

Máster Universitario Ingeniería Industrial

Trabajo Fin de Máster

Predicción del precio de la electricidad y análisis de los factores determinantes

Autor
Iñigo Mañueco Rubio

Directora
Sara Jiménez del Caso

Madrid
Agosto 2024

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

Predicción del precio de la electricidad y análisis de los factores determinantes

en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el curso académico 2023-2024 es de mi autoría, original e inédito y no ha sido presentado con anterioridad a otros efectos. El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido tomada de otros documentos está debidamente referenciada.

Fdo.: Iñigo Mañueco Rubio Fecha: 28/ 08/ 2024

Iñigo Mañueco Rubio

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO

Fdo.: Sara Jiménez del Caso Fecha: 28/ 08/ 2024





Máster Universitario Ingeniería Industrial

Trabajo Fin de Máster

Predicción del precio de la electricidad y análisis de los factores determinantes

Autor
Iñigo Mañueco Rubio

Directora
Sara Jiménez del Caso

Madrid
Agosto 2024

PREDICCIÓN DEL PRECIO DE LA ELECTRICIDAD Y ANÁLISIS DE LOS FACTORES DETERMINANTES

Autor: Mañueco Rubio, Iñigo.
Director: Jiménez del Caso, Sara.
Entidad Colaboradora: NESS Energy

Palabras clave: Mercado eléctrico, Predicción, OMIE, XGBoost, Hiperparámetros.

RESUMEN DEL PROYECTO

1. Introducción

A raíz de la pandemia causada por el COVID-19 y la posterior crisis energética, los costes y la volatilidad de la electricidad han aumentado respecto a los años anteriores. Esta situación ha generado la necesidad, tanto para productores como para consumidores, de conocer con antelación la evolución del precio de la electricidad tanto a corto como a largo plazo, con el objetivo de poder planificar sus estrategias de venta o compra de energía eléctrica.

A lo largo de este proyecto se han desarrollado tres bloques claramente diferenciados, cada uno de ellos enfocado en lograr los objetivos del proyecto, los cuales se resumen en conocer en mayor detalle el mercado eléctrico español y realizar un modelo de predicción del precio de la electricidad.

En el primer bloque se ha realizado un análisis del estado actual del sistema eléctrico español. Este análisis se basa en un estudio detallado de todos los programas presentados de forma secuencial hasta la obtención del P48, el cual recoge el compromiso que tiene cada generador y consumidor de vender o tomar energía de la red. Conocer en detalle este largo proceso ha sido fundamental para situar el precio que se pretende predecir.

El segundo bloque del proyecto se ha centrado en el desarrollo del modelo de predicción del precio de la electricidad. Para ello se han estudiado diferentes modelos de aprendizaje automático de forma teórica y, posteriormente se han desarrollado en la práctica, comparando los resultados entre ellos. Este bloque es el núcleo del proyecto, en el que se presentan los resultados de los modelos empleados utilizando métricas estandarizadas para compararlos.

En el último y tercer bloque se han examinado los factores que influyen en las

variables escogidas para la realización del modelo. Tras comprobar los resultados del segundo bloque y la relevancia de las variables escogidas para el modelo, se han analizado los factores que influyen en dichas variables para conocer mejor su impacto.

2. Metodología

De los tres bloques que se han presentado anteriormente, aquel en el que se ha seguido una metodología clara es en el segundo, ya que el primero de ellos se centra es aspectos teóricos y el tercero consiste en el análisis de los factores que influyen en las características del modelo.

Para la realización del modelo de predicción del precio de la electricidad se ha empleado el lenguaje de alto nivel de programación Python. Los pasos que se han llevado a cabo para la realización del modelo de predicción del precio de la electricidad han sido los siguientes:

1. Recopilación de los datos de las variables del modelo. Para la realización del modelo se han recopilado datos correspondientes a las características y a la variable objetivo desde el 1 de enero del 2015 hasta el 31 de mayo de 2024. Las características seleccionadas son el mix energético y la demanda de electricidad, mientras que la variable objetivo era el precio de la electricidad horario, resultado del Programa Diario Básico de Funcionamiento. Los datos de las características se han obtenido de ENTSO-e y los datos de la variable objetivo de OMIE.

2. Tratamiento de datos. Una vez reunidos todos los datos, se contaba con 82.500 observaciones, 27 características y 1 variable objetivo. Como consecuencia de la evolución del mercado eléctrico y su transición hacia la europeización de este, a partir del 23 de mayo de 2022, los datos obtenidos de las potencias se encontraban en valores cuarto-horarios. Ante la necesidad de disponer de todos los datos con el mismo formato y teniendo en cuenta que el valor de la variable objetivo es un valor horario, se calculó el promedio de los cuatro cuartos de hora para adaptar la potencia cuarto-horaria a una horaria.
   Por otro lado, aquellos datos de los que no se dispusiese información, se ha rellenado con el valor promedio del resto de datos de la misma característica.
   Finalmente, destacar que, debido a la transformación del mercado eléctrico tras la pandemia sufrida y la crisis energética, se decidió dividir el modelo en tres intervalos temporales con el objetivo de poder comparar los resultados obtenidos, diferenciando las peculiaridades y anomalías de cada periodo.

Los intervalos entre los que se dividió la ejecución del modelo fueron los siguientes:

- 2015-2019: aunque este intervalo no puede ser empleado para predecir el precio de la electricidad de los días futuros, el objetivo es conseguir una referencia y evidenciar la transformación sufrida en el mercado.
- 2020-2024: se ha seleccionado el periodo en el que se inicia la transformación del mercado eléctrico. Estos resultados sí que pueden ser empleados para predecir el precio de la electricidad de los días futuros.
- 2015-2024: este periodo contempla la totalidad de los datos recopilados. Su objetivo es comparar el resultado con el periodo anterior y de esta forma conocer cuál es el que brinda mejores resultados.

3. Desarrollo de los modelos. Para poder desarrollar los modelos, lo primero que se debía hacer es elegir un conjunto de entrenamiento y otro de test. Para ello se decidió implementar el Holdout method [30], por el cual se dividiese de forma aleatoria el conjunto de entrenamiento y el de test, pero estableciendo el 80 % de los datos para el entrenamiento y el 20 % para el test. Una vez realizada esta partición, se emplearon cinco tipos de modelos diferentes de aprendizaje automático supervisado por regresión para predecir el precio de la electricidad. Los tres primeros son modelos más sencillos que sirven para analizar la complejidad del problema y añaden una cota superior al error máximo admisible. Es decir, de haber conseguido resultados satisfactorios con los modelos Baseline, Linear y Linear Ridge hubiese implicado que se trata de un problema de resolución sencilla. En cuanto a los dos modelos restantes, son más sofisticados y permiten abordar problemas más complejos. Los modelos desarrollados se exponen a continuación.

   - Modelo Baseline. Se trata del modelo más simple de todo el proyecto. Para calcular la predicción realiza el promedio de los datos del conjunto de entrenamiento y ese es el valor con el que compara el test.
   - Modelo Linear. Se trata de un modelo de regresión lineal que se ajusta por mínimos cuadrados. El modelo consiste en aproximar la variable objetivo como una recta en el espacio de las características que siga la ecuación 1, donde x es la variable de las características e y es el valor de la predicción [28].

     $$y = b + \mathbf{m} \cdot \mathbf{x} \tag{1}$$

     El objetivo es minimizar la expresión correspondiente a la suma de errores cuadráticos medios sobre el conjunto de entrenamiento para calcular los $p + 1$ parámetros que mejor se ajustan a los mismos:

$$f(m,b) = \sum_{i=1}^{N_e} [t_i - b - \sum_{j=1}^{p} (m_j x_{ij})]^2 \tag{2}$$

- Modelo Linear Ridge. Se trata de un modelo muy similar al anterior en el que se añade un término a la expresión a minimizar que se corresponde con una regularización L2, cuyo objetivo es evitar que se produzca sobreentrenamiento, es decir, que el modelo se ajuste mejor en el entrenamiento que en el test [22]. Para ello, se debe minimizar la expresión siguiente.

$$f(m,b) = \sum_{i=1}^{N_e} [t_i - b - \sum_{j=1}^{p} (m_j x_{ij})]^2 + \lambda \cdot \sum_{j=1}^{p} m_j^2 \tag{3}$$

  .

- Modelo Random Forest. Se trata de un modelo más complejo que los tres analizados anteriormente, y sobre el que se han obtenido resultados con un error menor. Este modelo es una evolución de los árboles de decisión en el que en general se obtienen resultados más robustos [17]. Se trata de un bosque compuesto por varios árboles y que el valor final de la variable respuesta se calcula como la media de de la variable respuesta de cada árbol. En el caso de este modelo se debe llevar a cabo una búsqueda de hiperparámetros con anterioridad a la ejecución del modelo. Esta búsqueda de hiperparámetros consiste en ajustar el valor de algunos parámetros que no pueden ser entrenados durante la fase del entrenamiento y probarlos posteriormente en la parte del test [23].

- Modelo XGBoost. Se trata de un modelo que también está basado en árboles de regresión. La diferencia es que estos árboles se disponen secuencialmente de tal forma que la salida de cada uno de los árboles es utilizada para computar la salida total del modelo para cada iteración. Para entrenar este modelo se emplea un algoritmo iterativo que asocia una puntuación a cada una de las posibles configuraciones del árbol para cada nodo, y se toma la configuración que mejor resultado obtenga [11]. Este modelo, al igual que el anterior, requiere de realizar una búsqueda de hiperparámetros previa al entrenamiento.

4. Comparación entre modelos. Poder estudiar la eficacia del modelo se ha empleado la métrica estandarizada Mean Absolute Error (MAE), cuya expresión se muestra en la Ecuación 4.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - t_i| \tag{4}$$

Donde:

- N: hace referencia al número de observaciones de entrenamiento o de test.
- y_i: valor de la predicción.
- t_i: valor variable respuesta.

3. Resultados

En este proyecto se han obtenido resultados concluyentes en los bloques dos y tres de los definidos en la introducción del resumen. Los resultados del bloque dos se corresponden con los resultados de los modelos, mientras que los del bloque tres se corresponden con los factores que influyen en las características seleccionadas para el modelo.

En cuanto a los resultados del modelo, se muestra en la Figura 1 los MAE para los cinco modelos en los tres intervalos temporales comentados en el tratamiento de los datos tanto para los conjuntos de entrenamiento como de test.

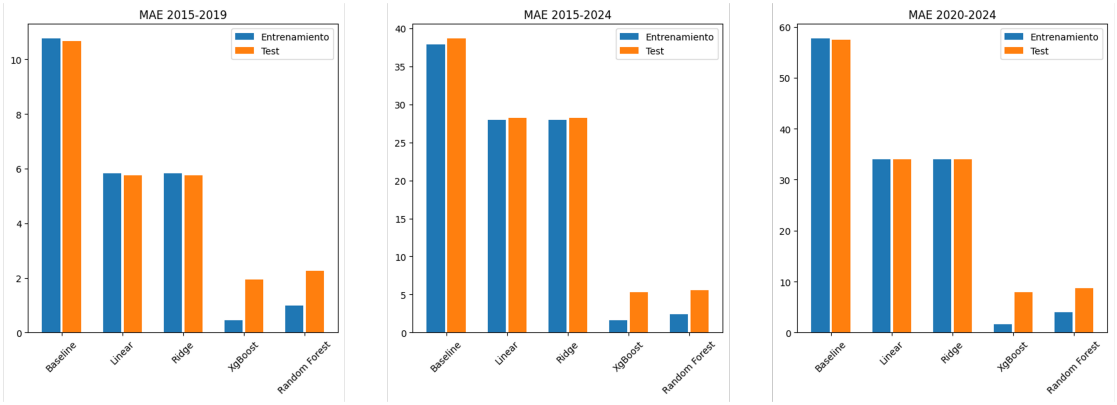

Figura 1: Comparación MAE cinco modelos. Fuente: elaboración propia

En esta figura se puede apreciar cómo se han obtenido los mejores resultados en el periodo 2015-2019, como era de esperar, ya que el mercado se encontraba mucho más estable con menor volatilidad antes de la aparición de la pandemia del COVID-19 y de la crisis energética. En cuanto a los otros dos periodos temporales, se han obtenido mejores resultados en todos los modelos en el periodo que

transcurre desde el 2015 hasta el 2024. Esto se puede deber a que el comportamiento del mercado en el 2024 es más similar al que se tenía en los años prepandémicos.

Poniendo énfasis en los modelos, como se comentaba anteriormente, los tres primeros no han obtenido un buen resultado en comparación con los otros dos debido a que son modelos muy simples y se emplean para establecer una referencia del error máximo admisible. En cuanto a los modelos Random Forest y XGBoost, los resultados en el periodo 2015-2024 han tenido resultados bastante satisfactorios tanto en el entrenamiento como en el test. El modelo que ha ofrecido un mejor comportamiento de los dos ha sido el XGBoost, lo cual es normal debido a que es un modelo más moderno y que suele ofrecer errores menores.

En la Tabla 1 se muestra un resumen de los errores obtenidos por intervalo de tiempo y modelo.

	Baseline		**Linear**		**Linear Ridge**		**Rand. For.**		**XGBoost**	
	Ent.	Test	Ent.	Test	Ent.	Test	Ent.	Test	Ent.	Test
2015 - 2019	10,75	10,67	5,82	5,77	5,82	5,76	0,98	2,25	0,45	1,94
2020 - 2024	57,67	57,47	33,99	34,06	33,99	34,06	4	8,76	1,58	7,87
2015 - 2024	37,87	38,63	27,93	28,21	27,93	28,21	2,38	5,55	1,6	5,26

Tabla 1: MAE (€/MWh) para los cinco modelos en los tres intervalos temporales

En la Tabla 1 se puede observar cómo el mejor resultado que se ha conseguido tiene un MAE de $5,26\frac{€}{MWh}$ en el Modelo XGBoost en el conjunto de test. En la Figura 2, se muestra la comparativa de las últimas 150 observaciones del periodo de test para el intervalo 2015-2024, en la que se muestra en rojo el valor de la predicción para cada hora y en azul el valor real de la electricidad en esa hora.

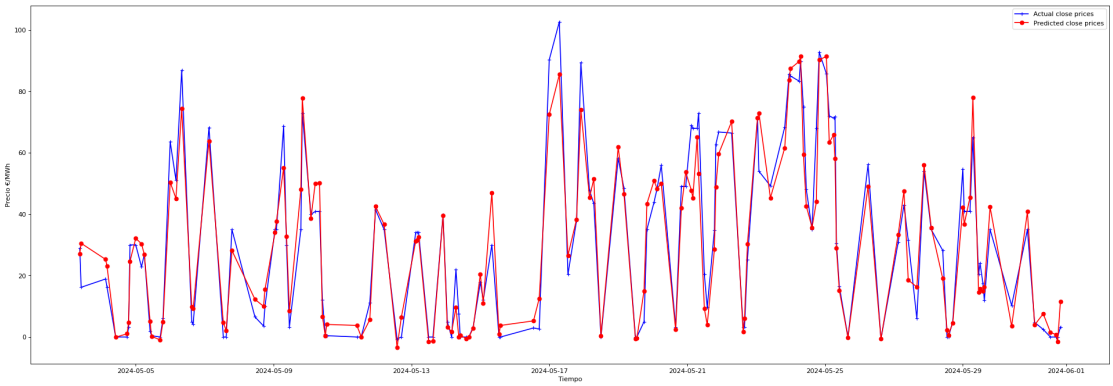

Figura 2: Comparación Modelo XGBoost 2015-2024. Fuente: elaboración propia

Una vez se ha demostrado la eficacia de los resultados obtenidos en el modelo,

estos han servido para confirmar que las características elegidas son correctas. Por ese motivo, y con el objetivo de conocer mejor su impacto en la precisión de las predicciones, se ha realizado un análisis de los factores que influyen en las características del modelo, la demanda y el mix de generación eléctrico. Las conclusiones obtenidas han sido las siguientes:

- La demanda de electricidad se encuentra claramente influenciada por factores como la temperatura, las festividades nacionales y los fines de semana. Además, hasta el año 2021 la demanda y la situación económica del país también se encontraban vinculadas. Sin embargo, por factores como la eficiencia energética, la digitalización y el autoconsumo, se rompió este vínculo entre esas dos características.

- En cuanto a los factores que influyen en el mix de generación eléctrico, destacan la climatología, los costes de producción asociados a cada tecnología, entre los que juegan un papel clave los costes de combustibles y los costes de la compra de los derechos de emisión (solo en ciertas tecnologías como en los ciclos combinados y cogeneración), y la disponibilidad de las plantas.

4. Conclusiones

Los modelos Random Forest y XGBoost desarrollados han resultado ser fiables a la hora de predecir el precio de la electricidad, obteniendo una mejora considerable con respecto a los tres modelos previamente desarrollados. El mejor modelo en cuanto a resultados en términos de MAE ha sido el XGBoost, que podría ser empleado por comercializadoras o productores para sus estrategias de venta o de compra. Las posibles limitaciones que presentan estos modelos es la incertidumbre de cómo se comportarían en el caso de que se produjese otro cambio en la tendencia del mercado como ya ocurrió años atrás. Como líneas futuras de investigación se propone utilizar como variables de entrada los factores que se han determinado que influyen en las características usadas en este proyecto y el uso de otros modelos de aprendizaje automático como redes neuronales.

5. Bibliografía

[11] Tianqi Chen y Carlos Guestrin. «Xgboost: A scalable tree boosting system». En: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016, págs. 785-794.

[17] Javier Jesús Espinosa-Zúñiga. «Aplicación de algoritmos Random Forest y XGBoost en una base de solicitudes de tarjetas de crédito». En: *Ingeniería, investigación y tecnología* 21.3 (2020).

[22] Álvaro Garnica Barco. «Regularización en problemas lineales de regresión: Una visión estadística y económica aplicada a datos de desarrollo humano». B.S. thesis. Universitat Politècnica de Catalunya, 2021.

[23] A. Géron. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, 2019, págs. 75-83. ISBN: 9781492032595. URL: `https://books.google.es/books?id=HnetDwAAQBAJ`.

[28] Ernestina Menasalvas. «Algoritmos de Machine Learning». En: (2021).

[30] Gonzalo Rodrigo Navarro Armijo. «La evolución y eficiencia del mix eléctrico ante la penetración de las energías renovables en Alemania, Canadá, Dinamarca, España, Francia, Italia, y Noruega». En: (2022).

ELECTRICITY PRICE PREDICTION AND ANALYSIS OF THE DETERMINING FACTORS

Keywords: Electricity market, Prediction, OMIE, XGBoost, Hyperparameters.

1. Introduction

Due to the COVID-19 pandemic and the subsequent energy crisis, electricity costs and volatility have increased compared to previous years. This situation has created the need for both producers and consumers to know in advance the evolution of electricity prices in both the short and long term, in order to be able to plan their strategies for the sale or purchase of electricity.

Throughout this project, three clearly differentiated blocks have been developed, each of them focused on achieving the project's objectives, which can be summarized as: to know in greater detail the Spanish electricity market and to develop a model for predicting the price of electricity.

In the first block, an analysis of the current state of the Spanish electricity system has been carried out. This analysis is based on a detailed study of all the programs presented sequentially until the P48 is obtained, which reflects the commitment of each generator and consumer to sell or take energy from the grid. Knowing in detail this long process has been fundamental to set the price to be predicted.

The second part of the project is focused on the development of the electricity price prediction model. For this purpose, different Machine Learning models have

been studied theoretically and then developed in practice, comparing the results among them. This block is the core of the project, in which the results of the models used are presented using standardized metrics for comparison.

In the last and third block, the factors influencing the variables chosen for the model were examined. After verifying the results of the second block and the relevance of the variables chosen for the model, the factors influencing these variables were analyzed to better understand their impact.

2. Methodology

Among the three blocks presented above, the one in which a clear methodology has been followed is the second, since the first block focused on theoretical aspects and the third block consisted of the analysis of the factors that influence the characteristics of the model.

The high-level programming language Python was used to develop the electricity price prediction model. The steps that have been carried out for the realization of the electricity price prediction model were the following:

1. Collection of data on the model variables. For the realization of the model, data corresponding to the characteristics and the target variable were collected from January 1st, 2015 to May 31st, 2024. The selected characteristics are the energy mix and electricity demand, while the target variable was the hourly electricity price, resulting from the Basic Daily Operating Schedule. The characteristics data were obtained from ENTSO-e and the target variable data from OMIE.

2. Data processing. Once all the data were collected, there were 82,500 observations, 27 characteristics and 1 target variable. As a consequence of the evolution of the electricity market and its transition towards the europeization of the market, as of May 23, 2022, the data obtained from the powers were in quarter-hourly values. Given the need to have all the data in the same format and taking into account that the value of the target variable is an hourly value, the average of the four quarter hours was calculated to adapt the quarter-hourly power to an hourly one.
   On the other hand, those data for which no information was available were filled in with the average value of the rest of the data of the same characteristic.
   Finally, it should be noted that, due to the transformation of the electricity market after the pandemic and the energy crisis, it was decided to divide the model into three time intervals in order to compare the results obtained,

differentiating the peculiarities and anomalies of each period. The intervals in which the model run was divided were as follows:

- 2015-2019: although this interval cannot be used to predict the price of electricity for future days, the objective is to obtain a reference and to show the transformation of the market.
- 2020-2024: the period in which the transformation of the electricity market begins has been selected. These results can be used to predict the price of electricity for future days.
- 2015-2024: this period contemplates the totality of the data collected. Its objective is to compare the result with the previous period and thus know which one provides the best results.

3. Model development. In order to develop the models, the first thing to do was to choose a training set and a test set. For this purpose, it was decided to implement the Holdout method [30], whereby the training and test sets were randomly divided, but with 80 % of the data being set for training and 20 % for testing. Once this partitioning was done, five different types of regression supervised machine learning models were used to predict the price of electricity. The first three are simpler models that serve to analyze the complexity of the problem and add an upper bound to the maximum admissible error. In other words, if satisfactory results had been obtained with the Baseline, Linear and Linear Ridge models, it would have implied that this is a simple problem to solve. As for the remaining two models, they are more sophisticated and allow us to tackle more complex problems. The models developed are presented below.

   - Baseline model. This is the simplest model of the whole project. To calculate the prediction, it averages the data of the training set and compares the test with this value.
   - Linear model. This is a linear regression model fitted by least squares. The model consists of approximating the target variable as a line in the feature space following equation [28], where x is the feature variable and y is the prediction value.

     $$y = b + \mathbf{m} \cdot \mathbf{x} \tag{5}$$

     The objective is to minimize the expression corresponding to the sum of mean squared errors over the training set of to compute the $p + 1$ best-fit parameters:

$$f(m,b) = \sum_{i=1}^{N_e} [t_i - b - \sum_{j=1}^{p} (m_j x_{ij})]^2 \tag{6}$$

- Linear Ridge model. This is a model very similar to the previous one, with the addition of a term to the expression being minimized, which corresponds to L2 regularization. The purpose of this is to prevent overfitting, meaning that the model performs better during training than in testing [22]. To achieve so, the following expression must be minimized.

$$f(m,b) = \sum_{i=1}^{N_e} [t_i - b - \sum_{j=1}^{p} (m_j x_{ij})]^2 + \lambda \cdot \sum_{j=1}^{p} m_j^2 \tag{7}$$

.

- Random Forest model. This model is more complex than the three previously analyzed and has yielded results with a lower error rate. It is an evolution of decision trees, generally producing more robust outcomes [17]. It is a forest composed of multiple trees, with the final value of the response variable calculated as the average of the response variable across all trees. For this model, a hyperparameter search must be conducted prior to model execution. This hyperparameter search involves adjusting the values of certain parameters that cannot be trained during the training phase and then proving them in the test phase [23].
- XGBoost model. This is a model that is also based on regression trees. The difference lies in the sequential arrangement of these trees, where the output of each tree is used to compute the overall model output in each iteration. To train this model, an iterative algorithm is employed that assigns a score to each possible tree configuration for each node, selecting the configuration that yields the best result [11]. This model, like the previous one, requires a hyperparameter search prior to training.

4. Comparison among models: in order to study the effectiveness of the model, the standardized metric Mean Absolute Error (MAE), whose expression is shown in Equation 8.

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - t_i| \tag{8}$$

Where:

- N: refers to the number of training or test observations.

- y_i: predictive value.
- t_i: response variable value.

3. Results

In this project, conclusive results have been obtained in blocks two and three of those defined in the introduction of the summary. The results in block two correspond to the model results, while those in block three correspond to the factors influencing the characteristics selected for the model.

As for the model results, the Figure 3 shows the MAE for the five models in the three time intervals discussed in the data treatment for both the training and test sets.

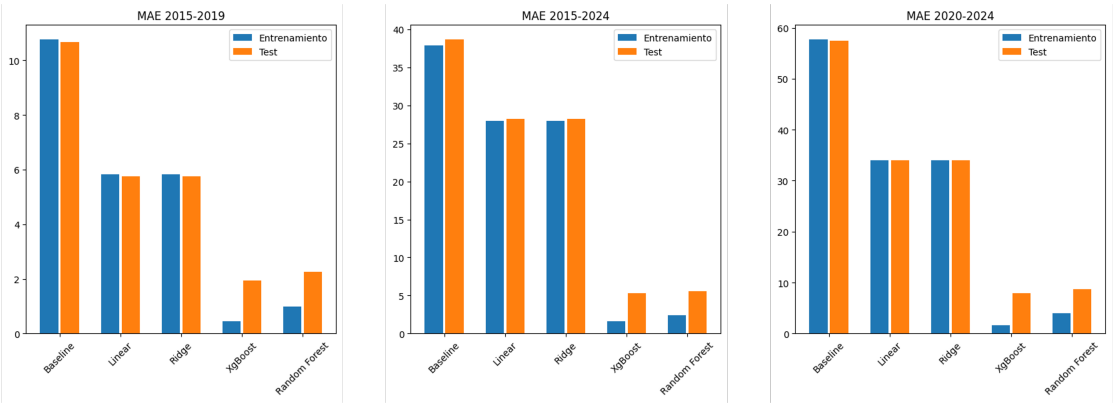

Figura 3: Comparison MAE among five models. Source: own elaboration

This figure shows how the best results have been obtained in the period 2015-2019, as expected, since the market was much more stable with less volatility before the appearance of the COVID-19 pandemic and the energy crisis. As for the other two time periods, better results were obtained in all models in the period from 2015 to 2024. This may be due to the fact that the market behavior in 2024 is more similar to that in the pre-pandemic years.

Emphasizing the models, as mentioned above, the first three have not obtained a good result compared to the other two because they are very simple models and are used to establish a reference of the maximum allowable error. As for the Random Forest and XGBoost models, the results in the period 2015-2024 have had quite satisfactory results in both training and testing. The model that has offered a better performance of the two has been the XGBoost, which is normal because it is a more modern model and usually offers lower errors.

A summary of the errors obtained by time interval and model is shown in the Table 2.

	Baseline		**Linear**		**Linear Ridge**		**Rand. For.**		**XGBoost**	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
2015 - 2019	10.75	10.67	5,82	5.77	5.82	5.76	0.98	2.25	0.45	1.94
2020 - 2024	57.67	57.47	33.99	34.06	33.99	34.06	4	8.76	1.58	7.87
2015 - 2024	37.87	38.63	27.93	28.21	27.93	28.21	2.38	5.55	1.6	5.26

Tabla 2: MAE (€/MWh) for the five models in the three time periods

In Table 2, it can be observed that the best result achieved is a MAE of $\frac{€5,26}{MWh}$ with the XGBoost model on the test set. Figure 4 illustrates the comparison of the last 150 observations from the test period for the 2015-2024 interval, showing the predicted values in red for each hour and the actual electricity values for that hour in blue.

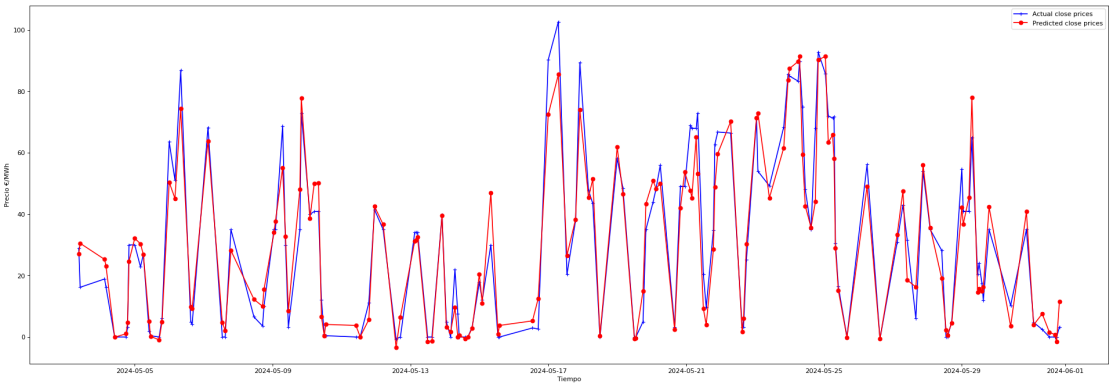

Figura 4: Comparison XGBoost model 2015-2024. Source: own elaboration

Once the effectiveness of the results obtained in the model has been demonstrated, these have served to confirm that the chosen characteristics are correct. For this reason, and in order to better understand their impact on the accuracy of the predictions, an analysis of the factors that influence the characteristics of the model, the demand and the electricity generation mix has been carried out. The conclusions obtained were as follows:

- Electricity demand is clearly influenced by factors such as temperature, national holidays and weekends. In addition, until 2021 demand and the country's economic situation were also linked. However, due to factors such as energy efficiency, digitalization and self-consumption, this link between these two characteristics was broken.

- As for the factors that influence the electricity generation mix, the following stand out: weather, the production costs associated with each technology, among which fuel costs and the costs of purchasing emission rights (only in certain technologies such as combined cycle and cogeneration) play a key role, and the availability of the plants.

4. Conclusions

The Random Forest and XGBoost models developed have proved to be reliable in predicting the price of electricity, obtaining a considerable improvement with respect to the three previously developed models. The best model in terms of MAE results has been the XGBoost, which could be used by traders or producers for their sale or purchase strategies. The possible limitations of these models is the uncertainty of how they would behave in the event of another change in the market trend, as happened years ago. As future lines of research it is proposed to use as input variables the factors that have been determined to influence the characteristics used in this project and the use of other machine learning models such as neural networks.

5. Bibliography

[11] Tianqi Chen y Carlos Guestrin. «Xgboost: A scalable tree boosting system». En: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016, págs. 785-794.

[17] Javier Jesús Espinosa-Zúñiga. «Aplicación de algoritmos Random Forest y XGBoost en una base de solicitudes de tarjetas de crédito». En: *Ingeniería, investigación y tecnología* 21.3 (2020).

[22] Álvaro Garnica Barco. «Regularización en problemas lineales de regresión: Una visión estadística y económica aplicada a datos de desarrollo humano». B.S. thesis. Universitat Politècnica de Catalunya, 2021.

[23] A. Géron. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, 2019, págs. 75-83. ISBN: 9781492032595. URL: `https://books.google.es/books?id=HnetDwAAQBAJ`.

[28] Ernestina Menasalvas. «Algoritmos de Machine Learning». En: (2021).

[30] Gonzalo Rodrigo Navarro Armijo. «La evolución y eficiencia del mix eléctrico ante la penetración de las energías renovables en Alemania, Canadá, Dinamarca, España, Francia, Italia, y Noruega». En: (2022).

Agradecimientos

Me gustaría que las primeras líneas de este Trabajo de Fin de Máster vayan dirigidas a agradecer a aquellas personas que han hecho posible finalizar una etapa tan bonita como la vivida en los últimos dos años.

Para empezar, quería agradecer a mi directora y compañera de trabajo, Sara Jiménez, la cual ha invertido una gran cantidad de su tiempo para que pueda sacar adelante este proyecto y de la que he aprendido una cantidad innumerable de cosas.

Al resto de mis compañeros del trabajo, que también me han ayudado en ciertos aspectos del proyectos gracias a sus conocimientos en el sector.

A mis compañeros y amigos, tanto del máster como del grado, que siempre hemos estado ahí para ayudarnos unos a otros en cualquier cosa que necesitásemos.

Por último a mi familia, mis padres, mi hermano y Ale, los que han creído en mí desde el principio y han sido mi apoyo fundamental sin los que nada de esto hubiese sido posible.

Índice general

Índice de figuras xxi

Índice de cuadros xxv

1. Introducción 1

2. Estado del arte 3
2.1. Evolución histórica sector energético español . . . 3
2.2. Agentes del mercado eléctrico . . . 5
2.3. Mercados eléctricos en España . . . 7
2.3.1. Mercados no organizados o bilaterales . . . 9
2.3.2. Mercados organizados . . . 10
2.4. Facturación al consumidor final. Tipologías de precios . . . 24
2.4.1. Término de potencia . . . 25
2.4.2. Término de energía . . . 26

3. Predicción del precio de la electricidad 29
3.1. Introducción al aprendizaje automático . . . 30
3.2. Algoritmos de aprendizaje automático . . . 34
3.2.1. Modelo Baseline . . . 34
3.2.2. Modelo Linear . . . 36
3.2.3. Modelo Linear Ridge . . . 40
3.2.4. Modelo Random Forest . . . 40
3.2.5. Modelo XGBoost . . . 44
3.3. Features consideradas en el modelo . . . 46
3.4. Resultados obtenidos . . . 50
3.4.1. Tratamiento de datos . . . 50
3.4.2. Resultados modelo Baseline . . . 51
3.4.3. Resultados modelo Linear . . . 55
3.4.4. Resultados modelo Linear Ridge . . . 59
3.4.5. Resultados modelo Random Forest . . . 63

3.4.6. Resultados modelo XGBoost 68
3.4.7. Resultados generales de los modelos 73

4. Factores clave en las features del modelo **75**
4.1. Factores que influyen en la demanda 75
4.1.1. Temperatura 77
4.1.2. Situación económica del país 79
4.1.3. Festividades y fines de semana 80
4.2. Factores que influyen en el mix eléctrico 82
4.2.1. Climatología 83
4.2.2. Costes asociados a cada tecnología 85
4.2.3. Disponibilidad de las plantas 92

5. Conclusiones y líneas futuras de investigación **97**

A. Alineamiento del Proyecto con los ODS **101**
A.1. ODS 7: Energía asequible y no contaminante 102
A.2. ODS 9: Industria, innovación e infraestructura 102
A.3. ODS 12: Producción y consumo responsables 103
A.4. ODS 13: Acción por el clima 103

Bibliografía **105**

Índice de figuras

1. Comparación MAE cinco modelos. Fuente: elaboración propia . . . VII
2. Comparación Modelo XGBoost 2015-2024. Fuente: elaboración propia VIII
3. Comparison MAE among five models. Source: own elaboration . . . XIV
4. Comparison XGBoost model 2015-2024. Source: own elaboration . . XV

2.1. Secuencia de mercados. Fuente: OMIE 9
2.2. Secuencia Programas hasta el Programa Diario Viable. Fuente: OMIE 12
2.3. Subastas intradiarias con anterioridad al día 13/06/2024. Fuente: OMIE 14
2.4. Subastas intradiarias con posterioridad al 13/06/2024. Fuente: OMIE 18
2.5. Secuencia subastas intradiarias posteriores al 13/06/2024. Fuente: OMIE 20
2.6. Proceso de Programación REE. Fuente: REE 21
2.7. Servicios de balance. Fuente: REE 23

3.1. Esquema del precio a predecir. Fuente: REE 30
3.2. Esquema modelos de aprendizaje. Fuente: elaboración propia 32
3.3. Ejemplo Baseline. Fuente: elaboración propia 36
3.4. Ejemplo Regresión Lineal. Fuente: elaboración propia 39
3.5. Ejemplo árbol de decisión. Fuente: elaboración propia 41
3.6. Algoritmo Random Forest [17] 43
3.7. Matriz Correlación Pearson Features Modelo. Fuente: elaboración propia 49
3.8. Comparación Modelo Baseline 2015-2019. Fuente: elaboración propia 52
3.9. MAE Modelo Baseline 2015-2019. Fuente: elaboración propia 53
3.10. Comparación Modelo Baseline 2020-2024. Fuente: elaboración propia 53
3.11. MAE Modelo Baseline 2020-2024. Fuente: elaboración propia 54
3.12. Comparación Modelo Baseline 2015-2024. Fuente: elaboración propia 54
3.13. MAE Modelo Baseline 2015-2024. Fuente: elaboración propia 55
3.14. Comparación Modelo Linear 2015-2019. Fuente: elaboración propia 56
3.15. MAE Modelo Linear 2015-201. Fuente: elaboración propia 56
3.16. Comparación Modelo Linear 2020-2024. Fuente: elaboración propia 57

3.17. MAE Modelo Linear 2020-2024. Fuente: elaboración propia 58
3.18. Comparación Modelo Linear 2015-2024. Fuente: elaboración propia 58
3.19. MAE Modelo Linear 2015-2024. Fuente: elaboración propia 59
3.20. Comparación Modelo Linear Ridge 2015-2019. Fuente: elaboración propia 60
3.21. MAE Modelo Linear Ridge 2015-2019. Fuente: elaboración propia . 61
3.22. Comparación Modelo Linear Ridge 2020-2024. Fuente: elaboración propia 61
3.23. MAE Modelo Linear Ridge 2020-2024. Fuente: elaboración propia . 62
3.24. Comparación Modelo Linear Ridge 2015-2024. Fuente: elaboración propia 62
3.25. MAE Modelo Linear Ridge 2015-2024. Fuente: elaboración propia . 63
3.26. Comparación Modelo Random Forest 2015-2019. Fuente: elaboración propia 65
3.27. MAE Modelo Random Forest 2015-2019. Fuente: elaboración propia 65
3.28. Comparación Modelo Random Forest 2020-2024. Fuente: elaboración propia 66
3.29. MAE Modelo Random Forest 2020-2024. Fuente: elaboración propia 66
3.30. Comparación Modelo Random Forest 2015-2024. Fuente: elaboración propia 67
3.31. MAE Modelo Random Forest 2015-2024. Fuente: elaboración propia 67
3.32. Comparación Modelo XGBoost 2015-2019. Fuente: elaboración propia 70
3.33. MAE Modelo XGBoost 2015-2019. Fuente: elaboración propia . . . 70
3.34. Comparación Modelo XGBoost 2020-2024. Fuente: elaboración propia 71
3.35. MAE Modelo XGBoost 2020-2024. Fuente: elaboración propia . . . 71
3.36. Comparación Modelo XGBoost 2015-2024. Fuente: elaboración propia 72
3.37. MAE Modelo XGBoost 2015-2024. Fuente: elaboración propia . . . 72
3.38. Comparación MAE cinco modelos. Fuente: elaboración propia . . . 73

