



UNIVERSIDAD PONTIFICIA COMILLAS

TRABAJO DE FIN DE MÁSTER

Modelos de Predicción de Restricciones Técnicas

Autor

Marcos Ventosa Pontes

Supervisores

David Soler Soneira

Andrés Fernández Ramos

29 de agosto de 2021

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título
Modelos de Predicción de Restricciones Técnicas en la ETS de Ingeniería - ICAI
de la Universidad Pontificia Comillas en el
curso académico 2020-2021 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.: Marcos Ventosa Pontes

Fecha: 28/08/2021



Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO

Fdo.: David Soler Soneira

Fecha: 28/ 08/ 2021



Andrés Fernández Ramos



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

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

El autor D.Marcos Ventosa Pontes

DECLARA ser el titular de los derechos de propiedad intelectual de la obra: Modelos de Predicción de Restricciones Técnicas, que ésta es una obra original, y que ostenta la condición de autor en el sentido que otorga la Ley de Propiedad Intelectual.

2º. Objeto y fines de la cesión.

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

3º. Condiciones de la cesión y acceso

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

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

4º. Derechos del autor.

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

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

5º. Deberes del autor.

El autor se compromete a:

- a) Garantizar que el compromiso que adquiere mediante el presente escrito no infringe ningún derecho de terceros, ya sean de propiedad industrial, intelectual o cualquier otro.
- b) Garantizar que el contenido de las obras no atenta contra los derechos al honor, a la intimidad y a la imagen de terceros.
- c) Asumir toda reclamación o responsabilidad, incluyendo las indemnizaciones por daños, que pudieran ejercitarse contra la Universidad por terceros que vieran infringidos sus derechos e

intereses a causa de la cesión.

- d) Asumir la responsabilidad en el caso de que las instituciones fueran condenadas por infracción de derechos derivada de las obras objeto de la cesión.

6º. Fines y funcionamiento del Repositorio Institucional.

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

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

Madrid, a 28 de agosto de 2021

ACEPTA

Fdo.....



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

--



UNIVERSIDAD PONTIFICIA COMILLAS

TRABAJO DE FIN DE MÁSTER

Modelos de Predicción de Restricciones Técnicas

Autor

Marcos Ventosa Pontes

Supervisores

David Soler Soneira

Andrés Fernández Ramos

29 de agosto de 2021

Resumen

Modelos de predicción de Restricciones Técnicas

Autor: Ventosa Pontes, Marcos

Directores: Soler Soneira, David y Fernández Ramos, Andrés

Entidad Colaboradora: Invesyde

Palabras clave: Restricciones Técnicas, Machine Learning, Predicción

Este trabajo pertenece a un proyecto más grande, cuyo objetivo final es el arbitraje entre mercados diario e intradiario. Para poder hacerlo es necesario predecir el Δ de precio entre ambos mercados. La predicción del volumen de restricciones técnicas debe ser clave para anticipar las diferencias entre estos mercados al calcularse las restricciones entre ambos mercados. El alcance del proyecto global se puede ver en la figura 1.

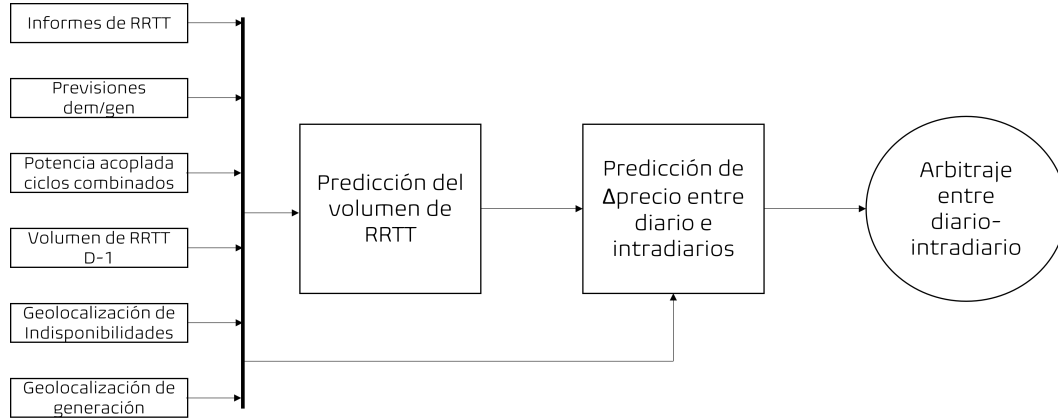


Figura 1: Proyecto global

Este trabajo se centra en la extracción, análisis y procesamiento de las variables explicativas que alimentaran tanto el modelo de predicción del volumen de restricciones como el modelo de predicción del Δ de precio entre diario e intradiario. Estas variables explicativas son los informes de Restricciones Técnicas que se publican cada día, las previsiones de demanda y generación, la potencia acoplada de los ciclos combinados, el volumen de Restricciones Técnicas del día anterior, la geolocalización de las Indisponibilidades de Red y una estimación de la geolocalización de la generación dependiente del clima. También se elaboran modelos de

Machine Learning como RF, KNN, RF y MLP usando gran parte de las variables explicativas, tal como se observa en la figura 2

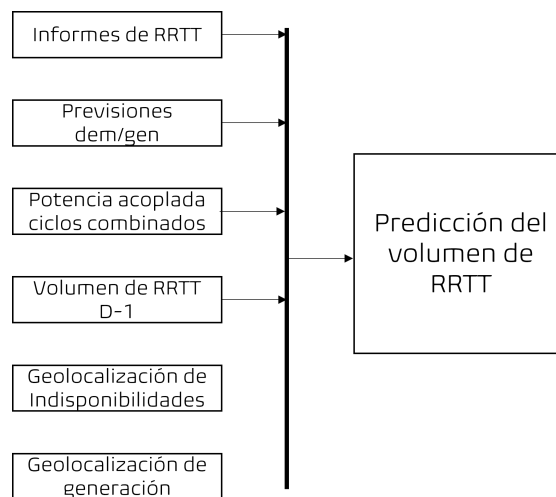


Figura 2: Alcance del proyecto

Los modelos entrenados usan *Online Learning*, técnica que va re-entrenando el modelo según pasa el tiempo y se puede usar cuando el tamaño del conjunto de datos aumenta secuencialmente, como en nuestro caso que cada día añadimos una fila a nuestro conjunto. De esta forma nuestros modelos para cada día predicen el valor de energía de restricciones de un día entrenándose con todos los días anteriores y para la predicción del día siguiente el modelo añade los datos del día anterior al conjunto de entrenamiento, se re-entrena y se hace la predicción.

La evolución de la variable a predecir (Energía por Restricciones Técnicas) se puede ver en la figura 3. Esta variable cuenta con una volatilidad muy alta, lo que hace que su predicción resulte más compleja.

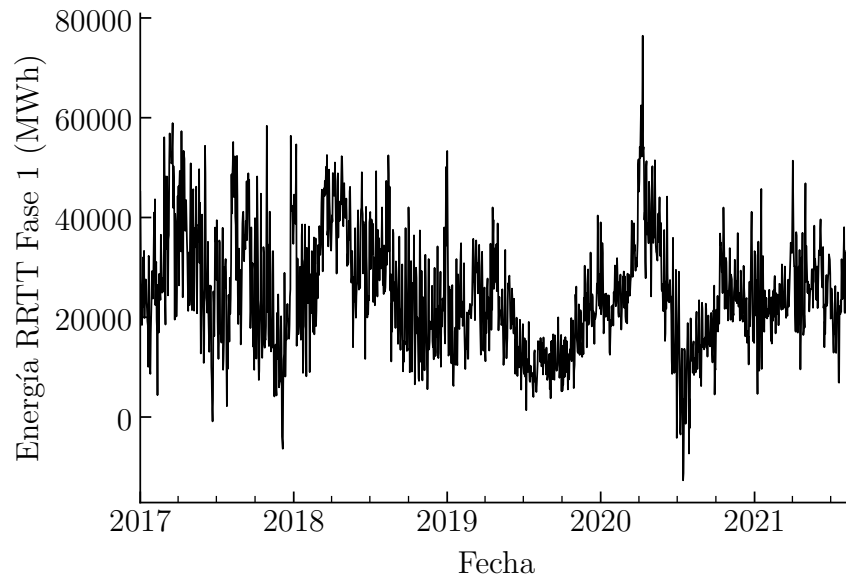


Figura 3: Energía restricciones técnicas Fase I (a subir) de 2017 a 2021

El mejor modelo obtenido es un *Random Forrest* cuyos resultados se pueden ver en la figura 4 y en la tabla 1.

	MAE	RMSE	R^2
Mejor modelo RF	4118.659761	5342.133502	0.203350

Tabla 1: Resultados del modelo RF

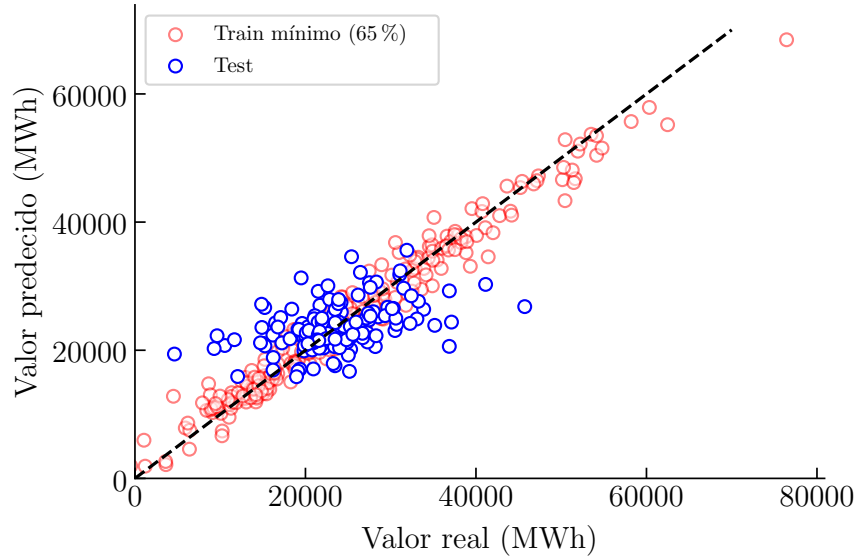


Figura 4: Resultados de modelo RF

En la figura 4 se puede observar que la varianza del conjunto de test es mucho menor que la varianza del conjunto de entrenamiento. Esto se debe a que el conjunto de entrenamiento cuenta con datos de 2020, que son anómalos debido a la pandemia. Este hecho ha afectado considerablemente a los resultados de los modelos entrenados. Pese a que el R^2 obtenido por el modelo pueda parecer bajo, es mucho mejor que los resultados de los *benchmarks* usados. Además, si el modelo se sigue usando y deja de depender de datos de 2020, los resultados probaran ser mucho mejores.

Como fruto del trabajo se puede concluir que el uso de las variables explicativas requieren de una gran cantidad de preprocesamiento para ser utilizadas en modelos de *Machine Learning*. También se puede afirmar que el uso de *Online Learning* mejora en gran medida los resultados obtenidos por los modelos.

Las variables explicativas de geolocalización de indisponibilidades y la estimación de la localización de la generación dependiente del clima no han sido introducidas en los modelos entrenados pero si han sido extraídas y preparadas. La introducción de estas variables geolocalizadas en el modelo deberían mejorar los resultados significativamente, puesto que las Restricciones Técnicas tienen una componente espacial muy grande.

Abstract

Technical Constrains forecasting models

Autor: Ventosa Pontes, Marcos

Directores: Soler Soneira, David y Fernández Ramos, Andrés

Collaborating Entity: Invesyde

Keywords: Technical Constrains, Machine Learning, Forecast

This work is a part of a larger project, the ultimate goal of which is arbitrage between day-ahead and intraday markets. In order to do so, it is necessary to forecast the Δ price between both markets. The volume forecast of technical constraints should be key to anticipate the differences between these markets considering that the Technical Constrains are calculated between these two markets. The scope of the overall project can be seen in the figure 5.

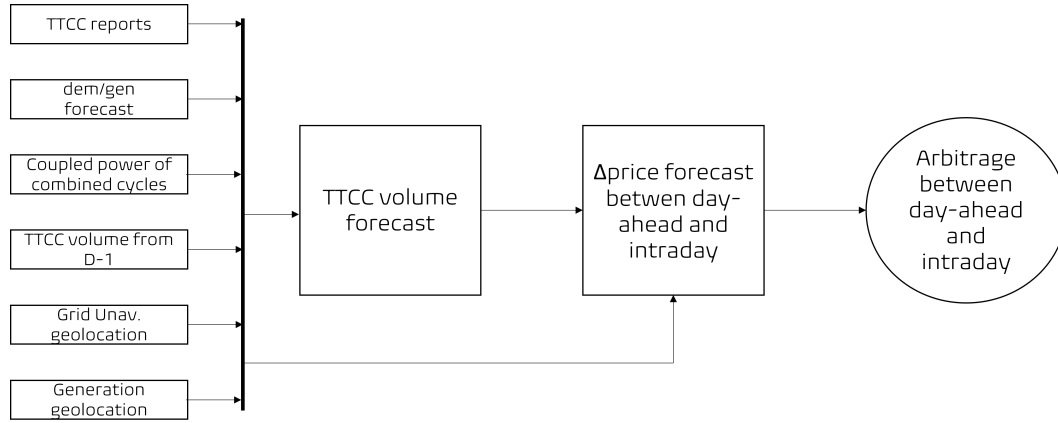


Figura 5: Full project

This work is focused on the extraction, analysis and processing of the explanatory variables that will feed both the constraints volume forecasting model and the daily and intraday price Δ forecasting model. These explanatory variables are the Technical Constrains reports published each day, the demand and generation forecasts, the coupled power of the combined cycles, the volume of Technical Constrains from the previous day, the geolocation of the Grid Unavailabilities and an estimation of the geolocation of the weather-dependent generation. Machine Learning models such as RF, KNN, RF and MLP are also developed using a large part of the explanatory variables, as shown in figure 6.

