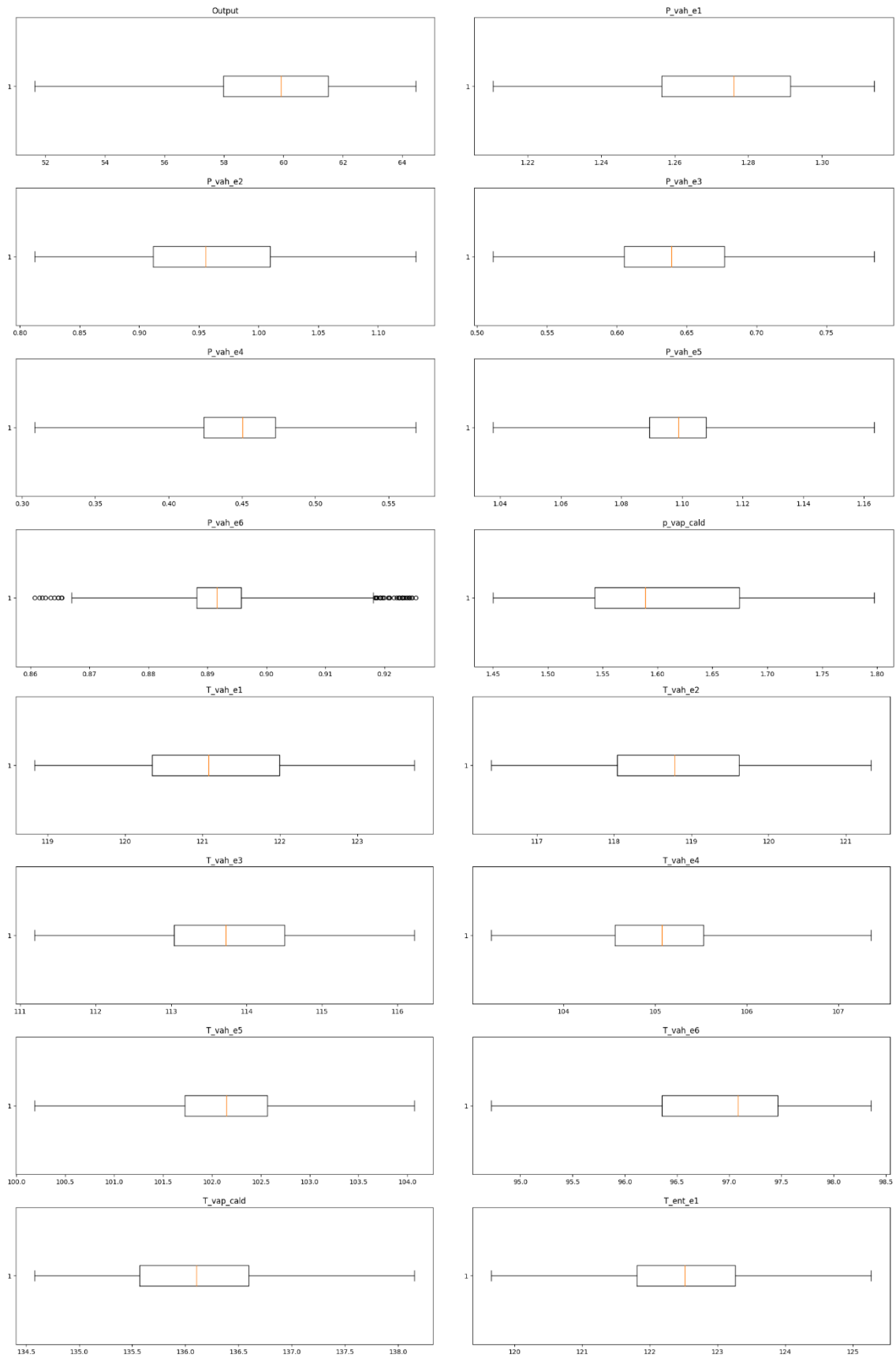


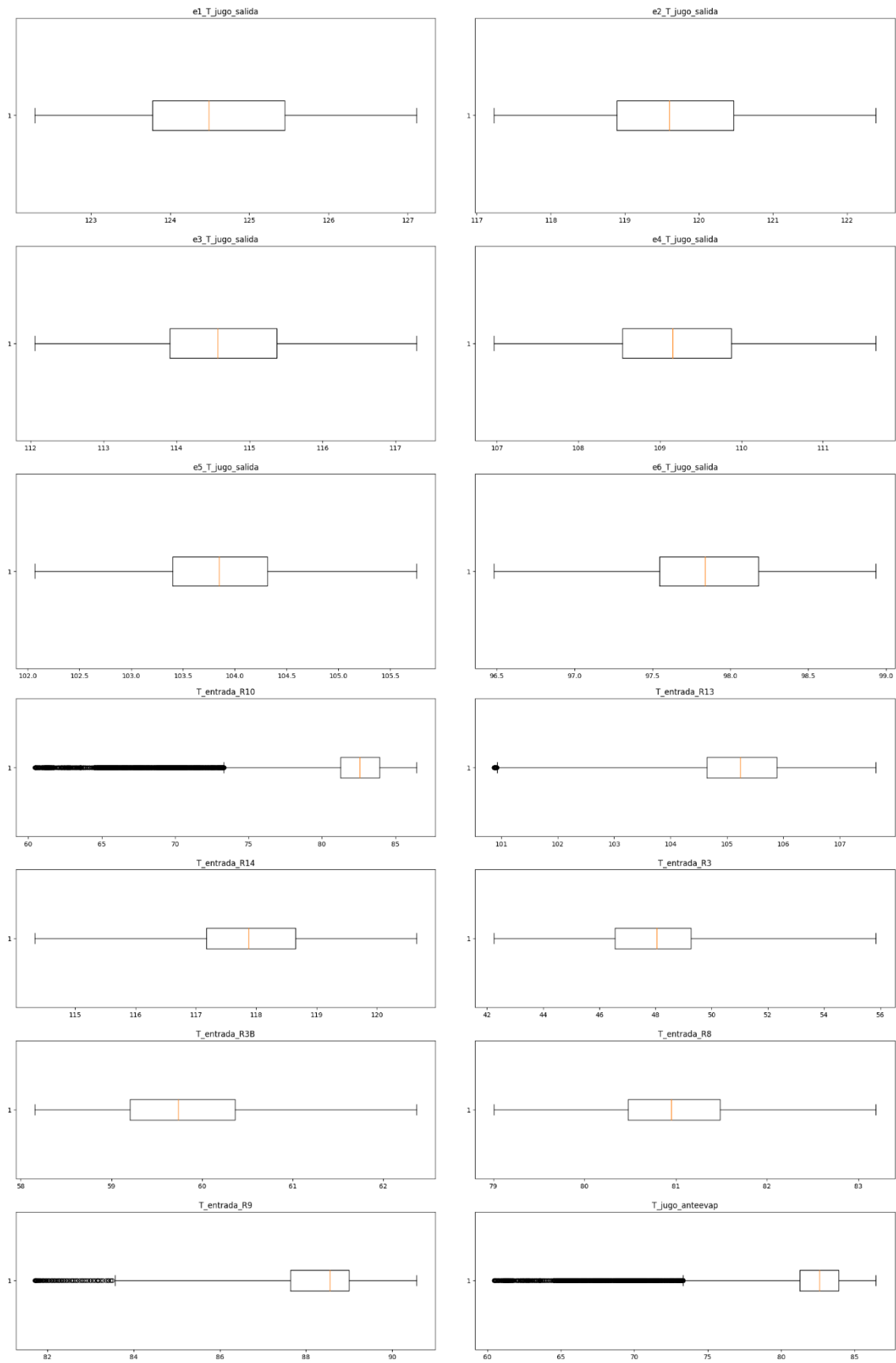


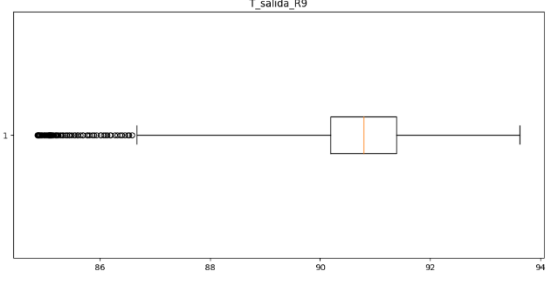