4.1. Curvas agregadas de oferta y demanda el 23/01/2024 en la hora 1. Fuente: OMIE 76
4.2. Comparación de la demanda frente a la temperatura en España en 2023. Fuente: elaboración propia. Datos demanda: REE. Datos temperatura: SIAR 78
4.3. Comparación de la demanda frente al PIB en España e 2023. Fuente: elaboración propia. Datos demanda: REE 79
4.4. Variación de la demanda en febrero. Naranja:fines de semana. Azul: Días laborables. Fuente: elaboración propia. Datos demanda: REE . 81
4.5. Variación de la demanda en diciembre. Naranja:fines de semana. Azul: Días laborables. Verde: Días festivos. Fuente: elaboración propia. Datos demanda: REE 81

4.6. Mix de generación eléctrica en 2023. Fuente: REE 83
4.7. Comparación de la generación eólica frente a la velocidad media del viento en Canarias en enero de 2023. Fuente: elaboración propia. Datos generación: REE. Datos velocidad viento: SIAR 84
4.8. Comparación de la generación fotovoltaica frente a la radiación solar media en Canarias en enero de 2023. Fuente: elaboración propia. Datos generación: REE. Datos radiación solar media: SIAR 85
4.9. Precio del uranio como combustible nuclear entre 2015 y 2023. Fuente: elaboración propia. Datos precios: Trading Economics 87
4.10. Precio gas natural PVB. Fuente: elaboración propia. Datos: MIBGAS 88
4.11. Comparación generación con gas natural frente al precio del gas. Fuente: elaboración propia. Datos generación: REE. Datos precio: MIBGAS 89
4.12. Emisiones de CO2eq por fuente de producción en España en 2023. Fuente: elaboración propia. Datos: REE. 90
4.13. Coste asociado a los derechos de emisión entre 2020 y 2021. Fuente: elaboración propia 92
4.14. Comparación potencia nuclear instalada, frente a potencia nuclear disponible y potencia media diaria demandada en 2023. Fuente: elaboración propia. Datos generación: REE 93
4.15. Comparación potencia nuclear instalada, frente a potencia nuclear disponible y potencia media diaria demandada entre abril y junio de 2023. Fuente: elaboración propia. Datos generación: REE 94
4.16. Mix de generación eólica, fotovoltaica e hidráulica entre abril y junio de 2023. Fuente: elaboración propia. Datos generación: REE 95
4.17. Mix de generación ciclo combinado y cogeneración entre abril y junio de 2023. Fuente: elaboración propia. Datos generación: REE 96

5.1. Comparación MAE cinco modelos. Fuente: elaboración propia . . . 98

A.1. Objetivos de Desarrollo Sostenible de las Naciones Unidas. Fuente: Web Naciones Unidas 101

Índice de cuadros

1. MAE (€/MWh) para los cinco modelos en los tres intervalos temporales VIII
2. MAE (€/MWh) for the five models in the three time periods XV

2.1. Secuencia subastas intradiarias con anterioridad al día 13/06/2024 . 13
2.2. Secuencias IDAs mercados intradiarios 18
2.3. Valor ATR término de potencia actualizado el 15/02/2024 para suministros ordinarios 26
2.4. Valor ATR término de energía actualizado el 15/02/2024 para suministros ordinarios 28

3.1. Ejemplo K-fold cross-validation con k=5 33
3.2. Ejemplo modelo Baseline 35
3.3. Ejemplo modelo Linear 38
3.4. Resultados Modelo Baseline en Entrenamiento y Test 55
3.5. Resultados Modelo Linear en Entrenamiento y Test 59
3.6. Resultados Modelo Linear Ridge en Entrenamiento y Test 63
3.7. Resultados hiperparámetros modelo Random Forest 64
3.8. Resultados Modelo Random Forest en Entrenamiento y Test 68
3.9. Resultados hiperparámetros modelo XGBoost 69
3.10. Resultados Modelo XGBoost en Entrenamiento y Test 73

4.1. Comparación cualitativa de las principales tecnologías de generación de electricidad en España 86
4.2. Paradas centrales nucleares en 2023 93

5.1. MAE (€/MWh) para los cinco modelos en los tres intervalos temporales 98
5.2. Errores porcentuales frente al promedio de mercado para los modelos Random Forest y XGBoost 99

Índice de abreviaturas

ATR Acceso de Terceros a la Red. xxv, 26–28

CNMC Comisión Nacional de los Mercados y la Competencia. 5

ENTSO-e European Network of Transmission System Operators for Electricity. iv, xi, 47

EU ETS European Union Emission Trading System. 91

GCT Gate Closure Time. 18, 19

GFS Good-For-Session. 17

GTD Good-Till-Date. 17

HVAC Heating Ventilation Air Conditioning. 77

ICP Interruptor de Control de Potencia. 25

IDA Intraday Auction. xxv, 17–19

LTS Local Trading Solutions. 16

MAE Mean Absolute Error. vi–ix, xiii–xvi, xxi–xxiii, xxv, 34–37, 39, 44, 45, 53–59, 61–63, 65–68, 70–73, 97–100

MIBEL Mercado Ibérico de la Energía. 7, 13, 18

MIC Mercado Intradiario Continuo. 16

NEMOs Nominated Electricity Market Operators. 15, 18, 19

ODS Objetivos de Desarrollo Sostenible. 100–103

OMIE Operador del Mercado Eléctrico de la Energía. iv, xi, xxi, xxii, 6, 8, 9, 12, 14, 16–18, 20, 27, 30, 47, 76

OMIP Operador del Mercado Ibérico - Polo Portugués. 8

PDBC Programa Diario Base de Casación. 12

PDBF Programa Diario Base de Funcionamiento. 12, 20, 22, 29

PDVD Programa Diario Viable. 12

PDVP Programa Diario Viable Provisional. 20, 21, 29

PHF Programas Horarios Finales. 13, 20, 21, 30

PHFC Programas Horarios Finales del Continuo. 17, 20, 21, 30

PIB Producto Interior Bruto. xxii, 79, 80, 99

PIBCA Programa Intradiario Básico de Casación Acumulada. 13

PIBCI Programa Intradiario Básico de Casación Incremental. 13

PNIEC Plan Nacional Integrado de Energía y Clima (2021-2030). 80

REE Red Eléctrica Española. xxi–xxiii, 6, 8, 21, 23, 30, 77–79, 81, 83–85, 89, 90, 93–96

RMSE Root Mean Square Error. 34

SIAR Sistema de Información Agroclimática para el Regadío. xxii, xxiii, 77, 78, 84, 85

STEEEP Support & Training for an Excellent Energy Efficiency Performance. 80

UNESA Unidad Eléctrica S.A.. 4

XBID Cross-Border Intraday. 13, 16

Capítulo 1

Introducción

Hoy en día, nos encontramos sumergidos en una de las mayores crisis energéticas que se ha vivido a lo largo de la historia. Los últimos sucesos ocurridos en el mundo como son una pandemia mundial originada por el COVID – 19 o la Guerra de Ucrania han puesto en cuestión el sistema eléctrico europeo y la capacidad de abastecer la demanda en condiciones normales. Además, nos encontramos en una constante transición energética, intentando emplear fuentes de energía que sean menos perjudiciales para el medio ambiente y desarrollando nuevas tecnologías de generación que marcarán el futuro de este sector en el mundo.

En este contexto de incertidumbre, un gran porcentaje de la población se ha visto afectada como consumidores de energía, subiendo el precio de las facturas hasta cifras desorbitadas tanto de la electricidad como del gas. En España, el promedio del mercado spot en 2020 llegó a situarse en un promedio de 33,96€/MWh, bastante por debajo del promedio de los años anteriores debido a la inactividad causada por la pandemia del COVID-19. Sin embargo, dos años después, se alcanzaron precios de electricidad 16 veces por encima del promedio de 2020 debido a la Guerra de Ucrania, entre otros.

Ante este problema de volatilidad del precio de la electricidad nos surgen preguntas como las que se exponen a continuación: ¿De qué depende realmente esta variación de los precios? ¿Va a continuar creciendo el precio de la electricidad? ¿Podemos predecir hacia dónde tiende el precio de la electricidad en el futuro?

En este proyecto se pretenden resolver todas estas dudas que se presentan acerca del futuro del precio de la electricidad y dar las claves acerca de qué variables pueden afectar a este.

Capítulo 2

Estado del arte

En este capítulo se analizará, en primer lugar, la evolución del sector eléctrico en España con el fin de comprender su situación actual. A continuación, se examinarán los agentes que participan en los mercados eléctricos hoy en día y se explicará el proceso por el cual se determinan los precios de la electricidad en España.

2.1. Evolución histórica sector energético español

La historia de la electricidad en España comienza en 1852, año en el que se considera que se empleó la electricidad con un sentido práctico, a ser utilizada para iluminar una botica de un farmacéutico en Barcelona. Desde esa fecha, se produjeron una serie de acontecimientos que conforman lo que fueron los inicios de la historia de la electricidad en el terriotorio español. En 1876 se fundó la primera empresa eléctrica de España, que se situó en Barcelona, cuyo nombre fue Sociedad Española de Electricidad. En el año 1885 se publicó un decreto cuyo acometido era regular la iluminación de los teatros. En 1901 se obtuvieron los primeros datos acerca de la potencia instalada a nivel nacional, indicando que el 39 % de la potencia instalada era de origen hidráulico, mientras que el 61 % restante de origen térmico. Debido a que toda la electricidad que se generaba en España era en corriente continua, no podía ser transportada largas distancias, y por lo tanto era necesario que el punto de generación se encontrase cerca del punto final de consumo [36].

En el siglo XX, con el desarrollo de la corriente alterna esta situación cambió, ya que se podía transportar la electricidad durante largos trayectos, lo que permitió alcanzar a finales de la década de 1920 una potencia instalada de 1.500MW, de la cual, el 81 % tenía origen hidroeléctrico. Esta potencia instalada se encontraba sobredimensionada con respecto a la demanda española, pero debido al continuo

crecimiento de la demanda eléctrica, junto con el estancamiento de la construcción de tecnologías de generación por la Guerra Civil Española, se acabó ajustando. Fue tal el aumento de la demanda que durante los años de postguerra no era posible abastecerse completamente. Con el objetivo de evitar dichos desajustes entre la potencia instalada y la energía consumida, en 1944 se fundó la empresa Unidad Eléctrica S.A. (UNESA), la cual se encargaba de las interconexiones entre los diferentes puntos de generación y consumo y que gestionaba el despacho económico de generación del país en su conjunto [7].

En enero de 1953 el gobierno estableció un sistema único de tarifas a nivel nacional, conocido como la Tarifa Tope Unificada, que se calculaba en función del coste de la electricidad. Esto permitió que se construyeran nuevas centrales que aumentasen la capacidad de producción total del país y se redujese la diferencia con la demanda [36].

En 1970, se había multiplicado por 12 la potencia instalada que existía en 1920, llegando a valores un poco inferiores a 1.800MW. Se habían desarrollado técnicas de generación que empleaban el fuel-oil como combustible, favorecidas por el bajo precio del petróleo en esos años. Además, en 1968 se incorporó la central nuclear de Zorita, convirtiéndose en la primera central nuclear del Sistema Eléctrico Español [36].

En 1979, tras la segunda crisis del petróleo, se intentó disminuir el consumo de petróleo, motivo por el que se pusieron en servicio entre 1980 y 1986 un total de cinco grupos nucleares, y se empezaron a desarrollar diferentes técnicas de cogeneración y energías renovables [36].

A finales de la década de los 80, España se encontraba sumergida en una importante deuda, por lo que se pretendió solventar dicha situación mediante un nuevo sistema de cálculo de las tarifas eléctricas, conocido como Marco Legal y Estable, el cual estuvo vigente entre 1988 y 1997. Este sistema se fundamentaba en parámetros como la retribución de amortizaciones, el pago por los costes de producción y distribución, compensaciones entre agentes y correcciones al finalizar el año [36].

En 1997, en España se aprobó la Ley del Sector Eléctrico (Ley 54/1997) que cumplía con los requisitos impuestos por la Directiva europea con normas comunes a los países de este continente para el mercado interior de la electricidad. El objetivo de dicha directiva europea era la liberalización y búsqueda de la competencia en el Sistema Eléctrico y perduró en España hasta 2013. Durante la primera mitad de

este periodo, aumentó más de un 30 % la demanda a nivel nacional y se produjo una reducción de los precios medios de la electricidad. Se encontraron dificultades para abastecer la demanda eléctrica en algunas ocasiones debido a la liberalización del sector buscada por Europa, así como por la ausencia de predicción del sistema regulatorio. En la búsqueda de dicha liberalización se llevaron a cabo una serie de acontecimientos que siguieron promoviendo el desarrollo del sector eléctrico. Entre los más importantes destaca la desintegración vertical del sector, en el que la misma empresa que se encargaba de generar la electricidad era la que la distribuía y finalmente vendía. Además, apareció la figura de las comercializadoras, intermediarios entre las distribuidoras y los clientes finales que se encargan de vender a estos la energía [2].

Sin embargo, en 2013 se reformó la Ley del Sector Eléctrico (Ley 24/2013) debido a los problemas producidos con la anterior ley como el déficit tarifario. El objetivo de esta ley era conseguir una mejora sustancial en la calidad y garantía del suministro eléctrico. Esta ley se encuentra vigente hasta la actualidad y es en la que se va a centrar el resto del documento [2].

2.2. Agentes del mercado eléctrico

Ante la búsqueda de la liberalización del mercado eléctrico con la Ley 54/1997, se llevó a cabo una desintegración vertical del sector. Antes de dicha ley, era la misma empresa la encargada de la generación, transporte y comercialización de la electricidad, es decir, eran empresas verticalmente integradas. Con la aparición de la Directiva Europea, se obligó a que fuesen diferentes empresas las encargadas de las distintas fases necesarias para obtener el suministro eléctrico [2]. De esta forma se establecieron los siguientes agentes en el mercado eléctrico, de los cuales algunos nacieron con la propia directiva y algunos con la Ley 24/2013 [29] [25]:

- **Reguladores**. En España, los reguladores del mercado eléctrico son la Administración del Estado y la Comisión Nacional de los Mercados y la Competencia (CNMC). Presentan como principal objetivo «velar por la competencia efectiva en los sistemas energéticos y por la objetividad y transparencia de su funcionamiento, en beneficio de todos los sujetos que operan en los mismos y de sus consumidores» [18].

- **Operadores**. Los operadores del sistema en España son los responsables de «la seguridad de abastecimiento de electricidad, del estado de disponibilidad adecuado de la red para la seguridad del sistema, de la correcta explotación del sistema de producción y transporte de acuerdo con los criterios de fiabilidad y seguridad, de que la red de transporte pueda satisfacer a largo

plazo la demanda de transporte de electricidad, así como de la fiabilidad de la misma» [6]. En España, los operadores del sistema son el Operador del Mercado Ibérico de Energía (OMIE) y Red Eléctrica Española (REE).

- **Generadores**. Los generadores de energía constituyen la primera fase del sistema eléctrico español. Ellos deben llevar a cabo la construcción de la estación, la generación de la electricidad y el mantenimiento de las centrales de producción de energía eléctrica. Estas se suelen clasificar en función de la fuente primaria empleada para la generación de la electricidad en: termoeléctricas, hidroeléctricas, nucleares, eólicas, solares termoeléctricas, solares fotovoltaicas y mareomotrices.

- **Transportistas**. Una vez que la electricidad ha sido generada, esta debe ser dirigida hasta grandes consumidores que se alimenten directamente en la red de transporte o hasta las subestaciones eléctricas. El transporte de electricidad comprende las líneas de transporte de alta tensión, las subestaciones eléctricas y las torres de alta tensión. De la gestión de todas las líneas y los flujos de corriente a través de ellas se encarga el operador del sistema Red Eléctrica Española.

- **Distribuidoras**: una vez que la electricidad ha sido transportada hasta las subestaciones eléctricas, las distribuidoras son las encargadas de hacer llegar la energía hasta los consumidores finales. Es decir, son las encargadas de construir, mantener y operar las instalaciones de electricidad de baja tensión en España.

- **Comercializadoras**: las empresas comercializadoras constituyen la última fase del sistema eléctrico español. Son las encargadas de vender la energía que las distribuidoras hacen llegar a los puntos de consumo final. Dentro de todas ellas, existen dos tipos [27]:

    1. Comercializadoras reguladas: son aquellas cuyo precio se encuentra regulado por el Estado. En este caso, los clientes de dichas comercializadoras pagan el precio que el Gobierno establece, el cual actualmente, se marca de forma horaria en el mercado spot de electricidad.
    2. Comercializadoras libres: son aquellas cuyo precio lo establecen las propias comercializadoras. Este tipo de comercializadoras ofrecen dos tipos de tarifas diferentes: tarifas fijas y tarifas indexadas. En las tarifas de precio fijo, este no varía a lo largo de la duración del contrato. Por otro lado, en las tarifas indexadas, el precio varía en función de cómo evoluciona el mercado spot. En la sección 2.4.2, se explica en mayor detalle las dos tipologías de tarifas.

Una vez conocidos todos los agentes que intervienen en el sistema eléctrico español, es importante conocer cómo se obtiene el precio diario de la electricidad y los diferentes mercados eléctricos que existen en España.

2.3. Mercados eléctricos en España

Se conocen como mercados eléctricos o mecanismos al conjunto de procedimientos que se deben llevar a cabo para que haya un acuerdo entre productores y consumidores, de tal manera que aquellos productores que quieran vender su energía abastezcan a aquellos consumidores que estén dispuestos a pagar el precio acordado con el productor [20]. Existe una única restricción en estos mercados, y es que la generación instantánea se iguale a la demanda instantánea, debido a que actualmente todavía no existen sistemas de almacenamiento de energía rentables que permitan almacenar esta.

Por lo tanto, son solo dos los agentes que participan en estos mercados eléctricos, los productores y los consumidores:

- **Productores**: cada uno de los productores puede ofertar todas las unidades de producción de las que disponga. Se conoce como una unidad de producción a cada una de las entidades físicas capaces de generar energía. Solo en algunos casos de plantas con una pequeña potencia instalada, una única unidad de producción puede estar formada por varias unidades físicas.

- **Consumidores**: normalmente son comercializadoras o grandes consumidores, que acuden al mercado eléctrico para comprar energía física en mercados diarios o cerrar posiciones en mercados a futuros.

Una vez conocidos los dos protagonistas que intervienen en los mercados eléctricos, se van a presentar a continuación los diferentes mecanismos en los que consumidores y productores acuerdan un precio de la energía, aunque luego serán explicados en mayor detalle en las secciones 2.3.1 y 2.3.2. Todos estos mecanismos componen lo que se conoce como el Mercado Ibérico de la Energía (MIBEL. Dentro de estos se puede realizar la siguiente diferenciación:

- **Mercados no organizados o bilaterales**: se acuerdan entre un productor y un consumidor o una comercializadora, un precio y una cantidad de energía definida para un periodo de tiempo determinado. Es un tipo de mercado a plazos. Existen dos tipos de contratos que se negocian dentro de los bilaterales [34]:

  - Contratos bilaterales físicos: son aquellos en los que los participantes acuerdan la compraventa de un total de electricidad determinado, en un periodo específico, para una localización en concreto [34].
  - Contratos bilaterales financieros: son aquellos en los cuales no implica la entrega física de la energía y se realiza a través de una liquidación financiera vinculada al precio de la electricidad [34].

- **Mercados organizados**:

  - Mercado diario: se trata de una subasta que presenta como objetivo la casación de la energía horaria entre productores y consumidores de las 24 horas del día siguiente a la subasta. La gestión de estos mercados recae sobre OMIE [33].
  - Mercado intradiario: dentro del mercado intradiario existen el mercado intradiario continuo y el mercado intradiario de subastas . Ambos posteriores al mercado diario, en los que consumidores y productores pueden comprar y vender de nuevo energía con el objetivo de ajustarse mejor a su capacidad de ofertar o demandar. Es decir, es empleado para corregir los posibles errores que existan en las previsiones. De la misma manera que con los mercados diarios, la gestión recae sobre OMIE [32].
  - Mercados de servicios de ajuste: para entender el motivo de la existencia del conjunto de mercados que componen el mercado de ajuste, se debe primero comprender que los mercados vistos hasta el momento se basan en predicciones acerca de la demanda. Con un objetivo parecido al de los intradiarios, los mercados de ajustes presentan como objetivo ajustar aún más la generación a la demanda y estudiar la viabilidad de los resultados obtenidos por los mercados diarios e intradiarios. La diferencia es que estos mercados se encuentran a una menor distancia temporal del momento real del intercambio de energía. Su responsable es REE.
  - Mercados a futuros: en estos mercados se produce una subasta de contratos futuros, que abarcan desde los días hasta los años próximos. Es el otro tipo de contrato a plazos. Estos mercados son organizados por OMIP (Operador del Mercado Ibérico – Polo Portugués).

La secuencia en la que se llevan a cabo estos mercados es el que se muestra en la Figura 2.1:

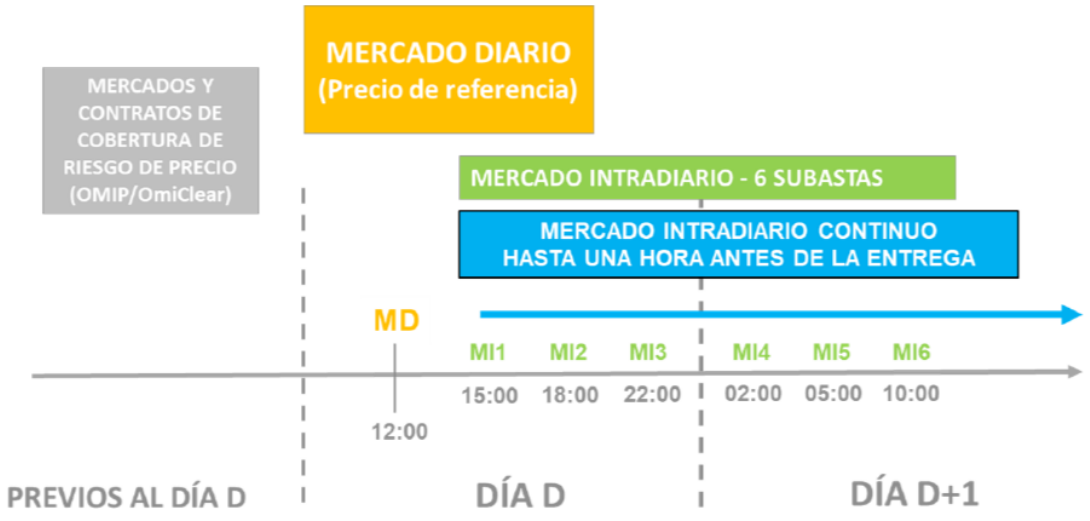

Figura 2.1: Secuencia de mercados. Fuente: OMIE

A continuación, se procede a desarrollar en mayor medida cada uno de los mercados expuestos anteriormente.

2.3.1. Mercados no organizados o bilaterales

Como se ha indicado anteriormente, los mercados bilaterales son un tipo de mercado no organizado en el que se oferta y demanda una energía que va a ser consumida en el futuro. Los resultados de los mercados bilaterales deben ser comunicados al operador del sistema una vez finaliza el mercado spot con el objetivo de que OMIE pueda contar con toda la energía que va a ser consumida en las 24 horas del día siguiente. Existen dos tipos de contratos bilaterales:

- **Contratos bilaterales físicos**: son aquellos en los que se realiza una liquidación física del contrato, es decir, los participantes acuerdan la compraventa de un total de electricidad, en un periodo específico, para una localización en concreto. Por este motivo, los resultados de los mercados bilaterales físicos deben ser comunicados al operador del sistema una vez finaliza el mercado spot con el objetivo de que OMIE pueda conocer toda la energía que va a ser consumida en las 24 horas del día siguiente. Existen dos tipos diferentes de contratos bilaterales físicos en función de la conexión física entre generador y consumidor:

  - Acuerdo bilateral físico on-site: se trata de aquellos acuerdos bilaterales en los que existe una conexión física entre generador y consumidor. Por

este motivo, suelen darse únicamente en el caso de que exista proximidad geográfica entre los dos puntos. En la situación de que el consumidor no sea una comercializadora y demande más electricidad de la que el generador es capaz de producir, debe firmar un acuerdo con una comercializadora para que pueda tener el suministro que necesita. Por otro lado, cuando el productor esté generando energía eléctrica y no toda esta esté siendo demandada por el consumidor, la planta podrá vender la energía en mercado o mediante otro acuerdo bilateral con un consumidor diferente [34].

- Acuerdo bilateral off-site o sleeved: se trata de aquellos acuerdos bilaterales en los que no existe una conexión física entre generador y consumidor. Tal y como ocurría en el caso anterior, en la situación de que en algún momento el generador no sea capaz de abastecer toda la energía demandada por el consumidor, esté la comprará directamente de la comercializadora [34].

- **Contratos bilaterales financieros**: son aquellos en los cuales el activo que se negocia es un producto financiero vinculado al precio de la electricidad, es decir, no existe una entrega de energía eléctrica. Por este motivo, no es necesaria una conexión física entre generador y consumidor. La forma de liquidación de este tipo de acuerdos se obtiene calculando la diferencia entre el precio fijo acordado en el contrato y el precio de mercado que se ha obtenido realmente. De esta forma, en el caso de que el precio de mercado fuera inferior al del acuerdo, el consumidor sería el que abonaría la diferencia al generador. En el caso contrario, se produciría el abono en sentido contrario [34].

2.3.2. Mercados organizados

Mercado diario

Tal y como se ha comentado en la introducción de la sección 2.3, el objetivo del mercado diario es llevar a cabo las transacciones de energía eléctrica para las 24 horas del día siguiente, a través de la casación de las ofertas de venta y de compra por parte de los agentes de mercado.

Para poder participar en este mercado, tanto compradores como vendedores están obligados a adherirse a las Reglas de Funcionamiento del Mercado de Producción de Energía Eléctrica [33].

Todas las unidades de producción, a excepción de las que han llegado a acuerdos bilaterales físicos, deben presentar ofertas por la parte de la energía no acordada

en el contrato bilateral físico. Por otro lado, aquellas figuras con potestad para poder acudir al mercado spot y presentar ofertas de adquisición son las comercializadoras. Los consumidores directos tienen la posibilidad de adquirir la energía eléctrica mediante un contrato bilateral [33], ya sea con una comercializadora o con un productor.

Todos los días del año a las 12:00 CET se produce la sesión del mercado diario, cuyo resultado son los precios y volúmenes de electricidad acordados para las 24 horas del día siguiente. Para casar el precio de cada hora, se agregan todas las ofertas de venta de forma creciente y las ofertas de compra de forma decreciente. El punto en el que se casan marca el precio al que se pagarán todas las ofertas de venta que hayan sido casadas (precio marginalista) [33].

Dichas ofertas de venta pueden ser de dos tipologías diferentes, simples o que incorporen condiciones complejas. Las ofertas simples incluyen únicamente el precio mínimo al que estarían dispuestos a vender un cierto volumen de energía. En cuanto a las ofertas de venta con condiciones complejas, existen las siguientes casuísticas [33]:

- **Condición de indivisibilidad**: fija en el primer intervalo temporal de cada hora un valor mínimo de funcionamiento. Es decir, se aseguran que no van a vender un volumen de energía menor a ese mínimo, a excepción de que se apliquen reglas de reparto con precios distintos de cero.

- **Gradiente de carga**: limita el volumen máximo de energía que una unidad de producción puede vender en función del volumen de energía casado en la hora anterior y la siguiente. Es decir, busca evitar variaciones elevadas de carga en las unidades de producción que, o no son capaces de efectuar físicamente, o no son rentables.

- **Condición de ingresos mínimos**: la suma de todos los ingresos de las horas que han sido casadas para esa unidad de producción debe superar un valor mínimo en euros, más una cantidad variable en euros por cada MWh casado.

- **Condición de parada programada**: en caso de que no se cumpla la condición de ingresos mínimos solicitada y por lo tanto la unidad de producción haya sido retirada de la casación, se puede realizar una parada programada de un tiempo máximo de tres horas. Es decir, dado que algunas unidades de producción no son capaces de parar su producción de forma instantánea, tienen un tiempo máximo de tres horas para ir decreciendo su producción de energía eléctrica, retribuyéndoles esa energía como una oferta simple.

Esta casación de ofertas de venta y compra de energía del mercado diario se resuelve mediante el algoritmo de casación Euphemia, el cual es capaz de considerar todas las condiciones anteriormente expuestas. El resultado de dicho mercado se conoce como Programa Diario Base de Casación (PDBC), obteniendo un desglose horario de la casación de las ofertas de producción y demanda de energía eléctrica, siendo el responsable de este el Operador del Mercado. Este programa es enviado al Operador del Sistema, que añade la información de los contratos bilaterales con entrega física realizados durante las 24 horas del día siguiente, obteniendo como resultado el Programa Diario Base de Funcionamiento (PDBF). Finalmente, tras comprobar el Operador del Sistema que sea técnicamente viable la solución propuesta, se obtiene el Programa Diario Viable (PDVD) [33].

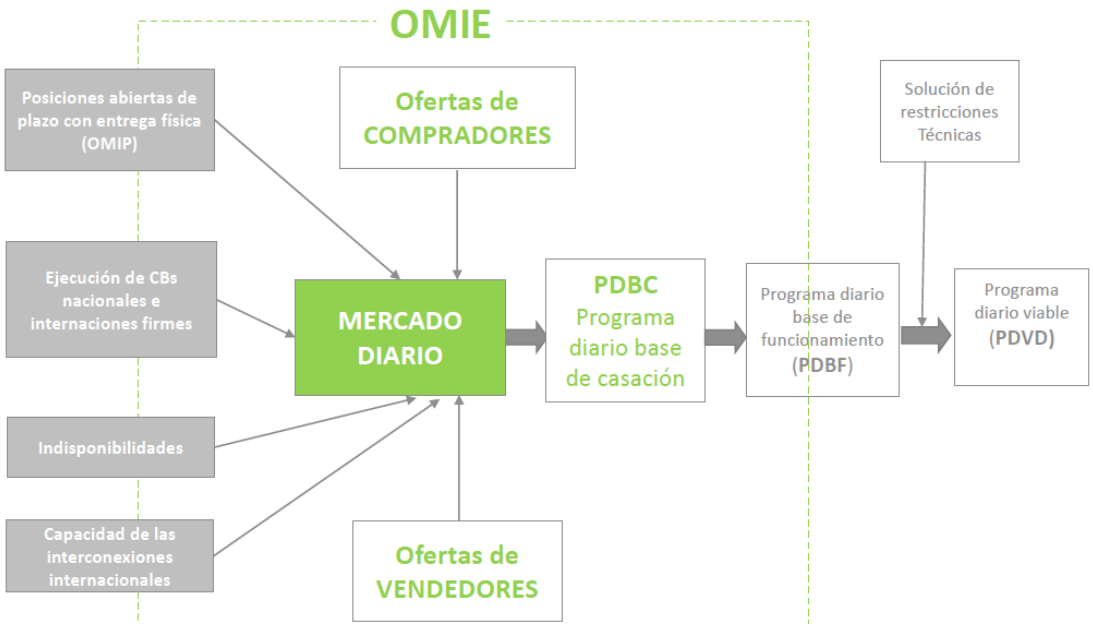

Figura 2.2: Secuencia Programas hasta el Programa Diario Viable. Fuente: OMIE

Mercado intradiario

El día 13/06/2024, y por lo tanto durante la elaboración del proyecto, el mercado intradiario cambió su funcionamiento. Por ese motivo, primero se expondrá de forma detallada cómo funcionaban los mercados con anterioridad al cambio, los motivos que impulsaron al cambio y en qué se vio afectado este.

La base del mercado intradiario es el PDVD y tiene por objetivo que tanto los compradores como los vendedores puedan ajustar sus ofertas de venta y de compra a lo que realmente son capaces de ofrecer o lo que necesitan en términos de

energía eléctrica. Para ello, antes del cambio del mercado intradiario se disponían de seis subastas en las que podían negociar energía en el ámbito MIBEL (España y Portugal) y un mercado continuo transfronterizo europeo, conocido como XBID [32].

El mercado intradiario: las sesiones de subastas

Las seis sesiones de subastas intradiarias se organizan tal y como se muestra en la Tabla 2.1, gestionando así los precios de España y Portugal y las capacidades libres de las interconexiones de España-Portugal, España-Marruecos y España-Andorra . Tras cada sesión se obtiene el Programa Intradiario Básico de Casación Incremental (PIBCI) y el Programa Intradiario Básico de Casación Acumulada (PIBCA), publicados por el Operador del Mercado. Posteriormente, con la información del mercado diario, los intradiarios de subastas anteriores y el intradiario continuo, el Operador del Sistema publica los Programas Horarios Finales (PHF).

	SESIÓN 1	**SESIÓN 2**	**SESIÓN 3**	**SESIÓN 4**	**SESIÓN 5**	**SESIÓN 6**
Apertura de Sesión	14:00	17:00	21:00	1:00	4:00	9:00
Cierre de Sesión	15:00	17:50	21:50	1:50	4:50	9:50
Casación	15:00	17:50	21:50	1:50	4:50	9:50
PIBCA	15:07	17:57	21:57	1:57	4:57	9:57
PHF de los OSs	16:20	18:20	22:20	2:20	5:20	10:20
Horizonte de programación (Periodos horarios)	24 horas (1-24 D+1)	28 horas (21-24 D y 1-24 D+1)	24 horas (1-24 D+1)	20 horas (5-24 D)	17 horas (8-24 D)	12 horas (13-24 D)

Tabla 2.1: Secuencia subastas intradiarias con anterioridad al día 13/06/2024

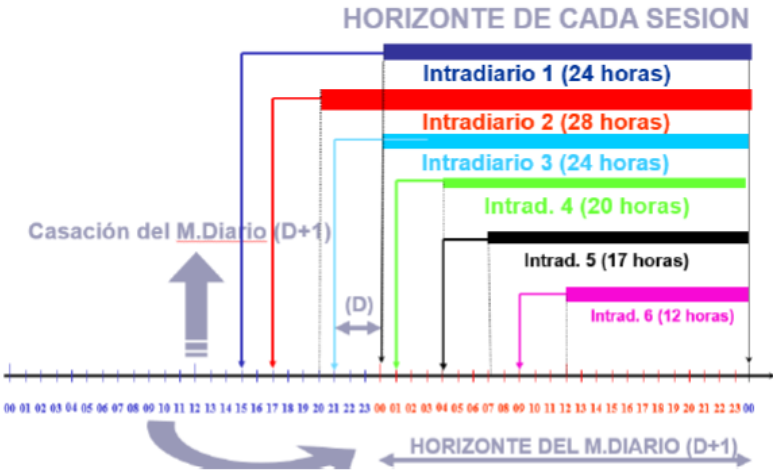

Figura 2.3: Subastas intradiarias con anterioridad al día 13/06/2024. Fuente: OMIE

Una característica peculiar de los mercados intradiarios es que solo pueden participar aquellos agentes (compradores como vendedores) que hayan participado previamente en el mercado diario, habiendo conseguido casar alguna oferta o no. De esta forma, en este mercado un agente que actuaba como comprador del mercado diario puede ser que en este actúe como comprador, porque quiera adquirir un mayor volumen de energía, o como vendedor, porque considere que va a necesitar menos energía de la comprada en el mercado diario. Lo mismo ocurre con los agentes que actuaban como vendedores en el mercado diario [32].

Al igual que ocurría en el mercado diario, el mercado intradiario de subastas es un mercado marginalista, en el que las ofertas de venta o adquisición del mercado intradiario pueden ser simples o complejas. Las ofertas simples son ofertas de adquisición de uno a cinco tramos, presentadas para cada periodo horario. En cuanto a las ofertas complejas, existen las diferentes alternativas [32]:

- **Gradiente de carga e ingresos mínimos**: se trata de las mismas condiciones que las descritas en el mercado diario.
- **Aceptación completa en la casación del tramo primero de la oferta de venta**: se trata de una condición por la cual, el agente que acude a la subasta a vender energía, solo vende dicha energía si le compran la totalidad de lo que ofrece (condición de aceptación completa), durante un perfil de tiempo específico. En caso de que estas dos condiciones no se cumplan, no se venderá la energía.
- **Aceptación completa en cada hora en la casación del tramo primero de la oferta de venta**: esta condición es similar a la anterior. La condición

de aceptación completa es exactamente igual, es decir, solo se casará la oferta si se vende la totalidad de la energía que se ofrece. Sin embargo, en este caso el perfil de la oferta se divide en las horas que dure la oferta, de tal manera que puede ser que case durante alguna de las horas y en otras no.

- **Condición de mínimo número de horas consecutivas de aceptación completa del tramo primero de la oferta de venta**: esta condición es similar a la aceptación completa en la casación del tramo primero de la oferta de venta. En este caso, tanto los agentes que acuden al mercado a vender como aquellos que acuden a comprar, imponen una condición que se centra en el número de horas consecutivas que necesitan que se acepten para que se pueda casar la oferta. Por otro lado, es imperativo que se cumpla la condición de aceptación completa.

- **Energía máxima**: condición que limita el volumen máximo de energía que puede un agente que acude al mercado comprar o vender. Esto se emplea debido a que no se sabe a qué horas será más rentable vender la energía por la volatilidad de los precios.

Una vez se han presentado todas las ofertas de compra y de venta de energía eléctrica, el operador de mercado es el encargado de realizar la casación de estas. En el caso de que exista congestión interna en la interconexión entre España y Portugal, se repetiría el proceso realizándose lo que se conoce como Market Splitting, es decir, se obtiene un precio diferente para cada zona del mercado Ibérico, con el objetivo de evitar dicha congestión [32].

El mercado intradiario continuo

Por otro lado, tal y como se comentó en la introducción de la sección 2.3, el mercado intradiario continuo permite negociar entre agentes que se encuentren regulados por el mismo mercado o entre mercados de diferentes países, siempre y cuando haya disponible capacidad interzonal. Este mercado es capaz de acercarse mucho más que las subastas al momento del consumo, es decir, se puede negociar energía eléctrica hasta una hora antes del momento de entrega de esta. Este mercado se encuentra gestionado por los diferentes Nominated Electricity Market Operators (NEMOs), o en español, los Operadores de Mercado de Europa. El valor añadido que aporta este mercado a los agentes es la capacidad de que puedan gestionar sus desbalances de energía con una mayor liquidez debido a su ámbito internacional [32].

La base que sustenta el mercado intradiario continuo es un sistema informático

que se alimenta de los resultados aportados por los Operadores de Mercado a nivel local y la información necesaria acerca de las capacidades de interconexiones transfronterizas que comunican los operadores de sistema [32].