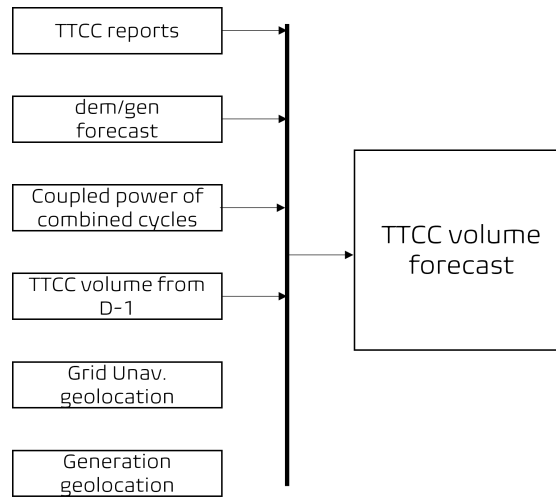


Figura 6: Scope of this proyect

The trained models use Online Learning, a technique that re-trains the model over time and can be used when the size of the data set increases sequentially, as in our case where each day we add a row to our set. In this way our models forecast the Technical Constrains energy volume of a day by training with all the previous days and for the next day's forecast the model adds the previous day's data to the training set, re-trains and then makes the forecast.

The evolution of the variable to be forecasted (Energy assigned to Technical Constraints) can be seen in figure 7. This variable has a very high volatility, which makes its forecast more complex.

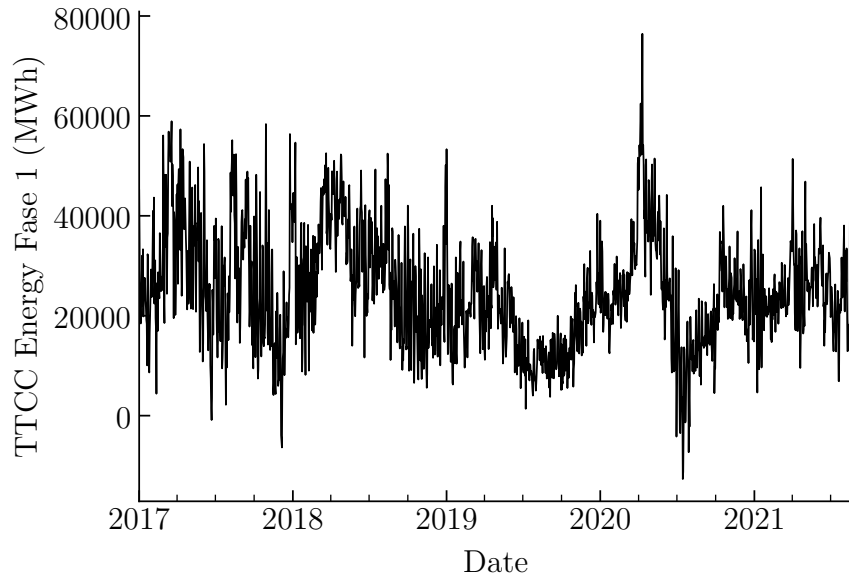


Figura 7: Energy assigned to Technical Constrains in Fase I from 2017 to 2021

The best model obtained is a Random Forrest”whose results can be seen in the figure 8 and in the table 2.

	MAE	RMSE	R^2
RF model	4118.659761	5342.133502	0.203350

Tabla 2: Results from RF model

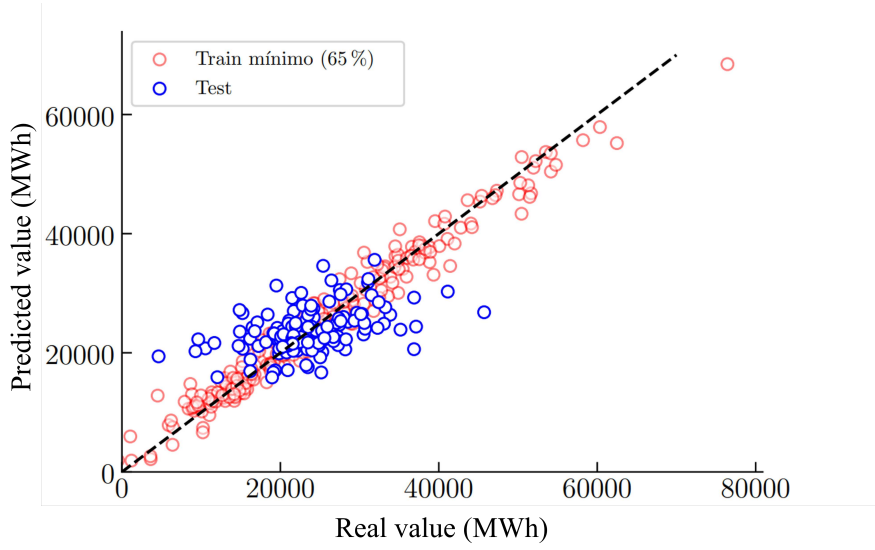


Figura 8: Results from RF model

In Figure 4 it can be seen that the variance of the test set is much smaller than the variance of the training set. This is because the training set has 2020 data, which are anomalous due to the pandemic. This fact has significantly affected the results of the trained models. Despite the fact that the R^2 obtained by the model may seem low, it is much better than the results of the used benchmarks. Furthermore, if the model continues to be used and no longer relies on 2020 data, the results will prove to be much better.

As a result of the work, it can be concluded that the use of explanatory variables requires a large amount of preprocessing to be used in Machine Learning models. It can also be affirmed that the use of Online Learning greatly improves the results obtained by the models.

The explanatory variables of geolocation of Grid Unavailabilities and the estimation of the location of weather-dependent generation have not been introduced in the trained models but have been extracted and prepared. The introduction of these geolocated variables in the model should improve the results significantly, since the Technical Constraints have a very large spatial component.

Índice general

1. Introducción	1
2. Estado de la cuestión	3
2.1. Previsión de precios del mercado eléctrico	3
2.2. Predicción del volumen de RRTT	4
2.3. Predicción de precios del mercado intradiario	5
3. Informes diarios de Restricciones Técnicas del PDBF	7
3.1. Conversión a tablas	7
3.2. Análisis de los datos de Restricciones Técnicas	9
3.2.1. Fecha	9
3.2.2. Zona	10
3.2.3. Escenario	11
3.2.4. Contingencias	12
3.2.5. Tipo	13
3.2.6. Nivel	14
3.2.7. Localización	15
3.3. Análisis de los datos de Indisponibilidades de Red	16
3.3.1. Fecha	16
3.3.2. Indisponibilidades	17
3.3.3. Inicio - Final	19
3.3.4. Tipo	21
3.3.5. Reposición Diario - Reposición de Emergencia	21
3.3.6. Observaciones	22
4. Geolocalización de Indisponibilidades de Red	23
4.1. Extracción de datos geográficos de nudos y líneas	23
4.1.1. Scraping del mapa de red de ENTSO-E	24
4.1.2. Estructura de los datos	26
4.2. Análisis de los datos	29
4.2.1. Visualización de los datos del mapa de ENTSO-E	29

4.2.2.	Precisión de la geolocalización de los nudos y líneas	30
4.3.	Geolocalización de Indisponibilidades de Red	41
5.	Geolocalización de la generación en España	47
5.1.	Extracción de datos geográficos de instalaciones de generación . . .	47
5.1.1.	Scrapping del mapa de red de ESIOS	48
5.2.	Análisis de los datos obtenidos	53
5.2.1.	Variables	53
5.2.2.	Simplificación de los datos	61
5.2.3.	Fecha real de los datos	62
5.2.4.	Visualización de los datos	62
5.3.	Cálculo de generación con geolocalización	63
5.3.1.	Cálculo de generación de tecnologías independientes del clima	64
5.3.2.	Cálculo de generación de tecnologías dependientes del clima	66
6.	Nudos con alta indisponibilidad en 2020	89
6.1.	Programas de Indisponibilidades de Red	91
6.1.1.	P25	91
6.1.2.	P50	92
6.1.3.	P90	94
6.2.	Cambios en los nudos con Indisponibilidades de Red bajo condicio- nes de generación extremas	95
6.2.1.	Escenarios de generación eólica	95
6.2.2.	Escenarios de generación solar fotovoltaica	102
7.	Herramienta de visualización de indisponibilidades de red y ge- neración	109
8.	Modelos de predicción de Restricciones Técnicas	113
8.1.	Variable de salida	113
8.2.	Variables de entrada	115
8.2.1.	Informes de Restricciones Técnicas	115
8.2.2.	Desplazamiento de los valores de un día	119
8.2.3.	Días laborales	120
8.2.4.	Ciclos combinados acoplados	120
8.2.5.	Previsiones	120
8.2.6.	Valor anterior	126
8.3.	Modelos	126
8.3.1.	Regresión Lineal	126
8.3.2.	K-Nearest-Neighbors	127
8.3.3.	Random Forrest	127

8.3.4. Perceptrón multicapa	128
8.4. Preprocesamiento	129
8.4.1. Normalización/Estandarización	130
8.4.2. Valores incrementales	130
8.5. División del conjunto de entrenamiento y de test	130
8.6. Resultados	131
8.6.1. Regresión Lineal	133
8.6.2. K-Nearest-Neighbors	137
8.6.3. Random Forest Regressor	141
8.6.4. Perceptrón multicapa	145
8.6.5. Tabla de Resultados	150
8.6.6. Conclusiones	150
9. Conclusiones	153
10.Alineación con los Objetivos de Desarrollo Sostenible (ODS)	155
11.Anexos	157
11.1. Anexo I: Informe de restricciones	157
11.2. Anexo II: Mapa de nudos y lineas	164
11.3. Anexo III: Nudos transformados y excluidos de las transformación de nudos de Madrid	165
11.4. Anexo IV: Nudos problematicos	165
Bibliografía	167

Índice de figuras

1.	Proyecto global	III
2.	Alcance del proyecto	IV
3.	Energía restricciones técnicas Fase I (a subir) de 2017 a 2021	V
4.	Resultados de modelo RF	VI
5.	Full project	VII
6.	Scope of this project	VIII
7.	Energy assigned to Technical Constrains in Fase I from 2017 to 2021	IX
8.	Results from RF model	X
3.1.	Histograma de restricciones por día	10
3.2.	Histograma de número de veces que se repite una localización . . .	16
3.3.	Histograma de indisponibilidades por día	17
3.4.	Histograma de número de veces que se repite una indisponibilidad .	19
3.5.	Histograma de hora de inicio de indisponibilidad	20
3.6.	Histograma de hora de final de indisponibilidad	21
4.1.	Datos geográficos de las líneas de España	29
4.2.	Datos geográficos de los nudos de España	30
4.3.	Datos geográficos de los nudos y líneas de España	30
4.4.	Nudos con su posición real la posición de ENTOSOE	31
4.5.	Nudos de Madrid con su posición real y la posición con la que con- tamos)	33
4.6.	Nudos de Madrid con su posición real y la posición con la que con- tamos después de aplicarles una transformación	34
4.7.	Área de detección de los nudos de Madrid (círculo de 140km de radio)	36
4.8.	Localización original de los nudos de Madrid	37
4.9.	Localización de los nudos de Madrid después de aplicarles una trans- formación	38
4.10.	Área de detección de las líneas de Madrid (círculo de 140km de radio)	39
4.11.	Mapa con la localización original de la líneas	40
4.12.	Mapa con la localización de la líneas después de aplicar una trans- formación alrededor de Madrid	40

4.13. Mapa de calor de las indisponibilidades durante 2020	45
4.14. Mapa de calor de las indisponibilidades durante 06-2020	46
4.15. Mapa de calor de las indisponibilidades durante 09-2020	46
5.1. Mapa de instalaciones de generacion de Esios	48
5.2. Foto PNG de un trozo del fondo del mapa de generadores	49
5.3. Mapa en el que se ven las cuadrículas usadas por Carto para un determinado nivel de zoom	50
5.4. Foto PNG de un trozo del mapa de generadores con la representa- ción de las instalaciones	51
5.5. Mapa de los municipios de España	52
5.6. Histograma de los MW en instalaciones de generación por tecnología	57
5.7. Histograma del número de instalaciones de generación agrupadas por tecnología	58
5.8. Histograma de número de veces que se repite un municipio	60
5.9. Mapa de calor de la potencia instalada en España	63
5.10. Generación nuclear durante el año 2020	65
5.11. Histograma de la velocidad media del viento en las estaciones me- teorológicas de España durante 2019, 2020 y parte de 2021	70
5.12. Media de la velocidad del viento en España durante 2019, 2020 y parte de 2021	71
5.13. Histograma de la media de la velocidad media del viento en España durante 2019, 2020 y parte de 2021	72
5.14. Histograma de las horas de insolación en España durante 2019, 2020 y parte de 2021	73
5.15. Horas de insolación media por día en España durante 2019, 2020 y parte de 2021	74
5.16. Histograma de la media de horas de insolación en España durante 2019, 2020 y parte de 2021	75
5.17. Generación eólica durante 2020 y parte de 2021	76
5.18. Diagrama de dispersión de la generación eólica en función de la velocidad media del viento en España durante 2020 y parte de 2021	77
5.19. Mapa de calor de la generación eólica durante 2020	79
5.20. Generación solar fotovoltaica durante 2019, 2020 y parte de 2021 . .	80
5.21. Diagrama de dispersión de la generación solar fotovoltaica en fun- ción de la horas de insolación medias en España durante 2019, 2020 y parte de 2021	81
5.22. Mapa de calor de la generación fotovoltaica durante 2020	83
5.23. Generación solar térmica durante 2020 y parte de 2021	84

5.24. Diagrama de dispersión de la generación solar térmica en función de la horas de insolación medias en España durante 2020 y parte de 2021	85
5.25. Mapa de calor de la generación térmica durante 2020	87
6.1. Histograma del porcentaje del tiempo que un nudo tiene indisponibilidades asociadas	91
6.2. Mapa de calor de los nudos del P25	92
6.3. Mapa de calor de los nudos del P50	94
6.4. Mapa de calor de los nudos del P90	95
6.5. Top 20 % días de generación eólica durante 2020	96
6.6. Distribución en el año del top 20 % días de generación eólica durante 2020	97
6.7. Bot 20 % días de generación eólica durante 2020	100
6.8. Distribución en el año del bot 20 % días de generación eólica durante 2020	101
6.9. Top 20 % días de generación fotovoltaica durante 2020	103
6.10. Distribución en el año del top 20 % días de generación solar fotovoltaica durante 2020	104
6.11. Bot 20 % días de generación fotovoltaica durante 2020	106
6.12. Distribución en el año del bot 20 % días de generación fotovoltaica durante 2020	107
7.1. Herramienta de visualización de indisponibilidades de red y generación	110
8.1. Energía restricciones técnicas Fase I (a subir) durante 2020 y parte de 2021	114
8.2. Energía restricciones técnicas Fase I (a subir) de 2017 a 2021 . . .	115
8.3. Distribución de las horas con restricción al día en diferentes zonas .	117
8.4. Previsión de generación eólica frente a generación real	122
8.5. Previsión de generación fotovoltaica frente a generación real	123
8.6. Generación programada nuclear frente a generación real	124
8.7. Previsión de demanda frente a demanda real	125
8.8. Ejemplo de regresión lineal	127
8.9. Árbol de decisión [50]	128
8.10. Perceptron multicapa	129
8.11. Resultados del <i>naive approach</i>	132
8.12. Resultados de aproximar por la media del conjunto de train	133
8.13. Resultados de modelo de regresión lineal usando variables zona y tipo agrupadas	134