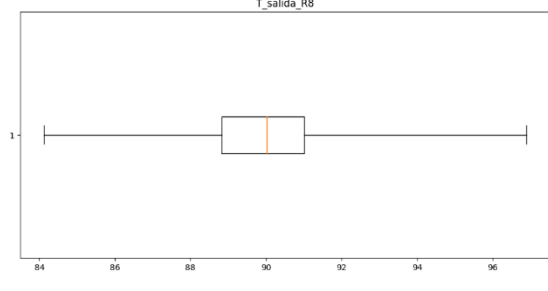
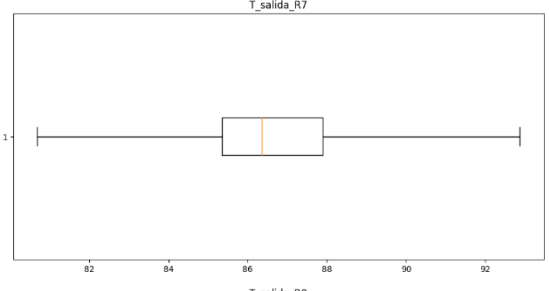
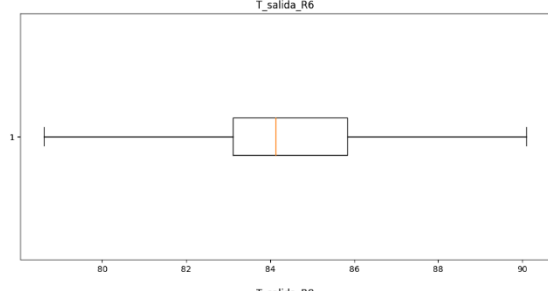
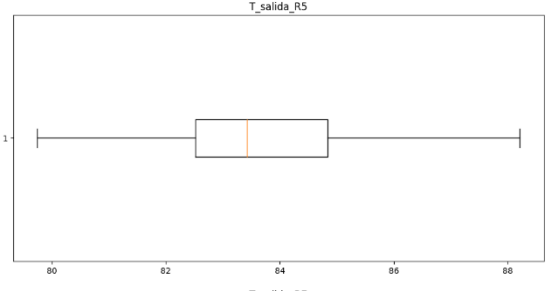
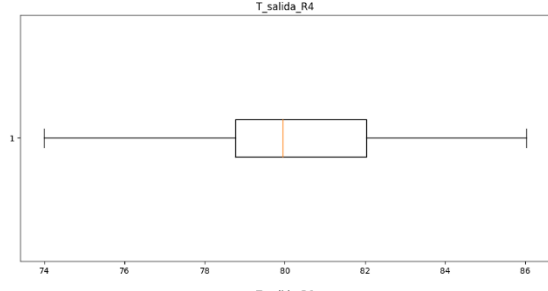