En el caso de España y Portugal, OMIE cuenta con una plataforma conocida como Local Trading Solution (LTS) en la que los agentes tienen disponibles todas las ofertas que no han sido casadas en cualquier parte de Europa que participe en el XBID. La negociación de los contratos se realiza mediante rondas, intervalos de tiempo en los que se presentan las diferentes ofertas de compra. Las rondas se concatenan unas con otras, de manera que el inicio de una coincide con el final de la anterior. Para que una oferta pueda ser casada, deben cumplir una serie de requisitos que son validados por el sistema, como cumplir con las condiciones de aceptación o disponer de unas garantías como mínimo iguales al valor de la compra que se desea realizar. Además, también se validarán las restricciones técnicas enviadas por el Operador del Sistema [32].

Una vez el Operador del Mercado ha validado que la oferta de compra es correcta, esta se envía al Mercado Intradiario Continuo (MIC) y puede tener las siguientes condiciones de ejecución [32]:

- **None(NON)**: se trata de una condición de ejecución en la que se indica el precio y cantidad de energía que se desea comprar o vender, que puede ser de manera inmediata o parcial. La cantidad de energía que no ha sido casada se incluye en el libro de ofertas. Si no son casadas al cierre del contrato, se cancela.

- **Inmediate Or Cancel (IOC)**: se trata de una condición de ejecución en la que se indica el precio y cantidad de energía que se desea comprar o vender. En este caso a diferencia del anterior, la cantidad de energía que no ha sido casada no se incluye en el libro de ofertas, sino que se cancela.

- **Fill or Kill (FoK)**: se trata de una condición de ejecución en la que o bien se casa la totalidad de la energía ofertada o esta es eliminada.

- **Iceberg**: se trata de una condición de ejecución similar a la anterior pero con la diferencia de que la nueva oferta que se genera una vez ha sido casado el primer porcentaje de energía no se hace al mismo precio que la primera:

    - En el caso de las ofertas de compra, el precio sería el precio original más un incremento negativo, es decir, se ofertaría a un precio menor.
    - En el caso de las ofertas de venta, el precio sería el precio original más un incremento positivo, es decir, se ofertaría a un precio mayor.

Una condición diferente que se puede aplicar tanto a las ofertas de compra como de venta es la validez de las ofertas[32]:

- En el caso de las ofertas Good-For-Session (GFS), estas se encuentran vigentes hasta el final del contrato. Por defecto, las ofertas tienen esta condición.
- Por otro lado, las ofertas Good-Till-Date (GTD) son válidas hasta un momento concreto establecido por el agente que realiza la oferta pero siempre anterior al final del contrato.

Una vez finaliza cada ronda, en el caso de España y Portugal, el Operador de Mercado, OMIE, publica los resultados incrementales y acumulados y los comunica al Operador del Sistema, con el objetivo de que validen de nuevo las restricciones técnicas de las ventas y compras. Una vez han realizado dicha comprobación, envían a OMIE los Programas Horarios Finales del Continuo (PHFC) y comunican también el estado de las capacidades de las interconexiones [32].

El mercado intradiario con posteridad al 13 de junio de 2024

La información presentada en esta sección ha sido obtenida gracias al acceso de la entidad colaboradora de este proyecto a un seminario impartido por OMIE a las comercializadoras. Debido a la proximidad entre la fecha de esta resolución y el desarrollo del proyecto, no se han encontrado referencias públicas que den explicación a este cambio.

Tal y como se ha comentado anteriormente, el 13 de junio de 2024 se produjo un cambio en el mercado intradiario. La motivación de dicho cambio fue que las subastas del mercado intradiario tuviesen un alcance paneuropeo, adquiriendo el nombre cada una de las subastas de Intraday Auction (IDA). El 13 de junio a las 15:00h se esperaba que fuese el día que se produjese la primera IDA internacional que negociaba la energía a consumir durante las 24 horas del día 14 de junio de 2024, aunque finalmente fue cancelada por un mal funcionamiento de la puesta en marcha.

De esta forma, los únicos cambios que se han producido han sido en el mercado intradiario de subastas. La mayor diferencia es que se pasa de 6 subastas intradiarias a 3 subastas, en las cuales se cuenta con una hora de tiempo de recepción de ofertas en todas ellas.

	IDA 1	IDA 2	IDA 3
Apertura de sesión	14:00	21:00	09:00
Cierre de Sesión (GCT)	15:00	22:00	10:00
Horizonte de programación	24 horas (0-24 D+1)	24 horas (0-24 D+1)	12 horas (12-24 D+1)

Tabla 2.2: Secuencias IDAs mercados intradiarios

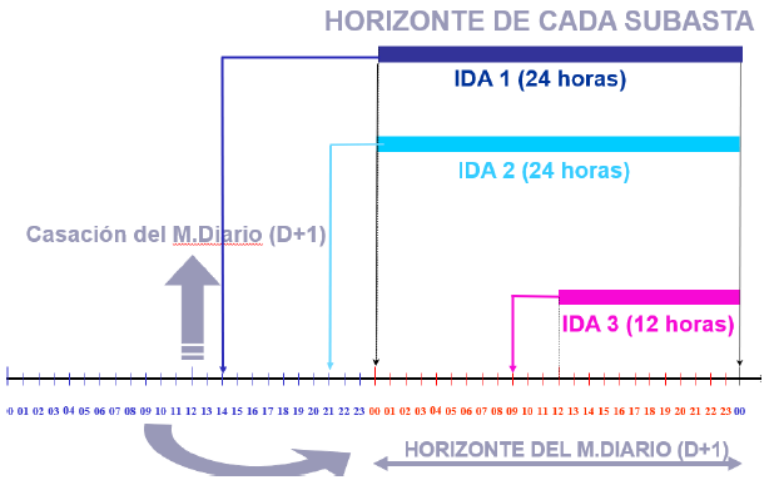

Figura 2.4: Subastas intradiarias con posterioridad al 13/06/2024. Fuente: OMIE

Al convertirse en subastas paneuropeas, la hora del cierre de cada una de las sesiones es la misma hora a nivel europeo, GCT (Gate Closure Time). Dicha hora es la referencia para determinar los tiempos de operación de las diferentes subastas IDA, que deben respetarse para un correcto funcionamiento del mercado. Estos tiempos de operación se dividen en:

- **Pre-Coupling**: el mercado intradiario continuo se suspende 20 minutos antes del GCT para aquellos contratos e interconexiones incluidos en el horizonte de la IDA. En el caso particular del MIBEL, solo se verán afectados aquellos que negocien contratos entre fronteras de España-Francia y España-Portugal, pudiendo negociar contratos dentro de España. Durante este periodo de tiempo, a nivel de la IDA, se llevan a cabo los siguientes procesos:

  1. Entre el GCT – 20' y el GCT - 15': durante estos 5 minutos se realiza el cálculo de las capacidades disponibles, enviando una primera versión en el GCT – 15' a todos los NEMOs, y ellos publicándola a los agentes una vez las tengan disponibles.
  2. Entre el GCT – 15' y el GCT – 5': este tiempo también es dedicado para calcular las capacidades disponibles. En el GCT – 15' se publica una

segunda versión, la cual puede ser igual o diferente de la primera. De la misma forma que en el periodo anterior, se le envía a todos los NEMOs dicha información, que a su vez la publican a los diferentes agentes.

3. Entre el GCT – 5' y el GCT: se trata de 5 minutos de contingencia por si hubiese algún problema en la publicación de las capacidades.

- **Coupling**: comprende los 20 minutos posteriores al GCT, con 10 minutos adicionales de contingencia en el caso de que ocurra algún problema. Durante este periodo de acoplamiento, el mercado intradiario continuo entre España-Portugal y España–Francia sigue suspendido, además de cualquier negociación a nivel nacional, pudiendo únicamente anular ofertas.

  1. Entre el GCT y el GCT + 6': durante los seis primeros minutos posteriores al GCT, los diferentes NEMOs se encargan de recopilar todas las ofertas, validarlas y anonimizarlas para enviarlas a la nube europea.
  2. Entre el GCT + 6' y el GCT + 13': la nube europea, una vez que disponga de la información de todas las capacidades y todos los ficheros de ofertas, cuenta con 7 minutos para casar todas las ofertas de venta y de compra mediante el algoritmo de Euphemia (el mismo empleado que para el mercado diario).
  3. Entre el GCT + 13' y el GCT + 17,5': una vez que se ha obtenido el resultado aportado por Euphemia con todas las ofertas de compra y venta (simples o complejas), los diferentes Operadores de Mercado cuentan con 4,5 minutos para realizar el portfolio y validar que el resultado aportado es correcto, estableciendo el nombre a todas las unidades de oferta que han sido casadas.
  4. Entre el GCT + 17,5' y el GCT + 20': se publicarán los resultados finales de la subasta, tanto para los Operadores del Mercado como para los agentes, que deben haber sido validados por todos los operadores de forma conjunta.

- **Post-Coupling**: una vez finalizada cada IDA (GCT+20), se restablece el mercado intradiario continuo con normalidad absoluta, con los resultados de la IDA ya integrados.

Se muestra en la Figura 2.5 la secuencia de los diferentes periodos que se producen para cada una de las 3 IDA del mercado intradiario de subastas europeas.

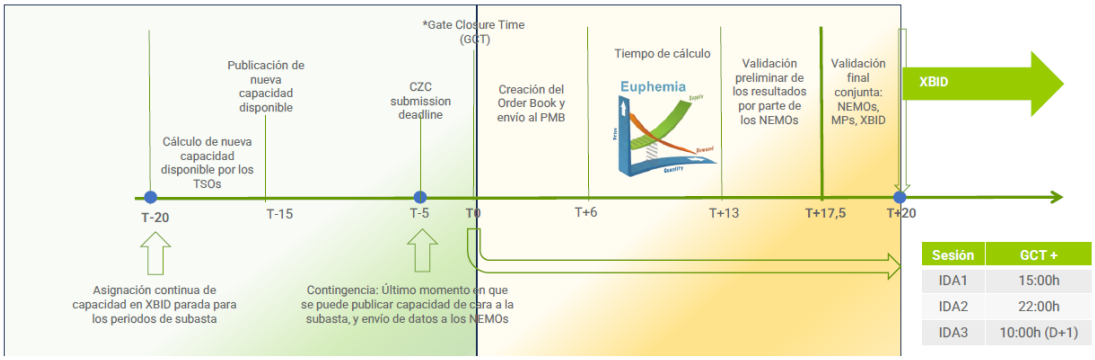

Figura 2.5: Secuencia subastas intradiarias posteriores al 13/06/2024. Fuente: OMIE

Mercado de servicios de ajuste

Tal y como se ha comentado anteriormente, los objetivos del mercado de servicios de ajuste, son fundamentalmente dos:

- En base a los resultados del mercado diario continuo, con el Programa Diario Base de Funcionamiento (PDBF), y de los mercados intradiarios, con los Programas Horarios Finales (PHF) y los Programas Horarios Finales del Continuo (PHFC), el Operador del Sistema mediante la solución de las restricciones técnicas, estudia la viabilidad técnica de la energía casada.

- Además, asegura siempre que generación y demanda sean iguales en todos los instantes de tiempo, mediante la asignación de reserva de regulación secundaria, los servicios de balance y el mercado de desvíos.

Con el objetivo de recapitular lo visto en los mercados anteriores y analizar cómo se interseca con los mercados de restricciones técnicas, se procede a destacar lo más importante para, de esta forma, poder tener una visión general de todos los mercados en su conjunto, ayudados por la Figura 2.6 aportada por Red Eléctrica.

El proceso de programación tiene su origen en el mercado diario. En base a los resultados de este, la información acerca de los mercados a plazos y los contratos bilaterales negociados para las 24 horas del día siguiente, se emite el PDBF. Este es enviado al Operador del Sistema (Red Eléctrica en el caso de España) el cual hace un análisis de red para cerciorarse de mantener los niveles de seguridad requeridos en las líneas y cuyo resultado es la solución de restricciones técnicas. El resultado económico (PDBF) junto con las soluciones técnicas tienen como resultado el PDVP. Además, en la tarde del día D, se hace la asignación de reserva de regulación secundaria para todo el horizonte del día D+1, cuyo objetivo es tener

una provisión que pueda mantener el equilibrio de generación y demanda.

El PDVP es la base del mercado intradiario, el cual, tras la casación de las ofertas de compra y venta, obtiene como resultados el PHF y el PHFC. Sobre estos dos programas se realiza un nuevo análisis técnico mediante una nueva solución de restricciones técnicas analizando la red, y de forma paralela se activan los servicios de balance para corregir posiciones de desvíos entre generación y demanda. Estos dos resultados, que se basan en los resultados del mercado intradiario, se registran en el P48, en el que se recoge el compromiso que tiene cada generador y consumidor de vender o tomar energía de la red.

Finalmente, el P48 es comparado con la medida final y se obtienen los desvíos como diferencia entre el compromiso y la realidad. Estos desvíos se liquidan al precio de desvíos.

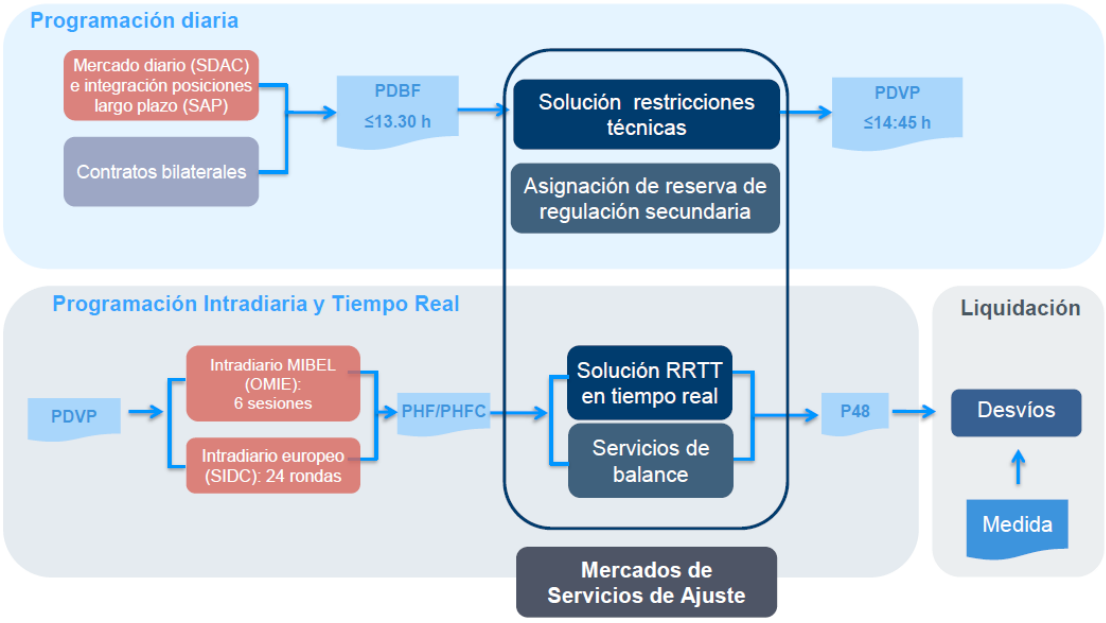

Figura 2.6: Proceso de Programación REE. Fuente: REE

A continuación, se explicarán en mayor detalle la solución de restricciones técnicas, la asignación de reserva de regulación secundaria y los servicios de balance.

Solución de restricciones técnicas

Según se indica en la disposición 15049 del BOE número 253 de 2019 P.O. 3.2., «Restricción técnica: Es cualquier circunstancia o incidencia derivada de la si-

tuación del sistema producción-transporte que, por afectar a las condiciones de seguridad, calidad y fiabilidad del suministro establecidas reglamentariamente y a través de los correspondientes procedimientos de operación, requiera, a criterio técnico del OS, la modificación de los programas de energía» [4].

En otras palabras, aquello que debe regular el Operador del Sistema es que se cumplan las condiciones de seguridad en aspectos como la frecuencia, la tensión y el nivel de carga en las líneas de la red de transporte, tanto en régimen permanente como tras contingencia. Su objetivo es corroborar que los programas establecidos sean viables desde la perspectiva técnica, y, en caso que no lo sean, buscar una solución que pueda garantizar la cobertura de la demanda prevista y el cumplimiento de los criterios de seguridad.

Uno de los principales criterios que se emplean para asegurarse de poder abastecer los consumos eléctricos en todo momento es garantizando el criterio N-1 y N-2. Estos criterios obligan a que, en el caso de que un componente de la red (criterio N-1) o dos componentes de la red (criterio N-2) fallen de forma simultánea, se pueda asegurar el suministro eléctrico a todos los puntos de la red de forma segura [26].

Como se ha destacado anteriormente, existen dos escenarios de aplicación de las restricciones técnicas [8]:

- **Horizonte diario**: se realiza un estudio de las 24 horas del día siguiente en conjunto en base al Programa Diario Base de Funcionamiento. Dentro del horizonte diario existen dos fases [8]:

    1. Fase 1: en esta primera fase, si se requiere más energía (oferta a subir), dicha cantidad adicional que vende el generador la cobra a precio de restricciones técnicas. Sin embargo, si se requiere menos energía generada (oferta a bajar), esa cantidad que se deja de generar la paga el generador al precio del mercado diario, ya que se considera una anulación.
    2. Fase 2: en esta segunda fase, el objetivo es equilibrar generación y demanda del programa PDBF, y la energía a subir o bajar se remunera al precio de restricciones técnicas.

- **En tiempo real**: utiliza como base el último programa disponible en el que se actualizan las previsiones de demanda, de generación, etc En este caso se analiza de forma continua la red para estudiar si se debe ajustar en tiempo real algún elemento. Debido a la cercanía del momento de consumo, cuenta con una única fase y el valor de la energía a producir debido a la restricción se paga al precio de la oferta [8].

Solución de servicios de balance

Los servicios de balance gestionan el equilibrio de generación y demanda mediante plataformas europeas en su mayoría (se espera que para 2025 ya se encuentre todo gestionado por plataformas europeas). Estos mercados también son marginalistas, y su objetivo es fijar un precio de desvío para aquellos productores que no producen la cantidad pactada a producir durante el resto de mercados.

Para calcular la reserva de energía necesaria, el Operador del Sistema tiene una estimación de la energía que se va a demandar. Por otro lado, también disponen de la información de la energía que se va a generar y consumir, como resultado de los programas.

En la Figura 2.7 obtenida de REE, se muestran los dos tipos de servicios de balance que existen: de activación manual y de activación automática.

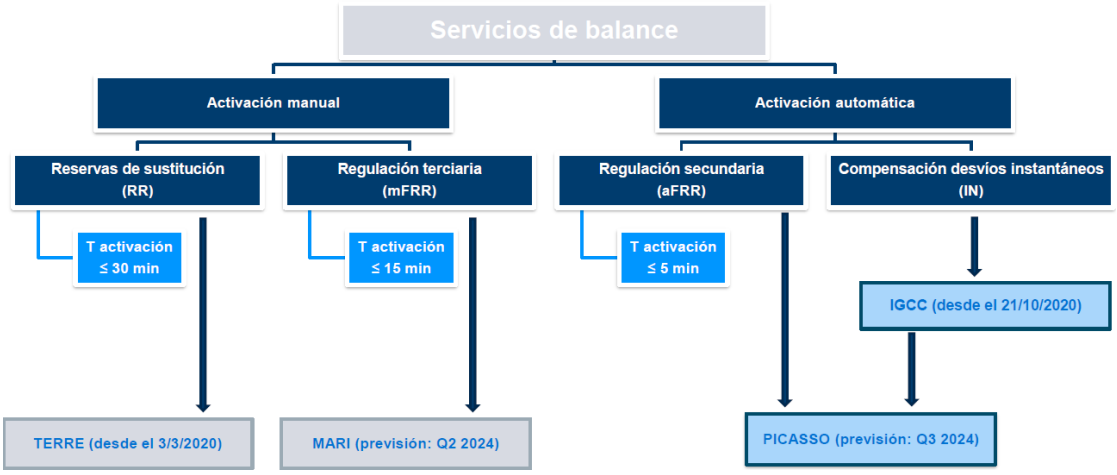

Figura 2.7: Servicios de balance. Fuente: REE

Los servicios de reserva de activación manual, a su vez se dividen en los de reservas de sustitución y de regulación terciaria [8]:

- **Reservas de sustitución (RR)**: también conocido como gestión de desvíos. Cuenta con un tiempo de activación de media hora para un tiempo de aplicación de una hora. El Operador del Sistema, al ver la diferencia entre su estimación y los resultados de los programas anteriores, se reserva esa cantidad de energía a emplear en el caso de que sea necesaria. Es el primer mecanismo que se emplea en el caso de que sea necesario.
- **Regulación terciaria (mFRR)**: cuenta con un tiempo de activación de

15 minutos. En este caso, se negocian pequeñas cantidades de energía para ajustar el desvío previamente comentado. El objetivo es restituir la reserva de la regulación secundaria que haya sido empleada.

En cuanto a la activación automática, cuenta con la compensación de desvíos instantáneos y la regulación secundaria [8]:

- **Compensación de desvíos instantáneos (IN)**: en el caso de que un país cuente con un desvío, y otro país cuente con otro desvío pero en sentido contrario, y se cumple que haya capacidad en la interconexión suficiente, estos desvíos son compensados. Esto cuenta con la ventaja de que evita activar recursos en tu propio país.

- **Regulación secundaria (aFRR)**:cuenta con un horizonte temporal con un intervalo que transcurre desde los 30 segundos hasta los 15 minutos. Se activa de forma automática y es la que permite cuadrar de forma perfecta generación y demanda.

2.4. Facturación al consumidor final. Tipologías de precios

A lo largo de la sección 2.3, se han explicado con detalle los distintos mercados que llevan a obtener el precio final de la electricidad. Sin embargo, para entender cómo estos precios se trasladan al consumidor, es esencial estudiar la estructura de una factura eléctrica. En esta sección, se analizarán los diferentes componentes que integran una factura eléctrica para comprender cómo se reflejan los precios del mercado en la factura final que llega a los consumidores de energía eléctrica.

Una factura eléctrica se encuentra compuesta por dos conceptos principales, el término de energía y de potencia. La suma de estos dos términos junto con otros conceptos como el bono social, las penalizaciones por demandar energía reactiva y los excesos de potencia, son multiplicados por el impuesto eléctrico. Adicionalmente, se suma el alquiler de equipos (cuando proceda si el contador es propiedad de la distribuidora), y otros conceptos que puede trasladar la distribuidora como derechos de acceso, conexión y/o actuación, forman en su conjunto la base imponible. Finalmente, aplicando el IVA correspondiente que se encuentre vigente en ese momento se conforma el precio a pagar por la factura eléctrica. A continuación, se procede a explicar de qué se compone el término de potencia y de energía.

2.4.1. Término de potencia

Todos los puntos de suministro tienen una potencia contratada, que es la potencia máxima que pueden demandar sin sufrir penalizaciones o que se activen las protecciones. Además, la potencia contratada junto con la tensión de los suministros son las que determinan la tarifa del suministro:

- **2.0TD**: suministros eléctricos que se encuentren en baja tensión con una potencia contratada menor o igual que 15kW[5]. Esta tarifa cuenta con dos periodos tarifarios en el término de potencia y tres en el término de energía. En función del modo que se tenga contratado, se puede tener:

  1. Modo 1: en el caso de que se supere la potencia contratada, el Interruptor de Control de Potencia (ICP) corta el suministro eléctrico.
  2. Modo 2: en el caso de que se supere la potencia contratada, se incurre en penalizaciones conocidas como excesos de potencia.

- **3.0TD**: suministros eléctricos que se encuentren en baja tensión con una potencia contratada mayor que 15kW[5]. Esta tarifa cuenta con seis periodos tarifarios en el término de potencia y seis en el término de energía. En el caso de que se supere la potencia contratada, se penaliza mediante excesos de potencia.

- **6.1TD**: suministros eléctricos que se encuentren en alta tensión, con una tensión mayor a 1kV y menor que 30kV[5]. Esta tarifa cuenta con seis periodos tarifarios en el término de potencia y seis en el término de energía. En el caso de que se supere la potencia contratada, se penaliza mediante excesos de potencia.

- **6.2TD**: suministros eléctricos que se encuentren en alta tensión, con una tensión mayor o igual a 30kV y menor que 72,5kV[5]. Esta tarifa cuenta con seis periodos tarifarios en el término de potencia y seis en el término de energía. En el caso de que se supere la potencia contratada, se penaliza mediante excesos de potencia.

- **6.3TD**: suministros eléctricos que se encuentren en alta tensión, con una tensión mayor o igual a 72,5kV y menor que 145kV[5]. Esta tarifa cuenta con seis periodos tarifarios en el término de potencia y seis en el término de energía. En el caso de que se supere la potencia contratada, se penaliza mediante excesos de potencia.

- **6.4TD**: suministros eléctricos que se encuentren en alta tensión, con una tensión mayor o igual a 145kV[5]. Esta tarifa cuenta con seis periodos tarifarios en el término de potencia y seis en el término de energía. En el caso

de que se supere la potencia contratada, se penaliza mediante excesos de potencia.

En función de la tarifa que se tenga contratada, los valores que se deben pagar por el término de potencia varían. Estos precios se calculan como el valor regulado conocido como Acceso de Terceros a la Red (ATR) más un margen, que puede ser cero, que añade la comercializadora. Los valores del ATR son, como se ha indicado anteriormente, conceptos regulados establecidos por los reguladores del mercado. En la Tabla 2.3, se muestran los valores de ATR para suministros ordinarios que se encuentran vigentes desde el 15/02/2024 para el término de potencia, compuestos por la suma de los cargos del sistema, los peajes de transporte y los peajes de distribución:

Término de potencia (€/kW año)						
Tarifa	**Periodo 1**	**Periodo 2**	**Periodo 3**	**Periodo 4**	**Periodo 5**	**Periodo 6**
2.0TD	25,391661	0,968852	-	-	-	-
3.0TD	15,713047	9,547036	4,658211	4,14256	2,285209	1,553638
6.1TD	24,414407	14,692911	11,328635	9,250764	1,727525	0,9679
6.2TD	15,403115	9,884764	6,439198	5,494646	1,062003	0,615925
6.3TD	12,287597	7,417845	5,901005	4,798116	1,000746	0,643682
6.4TD	8,197568	4,560304	3,48437	3,199933	0,517041	0,342328

Tabla 2.3: Valor ATR término de potencia actualizado el 15/02/2024 para suministros ordinarios

2.4.2. Término de energía

Una vez presentados todos los mercados, cuyo objetivo son los de casar un precio para cada hora del día, cerciorarse de la seguridad del suministro y que sea fiable y de calidad, se va a proceder a explicar los dos tipos de precios del término de energía que se pueden acordar como consumidores con las comercializadoras: precios fijos y precios indexados.

Los precios fijos son aquellos que, como su propio nombre indica, aquella parte del término de energía que no se encuentre regulado mantiene un valor constante durante toda la duración del contrato. En esta tipología de precio, las comercializadoras son capaces de asegurarse poder ofrecer dicho precio fijo porque, mediante un acuerdo bilateral con un productor o por una compra a futuros, son capaces de comprar la energía a un precio igual o menor al que la venden, o en su defecto asumiendo cierto riesgo por cómo prevén que vaya a evolucionar el mercado durante la duración del mercado comprando la energía en el mercado diario. Para

calcular el término de energía total que se factura al cliente, es necesario sumar el ATR a otros conceptos regulados, como los pagos por capacidad, la remuneración al Operador del Mercado y Operador del Sistema y las pérdidas de la red, así como el precio de compra de la energía y el margen de la comercializadora.

Por otro lado, los precios indexados son aquellos que para cada hora del día tienen un precio diferente en función del valor de casación del precio spot en OMIE. El cálculo del término de energía se determina mediante una fórmula la cual, en función de su estructura, se puede conoce como Pass Pool o Pass Through.

Los indexados Pass Through son aquellos que trasladan todos los conceptos de mercado directamente en la fórmula. Gracias a esto, la comercializadora disminuye en gran medida el riesgo al que vende la energía. Un ejemplo de fórmula indexada Pass Through es el siguiente:

$$TE = [(E_h + RRTTSSOO_h + RAD_h + PPCC_h + OM + OS + CO)(1 + P_h) + FNEE](1 + C) + ATR \tag{2.1}$$

Donde:

- TE: Término de energía
- E_h: Valor de casación de la energía horaria spot electricidad
- $RRTTSSOO_h$: Coste de cuotas y servicios de ajuste del sistema y del mercado horario.
- RAD_h: Respuesta activa de la demanda.
- $PPCC_h$: Pagos por capacidad
- OM: Retribución al operador del mercado.
- OS: Retribución al operador del sistema
- CO: Coste Operativo (margen que gana la comercializadora).
- P_h: Porcentaje de las pérdidas horarias de la red. Varía en función de la tarifa del suministro.
- $FNEE$: Fondo Nacional de Eficiencia Energética.
- C: Tasa Municipal.
- ATR: Acceso de Terceros a la Red, compuesto por peajes y cargos.

En cuanto a los indexados Pass Pool, los diferentes componentes del mercado se agrupan en coeficientes, que además contemplan un margen de riesgo. En otras palabras, una tipología de precio Pass Pool es menos transparente que una Pass Through. La fórmula de un Pass Pool es la siguiente:

$$TE = A \cdot E_h + B + ATR \tag{2.2}$$

- TE: Término de energía
- E_h: Valor de casación de la energía horaria spot electricidad
- A: Coeficiente
- B: Coeficiente
- ATR: Acceso de Terceros a la Red, compuesto por peajes y cargos.

El coeficiente A recoge los conceptos como la estimación de las pérdidas y la tasa municipal. Por otro lado, el coeficiente B se compone de todos los demás conceptos regulados de la fórmula del Pass Through (RRTTSSOO, RAD, PPCC, OM, OS, CO y FNEE) más un margen de riesgo. En el caso de que se actualicen alguno de los conceptos recogidos en el término B, las comercializadoras suelen guardarse el derecho de actualizar dicho valor.

A continuación, en la Tabla 2.4, se muestran los valores de los Accesos de Terceros a la Red para el término de energía de un suministro ordinario que, como se ha indicado anteriormente, varían en función de la tarifa. Estos se encuentran vigentes desde el 15/02/2024.

Término de energía (€/kWh)						
Tarifa	**Periodo 1**	**Periodo 2**	**Periodo 3**	**Periodo 4**	**Periodo 5**	**Periodo 6**
2.0TD	0,076974	0,027963	0,002752			
3.0TD	0,048443	0,030938	0,017361	0,010389	0,003561	0,002192
6.1TD	0,035204	0,021531	0,012716	0,008037	0,002112	0,001276
6.2TD	0,018115	0,011154	0,006183	0,004023	0,001049	0,000589
6.3TD	0,015516	0,009442	0,00565	0,003682	0,000894	0,000549
6.4TD	0,010701	0,006246	0,003845	0,002595	0,000388	0,000245

Tabla 2.4: Valor ATR término de energía actualizado el 15/02/2024 para suministros ordinarios

Capítulo 3

Predicción del precio de la electricidad

A lo largo del Capítulo 2, se ha llevado a cabo un análisis de la situación actual del sistema eléctrico español. Se ha abordado su evolución histórica, se han presentado a los principales agentes involucrados en él y se han explicado los diferentes mercados que determinan el precio final de la electricidad. Tomando este contexto como base, este segundo capítulo se centrará en la explicación de diferentes técnicas de Machine Learning (Aprendizaje Automático) que pueden ser empleadas para predecir el precio de la electricidad y, posteriormente, se presentarán los resultados obtenidos. Gracias a esta transición es posible vincular el conocimiento teórico del sistema eléctrico con herramientas de análisis de datos.

De esta forma, a lo largo del Capítulo 3 se va a proceder a exponer una breve introducción del Machine Learning, así como a desarrollar algunos de los algoritmos empleados para la predicción de precios del mercado eléctrico, explicar cuáles son las variables utilizadas para la realización de dicho modelo y las diferentes hipótesis tomadas. Por último, se van a mostrar y analizar los resultados obtenidos.

Como se ha visto a lo largo del proyecto, existen diferentes precios que se pueden tratar de predecir en el mercado eléctrico español:

- El primer precio que se puede predecir es el resultante del mercado diario, es decir, el PDBF, en el que se negocia el día D-1 a las 12:00AM el precio de las 24h del día siguiente.
- Tras la solución de restricciones técnicas emitidas por Red Eléctrica Española, se podría tratar de predecir el precio resultado del PDVP.
- Una vez finalizados los mercados intradiarios, tanto el continuo como los de

subastas, se podrían tratar de predecir los precios de los PHF y PHFC.

- Finalmente, el último precio que se podría tratar de predecir sería el correspondiente al de la solución de Restricciones Técnicas y Servicios de Balance, es decir, el P48.

Teniendo lo anteriormente expuesto en cuenta, es de gran importancia tener presente que el precio que se va a predecir en este proyecto es el precio del mercado diario publicado por OMIE, que es subastado todos los días a las 12:00 AM, en el que se obtienen los precios de las 24 horas del día siguiente y que es el resultado del Programa Diario Básico de Funcionamiento. En la Figura 3.1 se sitúa el precio a predecir dentro del esquema de REE.

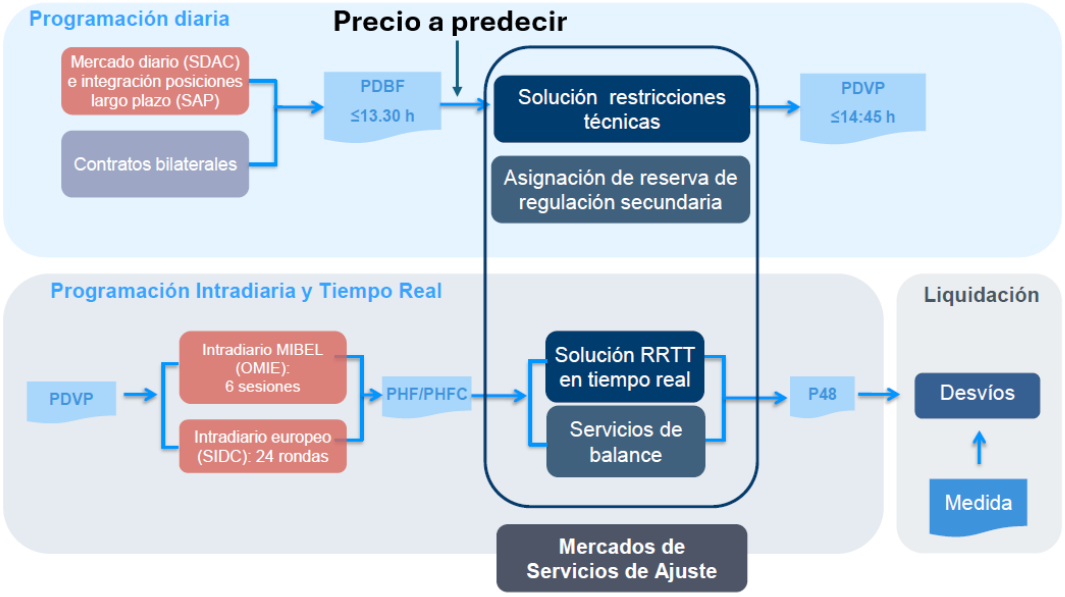

Figura 3.1: Esquema del precio a predecir. Fuente: REE

3.1. Introducción al aprendizaje automático

«El aprendizaje automático o Machine Learning es una rama de los algoritmos computacionales diseñado para imitar la inteligencia humana mediante el aprendizaje del entorno que los rodea» [12].

En otras palabras, el Machine Learning es una rama de la inteligencia artificial

cuyo objetivo es que las máquinas aprendan patrones asociados a las distribuciones de las que proviene un conjunto de datos. Se considera que una máquina es capaz de aprender cuando mejora los resultados a medida que dispone de más datos mediante un proceso conocido como entrenamiento [40].

En base a los datos de los que se disponen y los objetivos de los aprendizajes, existen las siguientes ramas dentro del Machine Learning:

- **Aprendizaje supervisado**: en los modelos de aprendizaje supervisado se relacionan una serie de variables de entrada o inputs para predecir el valor de una variable objetivo, llamada output. A los inputs se les conocen como features (características), y se agrupan en una matriz, la cual cuenta con N filas (número de observaciones) y p columnas (número de características). Por otro lado, a la variable objetivo se la conoce como atributo o variable respuesta, y se agrupa en un vector columna Y, que cuenta con n filas, que contienen los valores de la variable objetivo asociado a cada observación [40].

  En otras palabras, el aprendizaje supervisado corresponde con aquellos modelos en los que se disponen de los valores de las features y se utilizan para predecir los valores de la variable respuesta. Este tipo de aprendizajes se emplea para construir a partir de los datos de entrenamiento un modelo de predicción [40].

- **Aprendizaje no supervisado**: por otro lado, aquellos modelos en los que no se tiene ninguna variable objetivo, se denominan como modelos de aprendizaje no supervisado. Este tipo de aprendizaje automático suelen ser empleados para tareas de agrupamiento, así como para la generación de datos sintéticos [40].

Este documento se centrará en el aprendizaje automático supervisado debido a que el objetivo es la predicción del precio de la electricidad a partir de un conjunto de variables. De esta forma, una subclasificación dentro de los modelos de aprendizaje supervisado en función del tipo de variable objetivo es:

- **Regresión**: la variable respuesta es una variable cuantitativa que toma valores en un conjunto continuo.

- **Clasificación**: la variable respuesta puede pertenecer a un número finito de clases. Se divide en clasificación binaria (2 clases) y clasificación multiclase (más de dos clases).

El la Figura 3.2 se muestra un esquema de los tipos de aprendizaje dentro del Machine Learning.



Figura 3.2: Esquema modelos de aprendizaje. Fuente: elaboración propia

Dado que el precio de la electricidad se puede modelizar como una variable continua, se centrará la explicación en los modelos de aprendizaje automático supervisado por regresión, las fases que se deben llevar a cabo para obtener las predicciones del modelo son las siguientes [31]:

1. **Obtención de los datos** correspondientes a las variables del modelo, de todas las características y de la variable respuesta. Cuanto mayor cantidad de datos se disponga y mayor sea la calidad de los datos, mejores resultados en general se podrán obtener en el modelo.

2. **Tratamiento de los datos**. Convertir todos los datos al mismo formato, eliminar datos que sean incompletos o incorrectos. En el caso del modelo desarrollado, en aquellas situaciones en los que no se dispusiese de algún dato se ha rellenado con el promedio o mediana del resto de datos de esa característica.

3. **Dividir el conjunto de datos entre entrenamiento y test**. Un porcentaje de los datos sirve para entrenar el modelo, conociendo a estos datos como conjunto de entrenamiento, y un segundo porcentaje de los datos que se emplean para comprobar la eficiencia del modelo, cuyo nombre es conjunto de test. En general, para cerciorarse de que la partición de los datos es independiente del resultado del modelo, se emplea una validación cruzada o cross validation[22]. A la hora de llevar a cabo la validación cruzada, existen diferentes alternativas:

    - Holdout method: este tipo de validación cruzada divide aleatoriamente los datos en las categorías de entrenamiento y de test. El porcentaje de

partición de cada uno de los conjuntos ha de ser prefijado. Como desventaja de este método es que, al hacerlo de forma aleatoria, puede ser que no se detecten patrones escondidos en alguna de las dos categorías [31].