8.14. Resultados de modelo de regresión lineal usando variables zona y localización agrupadas	135
8.15. Resultados de modelo de regresión lineal usando el semáforo de restricciones	136
8.16. Resultados de modelo de regresión lineal usando el semáforo de restricciones y variables incrementales	137
8.17. Resultados de modelo KNN usando variables zona y tipo agrupadas	138
8.18. Resultados de modelo KNN usando variables zona y localización agrupadas	139
8.19. Resultados de modelo KNN usando el semáforo de restricciones . .	140
8.20. Resultados de modelo KNN usando variables incrementales	141
8.21. Resultados de modelo RFR usando variables zona y tipo agrupadas	142
8.22. Resultados de modelo RFR usando variables zona y localización agrupadas	143
8.23. Resultados de modelo RFR usando el semáforo de restricciones . . .	144
8.24. Resultados de modelo RFR usando variables incrementales	145
8.25. Resultados de modelo MLP usando variables zona y tipo agrupadas	146
8.26. Resultados de modelo MLP usando variables zona y localización agrupadas	147
8.27. Resultados de modelo MLP usando el semáforo de restricciones . .	148
8.28. Resultados de modelo MLP usando el semáforo de restricciones y variables incrementales	149
11.1. Mapa de nudos y líneas de REE de 2017 [54]	164

Índice de tablas

1.	Resultados del modelo RF	V
2.	Results from RF model	IX
3.1.	Datos de Restricciones Técnicas obtenido de REE	8
3.2.	Datos de Indisponibilidades de Red obtenido de REE (limitado a 4 resultados)	9
3.3.	Variable Zona y su frecuencia	10
3.4.	Variable Zona y su frecuencia después de solucionar problemas de consistencia	11
3.5.	Variable Escenario y nueva variable Horas	11
3.6.	Variable Contingencias y sus 20 valores más repetidos	12
3.7.	Diccionario de Contingencias	13
3.8.	Variable Tipo y su frecuencia	13
3.9.	Variable Tipo y su frecuencia después de aumentar sus valores . . .	14
3.10.	Variable Nivel y su frecuencia después de aumentar sus valores . . .	14
3.11.	Variable Localización y sus 20 valores más repetidos	15
3.12.	Variable Indisponibilidades y sus 20 valores más repetidos	18
3.13.	Variable tipo y su frecuencia	21
4.1.	Cuadrículas necesarias para obtener todos los datos de España . . .	26
4.2.	Distancia entre la posición real de 5 nudos y la posición con la que contamos	32
4.3.	Distancia de la posición de los nudos al centro de Madrid	33
4.4.	Distancia de la posición de los nudos al centro de Madrid después de aplicar la transformación	34
4.5.	Nudos de conexión	41
4.6.	Nudos con distinto nombre entre el mapa de ENTSO-E y la tabla de Indisponibilidades (limitado a 7 resultados)	42
4.7.	Nombres de nudos contenidos en otros (limitado a 7 resultados) . .	43
4.8.	Tabla de Indisponibilidades con los nudos en los que aparece (limitado a 5 resultados y quitando parte de las variables)	43

4.9. Tabla de Indisponibilidades con los nudos en los que aparece y el tiempo total en el día (limitado a 5 resultados y quitando parte de las variables)	44
4.10. Tabla de Indisponibilidades con su geolocalización y el tiempo total activa durante el año 2020 (limitado a 5 resultados)	44
5.1. Datos de instalaciones de generación obtenido de ESIOS (limitado a 5 resultados)	53
5.2. Variable tecnología y número de instalaciones que la tienen	54
5.3. Variable MW por tecnología y sus características (redondeado a 2 decimales)	55
5.4. Variable municipio y sus 20 valores más repetidos)	59
5.5. Datos de localización de los municipios de España [39]	60
5.6. Datos de instalaciones de generación después de añadir localización (limitado a 5 resultados)	61
5.7. Datos de instalaciones de generación separados por comunidad autónoma y tecnología (valores redondeados)	62
5.8. Tabla de producción de cada generador para cada día	64
5.9. Tabla de producción de centrales nucleares (limitado a 8 resultados)	66
5.10. Trozo 1 de tabla de datos de AEMET (limitado a 5 resultados)	67
5.11. Trozo 2 de tabla de datos de AEMET (limitado a 5 resultados)	67
5.12. Datos de AEMET que se utilizan (limitado a 5 resultados)	68
5.13. Tabla de estaciones meteorológicas y su localización (limitado a 5 resultados)	69
6.1. Nudos y su tiempo total con indisponibilidades durante el año 2020 (limitado a 10 resultados)	90
6.2. Nudos con más tiempo con indisponibilidades (limitado a 10 resultados)	90
6.3. Nudos de P50	93
6.4. Nudos de P90	94
6.5. Nudos con mayores cambios en escenario de generación eólica máxima	99
6.6. Nudos con mayores cambios en escenario de generación eólica mínima	102
6.7. Nudos con mayores cambios en escenario de generación fotovoltaica máxima	105
6.8. Nudos con mayores cambios en escenario de generación fotovoltaica mínima	108
8.1. Tabla del número de horas con Restricciones por Zona	116
8.2. Valor del semáforo en función del percentil	118
8.3. Tabla del semáforo de Restricciones por Zonas	118

8.4. Restricciones Técnicas agrupadas por Zona y Tipo (limitado a 5 columnas)	119
8.5. Restricciones Técnicas agrupadas por Zona y Localización (limitado a 5 columnas)	119
8.6. Variable laboral	120
8.7. Potencia acoplada de los ciclos combinados dividido en 10 zonas . .	120
8.8. División en 10 zonas de las unidades de programación	121
8.9. Previsión eólica dividida por Comunidad Autónoma (limitado a 5 columnas)	121
8.10. Previsión solar fotovoltaica dividida por Comunidad Autónoma (limitado a 5 columnas)	122
8.11. Resultados del <i>naive approach</i>	131
8.12. Resultados de aproximar por la media del conjunto de train	132
8.13. Resultados de modelo de regresión lineal usando variables zona y tipo agrupadas	134
8.14. Resultados de modelo de regresión lineal usando variables zona y localización agrupadas	135
8.15. Resultados de modelo de regresión lineal usando el semáforo de restricciones	136
8.16. Resultados de modelo de regresión lineal usando el semáforo de restricciones y variables incrementales	137
8.17. Resultados de modelo KNN usando variables zona y tipo agrupadas	138
8.18. Resultados de modelo KNN usando variables zona y localización agrupadas	139
8.19. Resultados de modelo KNN usando el semáforo de restricciones . .	140
8.20. Resultados de modelo KNN usando variables incrementales	140
8.21. Resultados de modelo RFR usando variables zona y localización agrupadas	141
8.22. Resultados de modelo RFR usando variables zona y localización agrupadas	142
8.23. Resultados de modelo RFR usando el semáforo de restricciones . .	143
8.24. Resultados de modelo RFR usando variables incrementales	144
8.25. Resultados de modelo MLP usando variables zona y tipo agrupadas	145
8.26. Resultados de modelo MLP usando variables zona y localización agrupadas	146
8.27. Resultados de modelo MLP usando el semáforo de restricciones . .	147
8.28. Resultados de modelo MLP usando el semáforo de restricciones y variables incrementales	148
8.29. Resultados de los modelos	150

Índice de ecuaciones

4.1. Ejemplo de llamada a la API de Mapbox para obtener datos de nudos y líneas	25
4.2. Estructura Geojson	26
4.3. Estructura Nudo	26
4.4. Estructura Línea	27
5.1. Ejemplo de respuesta de Carto con informacion de una instalación de producción	52
5.2. Fórmula para el cálculo de la generación para centrales con tecnologías independientes del clima	64
5.3. Fórmula para el cálculo de la generación para centrales con tecnologías dependientes del clima	66
5.4. Ecuación de la recta de generación eólica	77
5.5. Ecuación de la recta de generación eólica por tramos	78
5.6. Ecuación de la recta de generación fotovoltaica	81
5.7. Ecuación de la recta de generación fotovoltaica por tramos	82
5.8. Ecuación de la recta de generación eólica	85
5.9. Ecuación de la recta de generación fotovoltaica por tramos	86
6.1. Cálculo del incremento del porcentaje de tiempo con indisponibilidad en cada nudo	97
6.2. Cálculo de la diferencia del porcentaje de tiempo con indisponibilidad en cada nudo	98
8.1. Cálculo del Hueco Térmico	125
8.2. Ecuación del modelo de regresión lineal	126

Capítulo 1

Introducción

El proceso de programación diaria del Sistema Eléctrico Español incluye diferentes etapas sucesivas donde se definen distintos programas de generación (PDBF, PDVP, regulación secundaria, PHF/PHFC y P48) que garantizan el suministro y la seguridad del sistema. [1]

Existen dos agentes principales en el sistema eléctrico, el Operador del Mercado Eléctrico Español (OMIE) y Red Eléctrica Española (REE). OMIE es el encargado de cruzar las ofertas de generación con las de demanda a través de los diferentes mercados, mientras que REE garantiza el funcionamiento y la gestión técnica de la red como operador del sistema (OS).

Una vez que OMIE ha realizado una primera casación para cada hora del día siguiente, en el mercado diario se obtiene el Programa Diario Base de Funcionamiento (PDBF). Esta casación se envía al OS que corre un flujo de cargas donde se detectan las restricciones técnicas. Una restricción técnica es “cualquier circunstancia o incidencia derivada de la situación de la red de transporte o del sistema 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”. El proceso de solución de restricciones técnicas se realiza mediante la modificación de los programas de las Unidades de Programación (Fase I) y el posterior proceso de re-equilibrio generación-demanda (Fase II). Estas restricciones se recopilan en el Informe diario de Restricciones Técnicas del PDBF. Para solucionar las restricciones técnicas el OS introduce las modificaciones necesarias en el PDBF obteniendo el Programa Diario Viable Provisional (PDVP). Después se establecen los requerimientos de banda de regulación secundaria y reserva de regulación terciaria.

Con el objetivo de poder realizar ajustes, tanto oferta como demanda, pueden participar en las sesiones del mercado intradiario.

Dentro del mercado intradiario, podemos distinguir entre el mercado intradiario de subastas y el mercado intradiario continuo. El mercado intradiario de subastas tiene 6 sesiones y al igual que el mercado diario, es marginalista. El mercado intradiario continuo permite realizar ajustes hasta una hora antes y se beneficia de la liquidez en los mercados de otras zonas de Europa. [2]

La posibilidad de participar tanto en el mercado diario como en los mercados intradiarios, ofrece la oportunidad a los agentes de arbitrar para obtener beneficios aprovechando la diferencia en el precio de la energía entre ambos mercados. El estudio de las restricciones técnicas puede ofrecer información clave para anticipar esas diferencias.

El capítulo 1 consiste en una breve introducción al tema tratado. El capítulo 2 se centra en una revisión del estado del arte. El capítulo 3 se explican los informes de Restricciones Técnicas. En el capítulo 4 se extrae la localización de los nudos de la red y se analiza la localización geográfica de las indisponibilidades de Red. En el capítulo 5 se extrae la localización de la potencia instalada en España y se divide la generación. En el capítulo 6 se estudian los nudos mas problemáticos de la red en cuanto a Indisponibilidades. En el capítulo 7 se comenta la herramienta de visualización de indisponibilidades y generación que se ha desarrollado. En el capítulo 8 se muestran los modelos de *machine learning* entrenados con el objetivo de predecir el valor de energía de restricciones en fase I a subir y sus resultados. En el capítulo 9 se recogen las conclusiones del trabajo realizado.

Capítulo 2

Estado de la cuestión

2.1. Previsión de precios del mercado eléctrico

Debido al interés que genera este problema muchos grupos de investigación lo han abordado, pudiéndose agrupar las distintas aproximaciones en cinco grupos [3].

- Modelos multi-agente. Simulan el sistema como un conjunto de agentes heterogéneos que interactúan entre sí y se predice el precio casando la demanda y la generación de los agentes. Estos modelos son muy flexibles aunque muchos de sus parámetros tienen que ser definidos, lo cual presenta una dificultad. Podemos encontrar un ejemplo de la aplicación de estos modelos en el mercado [4].
- Modelos fundamentales. Usan relaciones físicas y económicas entre el precio y sus factores fundamentales. Estas relaciones se implementan a través de ecuaciones basadas en supuestos y por tanto estos modelos dependen fuertemente de la veracidad de esos supuestos. Se puede ver una aplicación de estos modelos al mercado noruego en [5].
- Modelos cuantitativos. Replican las principales características del precio. Estos modelos no son buenos para predecir con precisión los precios del mercado diario, sino que suelen usarse para la valoración de derivados o detección de picos en el precio [6].
- Modelos estadísticos. Predicen el precio tanto a través de una relación matemática entre el precio actual y los precios previos, así como usando variables exógenas. Su fácil interpretación hace que estos modelos resulten muy interesantes. Dentro de esta categoría podemos encontrar modelos simples, como el similar-day approach (también conocido como naive approach) [7]

o exponential smooting [8] que son comúnmente usados como *benchmark* y modelos más complejos como los modelos de regresión, modelos AR, modelos ARX. Hay algunas variables exógenas que mejoran la predicción como la previsión de la producción eólica [9]. Los modelos estadísticos suelen tener problemas con los picos en el precio y por tanto puede ser beneficioso filtrar los picos de nuestros datos [10], existen diferentes formas de hacerlo. Algunos ejemplos de aplicación de estos modelos a la predicción del mercado diario son [9][10].

- Modelos de computación inteligente. Se incluyen dentro de este grupo redes neuronales artificiales, SVM, algoritmos genéticos, etc. Su principal ventaja es la capacidad de adaptarse a sistemas no lineales. Suelen ofrecer mejor predicción de precios que los modelos estadísticos. Podemos encontrar ejemplos de estos modelos aplicados a la predicción del precio del mercado diario en [11] [9].

Se pueden encontrar propuestas con modelos híbridos que juntan dos o mas modelos de los grupos mencionados anteriormente [12].