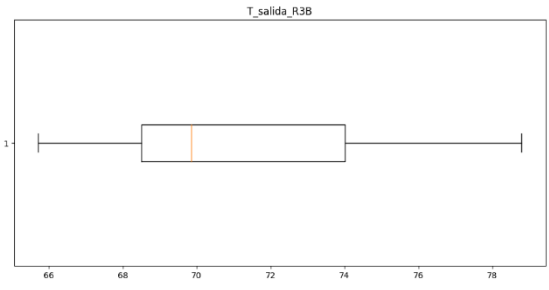
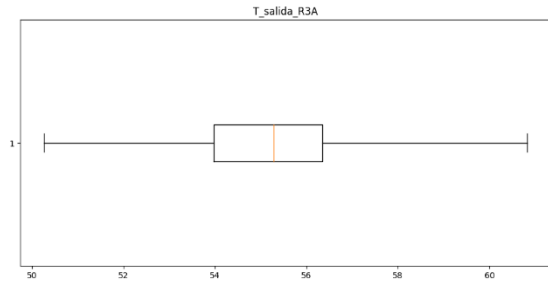
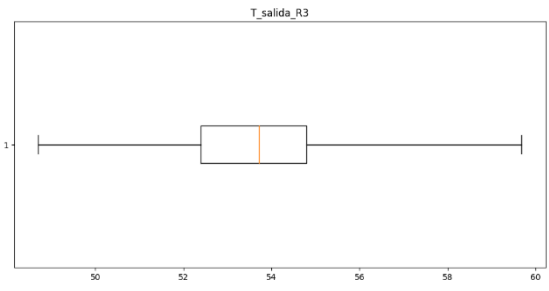