- K-fold cross-validation: en este tipo de validación cruzada se dividen los datos en K subconjuntos de la misma longitud también aleatorios. De todos estos subconjuntos, se emplean todos menos un subconjunto como entrenamiento y uno de ellos de validación. Se registra cuál ha sido el resultado tomando esta combinación y se hacen el resto de combinaciones restantes de tal forma que todos los subconjuntos sean una vez el de test. Una vez probadas las k combinaciones posibles, se calcula la media aritmética de las métricas de cada fold y de esta forma se obtiene un resultado más robusto [31].

	Subconj. 1	**Subconj. 2**	**Subconj. 3**	**Subconj. 4**	**Subconj. 5**	**Error**
k=1	Entren.	Entren.	Entren.	Entren.	Valid.	Error 1
k=2	Entren.	Entren.	Entren.	Valid.	Entren.	Error 2
k=3	Entren.	Entren.	Valid.	Entren.	Entren.	Error 3
k=4	Entren.	Valid.	Entren.	Entren.	Entren.	Error 4
k=5	Valid.	Entren.	Entren.	Entren.	Entren.	Error 5

Tabla 3.1: Ejemplo K-fold cross-validation con k=5

$$Error\ medio = \frac{\sum_{k=1}^{5} Error_k}{5} \tag{3.1}$$

En éste proyecto, para determinar qué observaciones se empleaban para el entrenamiento y cuáles para el test, se empleó inicialmente la alternativa K-fold cross-validation. Sin embargo, los resultados obtenidos no fueron los esperados obteniendo errores elevados. Por ese motivo, se decidió llevar a cabo el Holdout method. En este se decidió que el 80 % de las observaciones se empleasen para el conjunto de entrenamiento y el 20 % restante para el test. De esta forma se mejoraron considerablemente los resultados y son estos los que se han presentado.

4. **Entrenamiento del modelo**. Para ello se debe seleccionar un algoritmo que sea capaz de predecir los valores de la variable respuesta en el modelo. En el caso de este proyecto, se empleará el lenguaje de programación Python.

En él se desarrollarán varios algoritmos y se compararán los errores obtenidos entre ellos. Para la realización del modelo que predice el precio del mercado spot de electricidad se han empleado los siguientes cinco modelos:

- Baseline
- Linear
- Linear Ridge
- XGBoost
- Random Forest

A lo largo de la Sección 3.2 se procederá a explicar cuál es el funcionamiento y la lógica que existe detrás de cada uno de estos algoritmos de predicción.

5. **Análisis de los resultados obtenidos** que permita estudiar la eficacia del modelo utilizando métricas estandarizadas como el Mean Absolute Error (MAE) o Root Mean Square Error (RMSE).

3.2. Algoritmos de aprendizaje automático

Tal y como se ha comentado anteriormente, en todo modelo de aprendizaje automático supervisado por regresión, se determina un porcentaje de los datos que se emplean para el entrenamiento y otro porcentaje de los datos que se emplean para el test. En el caso del modelo llevado a cabo en este proyecto, independientemente de la serie temporal seleccionada (se verá más adelante que se han seleccionado diferentes series temporales), es el propio algoritmo el que selecciona el 80 % empleado para el entrenamiento y el 20 % restante para el test.

A continuación, se van a explicar los modelos empleados en este proyecto, así como los algoritmos de entrenamiento que tienen asociados.

3.2.1. Modelo Baseline

El modelo baseline es el más básico de todos los que se van emplear a lo largo del proyecto. Este algoritmo se empleará como cota superior del error máximo que se puede admitir en comparación con el resto de algoritmos.

A pesar de que existen diferentes variantes de este algoritmo, la que se ha empleado en este proyecto es la siguiente: durante el periodo de entrenamiento se ha calculado la media de los atributos, siendo ese el valor que se desea predecir. Posteriormente, se compara cada valor frente a dicha media para obtener todos los

errores y finalmente se calcula el MAE para el conjunto de entrenamiento y para el conjunto de test por separado.

$$MAE_e = \frac{1}{N_e}\sum_{i=1}^{N_e}|y_i - t_i| \tag{3.2}$$

$$MAE_t = \frac{1}{N_t}\sum_{i=1}^{N_t}|y_i - t_i| \tag{3.3}$$

Donde:

- N: número de observaciones. En el caso del MAE del entrenamiento, N_e hará referencia al número de observaciones del entrenamiento. En cuanto al MAE del test, N_t hará referencia al número de observaciones del test.
- y_i: valor de la predicción. En el caso de este modelo es el promedio del periodo de entrenamiento.
- t_i: valor variable respuesta.

A continuación, se va a mostrar un ejemplo para una mejor comprensión de dicho algoritmo:

	Feature 1	**Atributo**
Observación 1	1	15
Observación 2	2	10
Observación 3	3	20
Observación 4	4	12
Observación 5	5	18
Observación 6	6	20
Observación 7	7	16
Observación 8	8	14
Observación 9	9	13
Observación 10	10	17

Tabla 3.2: Ejemplo modelo Baseline

$$\overline{y_{entrenamiento}} = 15,62$$

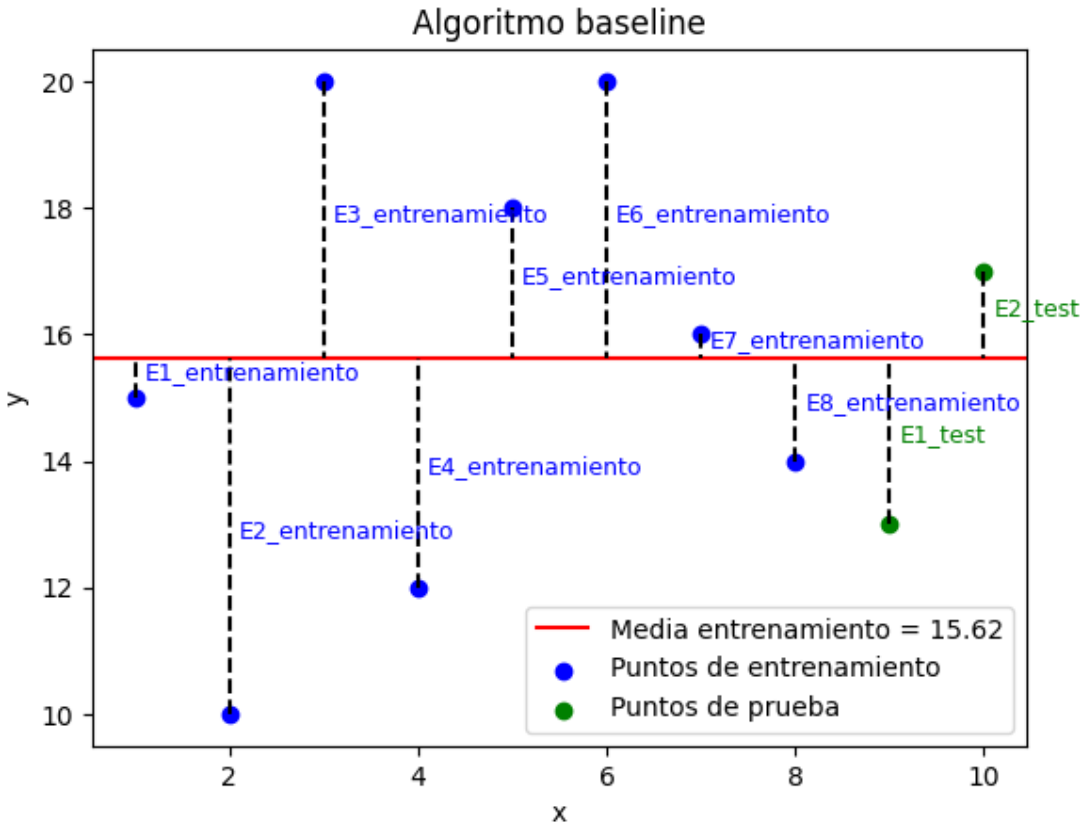

Figura 3.3: Ejemplo Baseline. Fuente: elaboración propia

$$MAE_{ent} = \frac{1}{8}\sum_{i=1}^{8}|15,62 - t_i| = 2,8$$

$$MAE_{test} = \frac{1}{2}\sum_{i=1}^{2}|15,62 - t_i| = \frac{1}{2}(|15,62 - 13| + |15,62 - 17| = 2$$

De esta forma, se obtendría que el MAE del modelo sería de 2,8 unidades para el entrenamiento y de 2 unidades para el test. Esto significa que de media se ha desviado 2,8 unidades en la predicción respecto a la realidad en el entrenamiento y 2 unidades en la predicción respecto a la realidad en el test.

3.2.2. Modelo Linear

El algoritmo linear es un algoritmo de regresión lineal que se ajusta por mínimos cuadrados. El modelo consiste en aproximar la variable objetivo como una recta en el espacio de las features que siga la siguiente ecuación [28]:

$$y = b + \mathbf{m} \cdot \mathbf{x} \tag{3.4}$$

Donde:

- y: es una variable escalar que contiene la variable objetivo para cada iteración.
- b: es la ordenada en el origen de la recta, un escalar.
- $\boldsymbol{m}$: es un vector que tiene registrada la pendiente de la recta, que tiene tantos elementos como features tiene el modelo.
- $\boldsymbol{x}$: es el vector de input para cada iteración.

De esta forma, en el entrenamiento se ajusta por mínimos cuadrados el valor de la pendiente y de la ordenada en el origen que se emplearán para la parte de test. Para ello se debe minimizar la expresión 3.5, correspondiente a la suma de errores cuadráticos medios sobre el conjunto de entrenamiento [37]:

$$f(m, b) = \sum_{i=1}^{N_e} [t_i - b - \sum_{j=1}^{p} (m_j x_{ij})]^2 \tag{3.5}$$

Para calcular dicho mínimo se debería hacer el gradiente de la función $f(m, b)$ respecto de cada una de los parámetros b, m_1, ..., m_p para tener $p + 1$ ecuaciones y poder calcular los $p + 1$ parámetros:

$$\begin{aligned} \frac{\partial f(m,b)}{\partial b} &= 0 \\ \frac{\partial f(m,b)}{\partial m_1} &= 0 \\ &\vdots \\ \frac{\partial f(m,b)}{\partial m_j} &= 0 \end{aligned}$$

Una vez se dispone de la recta de regresión, se pueden calcular los errores de cada uno de los puntos como el valor absoluto de la diferencia entre el valor real y el valor de la recta. Finalmente se calculará el promedio de dichos errores de forma separada para el entrenamiento y para el test para poder comparar con otros modelos (MAE).

A continuación, se va a mostrar un ejemplo para una mejor comprensión de dicho

algoritmo. Cabe destacar que en este ejemplo se va a considerar únicamente una feature para que pueda ser representado en $\mathbb{R}^2$, ya que al considerar 2 características se tendría que representar en $\mathbb{R}^3$ y a partir de tres features no sería posible representarlo. En este caso, m y x no serán vectores, sino escalares:

	Feature 1	**Atributo**
Observación 1	1	2,2
Observación 2	2	2,8
Observación 3	3	3,6
Observación 4	4	4,5
Observación 5	5	5,1
Observación 6	6	5,9
Observación 7	7	6,8
Observación 8	8	7,3
Observación 9	9	8,2
Observación 10	10	9

Tabla 3.3: Ejemplo modelo Linear

Calculando la ecuación de la recta de la regresión lineal de las 8 primeras observaciones que se correspondería al entrenamiento, se obtiene:

$$\frac{\partial f(m,b)}{\partial b} = 0; \quad -2 \cdot \sum_{i=1}^{N_e}[y_i - (mx_i + b)] = 0; \quad b = \frac{1}{N_e}\sum_{i=1}^{N_e} y_i - \frac{1}{N_e} m \sum_{i=1}^{N_e} x_i$$

$$b = \overline{y} - m\overline{x}$$

$$\frac{\partial f(m,b)}{\partial b} = 0; \quad 2 \cdot \sum_{i=1}^{N_e} x_i[y_i - (mx_i + b)] = 0; \quad \sum_{i=1}^{N_e} x_i[y_i - (mx_i + \overline{y} - m\overline{x})] = 0$$

$$m = \frac{\sum_{i=1}^{N_e} x_i(y_i - \overline{y})}{\sum_{i=1}^{N_e} x_i(x_i - \overline{x})}$$

De esta forma, sustituyendo los datos, y teniendo en cuenta que $\overline{x} = 4,5; \overline{y} = 4,775$

$$b = 2,19$$
$$m = 0,55$$

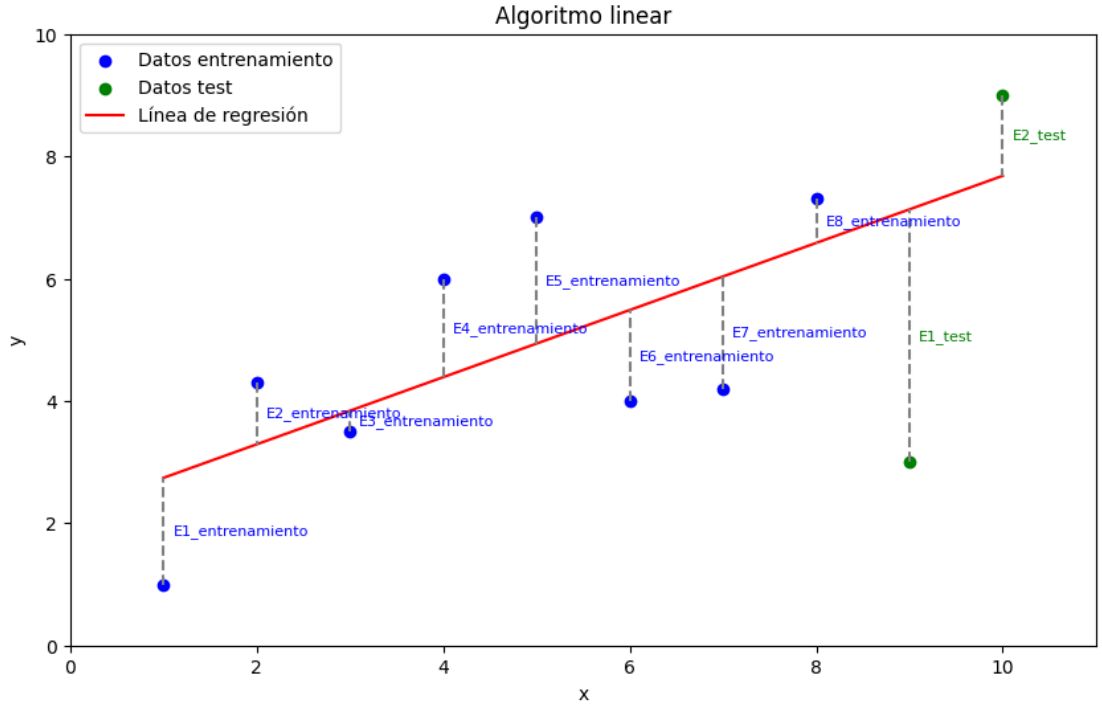

Figura 3.4: Ejemplo Regresión Lineal. Fuente: elaboración propia

Finalmente, se calcularía por separado la media de los errores del entrenamiento y del test utilizando la ecuación 3.3 explicada anteriormente.

$$MAE_{ent} = \frac{1}{8}\sum_{i=1}^{8}|(m \cdot x_i + b) - t_i| = 2,475$$

$$MAE_{test} = \frac{1}{2}\sum_{i=1}^{2}|(m \cdot x_i + b) - t_i| = 5,225$$

De esta forma, se obtendría que el MAE del modelo sería de 2,475 unidades para el entrenamiento y de 5,225 unidades para el test. Esto significa que de media se ha desviado 2,475 unidades en la predicción respecto a la realidad en el entrenamiento y 5,225 unidades en la predicción respecto a la realidad en el test.

3.2.3. Modelo Linear Ridge

Uno de los grandes problemas que se dan en los modelos de Machine Learning es que, al dividir el modelo en dos conjuntos de datos, uno para el entrenamiento y otro para el test, puede darse el caso de que se "sobreentrene" el modelo y que posteriormente no sea capaz de predecir con un error bajo la variable respuesta en el conjunto de datos del test. En otras palabras, el modelo se ajusta en exceso a las features de los datos de entrenamiento y no es capaz de predecir de forma correcta en los datos de test.

Este problema se conoce como sobreajuste u overfitting y puede ser afrontado mediante métodos de regularización [22]. La regularización consiste en añadir un término a la variable a minimizar que penalice a dicha variable y es de gran aplicación en modelos de aprendizaje continuo supervisado por regresión con el objetivo de minimizar el impacto que pueden tener algunas features sobre la variable respuesta cuando no deberían de ser influyentes y sin la regularización lo están siendo.

El algoritmo Linear Ridge es un algoritmo muy similar al algoritmo Linear previamente estudiado. La única diferencia es que el algoritmo Ridge añade un término a la función a minimizar que supone una regularización de este, conocido como regularización L2. Este término que añade, se calcula como la suma de los cuadrados de los coeficientes. El objetivo de esta regularización es que se penalicen los vectores de coeficientes con normas elevadas. Dichos coeficientes son la suma de las pendientes m_j elevadas al cuadrado y multiplicadas por un hiperparámetro (concepto explicado en la sección 3.2.4) que debe ser prefijado.

De esta forma, para el algoritmo Linear Ridge, el valor a minimizar es [22]:

$$f(m,b) = \sum_{i=1}^{N_e}[y_i - b - \sum_{j=1}^{p}(m_j x_{ij})]^2 + \lambda \cdot \sum_{j=1}^{p} m_j^2 \tag{3.6}$$

Debido a que una vez calculados los valores de los parámetros b, m_1, ..., m_j, el procedimiento es igual que en el caso del algoritmo linear, no se va a mostrar ningún ejemplo.

3.2.4. Modelo Random Forest

El algoritmo Random Forest es un algoritmo que se basa en un conjunto de árboles de decisión, por lo que, para entender de forma correcta cómo funcionan los modelos Random Forest se deben explicar primero conceptualmente primero los árboles de decisión.

Los árboles de decisión empleados para el problema de la regresión, también conocidos como árboles de regresión, están formados por los siguientes componentes [21]:

- **Nodos**: son aquellos puntos sobre los que se dividen los árboles. Cada nodo representa una característica.
- **Rama**: es la conexión entre nodos. De cada nodo parten o llegan las ramas. Cada rama representa una decisión que depende de los valores de esa característica.
- **Hoja**: se trata del final de una rama que no tiene un nodo posteriormente, es decir, es el resultado final de una serie de decisiones. Cada hoja representa el valor de cada predicción.

Para comprenderlo de una manera más gráfica se va a mostrar un ejemplo con tres características. Las características que se van a emplear son generación con tecnología A, generación con tecnología B y generación con tecnología C. El objetivo será predecir el precio de la electricidad para esa hora.

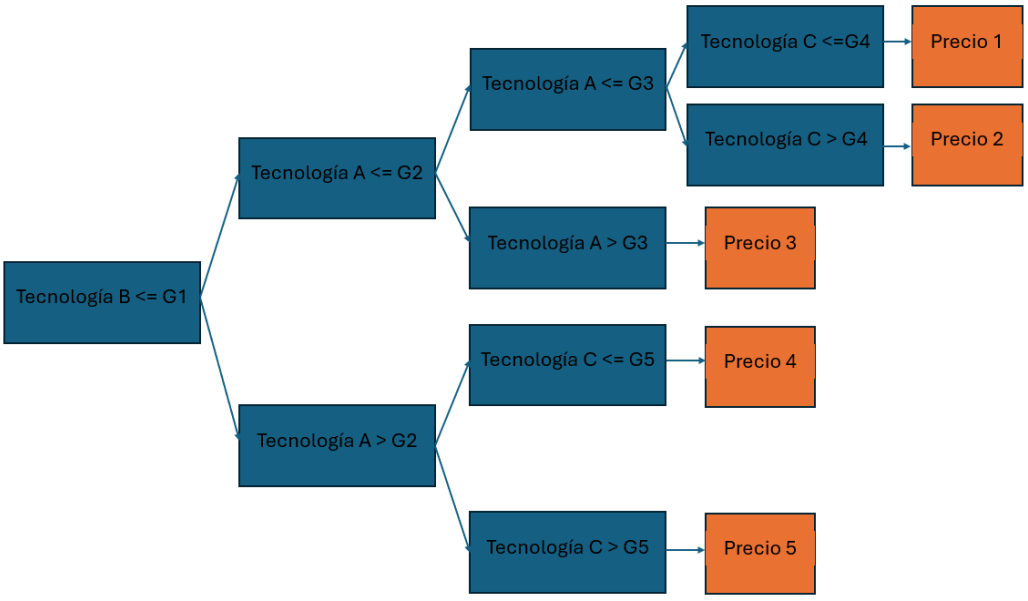

Figura 3.5: Ejemplo árbol de decisión. Fuente: elaboración propia

Para realizar la construcción y entrenamiento de un árbol de decisión asociado al problema de regresión se deben seguir los siguientes pasos [3]:

1. **Inicialización del nodo raíz**, al que llegan todos los elementos del conjunto de datos.

2. **Selección de características**: para cada uno de los nodos se contemplan las características disponibles y sus posibles divisiones en función de los valores que pueden tomar.

3. **Evaluación de las divisiones**: para cada uno de los nodos disponibles se selecciona aquella característica y división que minimice la varianza sobre los datos que llegan a los dos nuevos nodos generados.

4. **Recursividad**: para cada uno de los nodos generados volver al paso 2, salvo que se haya llegado a la condición de parada, las más comunes son:

    - Profundidad máxima: cuando el árbol alcanza la profundidad máxima permitida.
    - Número mínimo de ejemplos: el nodo contiene menos ejemplos que el umbral mínimo.

El principal problema que existe con los árboles de decisión es que, cuando se desea trabajar con una gran cantidad de datos y de características, no funcionan de forma tan precisa. Por ese motivo se emplea el modelo Random Forest o Bosque Aleatorio, el cual es una evolución de los árboles de decisión y obtiene resultados más robustos [17]. Este tipo de modelo está formado por una gran cantidad de árboles de decisión, los cuales se crean a partir de los siguientes pasos:

- Se crean un elevado número de árboles de decisión, donde cada árbol de decisión emplea un número de características m, menor que el número de características totales que se dispongan.

- Los árboles en este modelo crecen lo máximo posible hasta que se alcance la condición de parada.

Las observaciones de cada árbol que se emplean en el modelo Random Forest se elige mediante la técnica estadística Bootstrapping, existiendo la posibilidad que en diferentes árboles tengan la misma observación [17]. Además, hay una serie de observaciones que no se emplean en ningún árbol y se conocen como "out of the bag", las cuales son empleadas para validar el modelo. Cada árbol de regresión arrojará un valor diferente para la variable respuesta, siendo el valor final de esta normalmente la media de la variable respuesta de cada árbol, conocida como ensamblado.

En la siguiente Figura 3.6 se muestra un ejemplo con 10 observaciones:

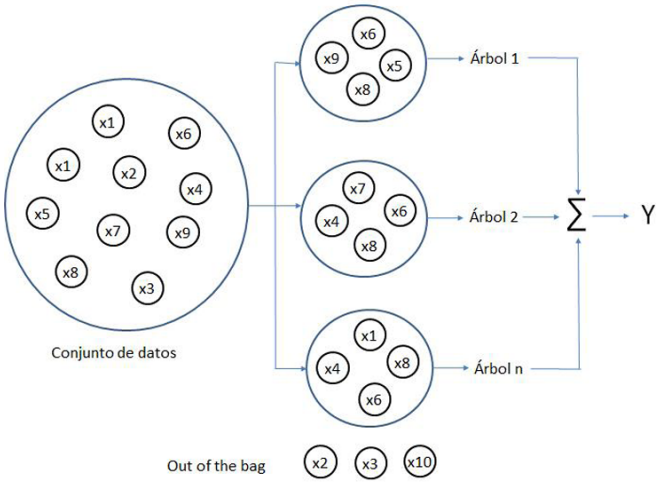

Figura 3.6: Algoritmo Random Forest [17]

Para poder llevar a cabo este modelo, al igual que ocurría con los tres vistos anteriormente, se deben ajustar los parámetros durante la fase del entrenamiento y probar posteriormente durante la fase del test. Sin embargo, a diferencia de los tres modelos anteriores hay una serie de parámetros que no pueden ser entrenados y que deben ser prefijados, los cuales se conocen como hiperparámetros [1]. Estos hiperparámetros deben ser elegidos antes de entrenar el modelo. Para obtener los valores de los hiperparámetros que mejor se ajustan al problema de estudio se lleva a cabo un proceso conocido como búsqueda de hiperparámetros. Este proceso consta de 4 fases [23]:

1. Definición de la malla: para cada uno de los hiperparámetros se definen todos los valores posibles que va a poder tomar el modelo.

2. Decidir si se utiliza Grid Search o Random Search, es decir, si se prueban todas las combinaciones posibles de hiperparámetros o si se fija un número máximo de combinaciones de parámetros a probar, elegidos de manera aleatoria. Este último acercamiento se suele utilizar cuando el número de posibles combinaciones de parámetros es muy alta o cuando se tiene una potencia computacional limitada.

3. Utilizar exclusivamente el conjunto de entrenamiento para comparar las combinaciones de hiperparámetros elegidas. La metodología que se suele utilizar es un modelo de Cross Validation, explicado anteriormente en la sección 3.1,

[1] Salvo en la regresión de Linear Ridge, que consta del hiperparámetro λ, para el cual se ha decidido no hacer una búsqueda de hiperparámetros, puesto que no es un modelo de interés para el trabajo

con cinco folds y una métrica de error estandarizada, en el caso de estudio, el MAE.

4. Los hiperparámetros elegidos, conocidos como mejores hiperparámetros, corresponden con los de la combinación que minimizan la métrica de error.

En el caso de este modelo los hiperparámetros elegidos son:

- **Simple imputer Strategy**: este hiperparámetro sirve para sustituir los valores nulos o que faltan por otros definidos que pueden ser: media, mediana, moda o un valor constante.
- **Random Forest regressor n stimators**: este hiperparámetro sirve para definir el número de árboles de regresión empleados en el modelo.
- **Random Forest regressor min samples split**: este hiperparámetro sirve para definir el mínimo número de observaciones que se necesitan para que un nodo se divida en ramas y no se convierta en una hoja.
- **Random Forest regressor max samples**: este hiperparámetro sirve para indicar el número de observaciones que se seleccionan por árbol, en porcentaje del total de observaciones del modelo.
- **Random Forest regressor max features**: este hiperparámetro sirve para indicar el número máximo de features que se emplean para buscar la mejor división de cada nodo. Es muy habitual emplear como alternativas la raíz cuadrada o el logaritmo en base dos del número de features.
- **Random Forest regressor max depth**: sirve para conocer cuál es la máxima profundidad de cada árbol.

3.2.5. Modelo XGBoost

El modelo XGBoost, abreviatura de Extreme Gradient Boosting, al igual que ocurría con el Random Forest, se basa en árboles de regresión. Estos árboles se disponen de forma secuencial de tal forma que la salida de cada uno de los árboles es utilizada para computar la salida total del modelo.

El procedimiento empleado para el entrenamiento del modelo es el siguiente [11]:

- Inicializar la predicción $y_i^{(0)} = \frac{1}{N_e}\sum_{i=1}^{N_e} t_i$, es decir, en la media del conjunto de entrenamiento. Donde el superíndice indica la predicción de la iteración i-ésima, en este caso la iteración 0.

- Para cada tiempo t, se debe encontrar la estructura del árbol de decisión que genera la salida $f_t(x_i)$, para cada x_i en el conjunto de entrenamiento que minimice la función de pérdida siguiente.

$$L^{(t)} = \sum_{i=1}^{N_e} |t_i - y_i^{(t-1)} - f_t(x_i)| + \Omega(f_t) \tag{3.7}$$

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \|\omega\|^2 \tag{3.8}$$

  El primer término de la ecuación 3.7 corresponde con la función de pérdida (MAE) entre la variable objetivo t_i y la predicción agregada al tiempo anterior $y_i^{(t-1)}$ más la salida del árbol $f_t(x_i)$, para cada elemento del conjunto de observaciones. Por otro lado, el segundo término de la ecuación 3.7 que se desarrolla en la ecuación 3.8 tiene un carácter regularizador: con el primer sumando se penaliza el número de hojas del árbol (T) y con el segundo se establece una regularización L2 sobre los pesos de las hojas del árbol.

  Encontrar la estructura óptima que induce a la salida que minimiza la Ecuación 3.7, es un problema no resoluble analíticamente en general, por lo que se acude a un algoritmo iterativo que asocia una puntuación a cada una de las configuraciones posibles del árbol para cada nodo, y se toma la configuración que mejor resultado obtenga. Al ser un problema de regresión, existen infinitas configuraciones para cada uno de los nodos, por lo que la estrategia utilizada consiste en tomar como valores posibles la media aritmética entre cada par de valores consecutivos de las features ordenadas.

- El algoritmo llega a su fin cuando se alcanza la tolerancia máxima permitida.

Se garantiza que tras cada iteración el algoritmo no empeora con respecto al anterior.

Tal y como ocurre con el modelo Random Forest, se debe realizar una búsqueda de hiperparámetros. En este caso, los hiperparámetros elegidos son los siguientes:

- **Simple imputer strategy**: tal y como se explicó en el capítulo anterior, este hiperparámetro sirve para sustituir los valores nulos o que faltan por otros definidos entre la media, mediana, moda o un valor constante.

- **Xgbregressor subsample**: es un hiperparámetro equivalente al Random Forest regressor max simples y se emplea para determinar el porcentaje de observaciones que se considerarán en cada árbol.

- **Xgbregressor n estimators**: es un hiperparámetro equivalente al Random Forest regressor n stimators y se emplea para determinar el número de árboles a construir en el modelo.
- **Xgbregressor min child weight**: este hiperparámetro sirve para controlar cuál es la suma mínima de ponderaciones de las observaciones que se requieren en un nodo.
- **Xgbregressor max depth**: es un hiperparámetro equivalente al Random Forest regressor max Depth y se emplea para controlar la profundidad del árbol.
- **Xgbregressor learning rate**: es un hiperparámetro que define la contribución de cada árbol al modelo. Aquellos que presenten valores más bajos requerirán de un mayor número de árboles.
- **Xgbregressor lambda**: es un hiperparámetro que se emplea para regularizar y evitar que se produzca sobreajuste. Es el mismo tipo de regularización que se empleaba en la Regresión Ridge (regularización L2) y que penaliza coeficientes grandes.
- **Xgbregressor gamma**: es un hiperparámetro que sirve para determinar si un nodo debe de seguir dividiéndose en más o si se convierte en una hoja. Para ello determina cuál es el valor mínimo del incremento de la función objetivo para que el nodo se particione o no.
- **Xgbregressor colsample bytree**: es un hiperparámetro equivalente a Random Forest regressor max features, y sirve para determinar el porcentaje de features a emplear en cada árbol.
- **Xgbregressor alpha**: es un hiperparámetro que se emplea para regularizar y evitar overfitting. Se trata de un hiperparámetro de regularización L1 y que además se emplea para eliminar algunas de las features. Esto es útil cuando se desea identificar cuáles son aquellas features más relevantes para el modelo.

3.3. Features consideradas en el modelo

Con el objetivo de asegurar precisión y fiabilidad de las predicciones, uno de los aspectos más importantes recae en la selección y el análisis de las features que se van a emplear. En el caso de este modelo, se han identificado dos conjuntos de variables clave: el mix energético y la demanda de electricidad. Los datos de

las variables que forman el mix energético de generación española y la demanda eléctrica han sido obtenidas del ENTSO-e, cuyas siglas provienen de European Network of Transmission System Operators for Electricity o Red Europea de Gestores de Redes de Transporte de Electricidad en español. Por otro lado, el precio de electricidad horario se ha obtenido de OMIE.

De esta forma, las features consideradas han sido las siguientes:

- **Tiempo**: la primera característica a considerar es el tiempo, compuesto por hora, día, año y hora coincidente con la franja horaria española. Esta sirve como referencia para el resto de características y nos permite colocar los datos de forma secuencial.
- **Mix energético**: se expone en MW, pero al ser una medida horaria coincide con la energía en MWh.
    - Generación de biomasa
    - Generación lignito/lignito fósil.
    - Generación fósil-carbón gas.
    - Generación gas fósil.
    - Generación hulla fósil.
    - Generación petróleo fósil.
    - Generación fósil pizarra.
    - Generación turba fósil.
    - Generación geotérmica.
    - Generación hidroeléctrica bombeada agregada.
    - Generación hidroeléctrica con acumulación por bombeo.
    - Generación hidroeléctrica de escorrentía y bombeo.
    - Generación hidráulica.
    - Generación marina.
    - Generación nuclear.
    - Generación solar.
    - Generación de residuos.
    - Generación eólica offshore.
    - Generación eólica onshore.
    - Previsión de generación solar.

  - Previsión de generación eólica offshore.
  - Previsión generación eólica onshore.
  - Generación otras tecnologías.
  - Generación otras tecnologías renovables.

- **Demanda**:se expone en MW, pero al ser una medida horaria coincide con la energía en MWh:
  - Previsión de demanda eléctrica.
  - Demanda eléctrica.

Por otro lado, tal y como se comentó en la introducción del Capítulo 3, la variable objetivo o variable a predecir es el precio resultante del Programa Diario Básico de Funcionamiento.

Teniendo en cuenta lo anteriormente expuesto, y antes de llevar a cabo los modelos de predicción de los explicados en la Sección 3.2, se llevó a cabo un estudio de la relación que tienen todas las variables del modelo, tanto las features como la variable objetivo. La métrica utilizada para medir la correlación entre variables es la matriz de correlación de Pearson. En ella se presentan todas las variables enfrentadas en las que las relaciones entre dos variables pueden tomar números entre el -1 y el 1, donde un valor de 1 indica que existe una correlación lineal positiva perfecta, un valor de 0 indica que no hay correlación lineal y un valor de -1 indica que hay correlación lineal negativa perfecta [38]. El resultado de la matriz de correlación de Pearson se puede apreciar en la Figura 3.7.

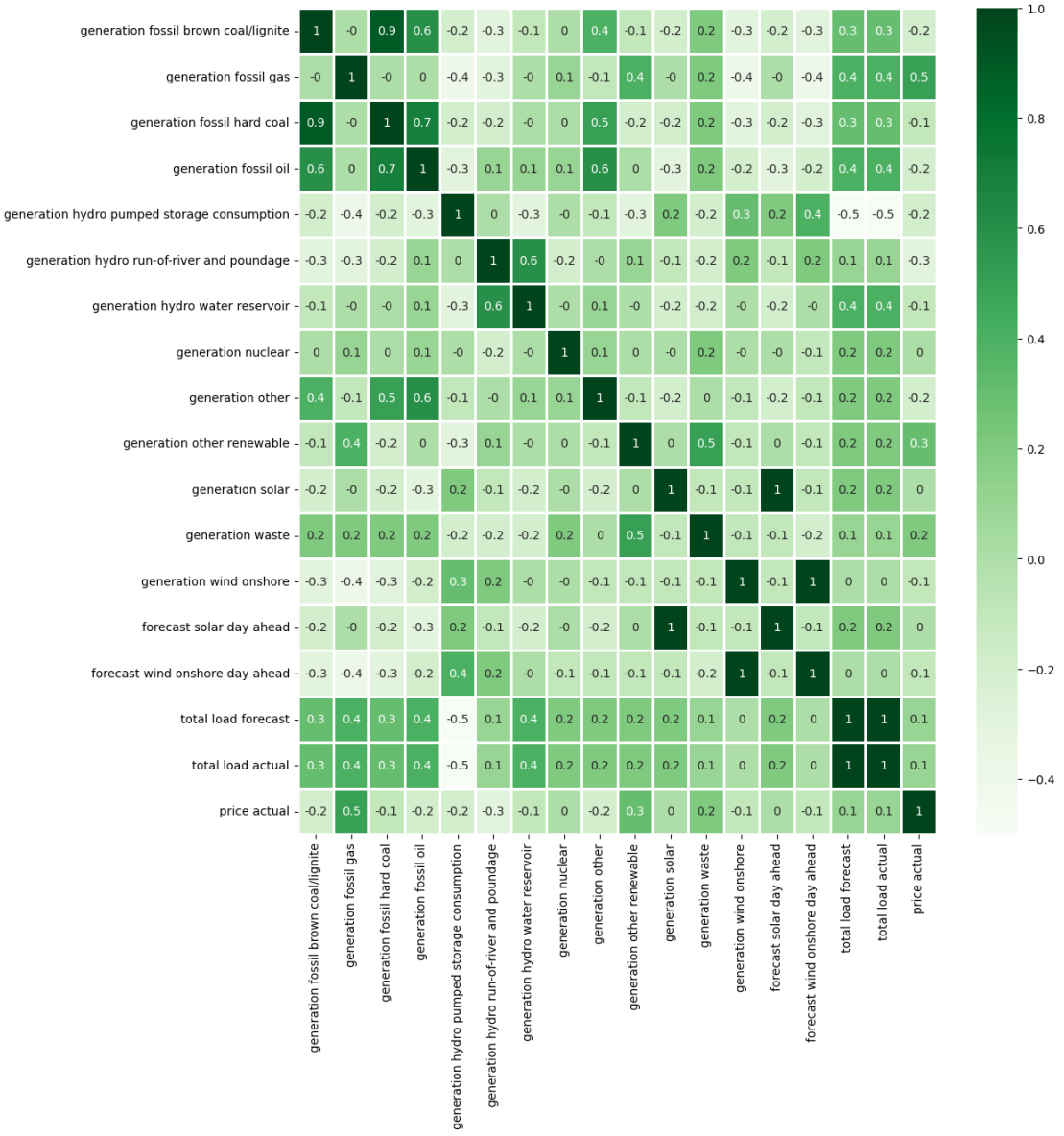

Figura 3.7: Matriz Correlación Pearson Features Modelo. Fuente: elaboración propia

Aquellas relaciones que presentan los valores más destacables son:

- La variable que tiene una mayor influencia en el precio de la electricidad es la Generation Fossil Gas, la cual presenta un coeficiente de correlación de 0,5 que implica que tienen una relación lineal notablemente presente. Esto quiere decir que cuando se genera electricidad con esta tecnología el precio sube, lo cual es comprensible debido a que el gas es una tecnología que suele

marcar el precio de la electricidad por ser un mercado marginalista y casar en punta.