2.2. Predicción del volumen de RRTT

El cálculo de restricciones técnicas es realizado por el OS con el objetivo de garantizar la viabilidad del programa de generación. Para realizar este cálculo se simulan flujos de red suponiendo muchos de los datos (previsión de demanda y su precisión, previsión de producción eólica y solar, indisponibilidades conocidas, etc.) para estimar como quedaría el Sistema Eléctrico con ese programa de generación y detectando posibles fallos [1].

Dentro de estos posibles fallos podemos encontrar: incumplimiento de las condiciones de seguridad en régimen permanente, insuficiente reserva de regulación secundaria o terciaria, insuficiente reserva de potencia adicional, etc. Los diferentes casos de estudio se realizan con PSS/E.

Podemos encontrar trabajos en los que se realizan predicciones tanto de la aparición de restricciones técnicas como de su volumen usando perceptrones multicapa (MLP). Este estudio se centra en las centrales térmicas por su mayor influencia en el mercado de restricciones y se hace una división en 13 zonas del territorio español [13].

2.3. Predicción de precios del mercado intradiario

Mientras que la predicción del precio del mercado diario ha sido ampliamente estudiada, hasta hace unos pocos años no había mucha literatura sobre la predicción del mercado intradiario. Este cambio de tendencia viene fomentado por el incremento de la generación renovable en el mix de generación eléctrica debido a las desviaciones que aparecen entre su producción acordada en el mercado diario y su producción real [14]. Este incremento debería mantenerse para cumplir así con los objetivos del Proyecto de Ley de Cambio Climático y Transición Energética (PLCCTE) [15].

Con estos mercados con un mix cada vez mas renovable la predicción es mucho más difícil de realizar, principalmente por la necesidad de incluir nuevas variables explicativas que aumentan la complejidad de los modelos sin garantizar mejoras en la predicción [16].

Algunos estudios muestran que la predicción del mercado intradiario mejora si se incluye el precio de la casación del mercado diario, y el precio de las sesiones del mercado intradiario anteriores [16].

Se ha intentado predecir tanto el mercado intradiario continuo como el mercado intradiario de subastas usando similares aproximaciones a las comentadas en 2.1 Previsión de precios del mercado eléctrico.

Podemos encontrar intentos de aplicar SARIMA, como en [17] al mercado intradiario alemán, intentos de usar redes neuronales [18] o intentos de aplicar estrategias multi-agente para encontrar estrategias óptimas [19].

Capítulo 3

Informes diarios de Restricciones Técnicas del PDBF

Como ya se ha comentado en el capítulo 1, cada día REE publica un Informe diario de Restricciones Técnicas del PDBF. Estos archivos se publican en formato PDF y se puede ver un ejemplo de éste en el anexo , estos PDFs son accesibles a través de ESIOS por los agentes del sistema (generadores y comercializadoras). Para el desarrollo de este trabajo se ha contado con los informes entre el 01/01/2020 y el 21/03/2021.

3.1. Conversión a tablas

Debido a que los archivos en formato PDF no se pueden procesar directamente, estos tienen que ser tratados previamente. Como se ve en informe de las dos tablas que aportan más información son la de Restricciones Técnicas de Seguridad y la de Indisponibilidades de Red. Las Restricciones Técnicas dependen del mercado ya que son fruto de una casación que no es posible técnicamente y las indisponibilidades son ajenas al mercado y ocurren cuando hay problemas técnicos, como una línea que no funcione, un nudo este en mantenimiento...

Existen diferentes alternativas *online* para convertir PDFs a tabla aunque la mayoría tienen problemas con documentos que contengan más información que una única tabla. En este caso, tenemos documentos con texto, gráficos y tablas. Aun así este PDF no necesita Optical Character Recognition (OCR) ya que las tablas están ordenadas y colocadas siempre de la misma forma y no están subidas como una imagen sino como texto plano. Pese a lo comentado previamente, hay alguna herramienta que hace la tarea con éxito [20]. Desgraciadamente, estas páginas requieren que el proceso se haga de forma manual, y las que tienen API son de

CAPÍTULO 3. INFORMES DIARIOS DE RESTRICCIONES TÉCNICAS DEL PDBF

pago.

Con el objetivo de buscar una alternativa que pudiera ser aplicable a cientos de informes de forma automática se han buscado librerías de *Python* que permitieran hacer esta tarea. Se ha probado *camelot* [21] y *tabula* [22], optando finalmente por *camelot*, ya que obtiene resultados más satisfactorios.

Pese a que esta librería hace un trabajo bueno, también ha sido necesario procesar ligeramente el texto de algunas columnas en ambas tablas para que el resultado fuera totalmente satisfactorio. Añadimos además a cada fila la fecha del informe, ya que es información de interés de cara a juntar los datos de todos los informes.

Después de aplicar esta conversión a todos los informes con los que se contaba, se han obtenido y guardado las tablas por separado de cada día en formato *CSV*.

En la tabla 3.1 se puede ver la tabla de Restricciones obtenida después de aplicar la conversión para el informe de 01/01/2020.

Fecha	Zona	Escenario	Contingencias	Tipo	Nivel	Localizacion
2020-01-01	Centro	H1 - H18	regular tension y seguridad en la zona	Control Tension	220 y 400 kV	Zona Centro
2020-01-01	Este	H1 - H24	regular tension y seguridad en la zona	Control Tension	220 y 400 kV	Zona Cataluna
2020-01-01	Levante	H1 - H24	regular tension y seguridad en la zona	Control Tension	220 y 400 kV	Zona Levante
2020-01-01	Levante	H21 - H24	l-400 kv eliana - plana	Sobrecargas	> 115 %	L-400 kV Gaussa - Plana
2020-01-01	Noroeste	H1 - H24	regular tension y seguridad en la zona	Control Tension	220 y 400 kV	Zona Galicia

Tabla 3.1: Datos de Restricciones Técnicas obtenido de REE

En la tabla 3.2 se puede observar la información de Indisponibilidades obtenida después de aplicar la conversión para el informe de 01/01/2020.

3.2. Análisis de los datos de Restricciones Técnicas

FECHA	INDISPONIBILIDADES	INICIO	FINAL	TIPO	REPO. DIARIA	REPO. EMERG	OBS.
2020-01-01	220 kV ADIF MONFORTE DEL CID MONFORTE DEL CID 1	2019-08-01 09:25:00	2020-01-13 23:59:00	P.S.	NO	NO	NaN
2020-01-01	220 kV CICERO SOLORZANO 1	2016-09-29 11:02:00	2021-12-31 23:59:00	P.S.	NO	NO	NaN
2020-01-01	220 kV CICERO SOLORZANO 2	2016-09-29 11:02:00	2021-12-31 23:59:00	P.S.	NO	NO	NaN
2020-01-01	220 kV SOTO DE RIBERA TRASONA	2015-10-09 12:49:00	2030-02-28 23:59:00	P.S.	NO	NO	NaN

Tabla 3.2: Datos de Indisponibilidades de Red obtenido de REE (limitado a 4 resultados)

3.2. Análisis de los datos de Restricciones Técnicas

La tabla 3.1 cuenta con 7 variables y 3562 filas. Ninguna de las variables es numérica, todas son texto o *Datetime*. A continuación se analizará cada variable de forma individual.

3.2.1. Fecha

La primera variable es *Fecha* (formato *Datetime*), y define el día de publicación del informe, sus valores van desde el 2020-01-01 al 2021-03-21 con una media de 8.0 datos por cada día.

En la figura 3.1 podemos ver el histograma de restricciones por día, su media y su función de densidad de probabilidad. Se podría aproximar por una distribución normal de media 8 y desviación típica de 2.

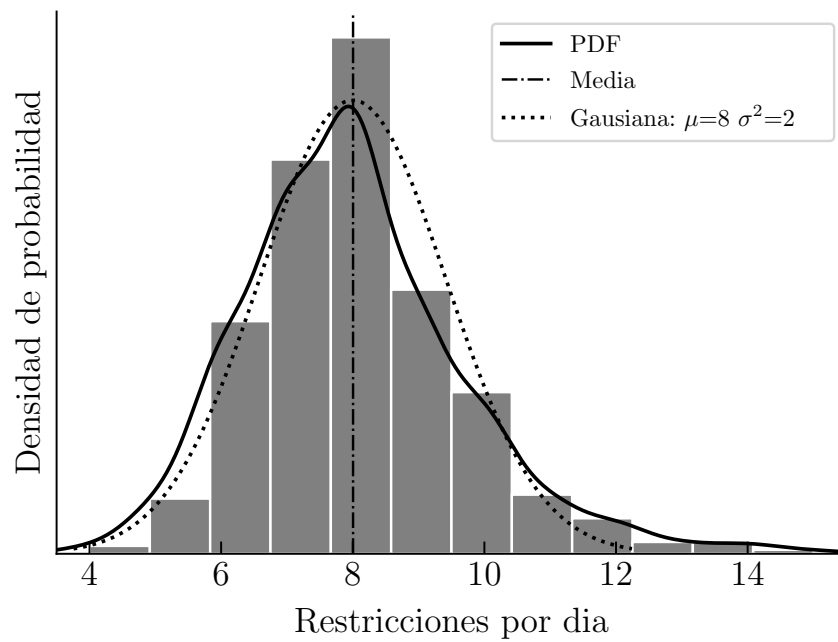


Figura 3.1: Histograma de restricciones por día

3.2.2. Zona

La segunda variable es *Zona*, y especifica el lugar de la restricción, puede tomar 10 valores distintos tal como se ve en la tabla 3.3.

Zona	Frecuencia
Noroeste	1088
Este	569
Levante	534
Norte	524
Centro	435
Sur	399
Levante Norte	7
Pais Vasco	3
Levante Sur	2
Levaante	1

Tabla 3.3: Variable Zona y su frecuencia

Se identifican problemas de consistencia en algunos de los valores. El más claro

es *Levaante*, que es un error al escribir *Levante*. Por otro lado, se puede ver que hay algunos valores que tienen muy poca frecuencia. Esto se puede deber de nuevo a problemas de consistencia; *Levante Norte* y *Levante Sur* muy probablemente deberían de estar recogidos en *Levante* y *Pais Vasco* en *Norte*. Estos problemas se solucionan fácilmente, obteniendo la tabla 3.4

Zona	Frecuencia
Noroeste	1088
Este	569
Levante	544
Norte	527
Centro	435
Sur	399

Tabla 3.4: Variable Zona y su frecuencia después de solucionar problemas de consistencia

3.2.3. Escenario

La variable *Escenario* contiene las horas en las que se produce la restricción. Por ejemplo *H1 - H18* informa de una restricción entre la hora 1 y la hora 18. Se pueden encontrar ejemplos de *Hx - Hy*, *Hx - Hr* en los casos que la restricción ocurra en más de un periodo de tiempo y algunos casos de *Hx* cuando la restricción ocurre solo durante una hora. La gran variedad de combinaciones hace que de una variable tan simple, podamos encontrar 448 valores distintos dentro de las 3562 filas. Para transformar esta variable en algo más procesable se crea la variable *Horas*, que contiene el número total de horas que una restricción está activa. Para calcularlo usamos el valor de *Escenario*, tal como se observa en la tabla 3.5.

	Escenario	Horas
Fecha		
2020-01-01	H21 - H24	4
2020-01-01	H1 - H24	24
2020-01-01	H1 - H18	18
2020-01-01	H1 - H24	24
2020-01-02	H1 - H7	7

Tabla 3.5: Variable Escenario y nueva variable Horas

3.2.4. Contingencias

La variable *Contingencias* contiene información sobre la restricción. Podemos encontrar 103 valores distintos en las 3562 filas. Los 20 valores más repetidos se pueden ver en 3.6. Se aprecia que los dos valores más frecuentes, aparecen en algo más del 80 % de los casos.

Contingencias	Frecuencia
regular tension y seguridad en la zona	2370
seguridad en la red de distribucion	620
imposibilidad evacuacion produccion pp. ee.	50
l-400 kv eliana - plana	48
l-220 kv aceca - pradillos	42
dc-400 kv aragon - asco	41
dc-400 kv adeadavila - aranuelo/almaraz - hinojosa	23
l-220 kv aceca - anover	21
situacion base	17
l-220 kv itxaso - orcoyen 1	16
l-400 kv plana - vandellos	16
dc-400 kv rocamora - pinilla/rocamora - penarrubia	14
l-200 kv aceca - anover	14
cn cofrentes + l-400 kv eliana - plana	11
dc-400 kv almaraz - villaviciosa	11
l-220 kv pinto - villaverde bajo	10
l-220 kv aceca - valdemoro	10
l-220 kv arganda - loeches	9
l-220 kv aceca - carroyuelas	9
l-220 kv els aubals - escatron	9

Tabla 3.6: Variable Contingencias y sus 20 valores más repetidos

En esta variable se pueden encontrar decenas de casos de inconsistencia. Pueden ser faltas de ortografía o formas diferentes de escribir abreviaturas/frases. También se pueden ver casos de Contingencias que contienen más de una Contingencia, como podría ser '*dc-400 kv adeadavila - aranuelo/almaraz - hinojosa*'. Para resolver todos estos casos, se ha elaborado un diccionario de contingencias, se pueden ver 3 ejemplos de este diccionario en la tabla 3.7.

Contingencia	Condiciones
'bp1 se-400 kv serna'	['serna', 'bp1', '400']
'dc-220 kv aena - hortaleza'	['aena', 'hortaleza', '220', ['dc-', 'c-']]
'c.n cofrentes'	[['cofrentes', 'grupo']]

Tabla 3.7: Diccionario de Contingencias

Las condiciones del diccionario funcionan como puertas lógicas. Para que a un texto se le asocie *'bp1 se-400 kv serna'*, tiene que contener las palabras de dentro de los corchetes, en este caso *'serna'*, *'bp1'* y *'400'* lo que convierte a las comas (,) que separan cada palabra en una puerta **AND**. Cuando encontramos unos nuevos corchetes dentro, las comas de dentro funcionan como puertas **OR**, de esta forma, si tengo *['a', 'b', ['c', 'd']]* es como decir que tiene que contener (**'a' AND 'b' AND ('c' OR 'd')**).

Al aplicar el diccionario a la tabla original se obtienen más filas que las originales, ya que a un texto como *'dc-400 kv bescano - vic/pierola - vic'* se le asignarán tanto *'dc-400 kv bescano - vic'* como *'dc-400 kv pierola - vic'* ya que contiene tanto estas palabras *['bescano', 'vic', '400', 'dc-']*, como estas *['pierola', 'vic', '400', 'dc-']*. Se pasan de las 3562 filas originales a 3661 filas.