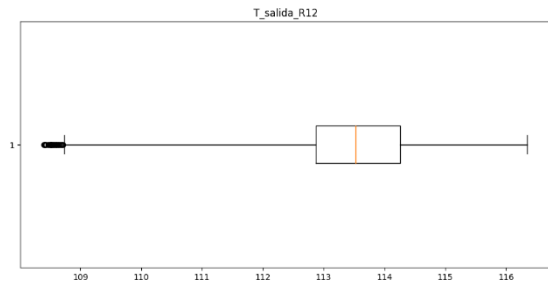
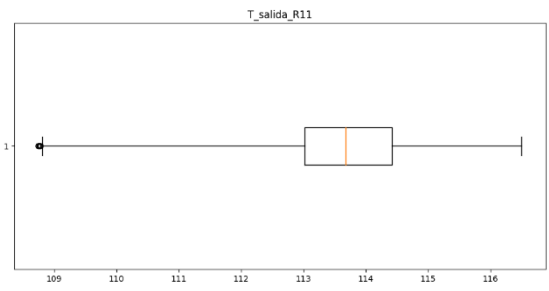
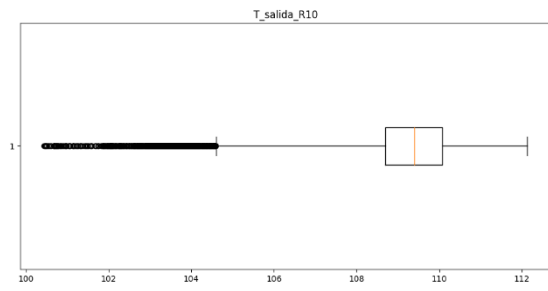


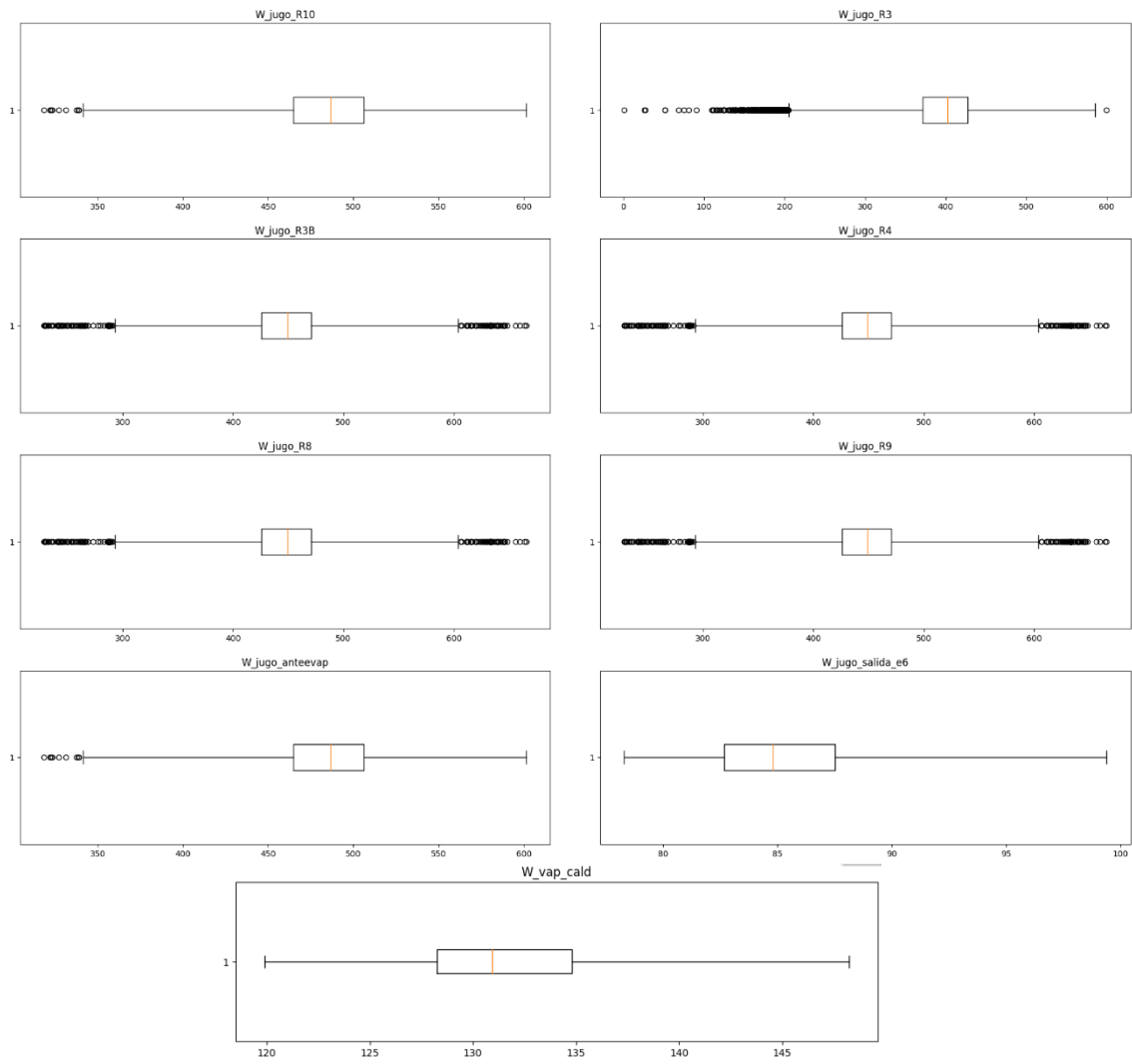


Anexo IV – Boxplots