- Por otro lado, se puede comprobar que, de las características que son predicciones, como la predicción de generación solar, generación eólica y demanda al compararlas frente a la realidad presentan un coeficiente de correlación de 1. Esto implica que las predicciones sobre la cantidad de energía eléctrica que se va a generar con dichas tecnologías son buenas y existe una correlación lineal positiva perfecta entre ellas.

Debido a que la correlación de Pearson mide únicamente la linealidad entre dos variables, puede ser que se arrojen resultados que parezcan no ser correctos. Esto puede ser el caso de la relación entre la generación solar y el precio actual, entre otros, la cual indica que es cero. Estas dos variables tienen relación, como veremos posteriormente, pero al ser no lineal este test de correlación no la refleja.

3.4. Resultados obtenidos

En esta sección del documento se van a presentar los resultados obtenidos a lo largo del proyecto. Para ello, primero se comentará el tratamiento de datos llevado a cabo antes de la programación de los diferentes modelos. Posteriormente, se comentarán modelo a modelo los resultados obtenidos y finalmente se concluirá determinando cuál es el modelo que presenta los resultados más robustos de todos.

Cabe destacar que, de los modelos expuestos en la Sección 3.2, el Baseline, el Linear y el Linear Ridge han sido empleados como cota superior del error máximo que se puede obtener. Es decir, se sabe de antemano que debido a la simplicidad de esos modelos los resultados no iban a ser lo suficientemente buenos, pero son de interés para fijar un límite superior del error que se podía obtener en los modelos Random Forest y XGBoost.

3.4.1. Tratamiento de datos

En los modelos de Machine Learning, uno de los primeros pasos con los que se debe proceder es con el tratamiento de las variables que se van a estudiar. En el caso de este proyecto, se ha trabajado con una base de datos que cuenta con más de 82.500 observaciones, 27 features y 1 variable objetivo. A pesar de que la fuente de la que se han obtenido los datos los presentaba de forma muy ordenada y contaba con un porcentaje muy bajo de datos que faltasen, ha sido una parte importante del esfuerzo disponer de todos estos en el formato correcto y uniforme durante todos los años.

Como consecuencia de la evolución del mercado eléctrico y su transición hacia la europeización de este, a partir del 23 de mayo de 2022, los datos obtenidos de las potencias se encontraban en valores cuarto-horarios. Ante la necesidad de disponer de todos los datos en el mismo formato y teniendo en cuenta que el precio de la electricidad es un valor horario, se debió de ajustar este valor cuarto horario a un valor horario calculado como el promedio de los cuatro cuartos de hora.

Por otro lado, como se ha comentado con anterioridad el modelo de predicción de datos se ha realizado entre los años 2015 y 2024. Sin embargo, debido a que el comportamiento del mercado eléctrico cambió de forma considerable en marzo del 2020 por la situación de la pandemia ocasionada por el COVID-19, se han hecho modelos con diferentes intervalos temporales con el objetivo de buscar el mejor resultado y buscar las posibles anomalías en el mercado durante periodos extraordinarios. Estas fueron las siguientes:

- Intervalo temporal desde 2015 hasta 2019, ambos incluidos: el mercado se encontraba estable durante los años previos a la pandemia. Este intervalo no puede ser empleado para predecir el precio de la electricidad de los años futuros(posteriores a 2024), pero se estudia para conocer la variación en el error que se obtiene al compararlo con el intervalo postpandemia, el cual representa un aumento considerable del error como se observará posteriormente.

- Intervalo temporal desde 2020 hasta 2024, ambos incluidos: cambia la situación del mercado eléctrico por circunstancias como la pandemia, la Guerra de Ucrania o el conflicto de Gaza, entre otros factores.

- Intervalo temporal desde 2015 hasta 2024, ambos incluidos: esta tercera división es el resultado de la unión de los dos anteriores y se estudia para comparar el resultado de este frente al del intervalo que transcurre desde 2020 hasta 2024.

Aquel intervalo entre los dos últimos que proporcione un error menor será el que se empleará para la predicción del precio de la electricidad.

3.4.2. Resultados modelo Baseline

El modelo Baseline es el modelo más sencillo que se ha empleado en este proyecto para predecir el precio de la electricidad. Cabe reiterar que este modelo es empleado simplemente para conocer el límite superior del error que no se debe superar en los siguientes modelos que se estudiarán y frente al que se compararán los errores

en los diferentes modelos. A continuación, se muestran los resultados obtenidos durante los tres periodos que se han indicado anteriormente.

2015-2019

Como se ha explicado con anterioridad en la Sección 3.2.1, este modelo calcula el promedio del precio de la electricidad en el periodo de entrenamiento y ese es el valor que emplea como predicción. En este caso, el valor de la predicción es un valor constante muy cercano a los 50€/MWh, y se compara frente a los valores reales del precio de la electricidad. Se muestra en la Figura 3.8 el resultado de las últimas 150 horas del periodo de test en las que se ha predicho el precio de la electricidad en el intervalo 2015-2019. En ella, como en el resto de figuras del documento que muestren esta comparación se muestra en rojo los valores de la predicción y en azul el valor de la variable objetivo, es decir, el precio real de la electricidad.

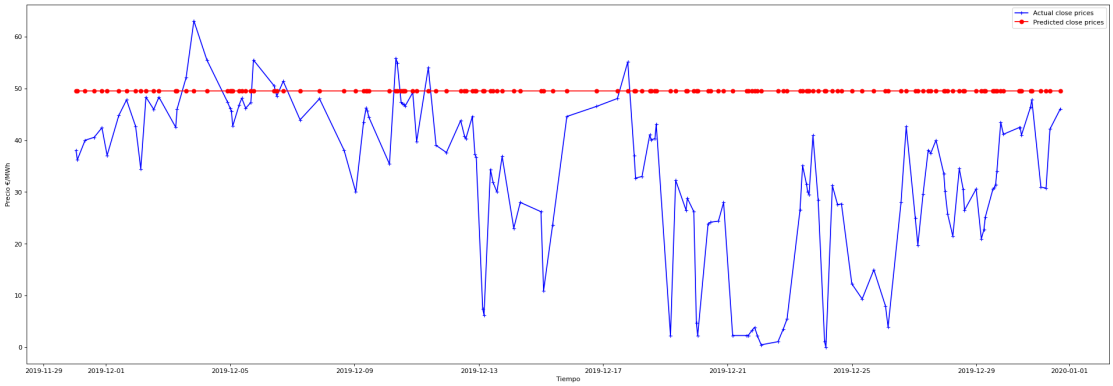

Figura 3.8: Comparación Modelo Baseline 2015-2019. Fuente: elaboración propia

En cuanto a los errores obtenidos, calculados como la diferencia en valor absoluto entre el valor de la predicción y el valor del precio de la electricidad, se muestran en la Figura 3.9.

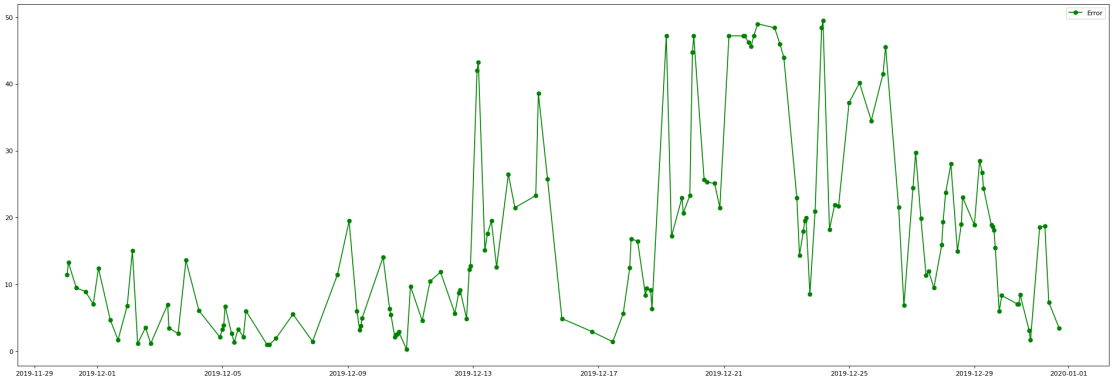

Figura 3.9: MAE Modelo Baseline 2015-2019. Fuente: elaboración propia

Finalmente, se obtiene un MAE de 10,75€/MWh para el periodo de entrenamiento y de 10,67€/MWh para el periodo de test. Estos resultados presentan un error bastante bajo debido a que la volatilidad de los precios es pequeña y el precio promedio durante el periodo de entrenamiento y el de test son bastante similares.

2020-2024

En cuanto a los resultados obtenidos para el periodo 2020-2024, como se puede observar en la figura siguiente, la volatilidad del precio de la electricidad es mucho mayor que en el periodo anterior y por lo tanto los datos se encuentran menos centrados en la media. Este hecho penaliza en gran medida los resultados de dicho modelo. Además, otra consideración importante de este periodo es que la media se encuentra casi al doble del valor que en el periodo anterior. Esto no penaliza al modelo en sí, pero muestra un cambio en la tendencia del mercado como se había comentado en anterioridad.

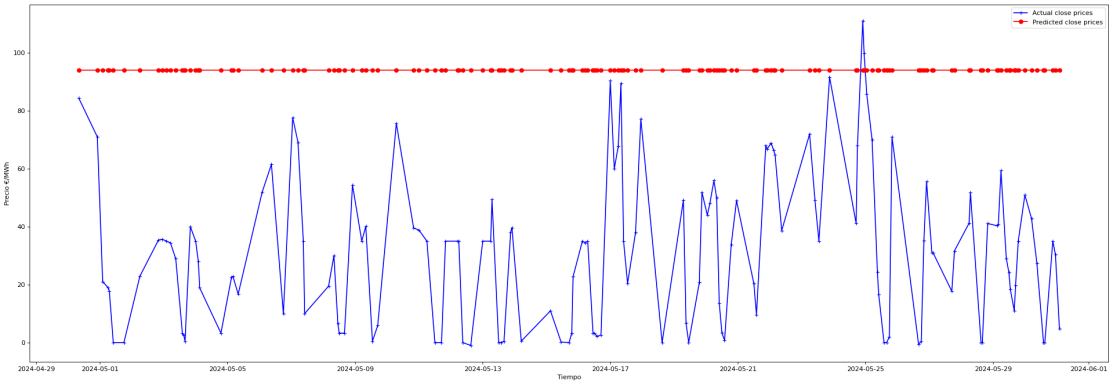

Figura 3.10: Comparación Modelo Baseline 2020-2024. Fuente: elaboración propia

En cuanto a los errores obtenidos, calculados como la diferencia en valor absoluto entre el valor de la predicción y el valor del precio de la electricidad, se muestran en la Figura 3.11.

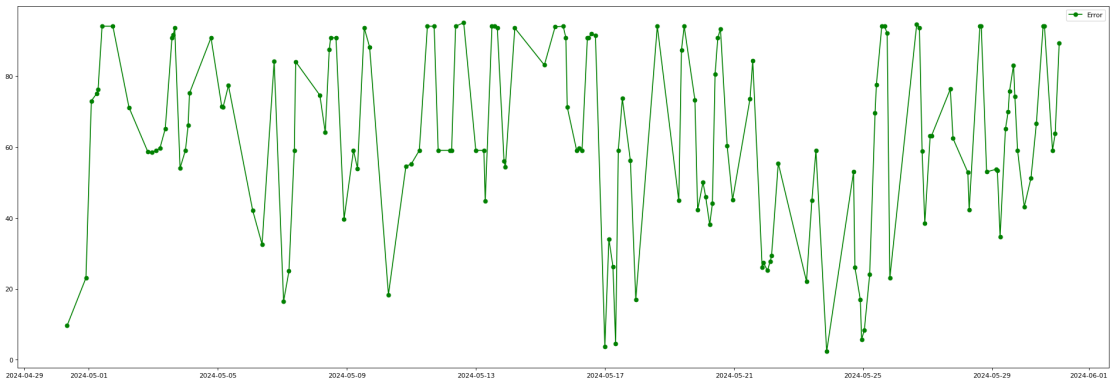

Figura 3.11: MAE Modelo Baseline 2020-2024. Fuente: elaboración propia

Se concluye con que, los resultados para el periodo 2020-2024 del modelo Baseline presentan un MAE de 57,67€/MWh para el periodo de entrenamiento y de 57,47€/MWh para el periodo de test.

2015-2024

Durante el último periodo de tiempo estudiado para el modelo Baseline se obtienen unos resultados mejores a los obtenidos con el periodo anterior. Un motivo por el cual ocurre esto es porque el año 2024 puede asimilarse más a los años prepandémicos que a los años de después de la pandemia, por lo que el entrenamiento se ajusta en mayor medida al periodo del test.

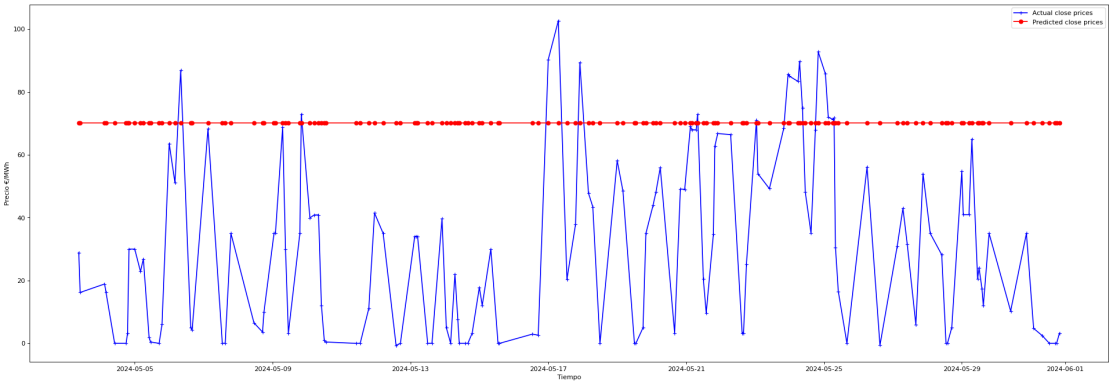

Figura 3.12: Comparación Modelo Baseline 2015-2024. Fuente: elaboración propia

En cuanto a los errores obtenidos, calculados como la diferencia en valor absoluto entre el valor de la predicción y el valor del precio de la electricidad, se muestran en la Figura 3.12.

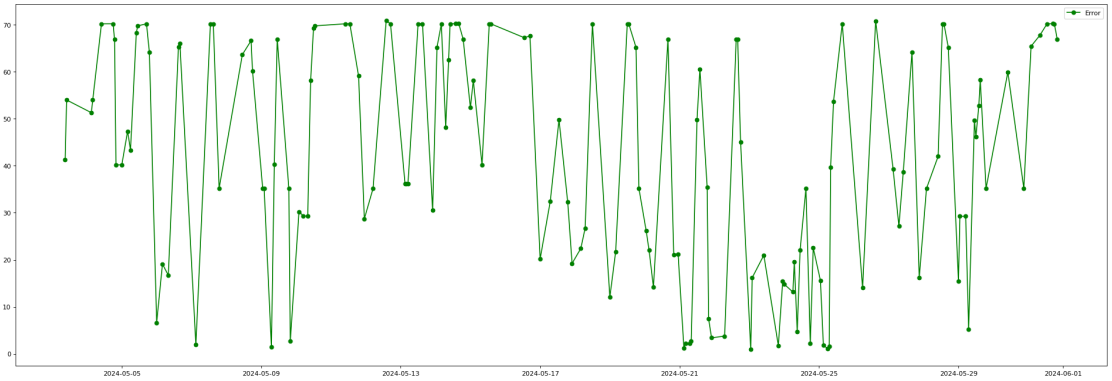

Figura 3.13: MAE Modelo Baseline 2015-2024. Fuente: elaboración propia

Para este periodo se obtiene un MAE de 37,87€/MWh en el periodo de entrenamiento y de 38,63€/MWh en el periodo de test.

Finalmente, en la Tabla 3.4 se recoge un resumen de los resultados de los tres periodos obtenidos para el periodo de entrenamiento y de test, estableciendo de esta forma el límite máximo de error admisible que se va a aceptar para el resto de modelos. En el caso que se obtuviese un error superior podría indicar que hay algo incorrecto en la programación o ejecución del modelo.

	Errores Baseline [€/MWh]	
	Entrenamiento	Test
2015 - 2019	10,75	10,67
2020 - 2024	57,67	57,47
2015 - 2024	37,87	38,63

Tabla 3.4: Resultados Modelo Baseline en Entrenamiento y Test

3.4.3. Resultados modelo Linear

El modelo Linear es un modelo que presenta resultados significativamente mejores que el modelo Baseline, a pesar de su sencillez. Este ha sido capaz de reducir los errores entre un 26,25 % y un 46,02 % frente a los resultados obtenidos en el modelo anterior. A continuación, se muestran los resultados obtenidos durante los tres periodos indicados anteriormente.

2015-2019

Como se ha explicado en la Sección 3.2.2, en el modelo lineal trata de predecir el precio de la electricidad mediante una función que se representa como una recta en $\mathbb{R}^p$, donde p es el número de features. Al calcular posteriormente el valor de esa función de regresión en cada uno de los puntos, es decir, sustituyendo por el valor de cada una de las 27 características en este caso, se obtiene el valor de la predicción y se representa en la curva roja.

En la Figura 3.14 se puede observar cómo el resultado del modelo se ajusta bastante mejor al precio real de la electricidad.

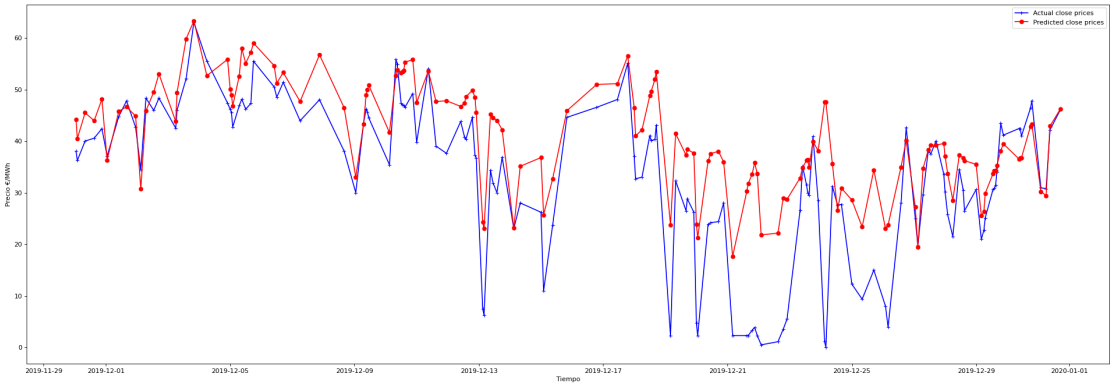

Figura 3.14: Comparación Modelo Linear 2015-2019. Fuente: elaboración propia

En cuanto a los errores obtenidos, calculados como la diferencia en valor absoluto entre el valor de la predicción y el valor del precio de la electricidad, se muestran en la Figura 3.15.

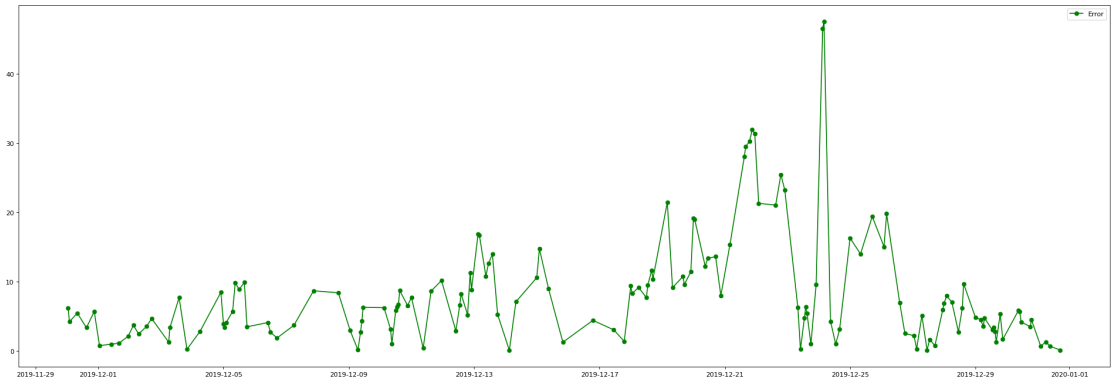

Figura 3.15: MAE Modelo Linear 2015-201. Fuente: elaboración propia

Como se puede observar en el gráfico, todos los errores para este periodo se encuentran muy constantes la mayoría de ellos sin superar la barrera de los 20€/MWh, si bien es cierto que existe un máximo que destaca por encima del resto y que se encuentra cerca de los 50€/MWh. Como objetivos de siguientes modelos se encuentra disminuir la media de errores a lo largo de todo el periodo pero sobretodo evitar que existan errores tan elevados como el comentado ya que serían bastante preocupantes en el caso de que se empleasen estos modelos para acudir a comprar o vender en el mercado eléctrico y haber realizado una estrategia basada en estos modelos.

Finalmente, para este modelo en este periodo se obtienen unos resultados con un MAE de 5,82€/MWh para el periodo de entrenamiento y de 5,77€/MWh para el periodo de test. Esto implica que se han reducido en un 45,86 % en el entrenamiento respecto al modelo Baseline y en un 46,02 % en el test.

2020-2024

Durante el periodo 2020-2024, al igual que ocurría en el modelo Baseline, los resultados empeoraron con respecto al primer periodo estudiado debido a los diferentes acontecimientos que afectaron de forma significativa al mercado. Sin embargo, los resultados fueron también mejores respecto al Baseline, reduciendo el MAE en un 41,06 % durante el periodo de entrenamiento y un 34,06 % durante el periodo de test. En la figura siguiente se muestra la comparativa entre la variable a predecir (rojo) y el precio real durante los últimas 150 horas de test (azul).

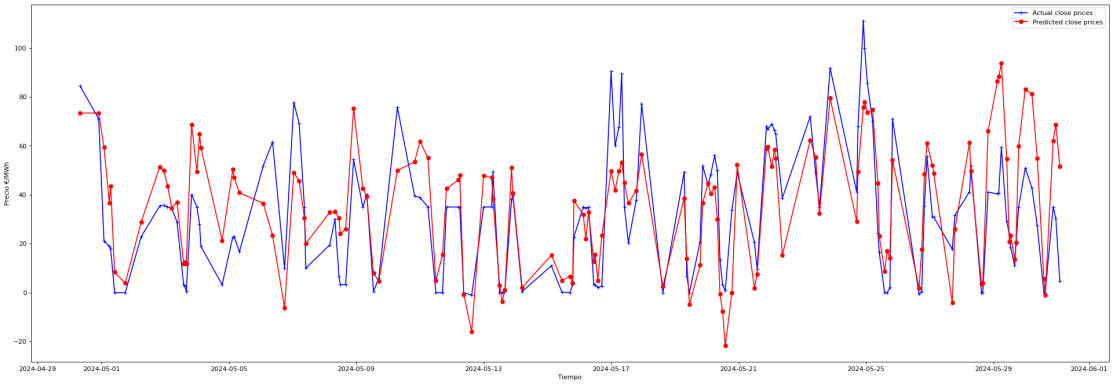

Figura 3.16: Comparación Modelo Linear 2020-2024. Fuente: elaboración propia

En cuanto a los errores obtenidos, calculados como la diferencia en valor absoluto entre el valor de la predicción y el valor del precio de la electricidad, se muestran en la Figura 3.17.

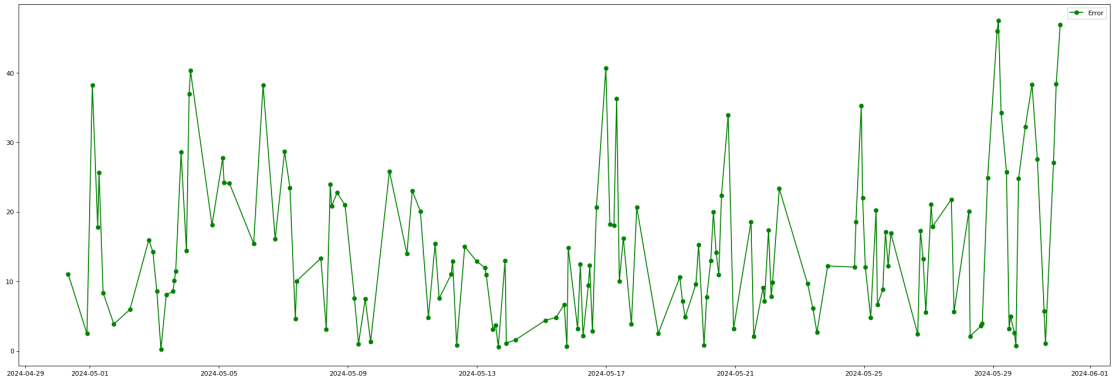

Figura 3.17: MAE Modelo Linear 2020-2024. Fuente: elaboración propia

De esta forma, se obtienen unos errores medios absolutos de 33,99€/MWh en el periodo de entrenamiento y de 34,06€/MWh durante el periodo de test.

2015-2024

Finalmente, durante el periodo completo de tiempo que transcurre desde 2015 hasta 2024, los errores medios absolutos llevan una tendencia similar a la observada en los dos periodos anteriores, en el que el MAE disminuye frente al otro modelo ya estudiado aunque en este caso en menor medida. En la Figura 3.18 se muestra la comparativa entre el valor de la predicción del precio de la electricidad (rojo) y el valor real de dicho precio (azul).

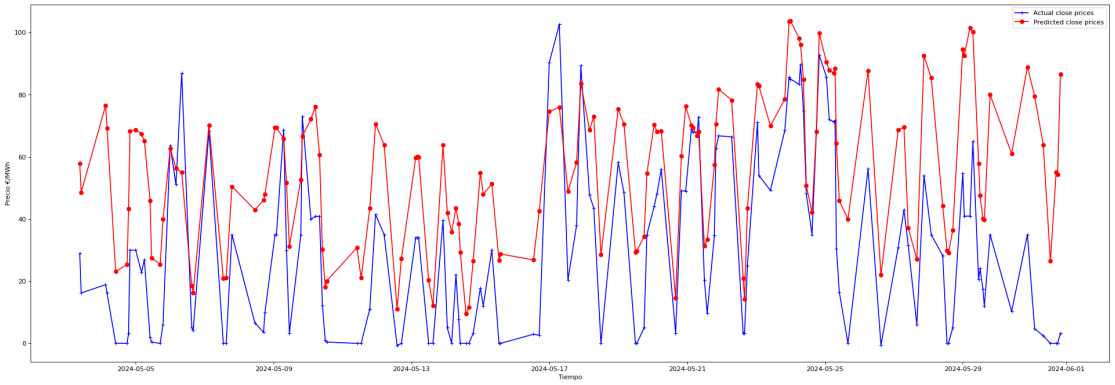

Figura 3.18: Comparación Modelo Linear 2015-2024. Fuente: elaboración propia

En cuanto a los errores obtenidos, calculados como la diferencia en valor absoluto entre el valor de la predicción y el valor del precio de la electricidad, se muestran a continuación en la Figura 3.19.

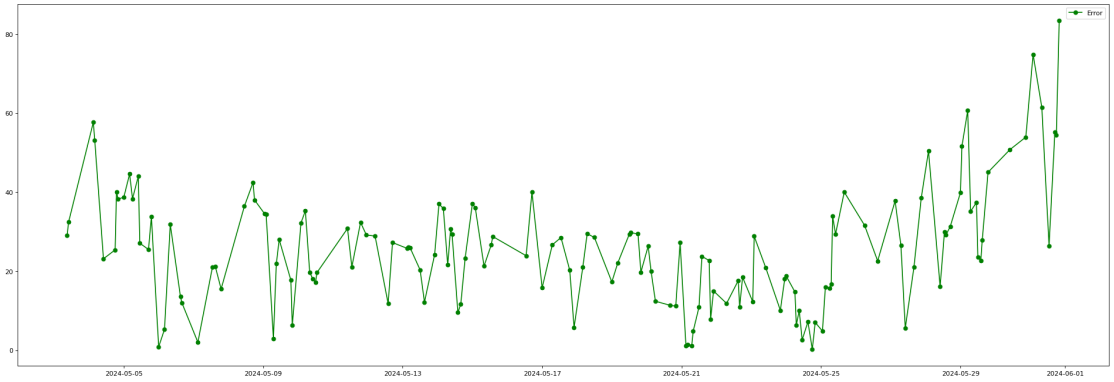

Figura 3.19: MAE Modelo Linear 2015-2024. Fuente: elaboración propia

De esta forma, se reduce en un 26,25 % el MAE con respecto al modelo baseline en el periodo de entrenamiento, obteniendo un valor de 27,93€/MWh y se reduce un 26,97 % el MAE en el periodo de test, obteniendo un valor de 28,21€/MWh.

Finalmente, en la Tabla 3.5 se expone un resumen con los resultados comentados previamente. Cabe matizar que, con este modelo se esperaban conseguir mejores resultados que con el modelo anterior, pero los resultados se deben seguir mejorando en los dos últimos modelos respecto a este de forma significativa.

	Errores Linear [€/MWh]	
	Entrenamiento	Test
2015 - 2019	5,82	5,77
2020 - 2024	33,99	34,06
2015 - 2024	27,93	28,21

Tabla 3.5: Resultados Modelo Linear en Entrenamiento y Test

3.4.4. Resultados modelo Linear Ridge

Tal y como se explicó en la Sección 3.2.3 del presente documento, el modelo Linear Ridge presenta una única diferencia frente al modelo Linear. Esta es que añade un término de regularización L2 cuyo objetivo es evitar que se produzca overfitting, es decir, que los errores obtenidos durante el periodo de entrenamiento sean mucho menores que durante el periodo de test.

No obstante, tal y como se ha mostrado en el modelo anterior, la diferencia de errores entre entrenamiento y test son mínimas. Por ese motivo, como se verá a lo

largo de este punto para los tres periodos no se apreciará apenas diferencia en los errores con y sin dicha regularización L2. Cabe destacar que, en este documento los resultados presentados de los errores se están redondeando al segundo decimal, pero el modelo los calcula con quince decimales. En el caso de haber mostrado todos los decimales se hubiese apreciado diferencia entre los resultados de ambos modelos, pero al redondear el segundo decimal el resultado parece ser el mismo.

A continuación, se muestran los resultados para los tres periodos indicados previamente.

2015-2019

Tal y como se ha comentado con anterioridad, los resultados de este periodo son muy similares a los del modelo Linear.

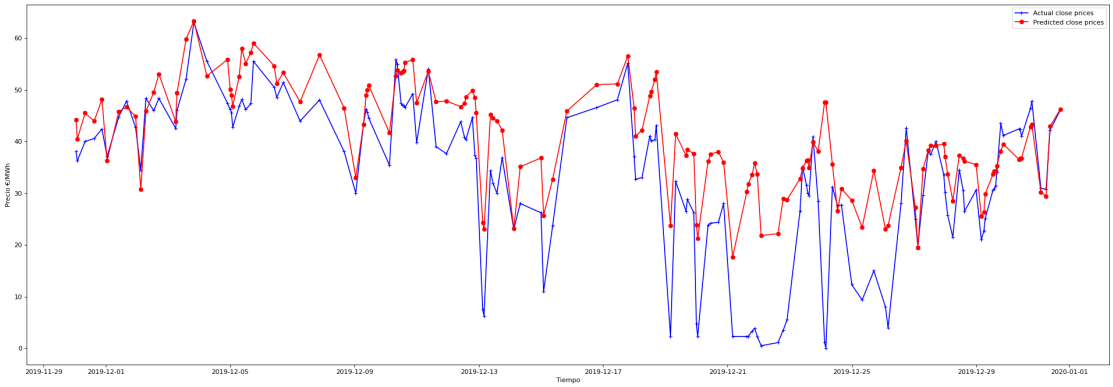

Figura 3.20: Comparación Modelo Linear Ridge 2015-2019. Fuente: elaboración propia

En cuanto a los errores obtenidos, calculados como la diferencia en valor absoluto entre el valor de la predicción y el valor del precio de la electricidad, se muestran en la Figura 3.21.

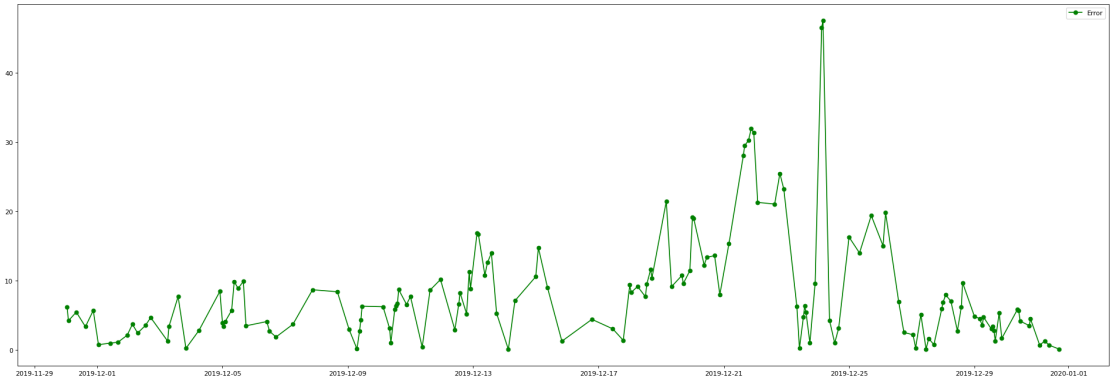

Figura 3.21: MAE Modelo Linear Ridge 2015-2019. Fuente: elaboración propia

Como se puede comprobar, la reducción L2 no ha servido para reducir el error cercano a los 50€/MWh comentado en el mismo periodo del modelo anterior.

De esta forma, se ha obtenido un MAE de 5,82€/MWh para el entrenamiento, lo que supone una reducción del 45,86 % con respecto al Baseline, y un MAE para el test de 5,76€/MWh, que implica una reducción del 46,02 %.

2020-2024

Del periodo que transcurre desde el 2020 hasta el 2024, se muestra en la Figura 3.22 la comparativa entre el precio real (azul) y el valor de la predicción para las últimas 150 horas del periodo de test.

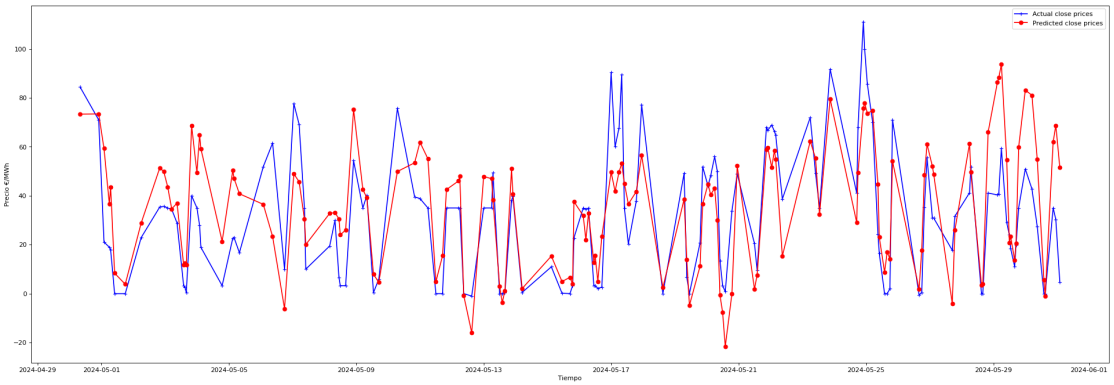

Figura 3.22: Comparación Modelo Linear Ridge 2020-2024. Fuente: elaboración propia

En cuanto a los errores obtenidos, calculados como la diferencia en valor absoluto entre el valor de la predicción y el valor del precio de la electricidad, se muestran en la Figura 3.23.

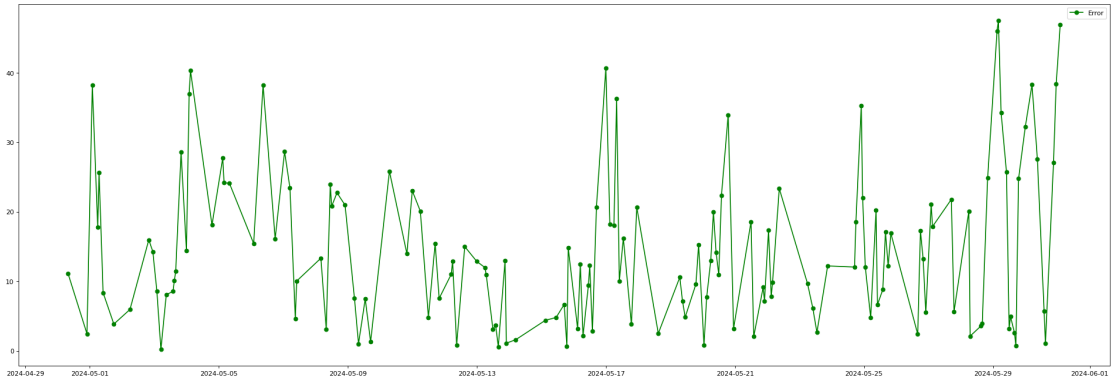

Figura 3.23: MAE Modelo Linear Ridge 2020-2024. Fuente: elaboración propia

En base a los resultados del modelo, se ha obtenido un MAE de 33,99€/MWh, lo que supone una disminución del 41,06 % con respecto al Baseline. En el periodo de test, se obtiene un MAE de 34,06€/MWh, consiguiendo así una reducción de 40,73 % frente al mismo modelo.

2015-2024

Durante el último periodo estudiado para este modelo, se presenta a continuación la Figura 3.24, que muestra la comparación entre los precios horarios del mercado eléctrico (azul) y los valores de las predicciones (rojo).

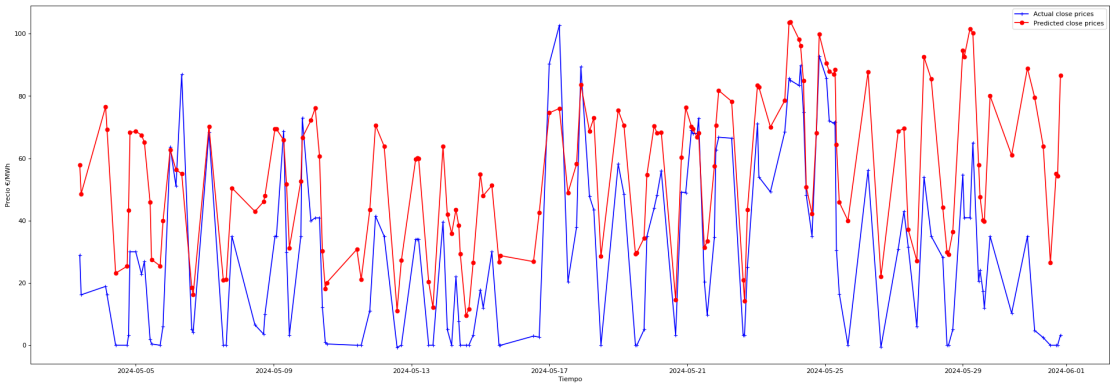

Figura 3.24: Comparación Modelo Linear Ridge 2015-2024. Fuente: elaboración propia

En cuanto a los errores obtenidos, calculados como la diferencia en valor absoluto entre el valor de la predicción y el valor del precio de la electricidad, se muestran en la Figura 3.25.