3.2.5. Tipo

La variable *Tipo* explica el tipo de contingencia. Los diferentes valores que esta variable adopta y sus frecuencias se encuentran en la tabla 3.8.

Tipo	Frecuencia
Control Tension	2397
-	686
Sobrecargas	478
Control de Tension	1

Tabla 3.8: Variable Tipo y su frecuencia

Se aprecia un pequeño error de consistencia al escribir *Control de Tensión* en vez de *Control Tension*. Por otro lado, se observa que hay 686 filas que no tienen nada (-) en esta variable. Para intentar solucionar esto, se han observado conjuntamente las variables de Contingencias y Tipo, viendo que hay mucha relación entre ambas y que determinadas *Contingencias*, podrían agruparse para formar un nuevo tipo y así dotar a la variable *Tipo* de mayor valor. De esta forma se crean 4 nuevos valores para la variable *Tipo*.

- *Evacuacion* cuando aparece la palabra *evacuacion* en las *Contingencias*.
- *Bombeo* cuando aparece la palabra *bombeo* en las *Contingencias*.
- *Desfase* cuando aparece la palabra *desfase* en las *Contingencias*.
- *Seguridad en la red de distribucion* cuando aparece *Seguridad en la red de distribucion* en las *Contingencias*.

Se pueden observar los resultados de aplicar esto a la variable Tipo en 3.9

Tipo	Frecuencia
Control Tension	2370
seguridad en la red de distribucion	620
Sobrecargas	478
Evacuacion	88
desfase	5
Bombeo	1

Tabla 3.9: Variable Tipo y su frecuencia después de aumentar sus valores

3.2.6. Nivel

La variable Nivel informa del nivel de tensión (kV) en caso de restricción de tensión y del porcentaje de sobrecarga (%) en los casos de sobrecarga. Se aprecian los diferentes valores que adopta esta variable y su frecuencia en la tabla 3.10.

Nivel	Frecuencia
220 y 400 kV	2356
-	686
> 115 %	450
221 y 400 kV	30
> 100 %	21
219 y 400 kV	12
> 110 %	6
> 115 %	1

Tabla 3.10: Variable Nively su frecuencia después de aumentar sus valores

Se puede ver que el número de veces que aparece '-' es igual que el número de veces que aparecía en la variable *Tipo* y esto se debe a que siempre que en una

aparece '-' en la otra se ponía lo mismo. De la misma forma que en las variables anteriores, se solucionan los errores de consistencia.

3.2.7. Localización

Esta es la última variable e informa de la localización de la restricción de forma más precisa que la variable Zona. Esta variable adopta 212 valores distintos y podemos ver sus 20 valores más repetidos en la tabla 3.11

Localizacion	Frecuencia
Red Distribucion Zona Asturiana	235
Zona Galicia	200
Pais Vasco	200
Cataluna	197
Galicia Norte	197
Zona Cataluna	195
Zona Levante	193
Zona Pais Vasco	189
Red Distribucion Asturias	174
Levante	168
Sur	158
Zona Centro	139
Zona Sur	122
Centro	120
Zona Asturiana	70
L-400 kV Gaussa - Plana	54
Asturias	54
Zona Catalana	47
L-220 kV Aceca - Anover	42
L-220 kV Aceca - Pradillos	37

Tabla 3.11: Variable Localización y sus 20 valores más repetidos

Esta variable cuenta con muy poca consistencia. En algunos casos se habla de zonas, en otros de comunidades y otras veces de líneas.

En la figura 3.2 podemos ver la distribución del número de veces que una indisponibilidad se repite. Pese a que algunas se repitan hasta más de 200 veces, como se ve en la tabla 3.2, un 81 % de los valores se repite menos de 10 veces, lo que hace que esta variable sea difícil de usar. La distribución puede aproximarse a una distribución de Poisson de λ de valor 3.5.

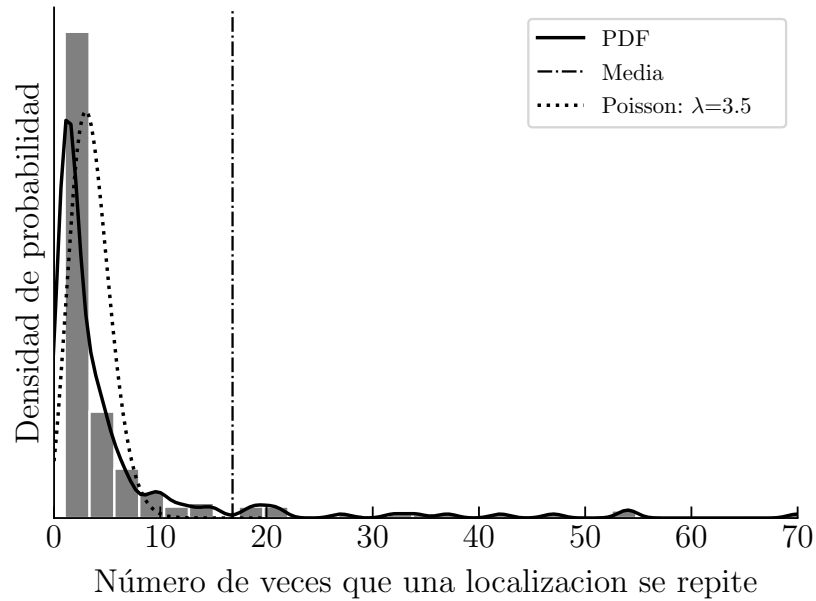


Figura 3.2: Histograma de número de veces que se repite una localización

3.3. Análisis de los datos de Indisponibilidades de Red

La tabla de Indisponibilidades 3.2 cuenta con 8 variables y 22334 filas. Esta tabla tiene muchas más filas que la tabla de Restricciones, debido a que a diario hay más indisponibilidades publicadas que restricciones.

3.3.1. Fecha

De igual forma que en el informe de restricciones la variable *Fecha* indica el día de la publicación del informe. Se puede ver la cantidad de indisponibilidades por día en 3.3, junto con su función de densidad de probabilidad y su media. Parece una distribución unimodal pero no se puede aproximar por una distribución normal.

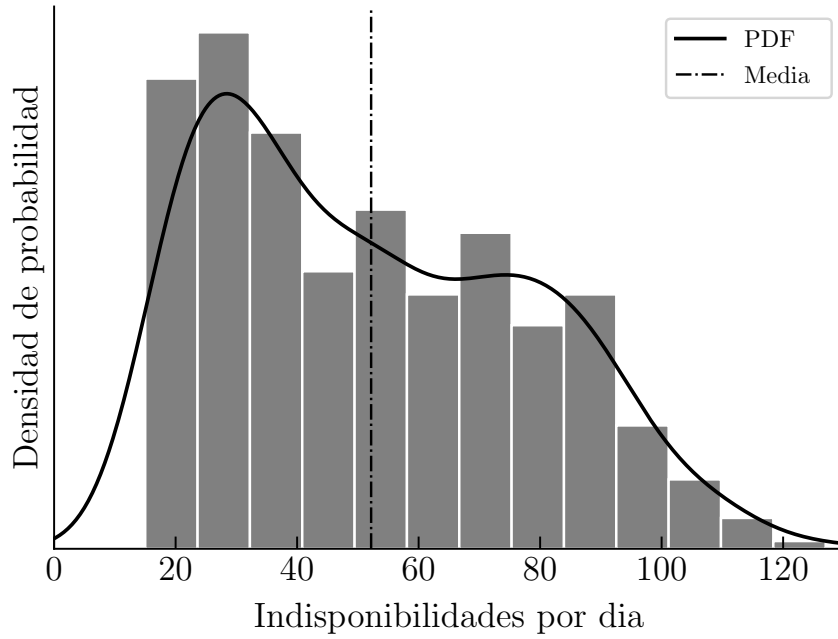


Figura 3.3: Histograma de indisponibilidades por día

3.3.2. Indisponibilidades

La variable Indisponibilidades define la indisponibilidad indicando su localización, tanto si se trata de una indisponibilidad que afecta a una línea (*220 kV SOTO DE RIBERA TRASONA*) como si afecta únicamente a un nudo (*REACTANCIA - LA LOMBA LOM24-RAT1*).

Podemos encontrar 2441 valores distintos de *Indisponibilidad*. Se puede ver la frecuencia que tienen los 20 más repetidos en la tabla 3.12.

CAPÍTULO 3. INFORMES DIARIOS DE RESTRICCIONES TÉCNICAS DEL PDBF

Indisponibilidades	Frecuencia
220 kV SOTO DE RIBERA TRASONA	428
REACTANCIA - GUADAME GMD33-RAT1	426
400 kV BALBOA SIDERURGICA BALBOA	424
BANCO DE CONDENSADORES - GALAPAGA GAL400-BCO1	416
220 kV SIERRA COSTERA VALDECONEJOS 1	377
REACTANCIA - MAGALLON MAG400-RPT1	282
220 kV PARQUE INGENIEROS VILLAVERDE 1	253
400 kV GUARDO VELILLA	247
REACTANCIA - LA LOMBA LOM24-RAT2	228
REACTANCIA - LA LOMBA LOM24-RAT1	227
400 kV MUDEJAR TERUEL 2	222
400 kV MUDEJAR TERUEL 1	222
JUEGO DE BARRAS - S.ELVIRA EVR220-BP2	221
JUEGO DE BARRAS - S.ELVIRA EVR220-BP1	221
TRANSFORMADOR DE POTENCIA - ALMARAZ ALZ-ATP1	208
220 kV CICERO SOLORZANO 1	145
220 kV BADALONA CANYET	143
220 kV CICERO SOLORZANO 2	142
JUEGO DE BARRAS - TALAVERA TAL220-BP2B	141
JUEGO DE BARRAS - TALAVERA TAL220-BP1A	139

Tabla 3.12: Variable Indisponibilidades y sus 20 valores más repetidos

Si observamos la figura 3.4 junto con la tabla 3.12, podemos identificar que, pese a que algunos valores se repiten hasta 400 veces, la mayoría se repite solo unas pocas veces. Cerca de 2000 de los 2441 valores distintos se repiten en menos de 10 informes. La gran cantidad de valores distintos y su limitada frecuencia hace que el uso de esta variable (y por extensión también esta tabla) en modelos de predicción parezca poco útil. Sin embargo, aportan información sobre localización muy útil, como se verá más adelante.

La distribución de probabilidad puede aproximarse a una distribución de Poisson de λ igual a 3.5.

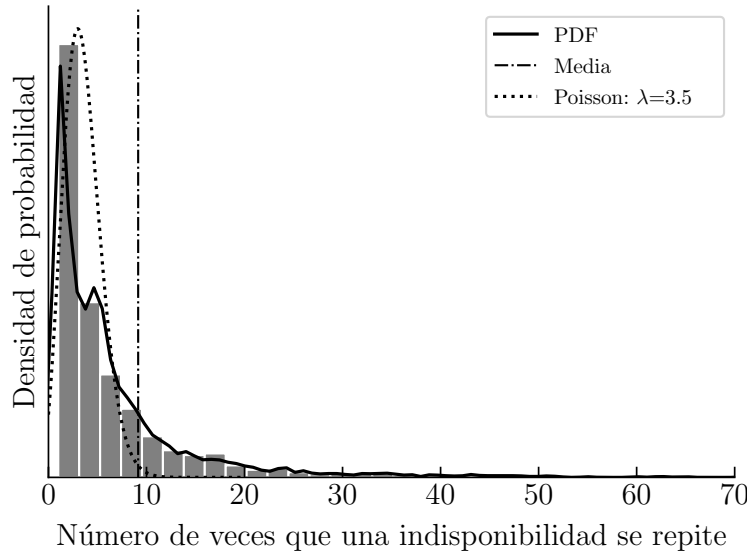


Figura 3.4: Histograma de número de veces que se repite una indisponibilidad

Al igual que la variable *Contigencias* en la tabla de Restricciones, esta variable tiene muchos problemas de consistencia. Desgraciadamente, la enorme cantidad de valores distintos hace que no sea posible hacer un diccionario para solucionarlos.

3.3.3. Inicio - Final

Estas variables indican el inicio y final de la indisponibilidad, indicando la fecha, hora y minutos.

Cabe destacar que en muchas ocasiones, la fecha final es solo aproximada y que una indisponibilidad que se repite muchos días seguidos puede cambiar de un día para otro su final.

Si observamos la hora de inicio de las indisponibilidades en la tabla 3.5 se aprecia que muy pocas empiezan antes de las 8 de la mañana y que gran parte (40 %) de las indisponibilidades comienzan a las 8 de la mañana, esto muy posiblemente se deba a que es el comienzo de la jornada laboral. La probabilidad de las indisponibilidades decrece poco a poco en las siguientes horas, habiendo muy pocas indisponibilidades que empiecen a partir de las 6 de la tarde. La media está en torno a las 10 de la mañana. Es claramente una distribución unimodal.

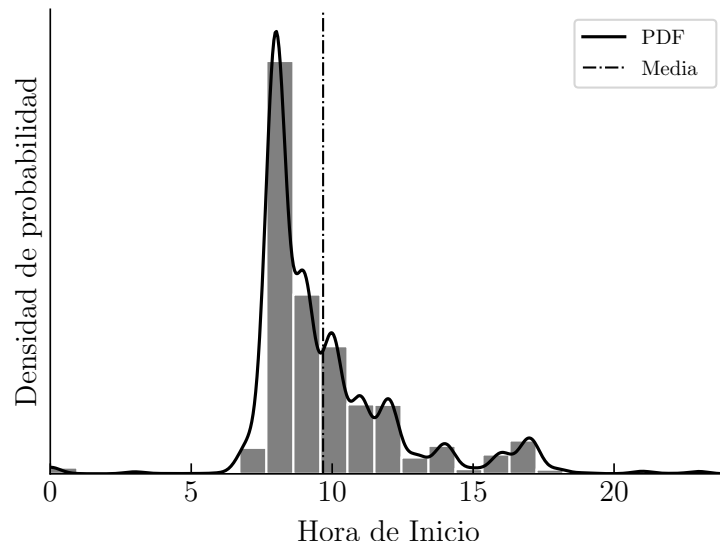


Figura 3.5: Histograma de hora de inicio de indisponibilidad

En la tabla 3.6 se puede ver el histograma de la hora del final de las indisponibilidades. A diferencia de la hora de inicio, la hora de final no parece seguir ninguna distribución, aunque cabe destacar que un alto porcentaje de las indisponibilidades terminan en torno a las 5 de la tarde (probablemente por los turnos de trabajo).