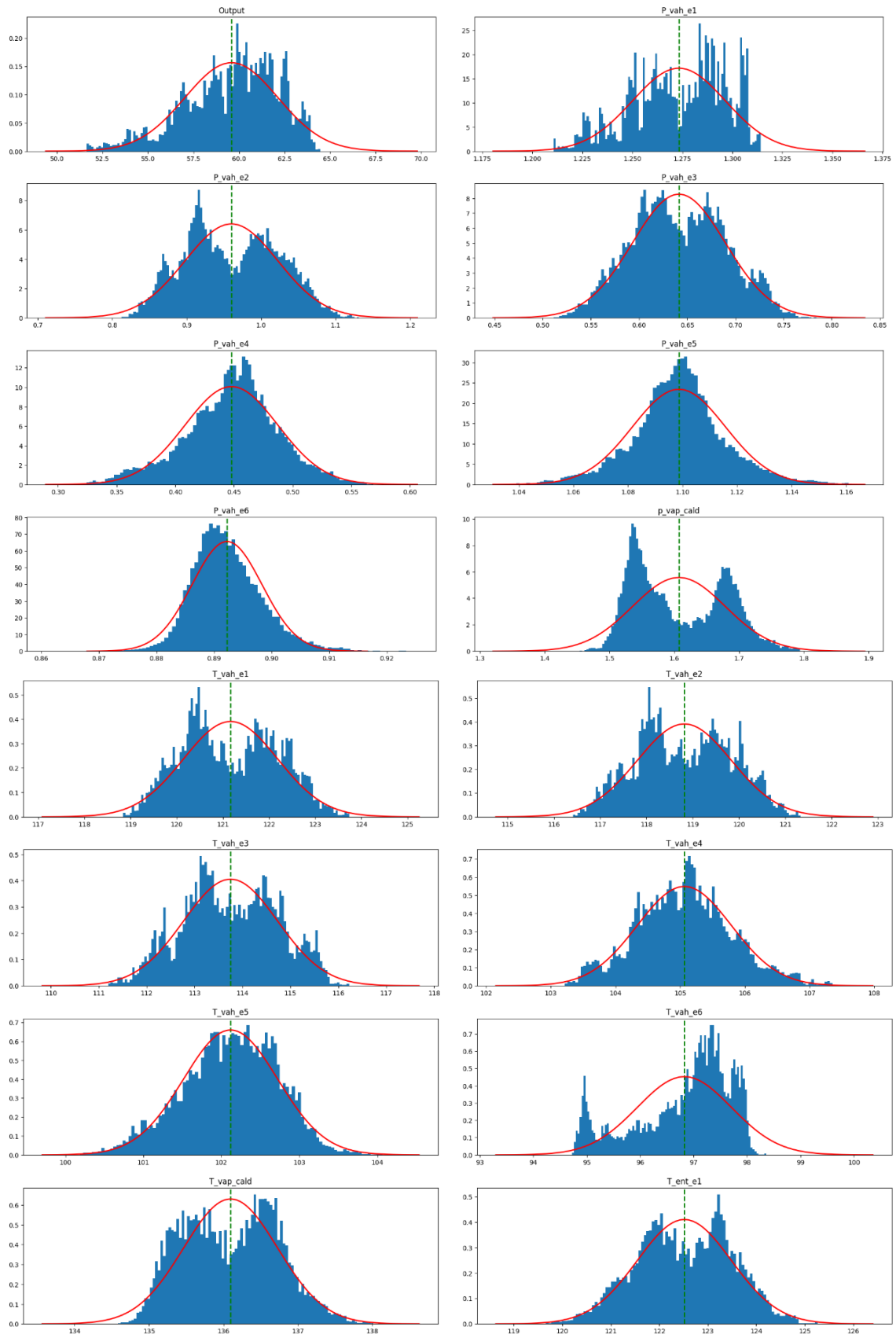


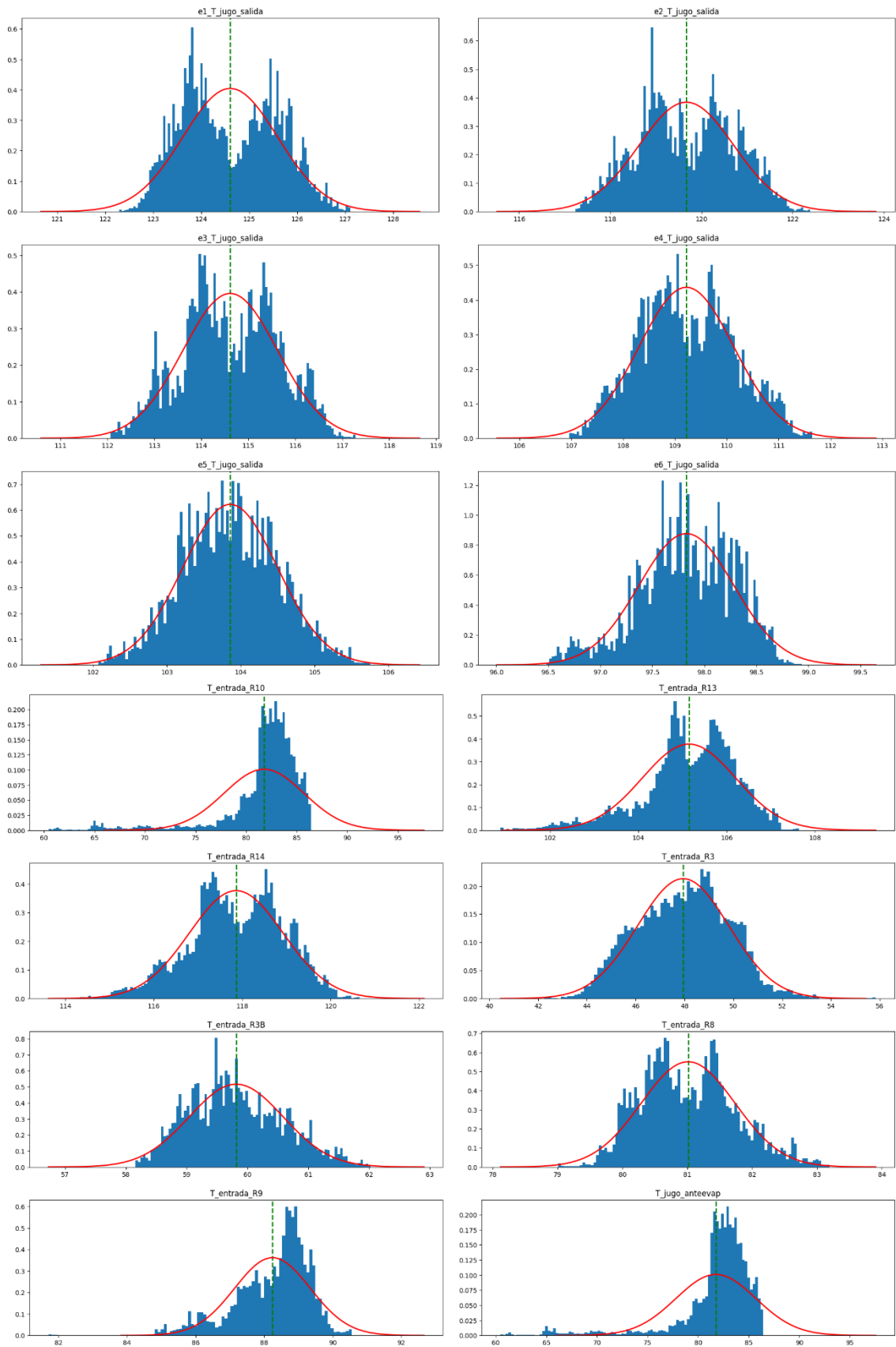


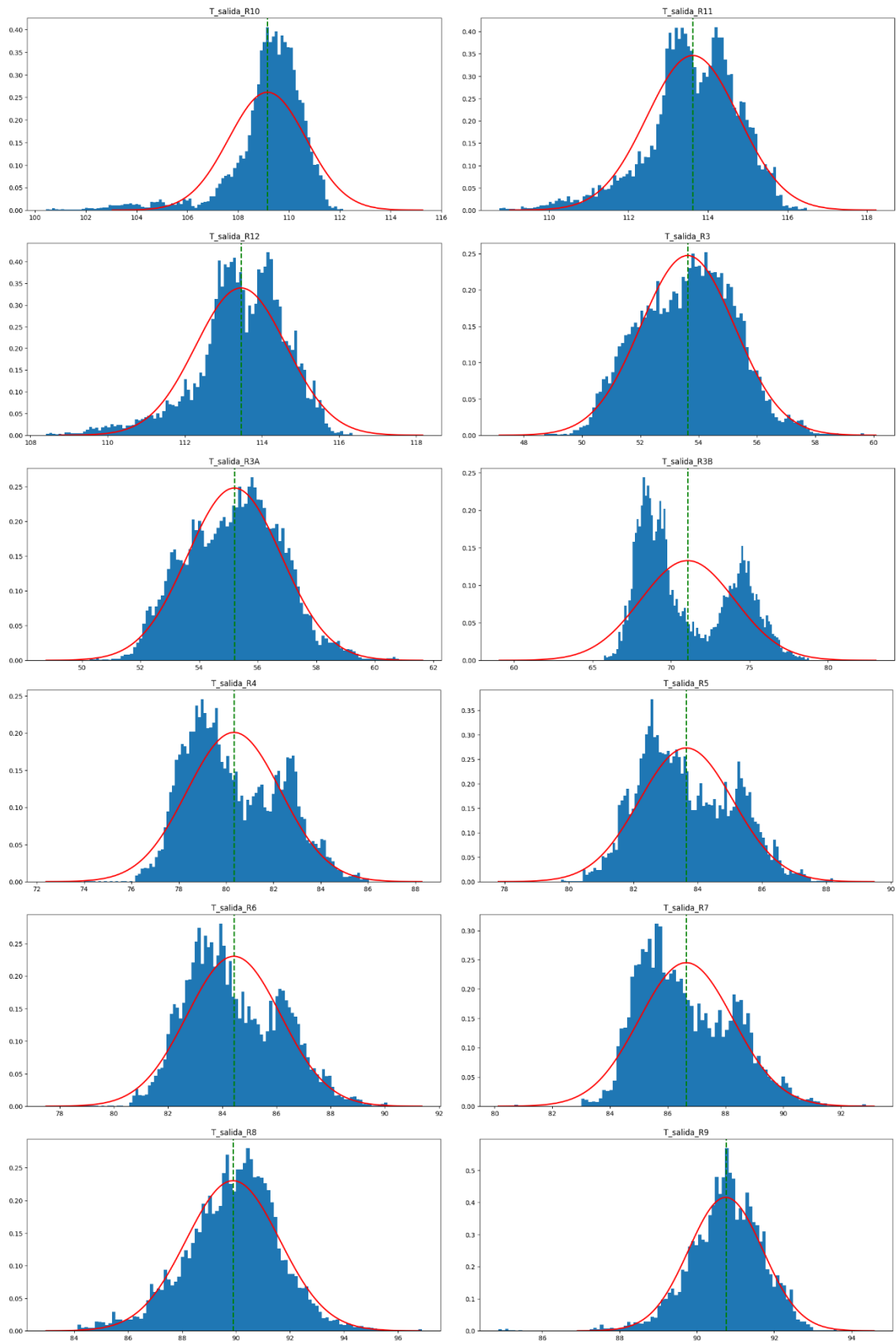