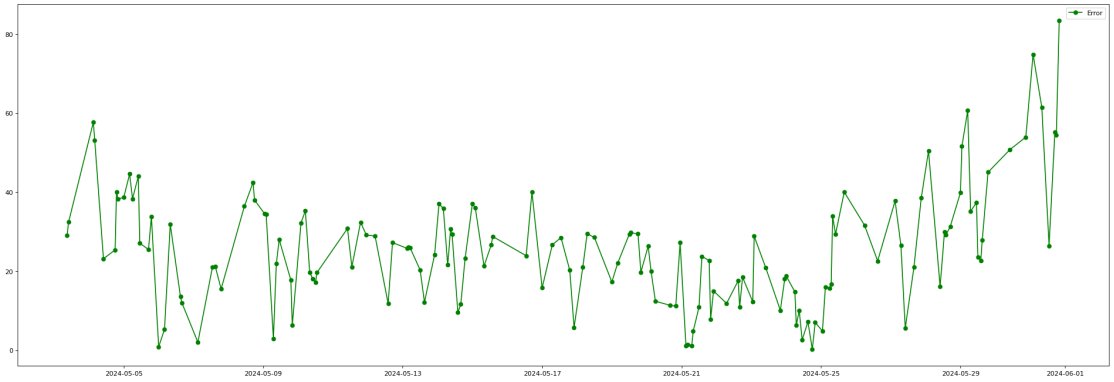

Figura 3.25: MAE Modelo Linear Ridge 2015-2024. Fuente: elaboración propia

Finalmente, durante el periodo de entrenamiento se ha obtenido un MAE de 27,93€/MWh, lo que supone un 26,25 % de reducción frente al modelo Baseline, mientras que en el periodo de test, el MAE es de 28,21€/MWh, que implica una reducción de 26,97 % frente al primer modelo estudiado.

A continuación, se muestra en la Tabla 3.6 el resultado de los MAE comentados a lo largo de este punto. Como se indicaba anteriormente, los resultados son muy similares que los del modelo lineal debido a la baja necesidad de regularización al tener resultados similares en entrenamiento y en test.

	Errores Linear Ridge [€/MWh]	
	Entrenamiento	Test
2015 - 2019	5,82	5,76
2020 - 2024	33,99	34,06
2015 - 2024	27,93	28,21

Tabla 3.6: Resultados Modelo Linear Ridge en Entrenamiento y Test

3.4.5. Resultados modelo Random Forest

Se considera el modelo Random Forest como el primer modelo en el que realmente se esperaba conseguir resultados capaces de ajustarse con un error mucho menor a los valores reales del precio de la electricidad.

Tal y como se ha comentado con anterioridad, el modelo Random Forest es el primer modelo de los desarrollados en este proyecto en el que se realiza una búsqueda de hiperparámetros. Para esto, en este modelo se ha decidido definir la siguiente malla de hiperparámetros:

- Simple imputer Strategy: en el caso de este modelo, los valores nulos o que falten tomarán el valor de la media o de la mediana.
- Random Forest regressor n stimators: en el caso de este modelo, el hiperparámetro n stimators puede tomar valores entre 25 y 200 de 10 en 10.
- Random Forest regressor min samples split: en el caso de este modelo, el hiperparámetro min samples split puede tomar valores entre 2 y 5 de 1 en 1.
- Random Forest regressor max samples: en el caso de este modelo, el hiperparámetro max samples split puede tomar valores entre 0,2 y 1 de 0,1 en 0,1.
- Random Forest regressor max features: en el caso de este modelo, el hiperparámetro max features podrá tomar como posibles valores el logaritmo en base dos o la raíz cuadrada del número de características.
- Random Forest regressor max depth: en el caso de este modelo, el hiperparámetro max depth puede tomar valores entre 5 y 35 de 5 en 5.

Al realizar la búsqueda de hiperparámetros se empleó la estrategia Random Search, limitando el número de combinaciones posibles a 100. De esta forma, los resultados fueron los siguientes:

Hiperparámetro	**2015-2019**	**2020-2024**	**2015-2024**
Simple imputer Strategy	Mediana	Media	Media
R. F. regressor n estimators	155	165	105
R. F. regressor min samples split	3	2	3
R. F. regressor max samples	0,9	0,9	0,9
R. F. regressor max features	log 2	sqrt	log 2
R. F. regressor max depth	25	20	25

Tabla 3.7: Resultados hiperparámetros modelo Random Forest

Una vez fijados estos parámetros, se llevó a cabo la ejecución del modelo en sí. Diferenciando en los mismos tres periodos que en el resto de modelos, se obtuvieron los resultados que se muestran a continuación.

2015-2019

A lo largo de este periodo los resultados mejoran considerablemente con respecto a los obtenidos en cualquiera de los tres modelos anteriores. En la Figura 3.26

se muestra la predicción realizada (rojo) de forma horaria frente a lo que ocurrió realmente durante las últimas 150 horas del periodo de test.

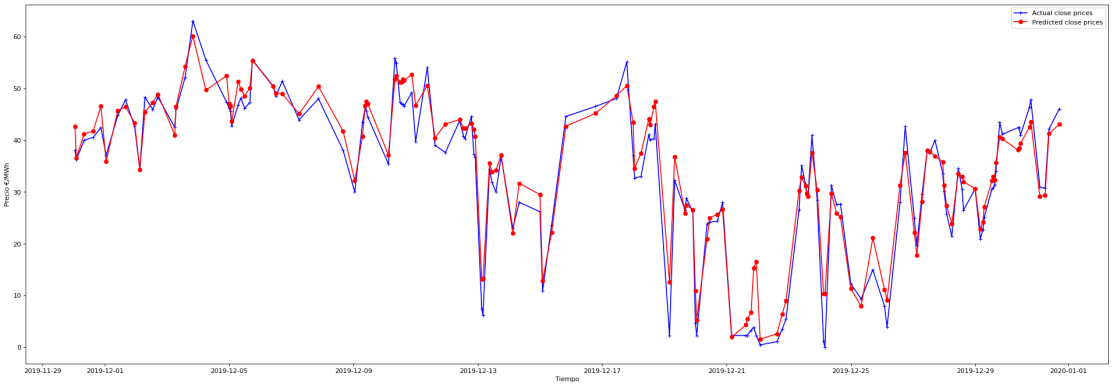

Figura 3.26: Comparación Modelo Random Forest 2015-2019. Fuente: elaboración propia

En cuanto a los errores obtenidos, calculados como la diferencia en valor absoluto entre el valor de la predicción y el valor del precio de la electricidad, se muestran en la Figura 3.27.

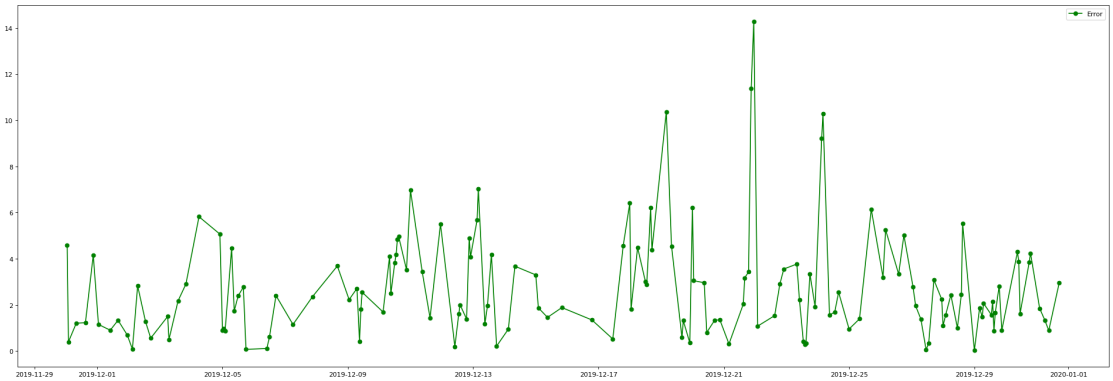

Figura 3.27: MAE Modelo Random Forest 2015-2019. Fuente: elaboración propia

Los resultados obtenidos para este modelo para el periodo de entrenamiento muestran un MAE de 0,98€/MWh lo que supone un descenso del 90,88 % frente al modelo Baseline, mientras que en el periodo de test se obtiene un valor de 2,25€/MWh que implica un descenso del 78,91 %.

2020-2024

En el periodo de tiempo que transcurre desde el 2020 hasta el 2024, se obtienen unos resultados con un error mayor que en el periodo anterior. Sin embargo, tal

como se mostrará posteriormente, la mejora porcentual frente al primer modelo es mayor que para el periodo 2015-2019.

En la Figura 3.28 se muestra la comparativa entre los precios reales y los valores de las predicciones obtenidos durante los últimos días del periodo de test.

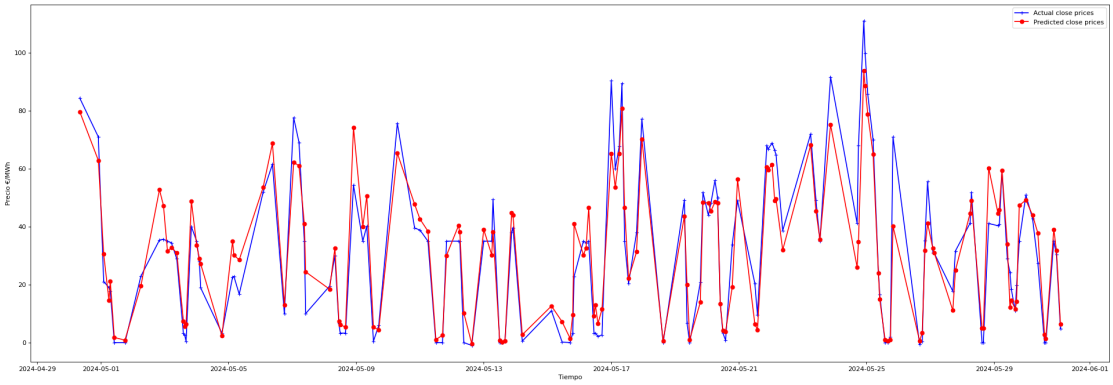

Figura 3.28: Comparación Modelo Random Forest 2020-2024. Fuente: elaboración propia

En cuanto a los errores obtenidos, calculados como la diferencia en valor absoluto entre el valor de la predicción y el valor del precio de la electricidad, se muestran en la Figura 3.29.

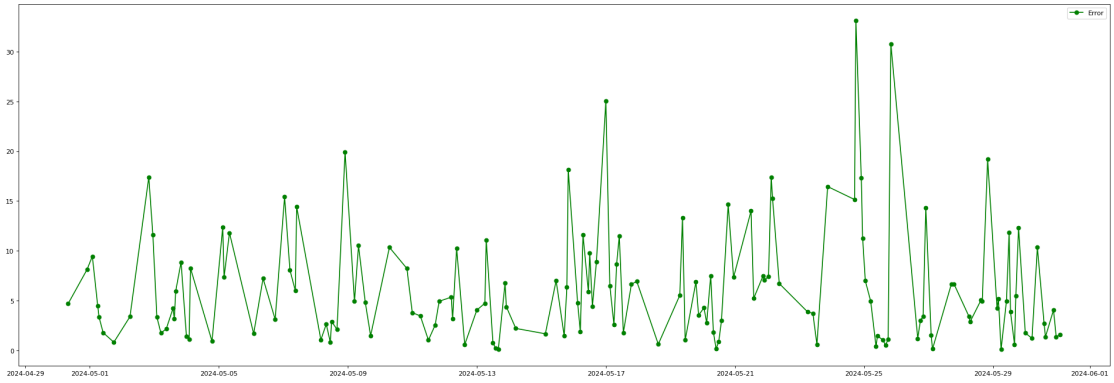

Figura 3.29: MAE Modelo Random Forest 2020-2024. Fuente: elaboración propia

En el caso de este periodo temporal, el MAE obtenido para el entrenamiento es de 4€/MWh, lo que supone una mejora con respecto al modelo Baseline de un 93,06 %. Por otro lado, para el periodo de test el MAE obtenido es de 8,76€/MWh, obteniendo así una mejora del 84,76 %.

2015-2024

Tomando observaciones que transcurren entre el 2015 y el 2024, al igual que ocurría en los modelos anteriores, permite obtener mejores resultados que seleccionando el periodo anterior que solo transcurre entre el 2020 y el 2024.

En la Figura 3.30 se muestra la diferencia entre los valores horarios de predicción y reales de precio de la electricidad.

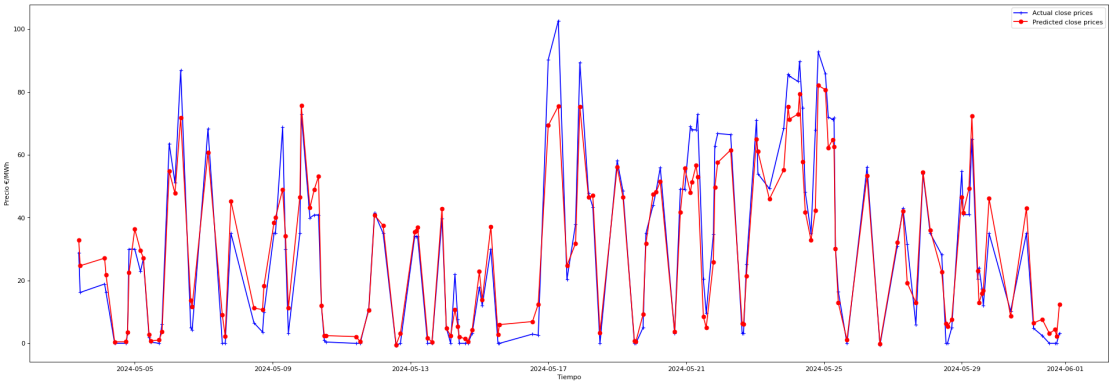

Figura 3.30: Comparación Modelo Random Forest 2015-2024. Fuente: elaboración propia

En cuanto a los errores obtenidos, calculados como la diferencia en valor absoluto entre el valor de la predicción y el valor del precio de la electricidad, se muestran en la Figura 3.31.

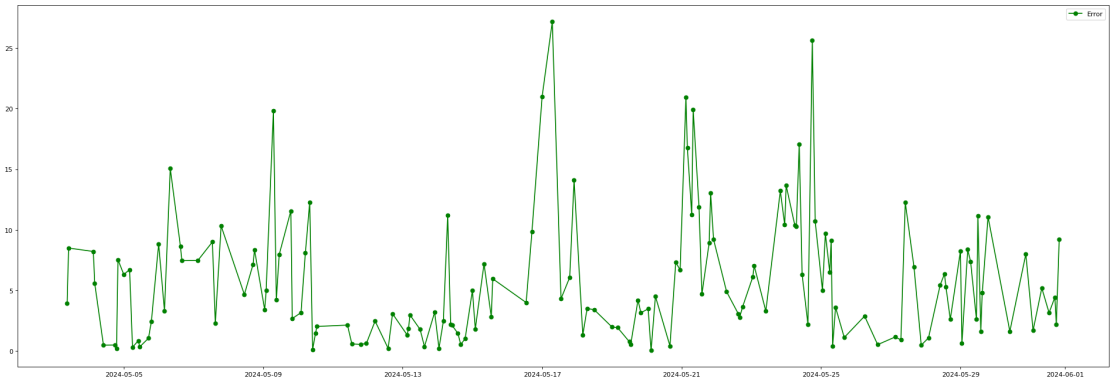

Figura 3.31: MAE Modelo Random Forest 2015-2024. Fuente: elaboración propia

En base a los resultados del modelo, se ha obtenido un MAE de 2,38€/MWh, lo que supone una disminución del 93,72 % con respecto al Baseline. En el periodo de test, se obtiene un MAE de 5,55€/MWh, consiguiendo así una reducción de

85,63 % frente al mismo modelo.

En la Tabla 3.8 se dispone de un resumen de los MAE del modelo Random Forest para los tres periodos anteriormente comentados. En ella se puede apreciar una mejora significativa en los resultados frente a los modelos anteriores.

	Errores Random Forest [€/MWh]	
	Entrenamiento	Test
2015 - 2019	0,98	2,25
2020 - 2024	4,00	8,76
2015 - 2024	2,38	5,55

Tabla 3.8: Resultados Modelo Random Forest en Entrenamiento y Test

3.4.6. Resultados modelo XGBoost

Al igual que ocurre con el modelo Random Forest, en el modelo XGBoost se debe llevar a cabo una búsqueda de hiperparámetros previa a la obtención de la previsión del precio de la electricidad. Para esto, en este modelo se ha decidido definir la siguiente malla de hiperparámetros:

- Simple imputer strategy: en el caso de este modelo, los valores nulos o que falten tomarán el valor de la media o de la mediana.
- Xgbregressor subsample: en el caso de este modelo, el hiperparámetro Xgbregressor subsample puede tomar valores entre 0,8 y 1 de 0,1 en 0,1.
- Xgbregressor n estimators: en el caso de este modelo, el hiperparámetro Xgbregressor n estimators puede tomar valores entre 100 y 300 de 100 en 100.
- Xgbregressor min child weight: en el caso de este modelo, el hiperparámetro Xgbregressor min child weight puede tomar valores entre 1 y 5 de 2 en 2.
- Xgbregressor max depth: en el caso de este modelo, el hiperparámetro Xgbregressor max depth puede tomar los valores 3, 5, 7 y 10.
- Xgbregressor learning rate: en el caso de este modelo, el hiperparámetro Xgbregressor learning rate puede tomar los valores 0,01, 0,1 y 0,2.
- Xgbregressor lambda: en el caso de este modelo, el hiperparámetro Xgbregressor lambda puede tomar los valores 0, 0,1 y 0,5.

- Xgbregressor gamma: en el caso de este modelo, el hiperparámetro Xgbregressor gamma puede tomar los valores 0, 0,1 y 0,2.
- Xgbregressor colsample bytree: en el caso de este modelo, el hiperparámetro Xgbregressor colsample bytree puede tomar los valores 0,8, 0,9 y 1.
- Xgbregressor alpha: en el caso de este modelo, el hiperparámetro Xgbregressor alpha puede tomar los valores 0, 0,1 y 0,5.

Al realizar la búsqueda de hiperparámetros se empleó la estrategia Random Search, limitando el número de combinaciones posibles a 100. De esta forma, los resultados fueron los siguientes:

Hiperparámetro	2015-2019	2020-2024	2015-2024
Simple imputer Strategy	Mediana	Media	Media
Xgbregressor subsample	0,8	0,8	1
Xgbregressor n estimators	300	300	300
Xgbregressor min child weight	5	5	3
Xgbregressor max depth	10	10	10
Xgbregressor learning rate	0,1	0,1	0,1
Xgbregressor lambda	0,5	0,1	0
Xgbregressor gamma	0,2	0	0,2
Xgbregressor colsample bytree	0,9	0,9	0,8
Xgbregressor alpha	0	0	0,5

Tabla 3.9: Resultados hiperparámetros modelo XGBoost

Una vez fijados estos parámetros, se llevó a cabo la ejecución del modelo en sí. Diferenciando en los mismos tres periodos que en el resto de modelos, se obtuvieron los resultados que se muestran a continuación.

2015-2019

Los mejores resultados de todo el proyecto se obtienen mediante el modelo XGBoost en el intervalo temporal en cuestión. La comparativa entre los valores del precio de la electricidad real (azul) frente a sus predicciones (rojo) se muestran en la Figura 3.32

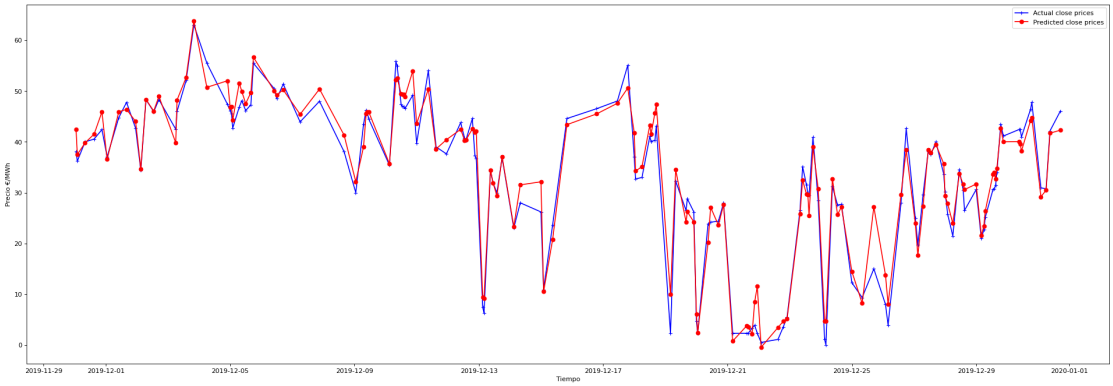

Figura 3.32: Comparación Modelo XGBoost 2015-2019. Fuente: elaboración propia

En cuanto a los errores obtenidos, calculados como la diferencia en valor absoluto entre el valor de la predicción y el valor del precio de la electricidad, se muestran en la Figura 3.33.

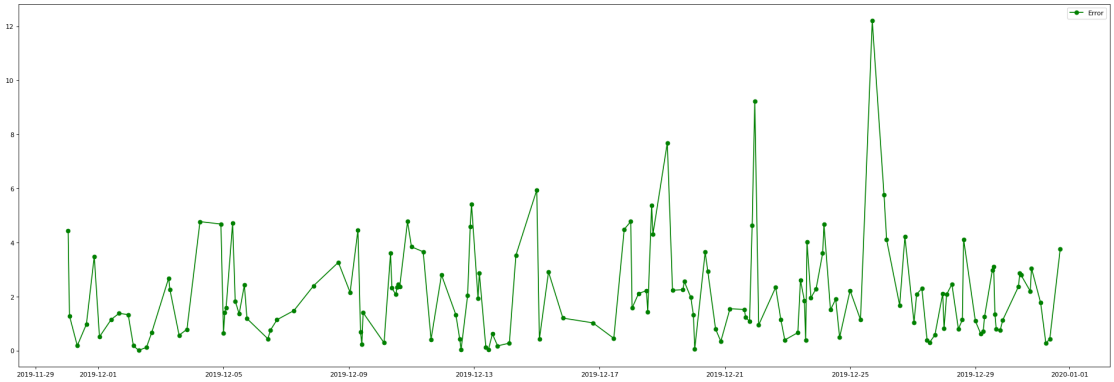

Figura 3.33: MAE Modelo XGBoost 2015-2019. Fuente: elaboración propia

Durante este periodo, el valor del MAE para el periodo de entrenamiento es de 0,45€/MWh, lo que supone una reducción de este frente al primer modelo estudiado (Baseline) de un 95,81 %. Por otro lado, en el periodo de test, el MAE obtenido es de 1,94€/MWh, reduciendo así el error frente al modelo Baseline en un 81,82 %.

En este modelo, a pesar de haber empleado hiperparámetros que tratasen de regularizar y evitar así que se produzca overfitting, es algo que puede mejorarse. Como se puede observar en los resultados comentados hasta el momento y en los que se comentarán durante los otros dos periodos temporales, el error obtenido en el entrenamiento es considerablemente menor que en el test.

2020-2024

En el periodo de tiempo que transcurre desde el 2020 hasta el 2024, se obtienen unos resultados con un error mayor que en el periodo anterior. De esta forma se ha cumplido en los cinco modelos estudiados que el MAE de este intervalo ha sido el mayor de los tres por el cambio en la tendencia del mercado ya comentado en el documento.

A continuación, se muestra la comparativa entre los precios reales de la electricidad (azul) y el valor de las predicciones de dicha variable objetivo (rojo).

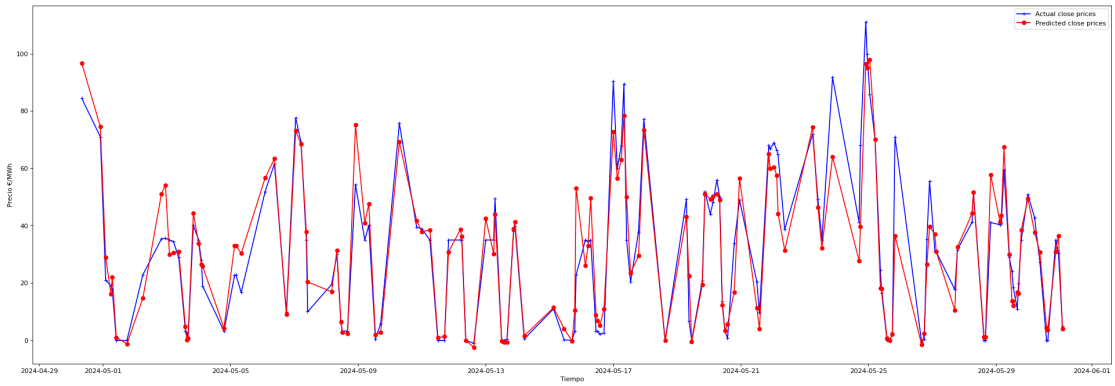

Figura 3.34: Comparación Modelo XGBoost 2020-2024. Fuente: elaboración propia

En cuanto a los errores obtenidos, calculados como la diferencia en valor absoluto entre el valor de la predicción y el valor del precio de la electricidad, se muestran en la Figura 3.35.

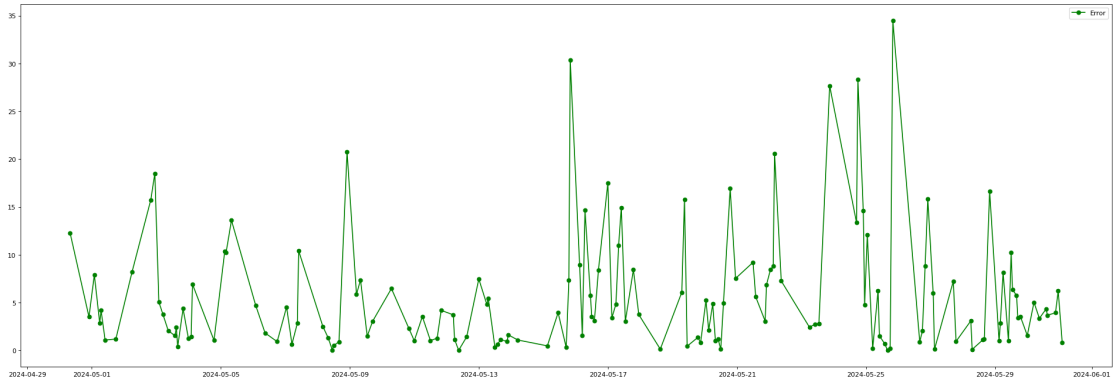

Figura 3.35: MAE Modelo XGBoost 2020-2024. Fuente: elaboración propia

En base a los resultados mostrados, se ha obtenido un MAE de 1,58€/MWh para el periodo de entrenamiento, lo que supone una mejora de un 97,26 % frente al modelo

Baseline. En cuanto al periodo de test, el MAE obtenido asciende a 7,87€/MWh, obteniendo así una reducción de este de un 86,31 %. Tal y como se destacaba previamente, este modelo también sufre de overfitting.

2015-2024

Finalmente, para el último intervalo estudiado del modelo XGBoost, se consiguen los mejores resultados de todo el proyecto que puedan ser realmente empleados para predecir el precio de la electricidad. En la Figura 3.36, se muestra la comparativa de las últimas 150 horas del periodo de test del intervalo 2015-2024, en la que se muestra en rojo el valor de la variable a predecir para cada hora y en azul el valor real de la electricidad en esa hora.

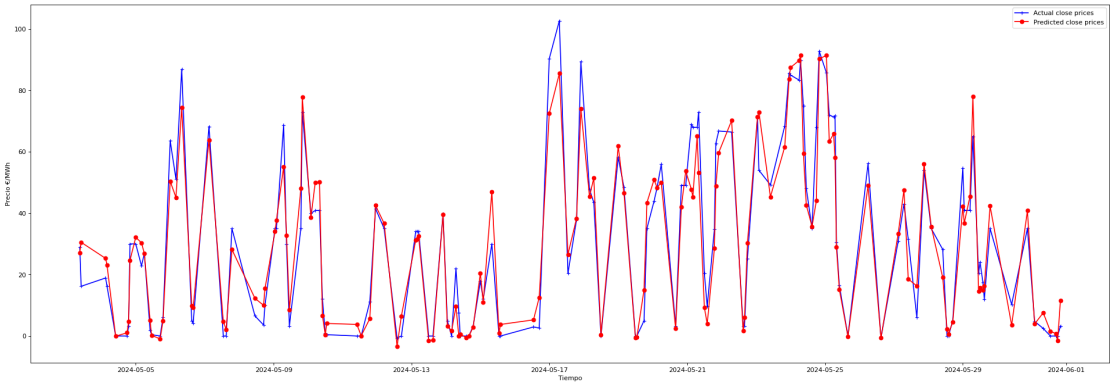

Figura 3.36: Comparación Modelo XGBoost 2015-2024. Fuente: elaboración propia

En cuanto a los errores obtenidos, calculados como la diferencia en valor absoluto entre el valor de la predicción y el valor del precio de la electricidad, se muestran en la Figura 3.37.

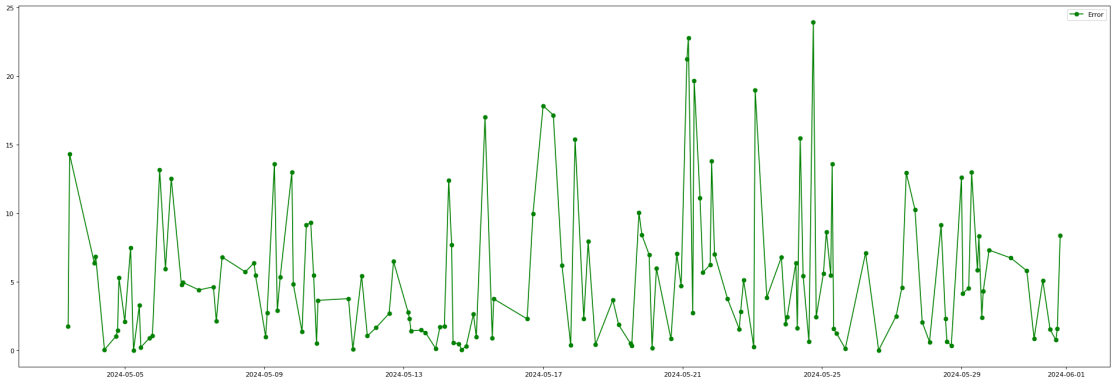

Figura 3.37: MAE Modelo XGBoost 2015-2024. Fuente: elaboración propia

Para finalizar, el MAE obtenido con el modelo XGBoost entre los años 2015 y 2024 para el periodo de entrenamiento es de 1,6€/MWh, alcanzando así una reducción del 95,78 % frente al modelo Baseline. Por otro lado, en el periodo de test se alcanza un MAE de 5,26€/MWh, obteniendo una reducción del 86,38 % frente al Baseline.

A continuación, en la Tabla 3.10 se muestra un resumen con los resultados de los errores medios absolutos obtenidos en el modelo XGBoost durante los tres intervalos temporales estudiados.

	Errores XGBoost [€/MWh]	
	Entrenamiento	Test
2015 - 2019	0,45	1,94
2020 - 2024	1,58	7,87
2015 - 2024	1,60	5,26

Tabla 3.10: Resultados Modelo XGBoost en Entrenamiento y Test

3.4.7. Resultados generales de los modelos

Una vez se han estudiado los resultados de los cinco modelos, se puede proceder a determinar una serie de conclusiones que ayuden a saber cuál ha sido el mejor modelo.

Para empezar, se muestra una gráfica que representa los errores de todos los modelos para cada intervalo temporal diferenciando en estos los periodos de entrenamiento y los de test.

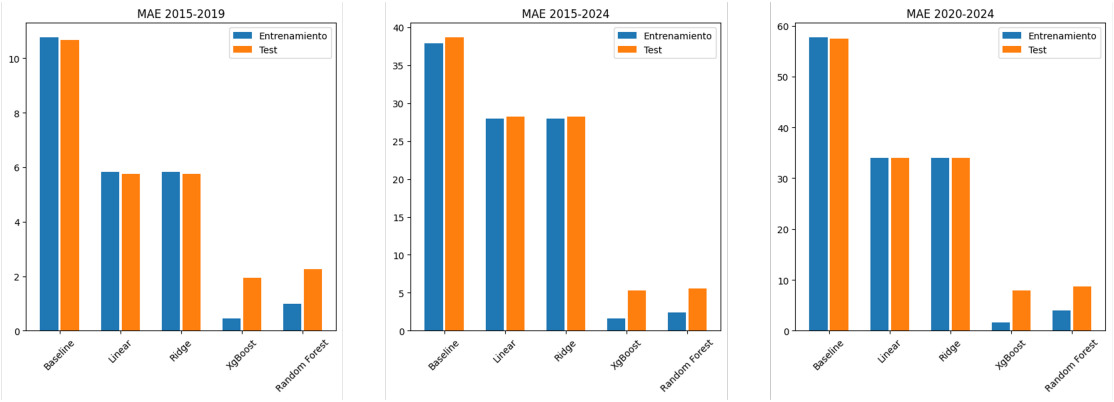

Figura 3.38: Comparación MAE cinco modelos. Fuente: elaboración propia

Tal y como se ha ido analizando en el documento, los resultados obtenidos para el periodo 2015-2019 en general han sido mucho mejores que durante los otros dos periodos. Tanto es así que, con modelos mucho más simples, en el primer intervalo se podían alcanzar predicciones con menores errores que en los otros dos periodos temporales. Sin embargo, como se adelantaba anteriormente, este periodo no puede ser empleado para predecir el precio del año 2024, o lo haría con un error muy grande debido a que se trata de una serie temporal y habría una gran cantidad de años entre el último dato del que se cuenta y el primero que se querría predecir.

En cuanto a los otros dos intervalos temporales, los resultados obtenidos en el intervalo que transcurre desde 2015 hasta 2024 proporciona en todos los modelos errores más bajos que en el periodo entre 2020 y 2024. Por ese motivo, se concluye que a la hora de predecir el precio es mejor no coger únicamente los valores posteriores al 2020, sino tener un histórico de datos lo más grande posible.

Poniendo el foco en los modelos empleados, tal y como se ha explicado en repetidas ocasiones a lo largo del proyecto, los tres primeros modelos (Baseline, Linear y Linear Ridge) se empleaban para ir asegurando una mejora continua de los resultados, siendo capaces de detectar así posibles errores. En cuanto a los otros dos modelos, Random Forest y XGBoost, el segundo de ellos ha sido aquel que ha proporcionado mejores resultados. Se trata de un modelo que es más moderno y que como se indicaba en la explicación de estos suele ofrecer menores errores. A destacar de estos dos modelos es que, tal y como se puede observar en la Figura 3.38, se produce overfitting por ser los errores del periodo de entrenamiento bastante menores que los obtenidos en el test.

Para finalizar este Capítulo 3, se concluye con que el modelo de predicción que se emplearía para predecir el precio, entre los estudiados, sería el XGBoost, debido a que este ha demostrado ser el más efectivo. Además, destacar que las features que han sido seleccionadas para desarrollar el modelo son correctas, tal y como evidencian los resultados obtenidos. Por otro lado, es importante aclarar que, a pesar de los resultados satisfactorios conseguidos, estos modelos pueden ser mejorados y por ese motivo en las conclusiones finales del proyecto y en las líneas futuras de investigación se propondrán una serie de alternativas para seguir mejorando los resultados de éste.

Capítulo 4

Factores clave en las features del modelo

En el Capítulo 3 del presente proyecto, se llevó a cabo un modelo de predicción del precio de la electricidad. Este sirvió, entre otras cosas, para validar la eficacia de las features empleadas. Una vez corroborado que esas características son válidas, a lo largo del Capítulo 4 se van a analizar aquellos factores que influyen en dichas features. De esta forma, se podrán obtener mejores predicciones en futuros modelos y entender mejor posibles cambios en tendencias de mercado como la sufrida a partir del año 2020.

4.1. Factores que influyen en la demanda

Tal y como se ha comprobado en la realización del modelo de predicción de precios, la demanda instantánea de electricidad es uno de los factores que afectan de forma directa al precio final de la electricidad. La forma que se obtienen los precios de la electricidad es casando la oferta y la demanda para cada hora del día, tal y como se muestra en el ejemplo de la Figura 4.1, correspondiente a la hora 1 del día 23/01/2024.

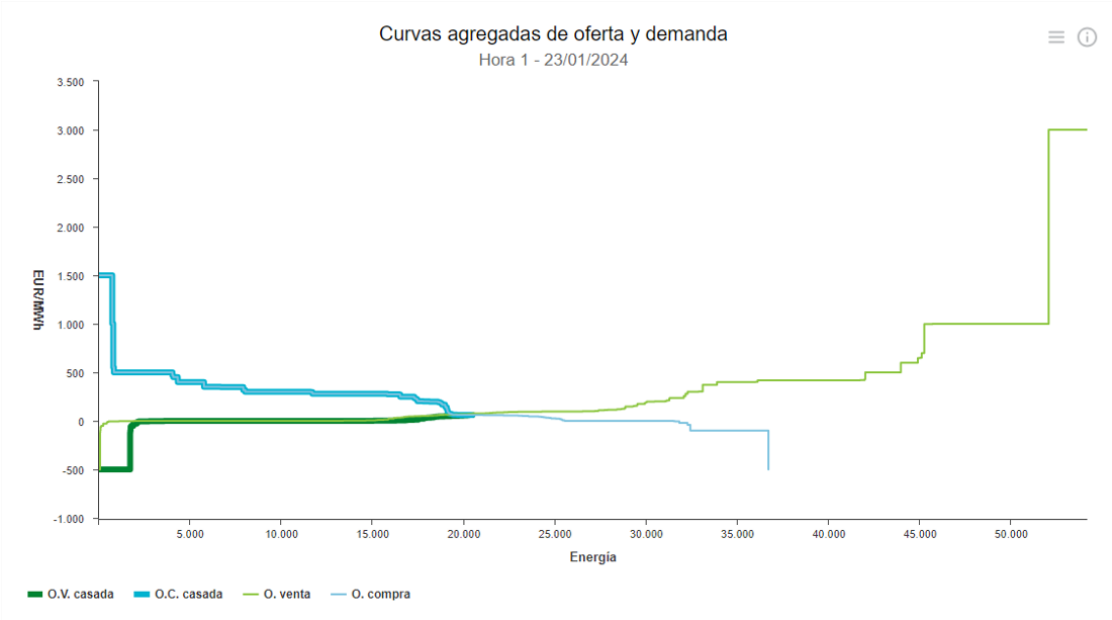

Figura 4.1: Curvas agregadas de oferta y demanda el 23/01/2024 en la hora 1. Fuente: OMIE

En la Figura 4.1, la curva azul representa la energía que los consumidores necesitan, así como el precio que están dispuestos a pagar por ella. Por otro lado, en verde se representa la energía que los generadores están dispuestos vender, con el precio al cual ofertan dicha energía. Se suman aquellas ofertas que tengan el mismo precio de venta, y se ordenan de forma ascendente. De la misma forma, se suman aquellas demandas que tengan el mismo precio de compra, pero en este caso se ordenan de forma descendente. El punto en el que cortan ambas curvas marca el precio al que se vende el total de la energía para esa hora. Es decir, a la izquierda de dicho punto queda el total de la energía que ha sido casada para esa hora.