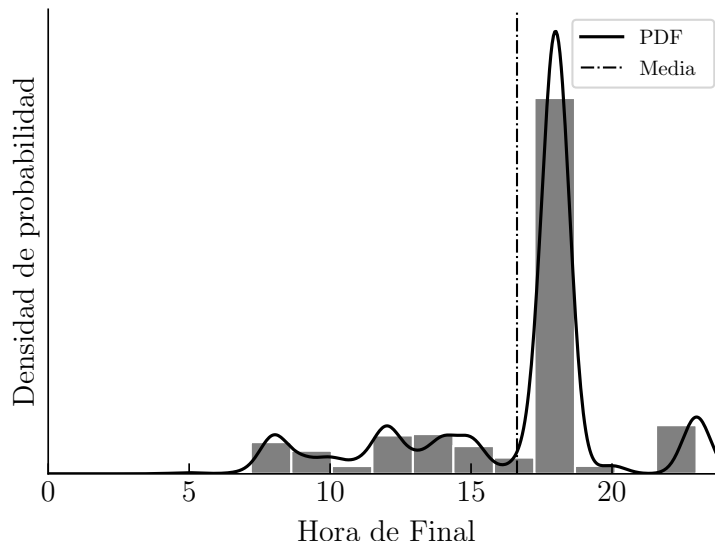


Figura 3.6: Histograma de hora de final de indisponibilidad

3.3.4. Tipo

Tipo	Frecuencia
P.A.	11888
P.S.	6848
C.P.	3412
CP	6
-	4
C.P	3

Tabla 3.13: Variable tipo y su frecuencia

Se aprecian errores de consistencia fácilmente solucionables.

3.3.5. Reposición Diario - Reposición de Emergencia

Estas variables indican la hora de la reposición o un 'NO' en caso de que no la haya.

3.3.6. Observaciones

Esta ultima variable incluye en ocasiones algún comentario sobre la indisponibilidad.

Se pueden encontrar 865 textos distintos en esta variable, pese a que solo un 32 % de las filas contengan información en esta variable. Esto implica que usarla de forma automática sea difícil.

Capítulo 4

Geolocalización de Indisponibilidades de Red

Como se ha comentado en la sección 3.2 y la sección 3.3, no se cuentan con datos geográficos ni de restricciones ni de indisponibilidades. Aun así, la tabla de indisponibilidades 3.2 permite una geolocalización muy precisa con su variable *indisponibilidades*, ya que aporta el nombre del nudo/nudos en el que ocurre la indisponibilidad, lo cual nos da más información que cualquiera de las variables de la tabla de restricciones 3.1.

4.1. Extracción de datos geográficos de nudos y líneas

Con el objetivo de usar esta información para localizar correctamente los datos de indisponibilidades se ha buscado extraer los datos de localización nudos y líneas. El nombre de los nudos de la red de transporte española, su tensión y la comunidad

autónoma a la que pertenecen son datos públicos y se pueden obtener en [23], pero la información geográfica exacta no es un dato publicado a pesar de que sí existe un mapa en PDF [24] con todos los nudos y líneas a fecha de 2018. Aun así, este mapa no se puede procesar para obtener las localizaciones de los nudos.

La Red Europea de Gestores de Redes de Transporte de Electricidad (ENTSO-E) es una asociación que promueve la cooperación de los operadores de las redes de transporte en Europa. A ella pertenecen 42 operadores de la red de transporte de 35 países distintos [25].

ENTSO-E ofrece en su página web un mapa interactivo con los datos de nudos

y líneas de los operadores que pertenecen a ella [26]. Este mapa tiene una función meramente ilustrativa y por tanto no indica la localización exacta de los nudos y líneas, que se explicará en mayor detalle en otras partes del trabajo. En su página web afirman que sus datos son del 01/01/2019. No obstante, pese a ofrecer un mapa interactivo, la información de este no es pública por motivos de licencia, lo cual complica el proceso de encontrar la localización exacta, aunque al mostrar la información en el mapa y ser este interactivo, es posible extraer parte de los datos de este sitio.

4.1.1. Scraping del mapa de red de ENTSO-E

Este mapa utiliza *Mapbox* para funcionar. *Mapbox* es un proveedor de mapas en línea para páginas web [27] parecido a Google Maps.

Si inspeccionamos la página web, se puede apreciar que el mapa se inserta en la página web con un *iframe*, que permite incrustar una página web dentro de otra. En este caso, la página que se incrusta es <https://transmission-system-map.entsoe.eu/>. Al explorar el contenido de esta, se puede ver que los nudos y líneas no están directamente insertados en el archivo HTML al tratarse de un mapa interactivo, y por tanto no se pueden extraer los datos mirando únicamente este archivo, como suele suceder en los casos de *web scraping*.

El siguiente paso consiste en analizar los recursos de la página web, buscando si en alguno de los archivos se encontraran los datos que usan el mapa interactivo. Sin embargo, no es el caso, en los archivos solo se encuentran los archivos *.html*, *.css* y *.js* que la página usa para funcionar. Se pueden encontrar archivos de la API de Mapbox. Todo esto solo deja lugar a una última opción.

La información de nudos y líneas se tiene que servir a través de peticiones a alguna API en tiempo real cuando cargamos el mapa o cuando nos movemos por él. Para encontrar estas llamadas, se analiza el tráfico de red de la página web. Ahí encontramos algunas llamadas iniciales y otras distintas según nos movemos por el mapa o hacemos zoom a la API de Mapbox que devuelven unos archivos *.vector.pbf*. Analizando las partes de llamadas, distinguimos entre

- API: todas las llamadas empiezan con `https://api.mapbox.com/v4/` ya que esta es la API de Mapbox.
- Tipo de dato: Aquí podemos distinguir entre dos tipos de datos distintos que se pueden pedir.

- `/entsoe.1nr7sorj,mapbox.mapbox-terrain-v2`: proporciona los datos de la topografía del mapa, como tierra/mar, ríos o carreteras que se ven en el mapa.
 - `/entsoe.0ol41nja`: aunque en principio no se pueda saber, estos datos deberían de ser los de nudos y líneas.
- Posición de los datos: los datos no se envían a la vez por motivos de eficiencia, sino que se proporcionan solo los que son necesarios en función de la posición del mapa y el zoom. Mapbox usa un sistema de cuadrícula Z,X,Y y por tanto, para indicar la posición se pasa un `/Z/X/Y.vector.pbf`. La coordenada Z indica el nivel de zoom (existen 22 niveles distintos[ref a la docu]) y las coordenadas X e Y indican la posición en el mapa, ya que tienen una conversión a latitud y longitud. Por último, `.vector.pbf` indica la extensión del archivo.
 - Token: por último, se envía un *token* de acceso como medida de seguridad. Al ser nuestro ordenador el que realiza la llamada, podemos cogerlo y hacer todas las llamadas que se requieran desde cualquier lado, ya que este *token* ni siquiera cambia entre sesiones, es siempre el mismo y por tanto no caduca.

`https://api.mapbox.com/v4/entsoe,0ol41nja/3/3/2.vector.pbf?access_token=pk.e...`

API Tipo Zona Token

Ecuación 4.1: Ejemplo de llamada a la API de Mapbox para obtener datos de nudos y líneas

Los archivos que se devuelven son formato `.vector.pbf` o `pbf` corresponde a *Protocol Buffer* [28]. Este protocolo permite comprimir de forma eficiente información estructurada. Estos archivos no pueden ser leídos directamente, ya que no son texto plano, sino archivos binarios. Para convertirlos en texto plano que podemos procesar usamos la librería de *mapbox vector-tile-py* para *Python* [29] que, pese a no estar ya mantenida, permite hacer fácilmente esta función.

Vector-tile-py permite convertir los archivos `.vector.pbf` en formato *geojson* [30] si se le proporciona la Z,X,Y del archivo. Los archivos *geojson* permiten almacenar fácilmente distintos tipos de datos geográficos estructurados siguiendo una estructura basada en *JSON*.

Después de usa *vector-tile-py* para transformar uno de los archivos `vector.pbf` en *geojson*.

Analizando el contenido de los diferentes *geojson* se ha comprobado que, si se quieren obtener todos los nudos y líneas de España, se necesitan las cuadrículas mostradas en 4.1. Estos datos también incluyen parte de los nudos y líneas de Francia y parte del Norte de África, pero son necesarios si queremos también los datos de las Islas Canarias y las Islas Baleares.

Cuadrículas
5/15/12
5/15/11
5/16/11
5/16/12

Tabla 4.1: Cuadrículas necesarias para obtener todos los datos de España

4.1.2. Estructura de los datos

El archivo *geojson* sigue la estructura de 4.2. El archivo es una lista de *Features*, que están contenidas dentro de "features".

```
{ "type": "FeatureCollection",
  "features": [ {} ... {} ]
}
```

Ecuación 4.2: Estructura Geojson

Tal como se suponía, */entsoe.0ol41nja* corresponde a los datos de nudos y líneas y */entsoe.1nr7sorj,mapbox.mapbox-terrain-v2* a datos de la topografía. A partir de ahora solo se trabajará con los datos de nudos y líneas. Podemos encontrar 2 tipos de *features*, líneas y nudos. En 4.3 podemos observar la estructura de los nudos.

```
{ "type": "Feature",
  "properties": {
    "OBJECTID": 33886,
    "Symbol": "Substation",
    "Country": "ES",
    "Name_Eng": "Almazan",
    "MW": -1,
    "Visible": 1,
    "TSO": " ",
    "Tie_line_s": 0,
```

```

    "Under_cons": 0,
    "visible_on_printed": 1,
    "layer": "points"
  },
  "geometry": {
    "type": "Point",
    "coordinates": [
      -2.56805419921875,
      41.399475357337565
    ]
  }
}

```

Ecuación 4.3: Estructura Nudo

En 4.4 podemos ver la estructura que siguen las líneas.

```

{
  "type": "Feature",
  "properties": {
    "OBJECTID": 27006,
    "Symbol": "220_kV_Transmission_Line_1_Circuit",
    "Text_": "_",
    "Tie_line": 0,
    "T9_Code": "0",
    "Visible": 1,
    "VoltageLev": "220-275",
    "Current_": "AC",
    "UnderConst": "0",
    "NumberOfCi": 1,
    "Undergroun": 0,
    "oneCircuit": 0,
    "ABR": "ES",
    "visible_on_printed": 1,
    "layer": "lines"
  },
  "geometry": {
    "type": "LineString",
    "coordinates": [
      [-8.536376953125, 43.305193797650546],
      [-8.445739746093748, 43.31718491566705],
      [-8.368835449218748, 43.33916248737746],
      [-8.297424316406248, 43.305193797650546]
    ]
  }
}

```

$$\left\{ \begin{array}{l} \left[-8.3441162109375, 43.167125915000284 \right] \\ \end{array} \right\}$$

Ecuación 4.4: Estructura Línea

A continuación, se enumeran las variables comunes a nudos y líneas mas relevantes.

- OBJETID: contiene un ID. Por motivos que desconocemos este ID no es realmente único y se pueden encontrar nudos y líneas que compartan ID. Se encuentra dentro de *properties*.
- Symbol: todos los nudos tienen *substation* en esta variable y todas las líneas contienen *Transmission Line* lo cual es perfecto para identificarlos. Se encuentra dentro de *properties*.
- Name_Eng: esta variable indica el nombre de la subestación.
- Coordinates: Los nudos contienen únicamente unas coordenadas geográficas que indican la latitud y longitud del nudo. Curiosamente se indica primero la longitud y después la latitud. En las líneas contiene una lista de coordenadas, que contienen los distintos puntos de una línea. Se encuentra dentro de *geometry*. El tipo de geometría de las líneas es *LineString* y en los nudos es *Point*.

Los nudos tienen las siguientes variables relevantes únicas:

- Country: Indica el país del nudo siguiendo el código ISO 3166 [31]. Para recoger todos los nudos de España simplemente hay que filtrar todos los nudos que contienen *ES* en esta variable. Se encuentra dentro de *properties*.

Las líneas tienen las siguientes variables relevantes únicas a ellas.

- VoltageLev: indica el nivel de tensión de la línea, pese a que este también se puede extraer de la variable *Symbol*. Se encuentra dentro de *properties*.
- Current_: indica si es una línea de corriente alterna *AC* o continua *DC*

Uniando los *features* de los 4 *geojson* (uno por cada cuadrícula del mapa 4.1) en uno solo obtenemos un único *geojson* con todos los datos. En estos datos se

cuenta con la información de 606 nudos de España, estos nudos no son todos los de España, como se verá más adelante. Es difícil determinar cuántas de las líneas de España se tienen, ya que los datos de líneas no cuentan con una variable que indique el país o países en el que se encuentra.

4.2. Análisis de los datos

4.2.1. Visualización de los datos del mapa de ENTSO-E

Para la visualización de los datos de nudos y líneas (y todos los datos geográficos a partir de este momento) se va a usar la API de Google Maps a través de la librería `jupyter-gmaps` de python [32].

En la figura 4.1 podemos ver los datos de las líneas puestas sobre el mapa. Las líneas de 220kV en color azul, las de 400kV en color rojo las de 150kV en color negro y las de DC en color morado.



Figura 4.1: Datos geográficos de las líneas de España

En la figura 4.2 podemos ver los datos de los nudos puestos sobre el mapa, representados con puntos rojos.



Figura 4.2: Datos geográficos de los nudos de España

Juntando las dos figuras anteriores en una sola obtenemos la figura 4.3, donde podemos ver los datos de las líneas y de los nudos juntos sobre el mapa.