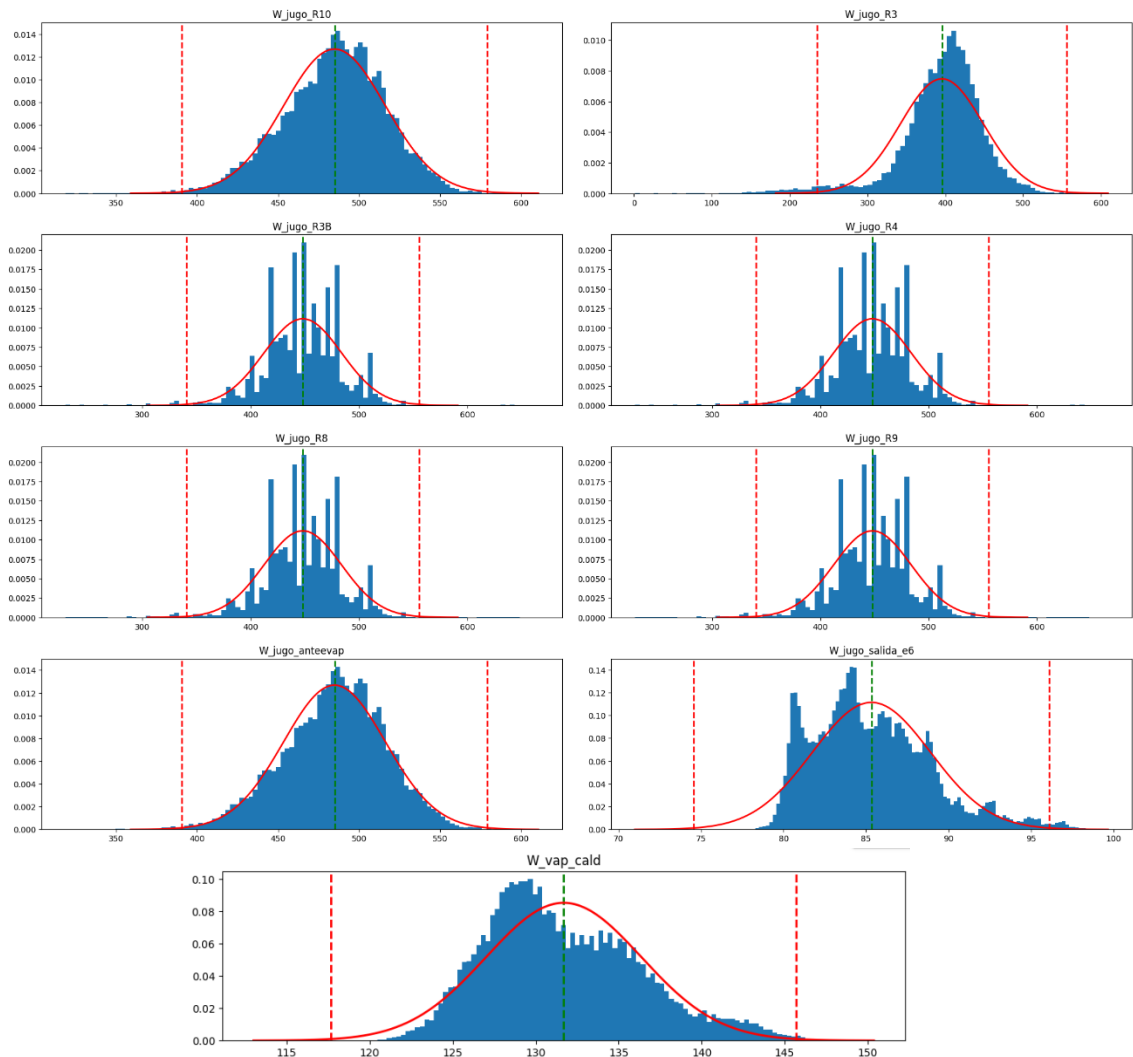


Anexo IV – Histogramas









Anexo V – Filtrado de Variables