Es por este motivo que se considera que la demanda es uno de los factores más significativos en el precio de la electricidad. Para el estudio de la demanda, se han identificado las siguientes variables que podrían influir en esta:

- **Temperatura**
- **Situación económica del país**
- **Festividades y fines de semana**

4.1.1. Temperatura

La temperatura es uno de los factores que más influye en la demanda. Algunas de las posibles razones por las que la demanda se ve tan influenciada en función de la temperatura que haga en el punto de consumo energético, se aprecian a continuación:

- La electricidad hoy en día se ha convertido en una de las principales fuentes para sistemas de HVAC. Cuando suben las temperaturas se emplean sistemas de aire acondicionado, ventiladores eléctricos, etc Por otro lado, cuando la temperatura es baja y se quiere aumentar en el interior de los edificios, también se emplean medios que demandan elevadas cantidades de energía eléctrica, como radiadores eléctricos, bombas de calor, sistemas de calefacción por suelos radiantes, etc Es por ello que, cuando las temperaturas son extremas, es decir, hay temperaturas elevadas o bajas, la demanda de electricidad se verá muy afectada.
- La eficiencia de algunos equipos que consumen elevados niveles de electricidad, pueden verse afectados debido a temperaturas elevadas, por lo que consumirán mayor energía para poder realizar el mismo trabajo.

Estos dos motivos son los que se consideran que más pueden afectar al aumento del consumo eléctrico debido a la temperatura.

A continuación, se muestra un análisis que se ha realizado en el que se observa cómo la temperatura afecta a la demanda de electricidad. Para realizar este análisis se han obtenido las temperaturas medias diarias de varias provincias repartidas por el territorio español, se ha calculado su media diaria y se ha comparado frente a la demanda en barras de control. Los datos de las demandas se han obtenido de la web de Esios de REE, y los datos de las temperaturas se han obtenido del Sistema de Información Agroclimática para el Regadío (SIAR). Las provincias de las que se han obtenido dichas temperaturas medias son:

- Provincia de Sevilla, estación meteorológica de Aznalcázar.
- Provincia de Huesca, estación meteorológica de Alcolea de Cinca.
- Provincia de Ávila, estación meteorológica de Losar del Barco.
- Provincia de Albacete, estación meteorológica de Albacete.
- Provincia de Madrid, estación meteorológica de Aranjuez.
- Provincia de Navarra, estación meteorológica de Arazuri.

- Provincia de Alicante, estación meteorológica de Almoradí.
- Provincia de Badajoz, estación meteorológica de Arroyo.
- Provincia de La Coruña, estación meteorológica de A Capela.
- Provincia de Murcia, estación meteorológica de Albanilla.

Comparando los datos de medias diarias de temperatura de las estaciones indicadas previamente, frente a los valores de demanda para el año 2023 entero, se obtienen los resultados que se muestran en la Figura 4.2.

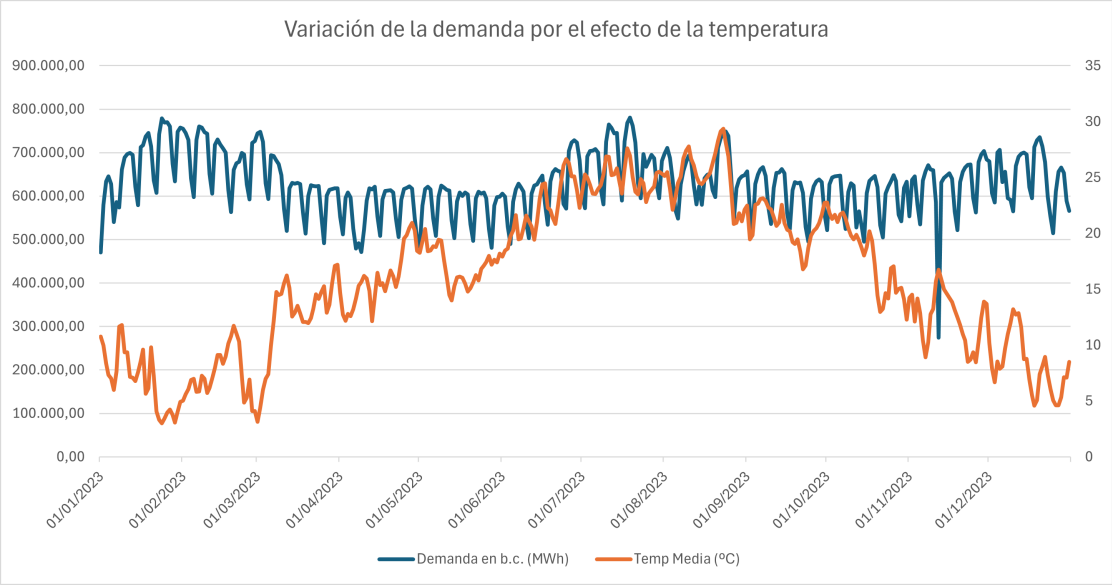

Figura 4.2: Comparación de la demanda frente a la temperatura en España en 2023. Fuente: elaboración propia. Datos demanda: REE. Datos temperatura: SIAR

En la figura anterior, se puede observar cómo durante los meses entre enero y marzo, en los cuales la temperatura media en el territorio español es menor, la demanda alcanza valores mayores que en los meses posteriores. Entre abril y junio, la demanda se estabiliza al alcanzar en el país temperaturas más estables. Durante los meses de verano, es decir, entre julio y septiembre en España se alcanzan las temperaturas medias más elevadas de todo el año y por ese motivo la demanda eléctrica aumenta de nuevo. Durante el otoño, la demanda se estabiliza de nuevo y finalmente durante los últimos meses del año 2023 se aprecia un nuevo ligero aumento de la demanda.

Puede resultar llamativa mínimo absoluto que se alcanza en la demanda el día

12/11/2023. El motivo de ese mínimo no se debe a la temperatura, ya que durante los días anteriores y posteriores se obtuvieron promedios de temperatura similares a los del día en cuestión y los promedios de las temperaturas máximas y mínimas también son similares. Se ha analizado también el resto de factores que se exponen en las siguientes secciones pero no se ha llegado a ninguna conclusión. Además, se ha hecho una investigación en publicaciones de noticias pero no se ha encontrado ninguna que explique esa caída en la demanda. Por este motivo, se considera que puede ser un error puesto que ya se han encontrado errores en otras ocasiones en los datos de REE.

4.1.2. Situación económica del país

A lo largo de los años, ha habido una gran relación entre la demanda de electricidad y el Producto Interior Bruto (PIB). En la Figura 4.3, se puede apreciar cómo hasta el año 2021 hay una correlación entre la demanda y el PIB en España, pero, sin embargo, a partir de este año se aprecia cómo el PIB sigue aumentando a medida que la demanda empieza a decrecer. Esto implica que, a pesar de que a lo largo de la historia han tenido una gran correlación, en los últimos años se ha perdido.

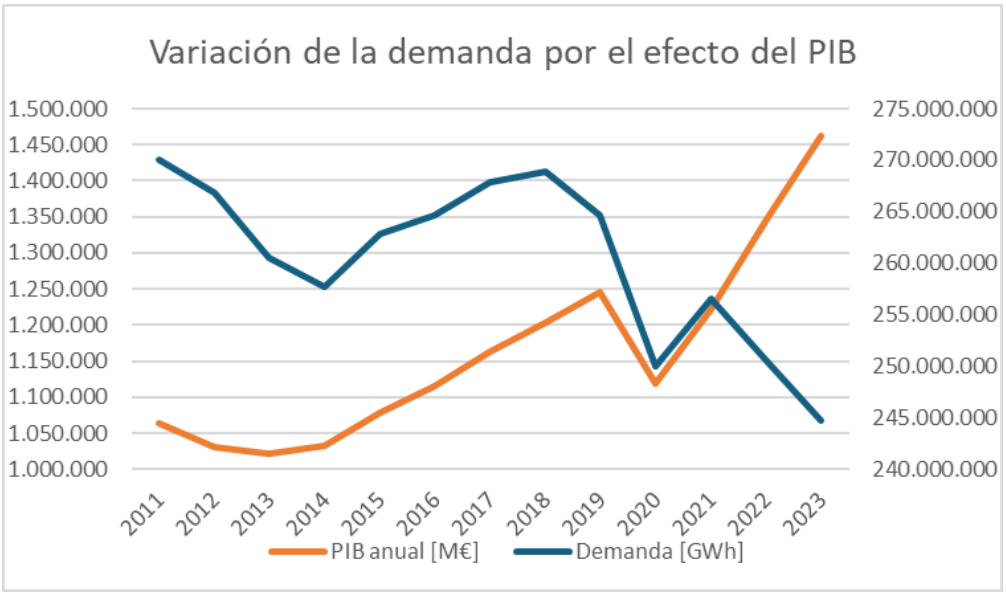

Figura 4.3: Comparación de la demanda frente al PIB en España e 2023. Fuente: elaboración propia. Datos demanda: REE

Los posibles motivos que hayan ayudado a romper esta correlación directa son:

- **Eficiencia energética**: la definición de eficiencia energética proporcionada desde un programa de la Unión Europea conocido como Support & Training

for an Excellent Energy Efficiency Performance (STEEEP), es la siguiente: "Se refiere a la utilización de tecnologías que requieren una menor cantidad de energía para conseguir el mismo rendimiento o realizar la misma función". Es decir, que son acciones que se emplean para conseguir reducir la cantidad de insumos de energía sin bajar el rendimiento. En los últimos años en la Unión Europea existen una serie de planes de eficiencia energética, como el Plan Nacional Integrado de Energía y Clima (2021-2030)(PNIEC), por el cual se establecen medidas de eficiencia energética en diferentes sectores.

- **Digitalización**: las tecnologías tienen la capacidad de aumentar la productividad de los recursos, logrando una mayor eficiencia con un aumento en el consumo de energía que es inferior al incremento en productividad obtenido [35]. Aunque la digitalización es intensiva en consumo eléctrico, y por tanto aumenta la demanda de electricidad también produce un aumento, en mayor medida, de la actividad económica y del PIB.

- **Autoconsumo**: según la Asociación de Empresas de Energías Renovables, la producción anual de energía generada por instalaciones de autoconsumo en España ha aumentado desde los 1.656 GWh que se produjeron en el año 2020, hasta los 7.262 GWh que se han producido en el año 2023. Esto implica que el autoconsumo es uno de los factores que ha provocado que haya disminuido la demanda eléctrica de la red.

4.1.3. Festividades y fines de semana

La demanda eléctrica se ve afectada por el día de la semana en el que se encuentre. Esta sigue un patrón semanal bastante constante durante los siete días de la semana: de martes a viernes son los días que mayor energía se demanda; los fines de semana la demanda disminuye; los lunes aumenta respecto al fin de semana pero no es tan elevada como el resto de días laborables porque la industria tarda en poner en marcha los procesos productivos [15].

En la Figura 4.4 se muestra un ejemplo de cómo disminuyó de forma considerada la demanda durante los fines de semana en el mes de febrero del año 2023. Se muestra en color azul los días laborables y en naranja los fines de semana.

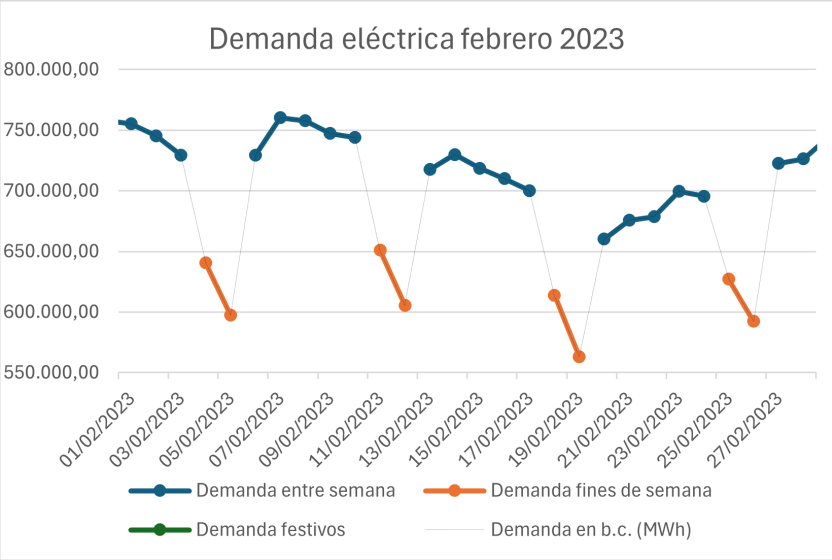

Figura 4.4: Variación de la demanda en febrero. Naranja:fines de semana. Azul: Días laborables. Fuente: elaboración propia. Datos demanda: REE

Un comportamiento similar tienen los días festivos, en los cuales baja también considerablemente la demanda. En las Figura 4.5 se observa el comportamiento de la demanda en un mes con varios días laborables, los cuales se muestran en verde:

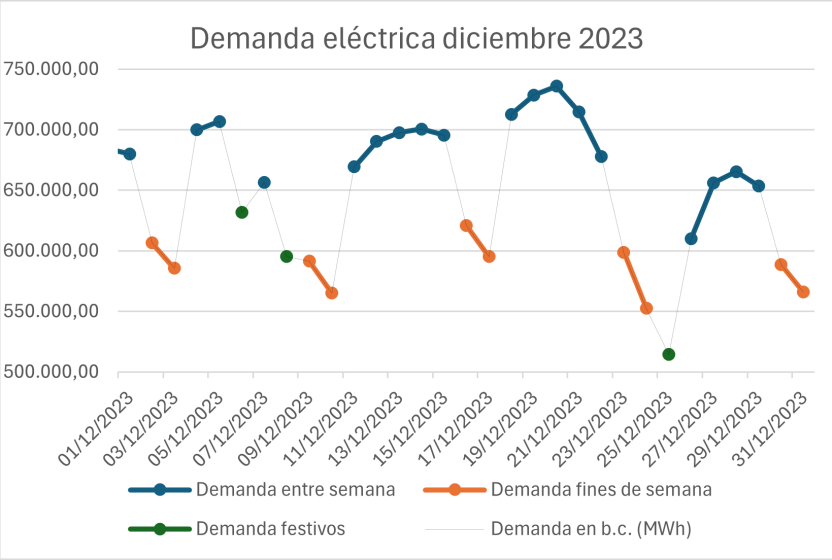

Figura 4.5: Variación de la demanda en diciembre. Naranja:fines de semana. Azul: Días laborables. Verde: Días festivos. Fuente: elaboración propia. Datos demanda: REE

4.2. Factores que influyen en el mix eléctrico

El mix eléctrico es el conjunto de tecnologías de generación que se emplean para la producción de energía eléctrica con el objetivo de cubrir la demanda en un país determinado [30]. Dentro de estas tecnologías no se considera aquella energía que se importa, únicamente aquella que es producida en el propio país.

La demanda eléctrica, tal y como se ha estudiado en la Sección 4.1, tiene una gran volatilidad, por lo que la generación debe ser lo suficientemente flexible para ajustarse a estos cambios, motivo por el que es preciso disponer de una variedad de tecnologías que lo permitan.

Según se define en el Anexo XI del RD 661/2007, dentro del mix de generación las tecnologías se pueden diferencias entre las gestionables y las no gestionables. En dicho Real Decreto se define a una tecnología de generación no gestionable como «aquella cuya fuente primaria que no es controlable ni almacenable y cuyas plantas de producción asociadas carecen de la posibilidad de realizar un control de la producción siguiendo instrucciones del operador del sistema sin incurrir en un vertido de energía primaria, o bien la firmeza de la previsión de producción futura no es suficiente para que pueda considerarse como programa»[19]. Son no gestionables los generadores de régimen especial que produzcan electricidad con tecnologías eólicas (onshore y offshore), solares (fotovoltaicas y térmicas), geotérmicas, undimotices (energía proveniente de las olas), mareomotrices (energía proveniente de las mareas), de las rocas calientes y secas, oceanotérmica, corrientes marinas y generadores hidráulicos fluyentes cuya potencia instalada sea menor o igual a 50 MW [19].

Debido a que la producción de dichas tecnologías de generación no gestionable depende de factores naturales que no son fáciles de prever cómo van a comportarse, es necesario que haya otras tecnologías que sean gestionables y, por lo tanto, capaces de dirigir cuándo deben o no producir energía. En este contexto, hay una serie de variables que afectan al mix de generación eléctrica cada día a cada hora. Las que se van a desarrollar en este documento son:

- **Climatología**
- **Costes asociados a cada tecnología**
- **Disponibilidad de las plantas**

4.2.1. Climatología

Tal y como se ha comentado en la introducción de la Sección 4.2, la climatología es uno de los principales factores que afectan a las tecnologías no gestionables y por lo tanto al mix de generación. Dentro de dichas tecnologías no gestionables existen dos que destacan por encima del resto en cuanto a su generación en España: la solar fotovoltaica y la eólica. Durante el año 2023, la energía eólica se ha consolidado como la principal fuente de producción de energía eléctrica a nivel nacional, suponiendo el 23,5 % de la generación total [13]. Por otro lado, la generación solar fotovoltaica ha sido la cuarta tecnología de generación que más energía ha producido durante este mismo año, con un 14 % de la generación total [13]. En la Figura 4.6, se muestra el mix de generación eléctrica de 2023 publicado por REE:

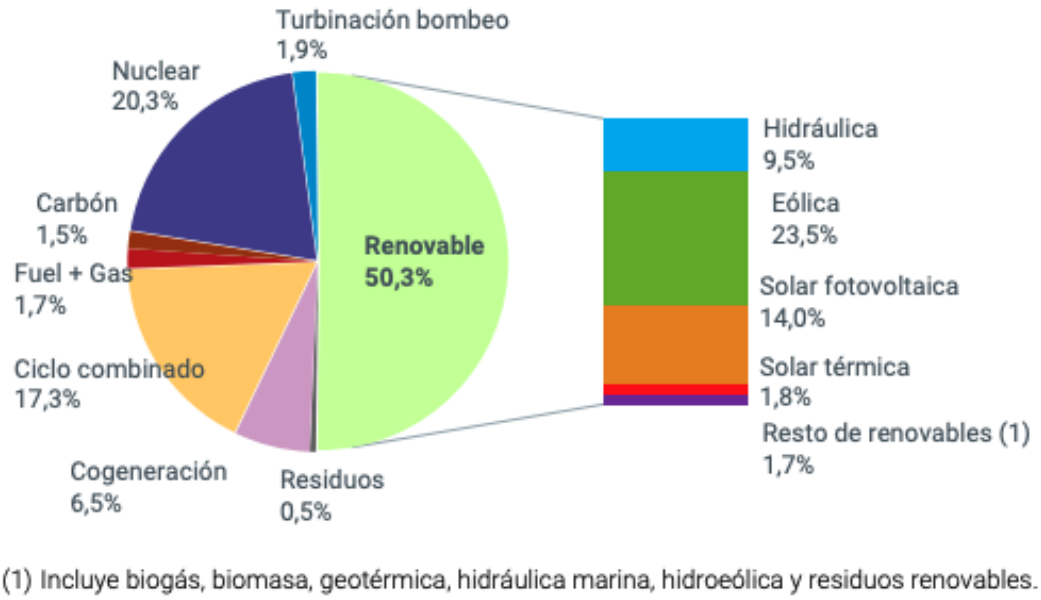

Figura 4.6: Mix de generación eléctrica en 2023. Fuente: REE

A lo largo de la Sección 4.2.1 se va a analizar por tanto el efecto que presenta la climatología sobre las dos tecnologías no gestionables más relevantes en España, la eólica y la solar fotovoltaica.

Para llevar a cabo el análisis de la influencia de un aspecto climatológico como es el viento, frente a la generación eólica, se ha recopilado la energía producida por dicha tecnología durante el 2023 en las Islas Canarias y se ha comparado frente a las velocidades medias del viento en dos puntos diferentes de la Comunidad Autónoma, como son Las Palmas y Tenerife. Los datos de las velocidades medias se

han obtenido del SIAR y los datos de la generación eólica de REE. Los resultados obtenidos para el mes de enero de 2023 se muestran en la Figura 4.7.

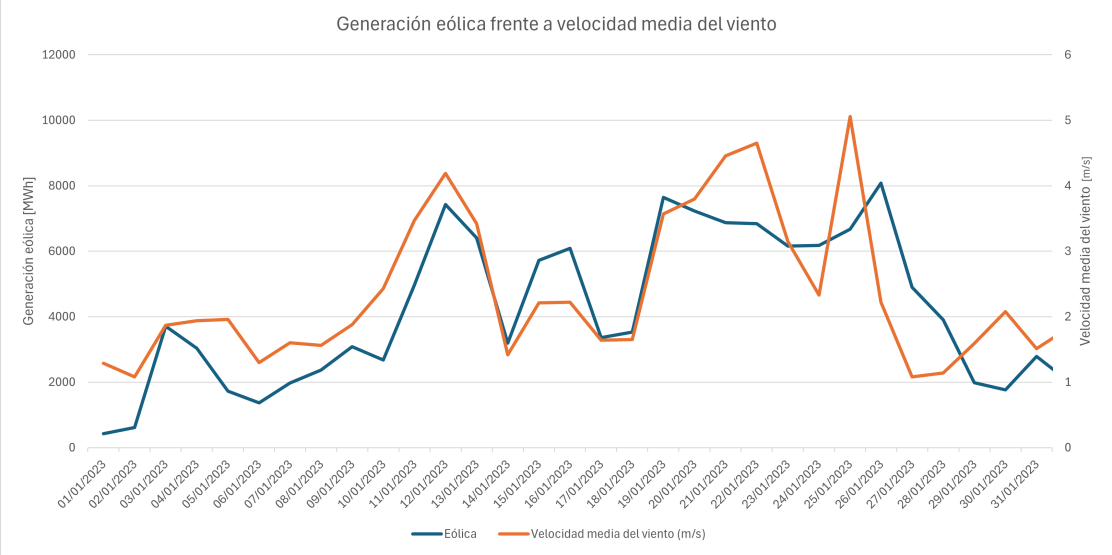

Figura 4.7: Comparación de la generación eólica frente a la velocidad media del viento en Canarias en enero de 2023. Fuente: elaboración propia. Datos generación: REE. Datos velocidad viento: SIAR

Se puede observar cómo existe una relación bastante directa entre la generación eólica y la velocidad media del viento. Esta relación no es perfecta debido a que hay otros factores que pueden influir en la generación eólica a parte de la velocidad media, como por ejemplo, las fluctuaciones del viento que pueden perjudicar a la hora de generar energía con esta tecnología. Es decir, cuanto más constante sea la velocidad del viento, más estable será la producción de energía.

Por otro lado, para el análisis de la influencia de otro aspecto climatológico, como es el sol, frente a la generación fotovoltaica se han recopilado los datos de radiación solar media durante todos los días de 2023 en las mismas dos provincias de Canarias que en la generación eólica, y se ha comparado frente a la generación fotovoltaica de toda la Comunidad Autónoma. Al igual que se ha comentado anteriormente, los datos de las radiaciones medias en MJ/m^2 se han obtenido del SIAR y los datos de la generación eólica de REE. Los resultados obtenidos para el mes de enero de 2023 se muestran en la Figura 4.8.

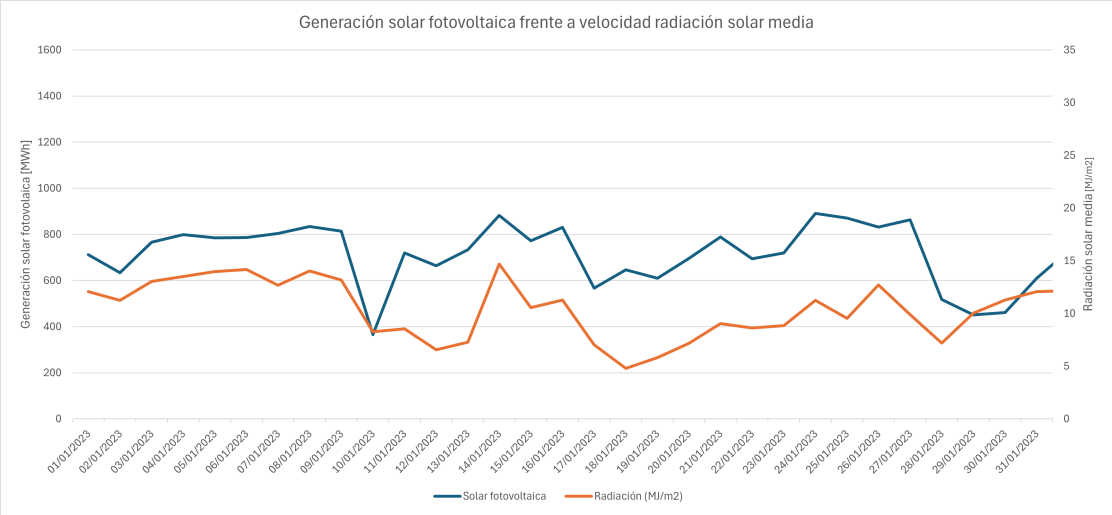

Figura 4.8: Comparación de la generación fotovoltaica frente a la radiación solar media en Canarias en enero de 2023. Fuente: elaboración propia. Datos generación: REE. Datos radiación solar media: SIAR

En este caso, se puede observar también una relación directa entre la generación fotovoltaica y la radiación solar. Al igual que ocurría en el caso de la generación eólica, esta relación no es perfecta ya que existen otros factores que influyen en la generación fotovoltaica, como es la inclinación de los rayos solares sobre los paneles fotovoltaicos y la temperatura que alcancen dichos paneles.

Una vez se ha estudiado la influencia de la climatología sobre la generación de energía eólica y solar fotovoltaica, se concluye que los factores climatológicos impactan de forma significativa en el mix de generación eléctrica.

4.2.2. Costes asociados a cada tecnología

El segundo de los factores, de los expuestos en la Sección 4.2, que influyen sobre el mix de generación son los costes de cada tecnología de generación asociados a la generación eléctrica. Cada fuente de producción presenta una estructura de costes diferente. Hay ciertas tecnologías de generación que cuentan con unos elevados costes de capital y unos bajos costes de operación, mientras que hay otras tecnologías que su situación es la contraria en la que las inversiones son menores pero sus costes de operación son altos. En la Tabla 4.1, se muestra una comparación cualitativa de las principales tecnologías de generación de electricidad en España [1] [10].

Tecnología	Coste inversión	Coste operación
Ciclo combinado	Bajo	Bajo
Carbón	Elevado	Medio
Nuclear	Elevado	Medio
Hidráulica	Muy elevado	Muy bajo
Eólica	Elevado	Muy bajo
Fotovoltaica	Medio[1]	Muy bajo
Cogeneración	Elevado	Medio

Tabla 4.1: Comparación cualitativa de las principales tecnologías de generación de electricidad en España

En un mix de generación ideal, aquellas tecnologías que tuviesen unos elevados costes de inversión pero unos bajos costes de operación deberían cubrir la carga base, en la que se mantiene un consumo bastante constante. Por otro lado, aquellas tecnologías con un coste de inversión bajo y de operación elevado deberían de cubrir aquello que se encuentre por encima de la carga base, es decir, la punta de la curva de carga. Sin embargo, se debe tener en cuenta que, además de los costes de operación, existen otros costes asociados a la producción de energía eléctrica que solo afecta a algunas de estas tecnologías de generación: los costes asociados a la compra de combustibles y los costes asociados a la compra de los derechos de emisión de CO_2. A continuación, se muestra un análisis de estos dos costes.

Costes de los combustibes

Este tipo de coste, como es evidente, no afecta a todas las tecnologías de generación expuestas anteriormente. Solo la nuclear, los ciclos combinados, la cogeneración y el carbón se ven impactados por estos costes. Este proyecto se centrará en la generación nuclear, la de los ciclos combinados y la cogeneración puesto que son aquellas que tienen un mayor impacto en el mix de generación eléctrica.

Energía nuclear

El principal combustible empleado para la generación de energía eléctrica en las centrales nucleares es el Uranio 235. Este es obtenido por lixiviación in situ, siendo necesario su procesado con el objetivo de reducir su tamaño y posteriormente realizar la molienda, obteniendo un polvo seco formado por uranio natural, cuya fórmula química es U_3O_8. Posteriormente debe ser convertido en hexafluoruro de

[1] Aunque la referencia de la que se ha obtenido esta información indicaba que este era muy elevado, el precio de los paneles fotovoltaicos desde 2007 se ha reducido en más de un 90 % y por ese motivo se ha bajado a medio [16]

uranio y finalmente enriquecido para obtener el uranio 235.

La evolución del precio del uranio durante los últimos años se puede apreciar en la Figura 4.9

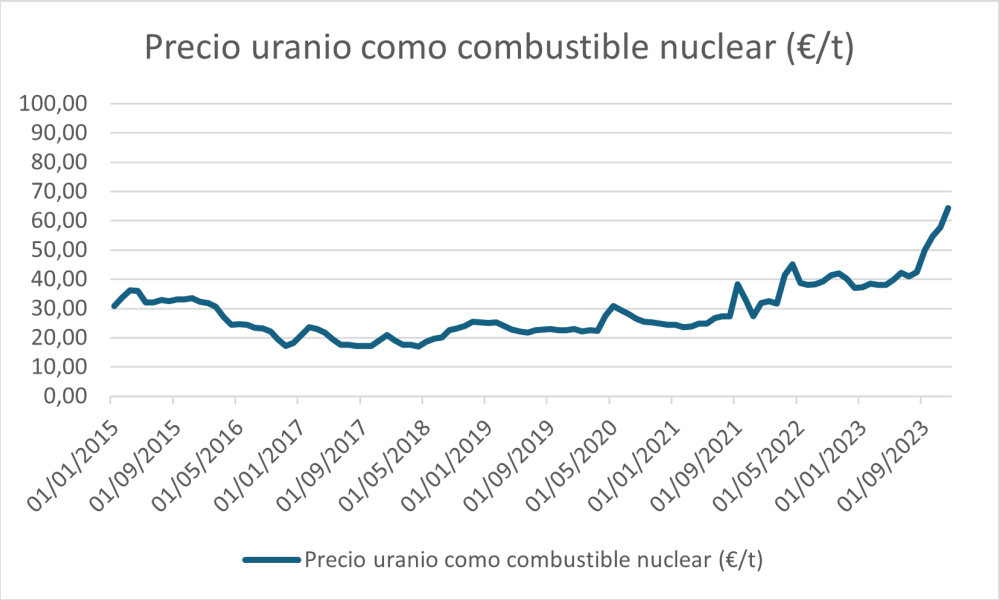

Figura 4.9: Precio del uranio como combustible nuclear entre 2015 y 2023. Fuente: elaboración propia. Datos precios: Trading Economics

Sin embargo, según la Sociedad Nuclear Española, el coste asociado al combustible empleado solo supone un 6 % de los costes de producción de energía proveniente de centrales nucleares. Es por ello que, a pesar de la gran volatilidad del precio del uranio, el coste de producción de electricidad no se muy afectado por el precio de este.

Energía de ciclos combinados y cogeneración

La principal materia prima empleada en los ciclos combinados y en la cogeneración es el gas natural. El gas natural es una mezcla homogénea de hidrocarburos entre los que destaca el metano con una proporción molar entre el 90 % y el 98 %, y donde el resto de hidrocarburos son etano, propano y butano, junto con otros componentes pesados [9]. Una de las principales ventajas que presenta el gas natural es su elevado poder calorífico, motivo por el que es empleado para la combustión en estas tecnologías de generación.

En Europa, existen diferentes mercados de gas natural como el PVB (España), TTF (Holanda), NBP (Reino Unido), etc. En el caso del PVB la evolución de los

precios de la cotización del Day Ahead entre los años 2021 y 2024 se pueden ver en la Figura 4.10.

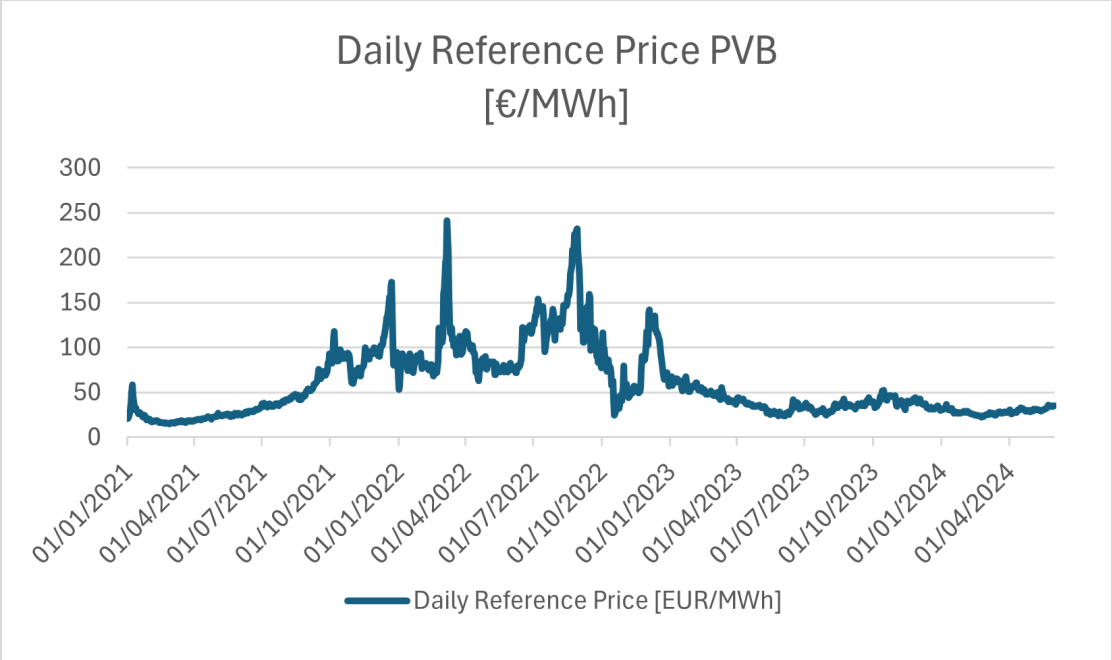

Figura 4.10: Precio gas natural PVB. Fuente: elaboración propia. Datos: MIBGAS

Para comparar el precio del gas natural frente a la generación a partir de este combustible, se ha representado en la Figura 4.11 los precios del PVB y la generación diaria de los ciclos combinados y la cogeneración del año 2022 en España, por ser el año con mayor volatilidad en los precios debido a la crisis energética. Sin embargo, el resultado de esta comparativa muestra que no existe una relación clara entre estas dos variables.

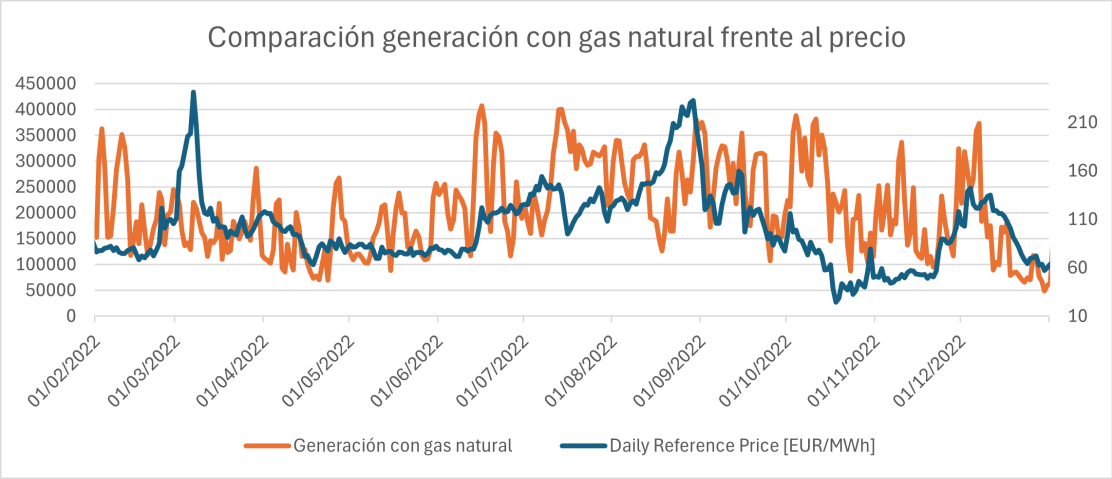

Figura 4.11: Comparación generación con gas natural frente al precio del gas. Fuente: elaboración propia. Datos generación: REE. Datos precio: MIBGAS

Los motivos que pueden explicarlo son:

- Hay una parte del gas, al igual que ocurre con la electricidad, que se compra en mercados a futuros, en los que los precios son diferentes a los mostrados en la Figura 4.11, los cuales son cotizaciones para el día siguiente.
- Los ciclos combinados son una tecnología que tienen una gran flexibilidad operativa, es decir, que son capaces de adaptarse a la demanda. Cuando una tecnología no gestionable no se comporta como estaba previsto, este tipo de tecnologías flexibles se ven ante la necesidad de aparecer para garantizar en todo momento el equilibrio entre generación y consumo. Con esto se concluye que este tipo de fuentes de producción son imprescindibles para ser capaces de cumplir dicho equilibrio, puesto que a pesar de los precios tan elevados, se seguía necesitando de ellos para abastecer de forma correcta toda la demanda eléctrica.

Costes de los derechos de emisión

Al igual que ocurría en la Sección 4.2.2, los costes asociados a la compra de derechos de emisión no afecta a todas las tecnologías de generación. En la Figura 4.12, se muestran las emisiones de CO2eq, diferenciadas por fuente de producción durante el año 2023. En ella se puede observar que las tres tecnologías que emitieron una mayor cantidad son los ciclos combinados, el carbón y la cogeneración.