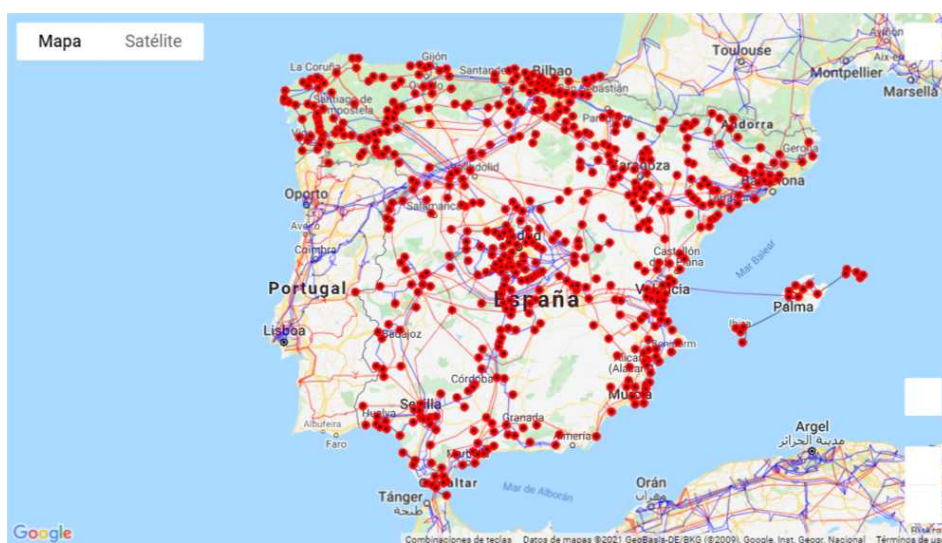


Figura 4.3: Datos geográficos de los nudos y líneas de España

4.2.2. Precisión de la geolocalización de los nudos y líneas

Si comparamos el mapa de líneas obtenido de ENTSO-E 4.1 con el mapa de REE de 2017 que se puede ver en 11.2, se puede apreciar que parece bastante

preciso excepto en la zona de Madrid. Esto se debe a que los datos se han obtenido de un mapa ilustrativo, y en este mapa, cuando hay muchos nudos o líneas cerca, se separan entre ellos para mejorar la visualización. Barcelona también es una zona con alta concentración de nudos, pero en este caso no aparecen deformados porque no se cuenta con los datos de muchos de los nudos pequeños de 220kV de esta zona, mientras que sí se cuentan con los de Madrid.

Para intentar cuantificar cómo de preciso es el mapa (exceptuando la zona de Madrid) se han cogido 5 nudos al azar, que no pertenezcan a Madrid cuya localización se haya podido encontrar en Google Maps y se han comparado con la localización de nuestros datos. Se puede encontrar la posición de los nudos con su posición real y la posición con la que contamos en la figura 4.4. Los puntos azules representan la posición real del nudo y los rojos la posición con la que contamos.

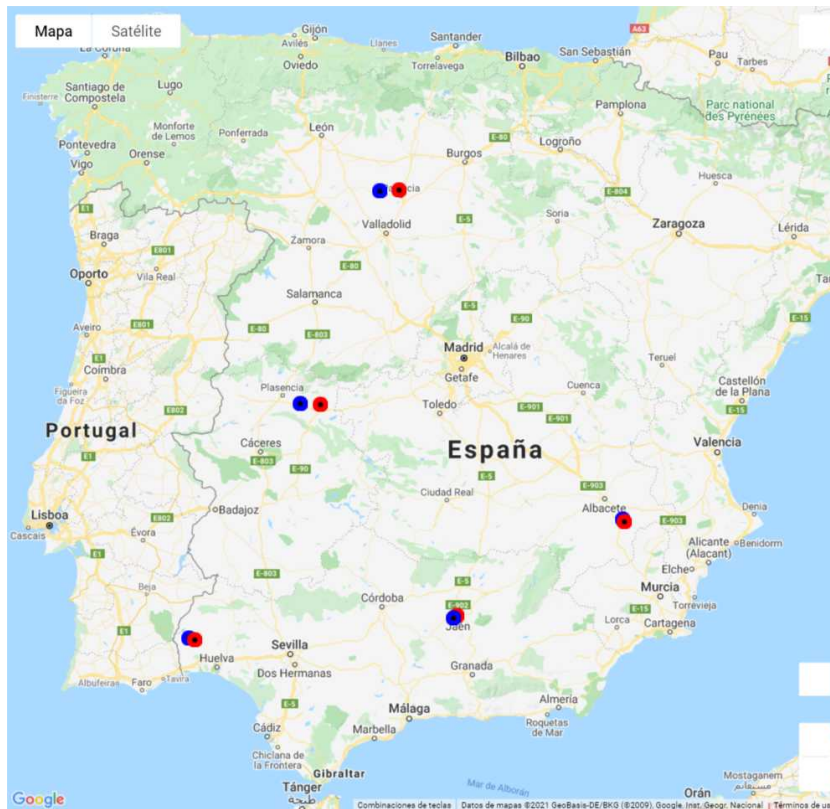


Figura 4.4: Nudos con su posición real la posición de ENTOSOE

Podemos ver una tabla con la distancia entre el nudo real y nuestro nudo en 4.2. La distancia media es 11.89 Km.

Nudo	Distancia
Arañuelo	22.588350
Olivares	5.250728
Puebla de Guzman	6.694026
Pinilla	3.797446
Grijota	21.157477

Tabla 4.2: Distancia entre la posición real de 5 nudos y la posición con la que contamos

A pesar de contar únicamente con 5 datos de distancia, son suficientes para hacer un test de hipótesis usando la T de Student con el objetivo de determinar a un nivel de significancia del 5% cual es la distancia media máxima (μ_{max}) que puede haber entre la posición de nuestros nudos y la real. Se asume por tanto normalidad en la distribución de las distancias. La hipótesis nula es por tanto.

$$H_0 : \mu \geq \mu_{max}$$

frente a la alternativa:

$$H_1 : \mu < \mu_{max}$$

Con una muestra de 5 y siendo un test de hipótesis de una cola, para rechazar la hipótesis nula con un nivel de significancia del 5%, $\mu_{max} \geq 20,648$ y por tanto se acepta que la media de las distancias entre las posiciones reales de los nudos y las posiciones con las que contamos es menor o igual de 20.64 km.

Para cuantificar la deformación de los nudos y líneas de Madrid, se han cogido 3 nudos cuya localización real se ha podido encontrar en Google Maps. Se puede ver la deformación de estos nudos en la figura 4.5. Se aprecia que estos nudos se deforman radialmente desde aproximadamente el centro de Madrid.



Figura 4.5: Nudos de Madrid con su posición real y la posición con la que contamos)

En la tabla 4.3 se observa la distancia entre la posición con la que contamos y la real, las diferencias en las distancias entre los nudos (tanto con la localización real como con la que disponemos) y el centro de Madrid. También se ve el ratio entre estas 2 distancias. Se puede apreciar que la media de los ratios está en 6.5.

Nudo	Distancia (km)			
	Entre posiciones	Al centro	Real al centro	Ratio
Majadahonda	89.49	105.64	16.63	6.35
Leganes	56.30	66.90	11.52	5.81
La Torrejilla	84.74	96.47	13.13	7.35
Media	76.85	89.67	13.76	6.50

Tabla 4.3: Distancia de la posición de los nudos al centro de Madrid

Podemos ver el resultado de acercar estos 3 nudos a razón de 6.5 en la figura 4.6. El resultado es bueno, se pueden observar las distancias entre las posiciones reales de los nudos y las posiciones con las que contamos en la tabla 4.4.

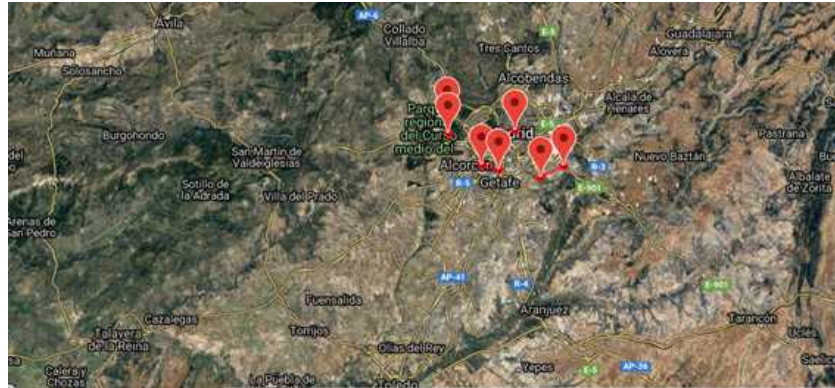


Figura 4.6: Nudos de Madrid con su posición real y la posición con la que contamos después de aplicarles una transformación

Nudo	Distancia (km)			
	Entre posiciones	Al centro	Real al centro	Ratio
Majadahonda	3.79	16.25	16.63	0.98
Leganes	4.19	10.29	11.52	0.89
La Torrecilla	6.16	14.81	13.13	1.13
Media	4.71	13.78	13.76	1.00

Tabla 4.4: Distancia de la posición de los nudos al centro de Madrid después de aplicar la transformación

En la tabla 4.4 se puede apreciar como la media de la distancia entre la posición real y la posición con la que contamos es tan solo 4.71 km. Al igual que en los nudos de fuera de Madrid se va a realizar un test de hipótesis. En este caso se cuentan solo con 3 datos de distancia, pero siguen siendo suficientes para hacer un test de hipótesis usando la T de Student. El objetivo es el mismo: determinar a un nivel de significancia del 5%, cual es la distancia media máxima (μ_{max}) que puede haber entre la posición de nuestros nudos y la real. Se asume normalidad en la distribución de las distancias, pero al estar aplicando una transformación a las posiciones podría no ser así.

La hipótesis nula es por tanto.

$$H_0 : \mu \geq \mu_{max}$$

frente a la alternativa:

$$H_1 : \mu < \mu_{max}$$

Con una muestra de 3 y siendo un test de hipótesis de una cola, para rechazar la hipótesis nula con un nivel de significancia del 5 %, $\mu_{max} \geq 5,94$ y por tanto se acepta que la media de las distancias entre las posiciones reales de los nudos y las posiciones con las que contamos para los nudos de Madrid es menor o igual de 5.94 km.

Es interesante observar como esta distancia es mucho menor que la que de los nudos de fuera de Madrid, pero, teniendo en cuenta que en la zona de Madrid hay mucha densidad de nudos, el resultado se vuelve comprensible.

No se puede aplicar esta transformación a todos los nudos que estén a menos de una cierta distancia de Madrid, ya que se estarían transformando también algunos que estaban bien colocados. Para solucionar esto, se han tomado todos los nudos que están a menos de 140 km del centro y de esta lista se han excluido a mano todos los que están bien colocados. Para definir los bien colocados se ha hecho comparando a mano nuestro mapa de nudos, con el de REE de 2017 11.2. Se puede ver el área de detección de los nudos alrededor de Madrid en la figura 4.7. Las listas de nudos transformados y excluidos se encuentran en el anexo 11.3



Figura 4.7: Área de detección de los nodos de Madrid (círculo de 140km de radio)

En la figura 4.8 se pueden ver los nodos de dsnkdnd en su localización original y en la figura 4.9 se pueden ver después de aplicarles la transformación a razón de 6.5.



Figura 4.8: Localización original de los nodos de Madrid



Figura 4.9: Localización de los nudos de Madrid después de aplicarles una transformación

Se puede apreciar que la localización de los nudos de la figura 4.9 se parece mucho a la localización de los nudos del mapa de REE de 2017 11.2, en el sentido de que todos están muy agrupados.

No es posible aplicar la misma transformación a las líneas y obtener la misma calidad de resultado, ya que las líneas pueden tener puntos que se tengan que acercar y otros que no. Además, las líneas no siempre empiezan en un nudo y terminan en otro, sino que en muchas ocasiones las líneas empiezan en un nudo pero terminan en un lugar donde no hay nudo, del que otras líneas parten para llegar a nudos.

Conociendo esto, la alternativa por la que se ha optado ha sido acercar todos los puntos de las líneas que estén a menos de 140 km del centro de Madrid, excluyendo por supuesto los puntos que coincidan con la localización de alguno de los nudos excluidos del anexo 11.3. Se puede ver las líneas afectadas por este área en la figura

4.10.



Figura 4.10: Área de detección de las líneas de Madrid (círculo de 140km de radio)

Se puede ver en la figura 4.11 las líneas en su localización original y en 4.12 se pueden ver después de aplicarles la transformación comentada.

CAPÍTULO 4. GEOLOCALIZACIÓN DE INDISPONIBILIDADES DE RED

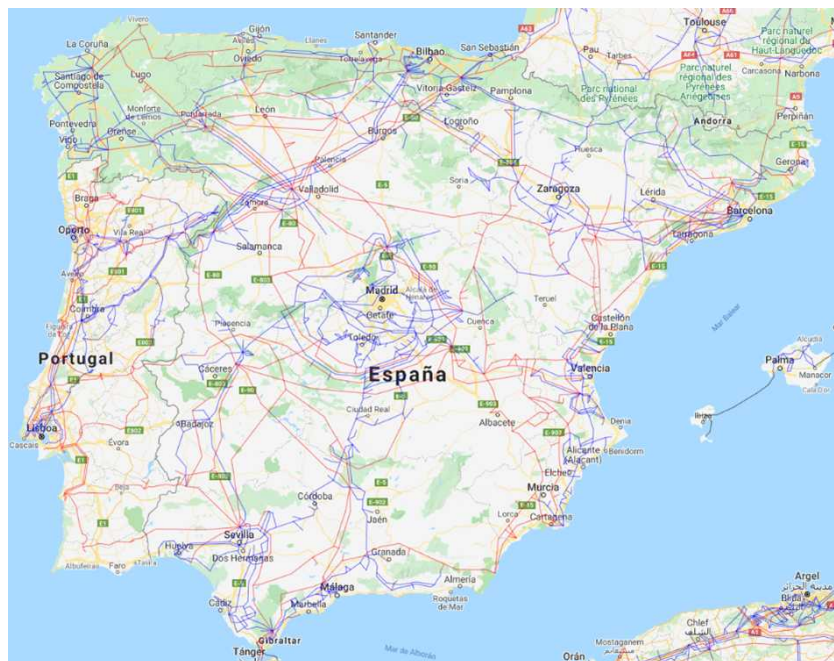


Figura 4.11: Mapa con la localización original de la líneas

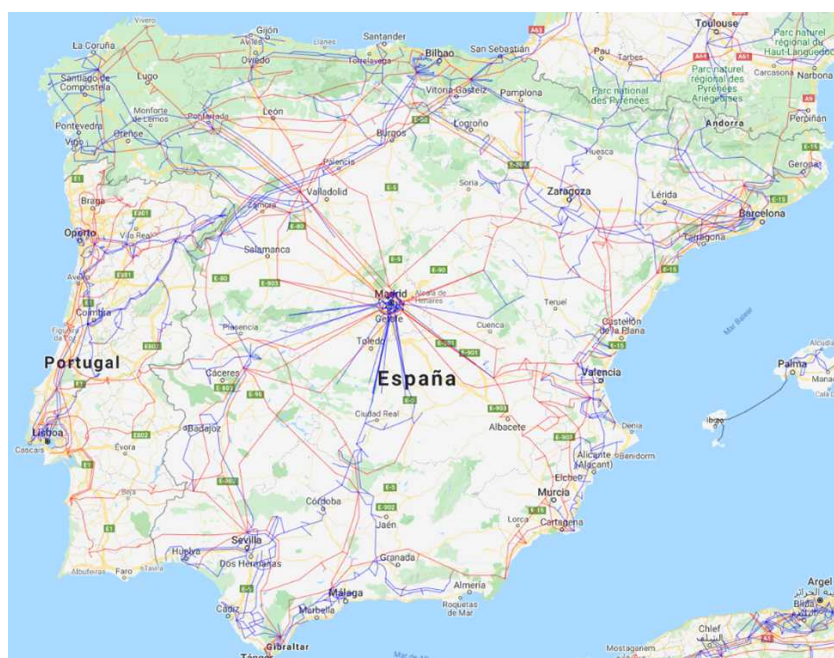


Figura 4.12: Mapa con la localización de la líneas después de aplicar una transformación alrededor de Madrid

Si se compara la figura de 4.12 con el mapa de REE de 2017 se puede apreciar que se parece más a lo real que lo que se tenía en la figura 4.11, pero que como ya se había comentado, el resultado no es perfecto, y hay ciertos puntos de las líneas que se han acercado cuando no deberían. Esto es muy claro en la línea que va desde La Mudarra a Almazán.