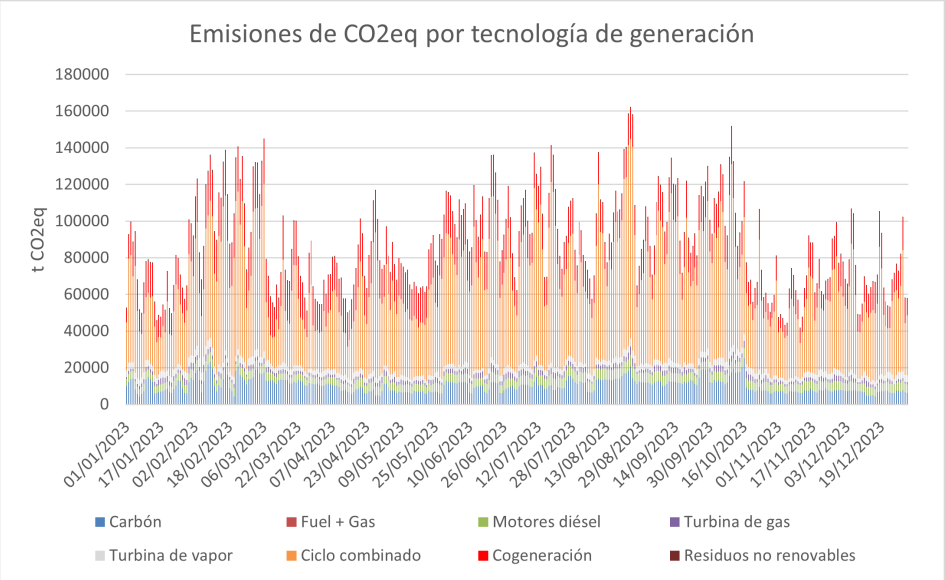

Figura 4.12: Emisiones de CO2eq por fuente de producción en España en 2023. Fuente: elaboración propia. Datos: REE.

Estas tecnologías, presentan la desventaja de tener que pagar una serie de derechos de emisión frente a otras que no emiten gases de efecto invernadero a la atmósfera a la hora de generar la electricidad. Para ello deben de acudir a los mercados del carbono. Estos son una herramienta empleada con el objetivo de reducir las emisiones de gases de efecto invernadero. Existen dos mercados principales de carbono:

- **Mercados regulados**: son los mercados en los que se impone una obligación de pagar por emitir. Se conocen también como sistemas de comercio de emisiones. Su funcionamiento se basa en un límite de emisiones establecido por alguna autoridad central que no puede superarse, siendo los derechos de emisión repartidos entre los diferentes sectores, ya sea de forma gratuita o mediante subastas.

- **Mercados voluntarios**: se tratan de mercados de acceso voluntario en los que se venden y compran "créditos de carbono".

Aunque originalmente la idea es que sean las propias empresas que contaminan las que paguen por el sobrecoste que implica emitir CO2eq a la atmósfera, está estudiado que realmente acaban trasladando estos importes en las facturas de electricidad y son los clientes finales los que pagan por ellos [14]. Este hecho implica que en las diferentes ofertas que hacen estas tecnologías en los mercados eléctricos incluyen un margen asociado al coste por estos derechos de emisión. Es por este

motivo por el que afecta el coste de los derechos de emisión al mix de generación, porque en aquellos momentos en los que estos tengan un precio más bajo los derechos, casarán una mayor cantidad de energía, mientras que cuando el precio suba, podrán casar menos energía por una oferta a un precio de venta más elevado [1].

Se muestra a continuación un ejemplo en el que se aprecia claramente cómo puede llegar a afectar los costes de los derechos de emisión sobre los costes de producción de un ciclo combinado. Los ciclos combinados tienen un factor de emisión de $0,37\frac{t\,CO_2}{MWh}$, y los costes de los derechos de emisión de CO_2 variaron entre diciembre de 2020 y junio de 2021 desde los $30,9\frac{€}{t\,CO_2}$ hasta los $52,8\frac{€}{t\,CO_2}$. Esto implica que los costes asociados a la producción de electricidad de los ciclos combinados en términos de derechos de emisión de CO_2 variaron un total de $8,103\frac{€}{MWh}$. Siguiendo el mismo procedimiento, se muestra en la Figura 4.13 la evolución de los costes asociados a los derechos de emisión para los ciclos combinados entre 2020 y 2023, resultado de la multiplicación de los precios de los derechos por el factor de emisión de los ciclos combinados. Los datos de las cotizaciones diarias de los derechos de CO_2 se han obtenido del Sistema Europeo de negociación de CO2 y se corresponden con los precios del mercado European Union Emissions Trading System (EU ETS), el cual es el mercado regulado actualmente vigente en la Unión Europea.

[1] Esta explicación se ha simplificado bajo el supuesto de que la demanda y las demás ofertas permanecen constantes en ambas situaciones, ya sea con un precio de los derechos de emisión más elevado o más bajo

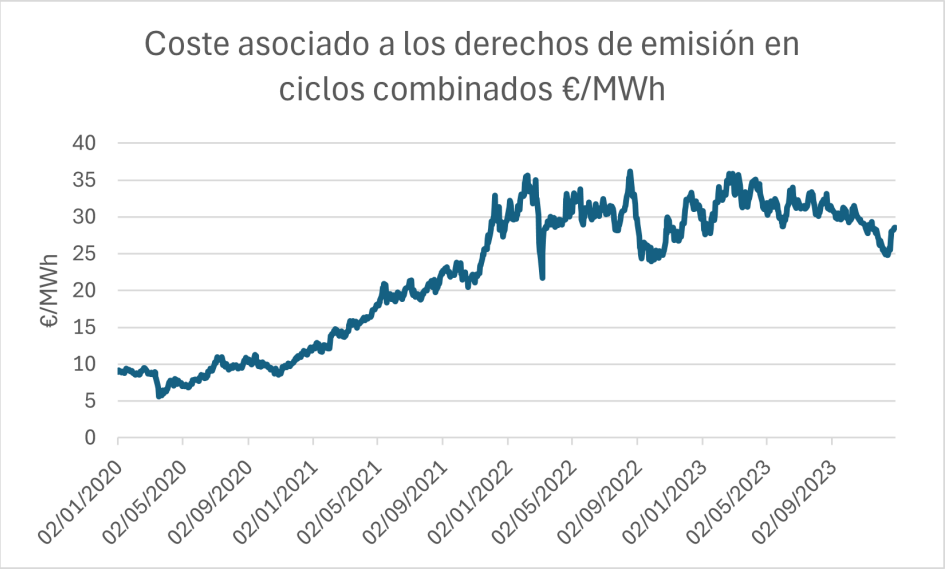

Figura 4.13: Coste asociado a los derechos de emisión entre 2020 y 2021. Fuente: elaboración propia

4.2.3. Disponibilidad de las plantas

La última de las variables que se va a analizar en este documento es la disponibilidad de las plantas de generación eléctrica para producir electricidad. Para ello, primero se va a definir lo que es la disponibilidad de una cierta tecnología de generación y posteriormente se estudiará un ejemplo de cómo, la disponibilidad de las centrales nucleares afectan al mix de generación.

La disponibilidad de operación se define como «el porcentaje del tiempo que una unidad está disponible para producir electricidad, ya sea que el sistema la necesite o no» [24]. Por otro lado, se presenta el concepto de potencia disponible como aquella que puede ser producida por un cierto tipo de tecnología en un instante determinado de tiempo. Cada tecnología cuenta a nivel territorial con una potencia instalada y una potencia disponible, la cual siempre es menor o igual que la potencia instalada.

Para llevar a cabo el análisis de cómo afecta la disponibilidad de las plantas de producción eléctrica al mix de generación, se va a estudiar cómo, cuándo la potencia disponible no es igual a la potencia instalada para ciertos tipos de tecnologías, hay otras que se ven afectadas y que deben producir en mayor cantidad. Este es el caso de las centrales nucleares, las cuales producen en índices de operación anuales

cercanos al 90 % según indica la empresa Nuclenor. Las centrales nucleares suelen disminuir su potencia disponible por mantenimientos o recargas de combustible. En la Figura 4.14, se muestra una gráfica con la comparativa entre la potencia instalada, la potencia disponible y la potencia media diaria demandada, calculada como la energía diaria producida entre las 24 horas del día.

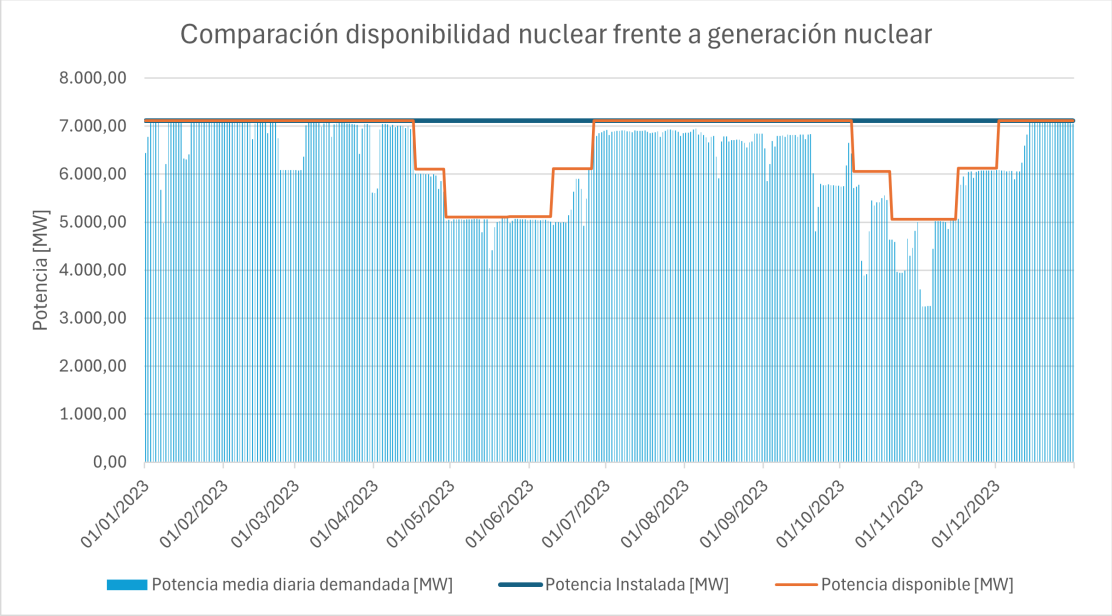

Figura 4.14: Comparación potencia nuclear instalada, frente a potencia nuclear disponible y potencia media diaria demandada en 2023. Fuente: elaboración propia. Datos generación: REE

Las variaciones de dichas potencias disponibles se deben a los motivos expuestos en la Tabla 4.2:

Centrales	Fecha inicio	Fecha fin	Motivo parada
Almaraz I	17/04/2023	23/05/2023	Recarga combustible
Ascó I	29/04/2023	15/06/2023	Recarga combustible
Ascó II	06/10/2023	10/11/2023	Recarga combustible
Cofrentes	23/10/2023	10/12/2023	Recarga combustible

Tabla 4.2: Paradas centrales nucleares en 2023

De esta forma, se observa cómo durante el año 2023, existen dos momentos diferentes en los que de forma simultánea hay dos centrales paradas. Durante estos periodos de tiempo en los que no se puede generar tanta energía como cuando toda la potencia instalada se encuentra disponible, la demanda debe ser cubierta con

otras fuentes de producción. Para una mejor visualización de los resultados, se va a aportar un mayor detalle del primer periodo afectado por las dos indisponibilidades simutáneas. Para ello, en la Figura 4.15, se muestra la comparación entre la potencia instalada, la potencia disponible y la potencia media diaria demandada, calculada como la energía diaria producida entre las 24 horas del día, durante los meses de abril a junio.

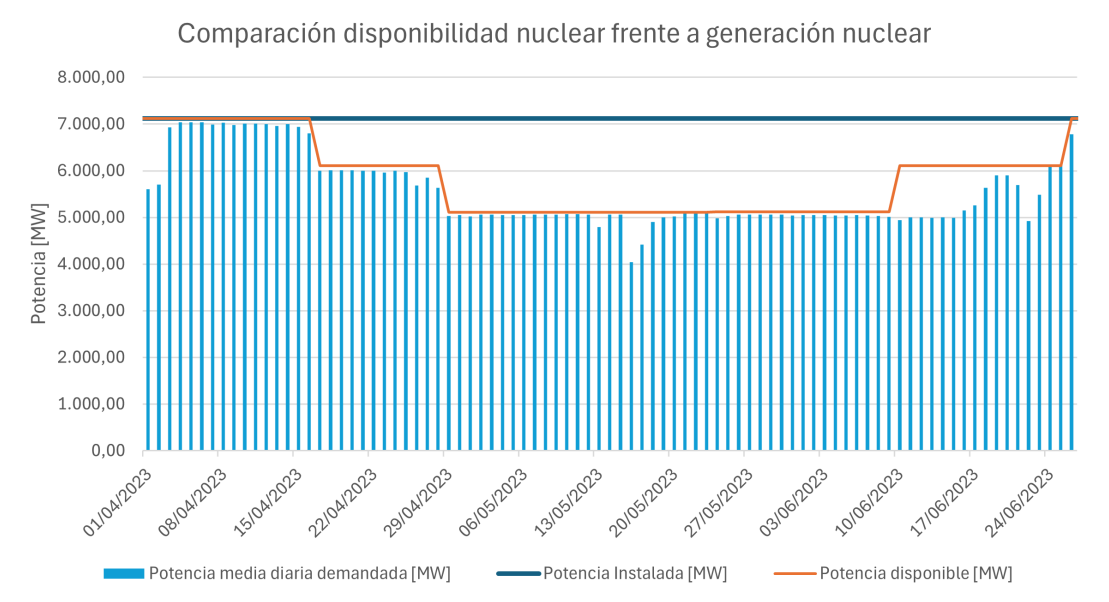

Figura 4.15: Comparación potencia nuclear instalada, frente a potencia nuclear disponible y potencia media diaria demandada entre abril y junio de 2023. Fuente: elaboración propia. Datos generación: REE

Tal y como se mostró en la Figura 4.6 de la Sección 4.2.1, las principales fuentes de generación en España son, además de la nuclear, la eólica, fotovoltaica, hidráulica, ciclo combinado y cogeneración. Por ese motivo, se va a estudiar a continuación cómo la disminución de la potencia disponible de la tecnología nuclear ha sido sustituida por estas otras fuentes de generación. En la Figura 4.16, se puede observar cómo, cuando las condiciones climatológicas lo han permitido, las tecnologías de generación renovables han podido compensar dicha reducción. Si bien es cierto que al ser tecnologías no gestionables (la hidraúlica lo es dependiendo de la potencia instalada), muy probablemente esta energía hubiese sido generada independientemente de si se hubiesen parado las nucleares o no.

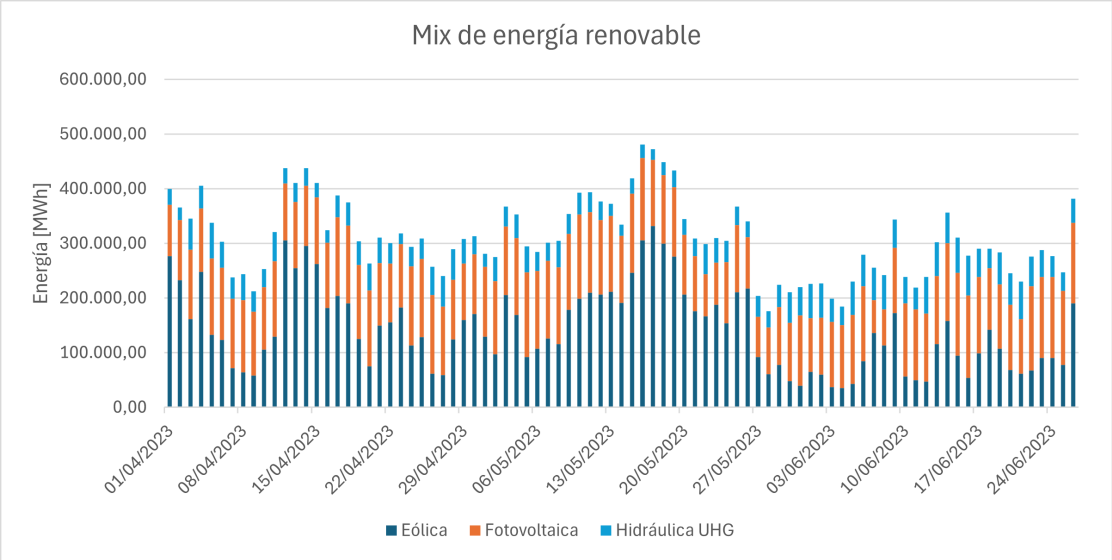

Figura 4.16: Mix de generación eólica, fotovoltaica e hidráulica entre abril y junio de 2023. Fuente: elaboración propia. Datos generación: REE

En la Figura 4.17 se aprecia cómo se ajustan las tecnologías de ciclos combinados y cogeneración, las cuales tienen en cuenta la disminución de la generación nuclear y la variabilidad de la renovable. De esta forma, se puede observar cómo durante la segunda mitad del mes de mayo de 2023, que hay una elevada generación renovable, disminuye la generación de los ciclos combinados y cogeneración, mientras que durante el mes de junio sucede lo contrario, disminuye la generación renovable y aumenta la de las tecnologías que emplean el gas como combustible.

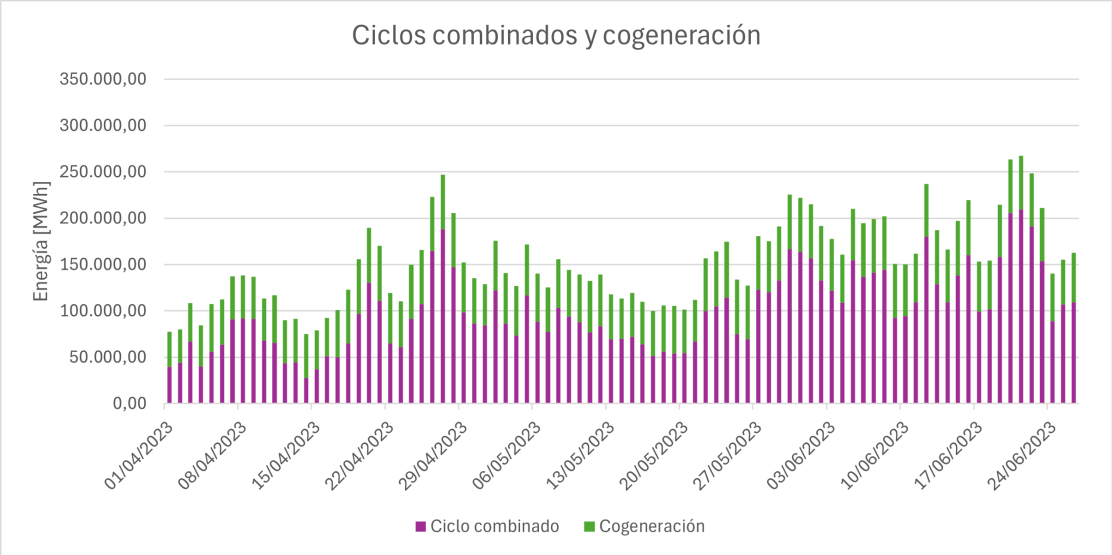

Figura 4.17: Mix de generación ciclo combinado y cogeneración entre abril y junio de 2023. Fuente: elaboración propia. Datos generación: REE

Se concluye que, se ha conseguido mostrar un ejemplo en el que la disponibilidad de ciertas tecnologías de generación afectan al mix eléctrico.

Capítulo 5

Conclusiones y líneas futuras de investigación

Los objetivos del presente Trabajo de Fin de Máster fueron fundamentalmente dos. El primero de ellos recaía en conocer más a fondo el mercado eléctrico español, estudiar la secuencia de los diferentes programas publicados por los operadores de mercado y de sistema, y entender cuál es la función de cada uno de ellos. Este conocimiento asentaba las bases para poder empezar a desarrollar el segundo de los objetivos, realizar un modelo de predicción del precio de la electricidad. De esta forma se ha conseguido integrar el campo de la ciencia de datos en el ámbito energético.

En cuanto a a los resultados obtenidos en el proyecto, de los cinco modelos de Machine Learning que se han empleado para predecir el precio de la electricidad, solo en dos de ellos se han conseguido resultados satisfactorios: Random Forest y XGBoost. No obstante esto era de esperar puesto que los modelos Baseline, Linear y Linear Ridge son más sencillos, teniendo una gran utilidad para analizar la complejidad del problema y añadir una cota superior al error máximo admisible. En otras palabras, si el problema a resolver hubiese sido sencillo, hubiesen aportado de la misma forma resultados buenos. En la Figura 5.1 se muestra la comparativa entre los diferentes modelos para los tres intervalos temporales estudiados. Además, en la Tabla 5.1 se encuentran los valores de los MAE representados en dicha figura en €/MWh.

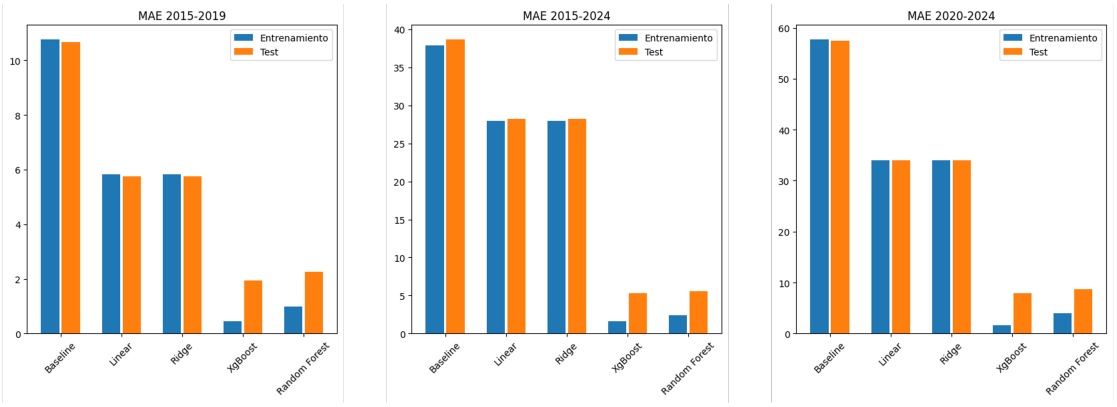

Figura 5.1: Comparación MAE cinco modelos. Fuente: elaboración propia

	Baseline		**Linear**		**Linear Ridge**		**Rand. For.**		**XGBoost**	
	Ent.	Test	Ent.	Test	Ent.	Test	Ent.	Test	Ent.	Test
2015 - 2019	10,75	10,67	5,82	5,77	5,82	5,76	0,98	2,25	0,45	1,94
2020 - 2024	57,67	57,47	33,99	34,06	33,99	34,06	4	8,76	1,58	7,87
2015 - 2024	37,87	38,63	27,93	28,21	27,93	28,21	2,38	5,55	1,6	5,26

Tabla 5.1: MAE (€/MWh) para los cinco modelos en los tres intervalos temporales

Haciendo hincapié en los resultados obtenidos en los modelos Random Forest y XGBoost, en ambos se cumple que los errores del intervalo temporal que transcurre de 2015 hasta 2024 son menores que en el que transcurre entre 2020 y 2024. Esto puede deberse a que la situación del mercado en 2024 se asemeja más a los años anteriores a la pandemia por el COVID-19 y la crisis energética. Comparando estos dos modelos entre sí, las diferencias son pequeñas para el conjunto de test, obteniendo en el periodo de 2015 a 2024 un resultado un 5,23 % mejor el XGBoost que el Random Forest. Por otro lado, puede llamar la atención la diferencia entre los errores obtenidos durante el intervalo temporal entre el 2015 y 2019 y los otros dos. Sin embargo, si se comparan de forma porcentual los MAE de cada intervalo frente a los valores promedios del precio de la electricidad en ese periodo, las diferencias son menores. Esto se puede observar en la Tabla 5.2.

	Promedio	MAE Random Forest		MAE XGBoost	
	€/MWh	Entrenamiento	Test	Entrenamiento	Test
2015 - 2019	49,44	1,98 %	4,55 %	0,91 %	3,92 %
2020 - 2024	94,02	4,25 %	9,32 %	1,68 %	8,37 %
2015 - 2024	70,75	3,38 %	7,89 %	2,27 %	7,48 %

Tabla 5.2: Errores porcentuales frente al promedio de mercado para los modelos Random Forest y XGBoost

Habiendo probado la validez del modelo y de las características seleccionadas para él, se realizó un análisis de los factores que influyen en estas variables de entrada. Para ello, se estudió la repercusión de factores como la temperatura, la situación económica del país y las festividades y fines de semana en la demanda; y de la climatología, los costes asociados a cada tecnología y la disponibilidad de las plantas al mix de generación eléctrica.

Este análisis concluyó en que la temperatura, las festividades y los fines de semana tienen un impacto directo sobre la demanda eléctrica. Se observó cómo cuando las temperaturas alcanzan valores extremos, tanto cálidos como fríos, la demanda eléctrica aumenta, mientras que cuando las temperaturas son medias, la demanda permanece estable. Por otro lado, durante los fines de semana y festivos se puede apreciar cómo la demanda sufre un descenso frente al resto de días de la semana. En cuanto a la situación económica del país, la relación entre ambas variables era clara antes del 2021, pero debido a factores como la eficiencia energética, la digitalización y el autoconsumo han quebrado dicha relación hasta llegar a un momento en el que el PIB español hasta entre 2021 y 2023 aumentó, mientras que la demanda decreció.

En cuanto a las factores que influyen en el mix de generación eléctrica, se concluyó que los tres factores estudiados tienen una repercusión sobre el mix energético. La climatología afecta en gran medida a tecnologías de generación renovable con aspectos como la velocidad media del viento o la radiación solar media diaria. Por otro lado, la indisponibilidad de algunas plantas obligan a otras a producir la energía que no puede ser generada por estas. Es el caso de las nucleares, que asumen una carga base constante y cuando no tienen disponible la potencia instalada total son otras tecnologías las encargadas de producir esa diferencia de energía. Finalmente, en cuanto a los costes asociados a cada tecnología, se ha estudiado la repercusión que tienen los costes de los combustibles y de la compra de derechos de emisión. Cuando estos costes son elevados, las tecnologías que deben pagar por ellos ofertan a precios más elevados, pudiendo casar una menor cantidad de energía.

Como cierre del proyecto, se ha desarrollado en el Anexo A el alineamiento del proyecto con los Objetivos de Desarrollo Sostenible (ODS) de las Naciones Unidas. En él se pueden encontrar los motivos por los que este Trabajo de Fin de Máster apoya algunos de los 17 ODS.

En cuanto a las líneas futuras de investigación para este Trabajo de Fin de Máster, se propone:

- Sustituir las características empleadas en este proyecto por los factores estudiados que las afectan. Este cambio añade complejidad al modelo, además que puede aportar resultados más robustos ya que se estarían considerando un mayor número de características y podría hacer frente de mejor manera a posibles cambios en la tendencia del mercado como ya ocurrió años atrás. No obstante, tener capacidad para acceder a todos esos datos puede suponer un reto adicional.

- Ampliar la malla de búsqueda de hiperparámetros. Tal y como se comentó cuando se explicó el concepto de búsqueda de hiperparámetros, esto es un compromiso entre mejora de resultados y carga computacional. Al aumentar la malla, son mayores las posibles combinaciones de hiperparámetros que podría probar el modelo, pudiendo conseguir resultados iguales o mejores a los logrados en este proyecto. Sin embargo, esto aumentaría el tiempo de compilación del modelo y los recursos del ordenador en el que se ejecute.

- Corregir el problema de overfitting obtenido en los modelos Random Forest y XGBoost. Aunque este problema no es muy significativo en el caso de los resultados de este modelo, se podría buscar otras vías para penalizar la diferencia entre los MAE obtenidos entre el conjunto de entrenamiento y de test.

- Emplear otros modelos de Machine Learning más modernos que pueden ofrecer resultados más robustos como las redes neuronales.

Para concluir, este proyecto sirve para cerrar una etapa de mucho trabajo y aprendizaje, y a su vez abre un gran abanico de puertas que se afrontan con entusiasmo y sobre las que se pretende continuar desarrollando proyectos como este. Este modelo será de aplicación a nivel empresarial en la entidad colaboradora que lo ha apoyado y se seguirá desarrollando con el objetivo de conseguir errores menores.

Apéndice A

Alineamiento del Proyecto con los ODS

Los líderes mundiales propusieron 17 Objetivos de Desarrollo Sostenible ODS con el propósito de erradicar la pobreza, proteger al planeta y asegurar prosperidad para todos el pasado 25 de septiembre de 2015. En la Figura A.1, se muestran los 17 ODS establecidos por las Naciones Unidas dentro de la Agenda 2030.



Figura A.1: Objetivos de Desarrollo Sostenible de las Naciones Unidas. Fuente: Web Naciones Unidas

El Trabajo de Fin de Máster que se ha llevado a cabo se encuentra alineado con

varios de los ODS mostrados anteriormente. A lo largo de este anexo se procede a analizar las aportaciones del proyecto a los ODS.

A.1. ODS 7: Energía asequible y no contaminante

El séptimo ODS pretende «garantizar el acceso a una energía, limpia y asequible, clave para el desarrollo de la agricultura, las empresas, las comunicaciones, la educación, la sanidad y el transporte» [39].

Mediante el modelo de predicción de precios desarrollado en este trabajo, se puede conocer con anterioridad cómo van a evolucionar los precios del mercado diario a corto plazo, pudiendo de esta forma hacer una planificación y gestión de los recursos energéticos, ya sea a nivel empresarial, doméstico, etc. En otras palabras, al conocer el comportamiento del mercado eléctrico en el corto plazo, es posible planificar los consumos eléctricos de tal forma que se vaya a consumir una mayor energía cuando los precios sean más bajos.

A.2. ODS 9: Industria, innovación e infraestructura

El noveno ODS pretende «construir infraestructuras resilientes, promover la industrialización sostenible y fomentar la innovación» [39].

En los últimos años el avance de las tecnologías ha permitido disponer de una gran cantidad de datos de forma instantánea. Esto facilita la utilización de técnicas avanzadas de Aprendizaje Automático para usos como el desarrollado en este proyecto, un modelo de predicción del precio de la electricidad. En estos modelos, cuanto mayor y de una calidad superior sea la información, el resultado va a ser mejor.

Por otro lado, el hecho de poder predecir el precio de la electricidad a corto plazo permite que, en relación con el ODS 7 previamente analizado, se puedan crear infraestructuras energéticas más inteligentes que sean capaces de decidir cuándo es el mejor momento para consumir electricidad y de esta forma obtener energía a un menor precio o más limpia.

A.3. ODS 12: Producción y consumo responsables

El duodécimo ODS prentende «garantizar modalidades de consumo y producción sostenibles, algo fundamental para sostener los medios de subsistencia de las generaciones actuales y futuras» [39].

Al haber realizado el análisis de las variables que influyen en el precio de la electricidad, se puede conocer cómo las diferentes tecnologías de generación tienen un impacto en el precio final de la electricidad. De esta forma, se pueden tomar decisiones a nivel gubernamental que permitan favorecer aquellas acciones que beneficien una energía más limpia, garantizando en todo momento un suministro eléctrico seguro, confiable y accesible.

A.4. ODS 13: Acción por el clima

El decimotercer ODS pretende «adoptar medidas urgentes para combatir el cambio climático y sus efectos» [39].

Tal y como se ha mostrado en la Sección 4.2.2, existen algunas tecnologías que emiten una gran cantidad de gases de efecto invernadero a la atmósfera. De hecho, según indica el Ministerio para la transición ecológica y el reto demográfico en un informe emitido en marzo de 2023, el sector de generación eléctrica es el cuarto sector que más gases de efecto invernadero emitió a la atmósfera en 2021 en España, contabilizando un 10,8 % de estos.

El conflicto que existe hoy en día con las tecnologías de generación que contaminan es la dependencia que se tiene para poder garantizar el suministro eléctrico a un precio razonable sin ellas. Si se quisiese reducir la emisión de gases de efecto invernadero se podría aumentar la potencia instalada de fuentes renovables, que no contaminan en la producción de electricidad, a costa de cerrar plantas de, por ejemplo, ciclos combinados. Sin embargo, esto podría incurrir en conflictos con el ODS 7, porque no se aseguraría el suministro eléctrico. Una de las posibles soluciones sería la aparición del almacenamiento de energía, de tal forma que, para aquellos momentos que no se pueda producir electricidad a partir de fuentes renovables, se emplee la energía almacenada en vez de emplear fuentes contaminantes.

Bibliografía

[1] International Energy Agency. *Power generation investment in electricity markets*. OECD Publishing, 2003.

[2] Luis Agosti, A Jorge Padilla y Alejandro Requejo. «El mercado de generación eléctrica en España: estructura, funcionamiento y resultados». En: *Economía industrial* 364 (2007), págs. 21-37.

[3] C.M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer New York, 2016, págs. 663-665. ISBN: 9781493938438. URL: https://books.google.es/books?id=kOXDtAEACAAJ.

[4] Boletín Oficial del Estado. *Disposición 15049 del BOE núm. 253 de 2019*. Oct. de 2019.

[5] Boletín Oficial del Estado. *Disposición General del BOE núm. 24 de enero de 2020*. Ene. de 2020.

[6] Boletín Oficial del Estado. *Orden TED/1253/2023*. Nov. de 2023.

[7] José Octavio Bonilla Navarro et al. «El mercado eléctrico». En: (2023).

[8] Alberto Carbajo. «Los mercados eléctricos y los servicios de ajuste del sistema». En: *Red Elétrica de España* (2007), págs. 55-62.

[9] G Carrillo y P Guerrero. «Cálculo de propiedades del gas natural». En: *Trabajo especial de grado en opción al título de Ingeniero Químico. Maracaibo* (2013).

[10] Fidel Castro-Rodríguez. «La configuración del mix tecnológico en un sistema eléctrico liberalizado». En: *Econ Ind* 364 (2007), págs. 75-86.

[11] Tianqi Chen y Carlos Guestrin. «Xgboost: A scalable tree boosting system». En: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016, págs. 785-794.

[12] Issam El Naqa y Martin J Murphy. *What is machine learning?* Springer, 2015.

[13] Red Electrica. «Informe del sistema eléctrico-informe resumen de energías renovables 2023». En: (2024).

[14] El Periódico de la Energía. *Qué son los derechos de emisión de CO2 y cómo afectan al precio de la luz*. 2021. URL: https://elperiodicodelaenergia.com/que-son-los-derechos-de-emision-de-co2-y-como-afectan-al-precio-de-la-luz/.

[15] Red Eléctrica Española. *Atlas de la demanda eléctrica española*, págs. 25-40.

[16] Cayetano Espejo-Marín y Ana Eulalia Aparicio-Guerrero. «La producción de electricidad con energía solar fotovoltaica en España en el siglo XXI». En: *Revista de Estudios Andaluces, 39, 66-93.* (2020).

[17] Javier Jesús Espinosa-Zúñiga. «Aplicación de algoritmos Random Forest y XGBoost en una base de solicitudes de tarjetas de crédito». En: *Ingeniería, investigación y tecnología* 21.3 (2020).

[18] Boletín Oficial del Estado. *Real Decreto 1339/1999*. Ago. de 1999.

[19] Boletín Oficial del Estado. *Real Decreto 661/2007*. Mayo de 2007.

[20] J.L.D. García y Universidad de Sevilla. Escuela Técnica Superior de Ingenieros. *Impacto de las tecnologías de generación en el mercado eléctrico [Recurso electrónico]: Proyecto Fin de Carrera*. 2013. Cap. 3. URL: https://books.google.es/books?id=DxryjgEACAAJ.

[21] Marta García Ruiz de León. «Análisis de Sensibilidad Mediante Random Forest». En: (2018).

[22] Álvaro Garnica Barco. «Regularización en problemas lineales de regresión: Una visión estadística y económica aplicada a datos de desarrollo humano». B.S. thesis. Universitat Politècnica de Catalunya, 2021.

[23] A. Géron. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, 2019, págs. 75-83. ISBN: 9781492032595. URL: https://books.google.es/books?id=HnetDwAAQBAJ.

[24] Paúl Roberto Gutiérrez Gómez. «Propuesta de definición de términos de confiabilidad para centrales de generación, líneas de transmisión y redes de distribución». B.S. thesis. QUITO/EPN/2000, 2000.

[25] Juan Yago Hernández-Canut Cano. «LA REALIZACIÓN DE LA LIBRE COMPETENCIA EN EL SECTOR ELÉCTRICO ESPAÑOL». En: (2019).

[26] Jaime Pilatásig Lasluisa. «Resiliencia de Sistemas Eléctricos de Potencia mediante la Conmutación de Líneas de Transmisión–Estado del arte». En: *I+ D Tecnológico* 16.2 (2020), págs. 67-77.

[27] CEOE Castilla y León. *Acceso al mercado y optimización de la factura eléctrica y de gas natural*. 2020.

[28] Ernestina Menasalvas. «Algoritmos de Machine Learning». En: (2021).

[29] Comisión Nacional de los Mercados y la Competencia. «¿Quién es quién en el sector eléctrico?» En: *La organización del sector eléctrico*.

[30] Gonzalo Rodrigo Navarro Armijo. «La evolución y eficiencia del mix eléctrico ante la penetración de las energías renovables en Alemania, Canadá, Dinamarca, España, Francia, Italia, y Noruega». En: (2022).

[31] Alejandro Nieto Jeux. «Algoritmos de aprendizaje automático: un estudio de su difusión y utilización». En: (2021).

[32] Dirección del Operador de Mercado. «Detalle del uncionamiento Mercado Intradiario». En: (), págs. 1-9.

[33] Dirección del Operador de Mercado. «Funcionamiento Mercado Diario». En: (), págs. 1-5.

[34] Consejo de Reguladores del MIBEL. «Power Purchase Agreements». En: (2024), págs. 1-26.

[35] Tilman Santarius, Johanna Pohl y Steffen Lange. «Digitalization and the decoupling debate: can ICT help to reduce environmental impacts while the economy keeps growing?» En: *Sustainability* 12.18 (2020), pág. 7496.

[36] Energía y Sociedad. «Manual de la energía». En: (2023), págs. 10-13.

[37] M.R. Spiegel. *Probabilidad y estadística*. Serie de compendios Schaum. McGraw-Hill Interamericana de España S.L., 2010, págs. 258-262. ISBN: 9786071502704.

[38] Richard Taylor. «Interpretation of the correlation coefficient: a basic review». En: *Journal of diagnostic medical sonography* 6.1 (1990), págs. 35-39.

[39] United Nations. *17 objetivos para transformar nuestro mundo*. 2024. URL: `https://www.un.org/sustainabledevelopment/es/sustainable-development-goals/`.

[40] Gema Valenzuela González et al. «Aprendizaje Supervisado: Métodos, Propiedades y Aplicaciones». En: (2022).