4.3. Geolocalización de Indisponibilidades de Red

Los datos geográficos de los nudos de REE comentados en 4.1 se pueden usar para localizar las indisponibilidades al contener estas el nombre del nudo en el que se encuentran.

El funcionamiento es el siguiente:

- Para cada valor distinto de la variable *Indisponibilidad* de la tabla de indisponibilidades 3.2 se busca si aparecen los nombres de los nudos del mapa de ENTSO-E, que están en la variable *Name_Eng*.
- Para evitar problemas, tanto el texto de la variable *Indisponibilidades* como la variable *Name_Eng* se convierten en minúsculas.
- Se utilizan solo los nudos de España aunque también se incluyen los nudos de conexión en las fronteras que se ven en la tabla 4.5.

Nudos de conexión
Baixas
Argia
Pragneres
Pocinho
FARDIOUA
Lindoso

Tabla 4.5: Nudos de conexión

Al llevar a cabo este proceso, se han identificado varios problemas:

- No se cuentan con los datos de todos los nudos de España: el mapa de ENTSO-E cuenta con datos de 606 nudos de España. España cuenta con

766 nudos en la actualidad (Dato calculado usando los datos de capacidad de acceso publicados el 2 de Agosto de 2021 [23]. Cogiendo los nudos que aparecen, eliminando el nivel de tensión y viendo el número de elementos únicos). Se ha visto que los nudos que principalmente faltan son nudos de 220kV de la zona de Barcelona aunque también faltan otros nudos por toda España.

Para solucionar este problema, manualmente se ha considerado que esos nudos pertenecen al nudo más cercano. Este proceso se ha realizado solo con los nudos que han aparecido alguna vez en la tabla de Indisponibilidades.

- Nudos abreviados o con distintos nombres en las diferentes fuentes: Se pueden observar 7 ejemplos de esto en la tabla 4.6.

Nombre en tabla de Indisponibilidades	Nombre en ENTSO-E
Riba-roja	Riba Roja
Archidona	Imora
Tajo encantada	Tajo de la Encantada
J.M.Oriol	J.M. de Oriol
Oriol	J.M. de Oriol
Espartal	El Espartal
Hoya Morena	N. Hoya Morena
Roda de Andalucia	La Roda de Andalucia

Tabla 4.6: Nudos con distinto nombre entre el mapa de ENTSO-E y la tabla de Indisponibilidades (limitado a 7 resultados)

Para solucionarlo se ha creado un diccionario de excepciones para que al buscar un determinado nudo, se busque también con todos los posibles nombres que se pueden encontrar.

- Nombres de nudos que contienen a otros: Se puede ver 7 ejemplos de esto en la tabla 4.7, aunque hay 27 casos. Algunos de los casos se han solucionado buscando el nombre de los nudos añadiéndoles un espacio en blanco delante y detrás. También hay que añadirlos al texto de indisponibilidades por los casos en los que el nombre del nudo aparece al final o al principio. Como consecuencia, también hay que buscar los nombres de los nudos con un guión delante o un guión detrás ya que en el texto de *indisponibilidades* cuando hay más de un nudo a veces aparece *nudo1-nudo2*. El resto de los casos se

4.3. Geolocalización de Indisponibilidades de Red

han tenido que solucionar añadiendo en el código condiciones particulares para estos casos.

Nudo	Contenido en
Puerto	Puerto Real
Puerto	Puerto de la Cruz
Puerto	Puertollano
Gurrea	Alcalá de Gurrea
Mora	Moraleja
Mora	Morata
Villanueva	Villanueva del Rey

Tabla 4.7: Nombres de nudos contenidos en otros (limitado a 7 resultados)

Una vez solucionados los problemas comentados se puede obtener la tabla 4.8, donde para cada indisponibilidad se tienen los nudos/nudo involucrados y la geolocalización (latitud y longitud). Para los datos de geolocalización se cogen los datos del Nudo 1.

FECHA	INDISPONIBILIDADES	Nudo 1	Nudo 2	Latitud	Longitud	Inicio	Final
2020-01-01	220 kV ACERIASA CACICEDO	Cacicedo	Aceriasa	43.456906	-3.930359	2019-12-19 09:29:00	2020-01-04 18:00:00
2020-01-01	220 kV CICERO SOLORZANO 1	Cicero	Solorzano	43.419014	-3.567810	2016-09-29 11:02:00	2021-12-31 23:59:00
2020-01-01	220 kV CICERO SOLORZANO 2	Cicero	Solorzano	43.419014	-3.567810	2016-09-29 11:02:00	2021-12-31 23:59:00
2020-01-01	220 kV SOTO DE RIBERA TRASONA	Tabiella	Soto de Ribera	43.548548	-6.006775	2015-10-09 12:49:00	2030-02-28 23:59:00
2020-01-01	400 kV BALBOA SIDERURGICA BALBOA	Sider. Balboa	Balboa	38.296404	-7.009277	2018-03-01 17:13:00	2020-12-21 12:00:00

Tabla 4.8: Tabla de Indisponibilidades con los nudos en los que aparece (limitado a 5 resultados y quitando parte de las variables)

A la tabla 4.8 se le ha añadido una nueva variable *Tiempo tot*, donde se calcula el tiempo total de la indisponibilidad usando la variable *Inicio* y la variable *Final* (estas variables no se ven en la tabla 4.8 aunque pertenezcan a ella por motivos de espacio). Al tener datos diarios, *Tiempo tot* solo calcula el tiempo que la indisponibilidad está activa en el día concreto que se publicó el informe (variable *FECHA*), sin importar que empezara días atrás o que parezca que va a continuar durante más días. Si esa indisponibilidad aparece también al día siguiente, se contará ese tiempo en la variable *Tiempo tot* de esa fila. Se puede ver el resultado en la tabla 4.9.

CAPÍTULO 4. GEOLOCALIZACIÓN DE INDISPONIBILIDADES DE RED

FECHA	INDISPONIBILIDADES	Nudo 1	Nudo 2	Inicio	Final	Tiempo tot (s)
2020-01-01	220 kV ACERIASA CACICEDO	Cacicedo	Aceriasa	2020-01-01	2020-01-02	86400.0
2020-01-01	220 kV CICERO SOLORZANO 1	Cicero	Solorzano	2020-01-01	2020-01-02	86400.0
2020-01-01	220 kV CICERO SOLORZANO 2	Cicero	Solorzano	2020-01-01	2020-01-02	86400.0
2020-01-01	220 kV SOTO DE RIBERA TRASONA	Tabiella	Soto de Ribera	2020-01-01	2020-01-02	86400.0
2020-01-01	400 kV BALBOA SIDERURGICA BALBOA	Sider. Balboa	Balboa	2020-01-01	2020-01-02	86400.0

Tabla 4.9: Tabla de Indisponibilidades con los nudos en los que aparece y el tiempo total en el día (limitado a 5 resultados y quitando parte de las variables)

La tabla 4.9 se puede agrupar por indisponibilidades distintas, sumando la variable *Tiempo tot* obteniéndose la tabla 4.10.

INDISPONIBILIDADES	Latitud	Longitud	Tiempo tot
220 kV ABADIANO MONDRAGON	43.070894	-2.612000	165600.0
220 kV ABADIANO SIDENOR	43.137070	-3.029480	565200.0
220 kV ABADIANO VITORIA	43.046813	-2.754822	1353600.0
220 kV ABRERA PUJALT	41.797936	1.266174	381600.0
220 kV ABRERA RUBI	41.537366	1.845703	36000.0

Tabla 4.10: Tabla de Indisponibilidades con su geolocalización y el tiempo total activa durante el año 2020 (limitado a 5 resultados)

Con la tabla 4.10 ya se pueden visualizar las indisponibilidades de REE en un mapa. En la figura 4.13 se pueden ver en un mapa de calor de las indisponibilidades durante el año 2020. Para extraer esta figura, se ha tenido en cuenta el tiempo total que una indisponibilidad está activa durante el año (variable *Tiempo tot*).

4.3. Geolocalización de Indisponibilidades de Red

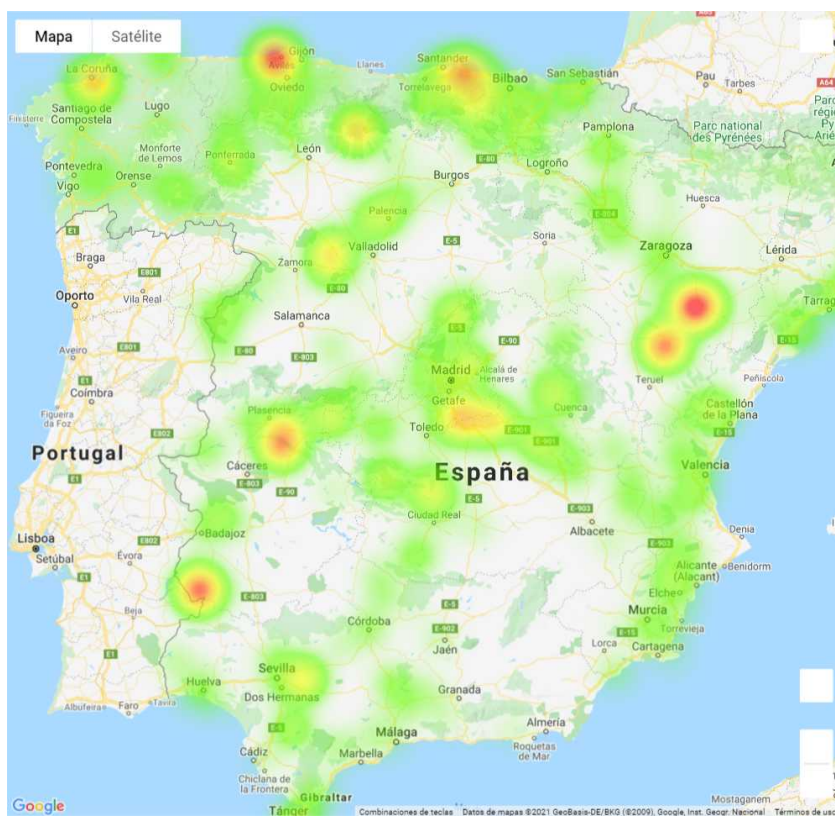


Figura 4.13: Mapa de calor de las indisponibilidades durante 2020

La distribución de indisponibilidades en el mapa cambia mucho entre los distintos meses, se puede ver claramente en la figura 4.14 y la figura 4.15, que muestran las indisponibilidades de junio y septiembre de 2020 respectivamente.

CAPÍTULO 4. GEOLOCALIZACIÓN DE INDISPONIBILIDADES DE RED

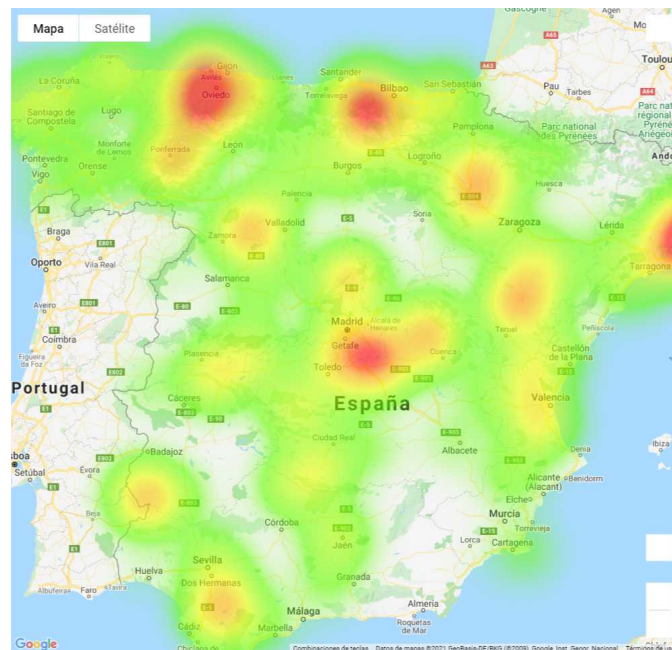


Figura 4.14: Mapa de calor de las indisponibilidades durante 06-2020

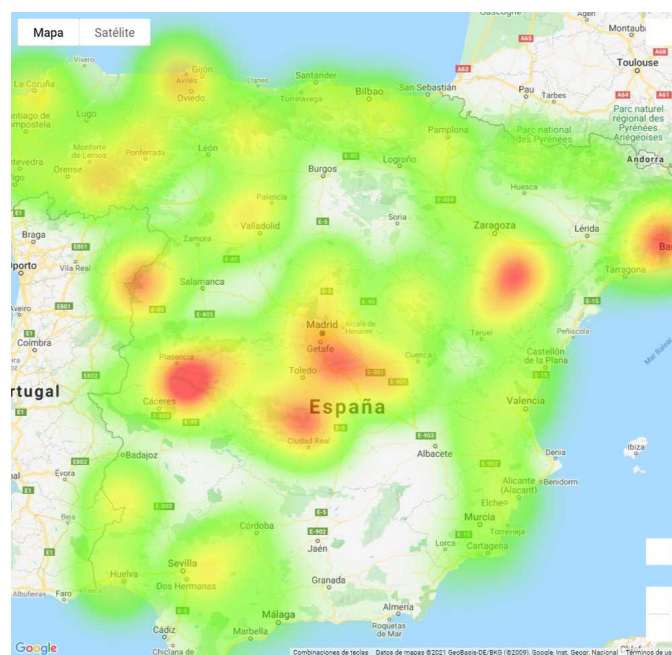


Figura 4.15: Mapa de calor de las indisponibilidades durante 09-2020

Capítulo 5

Geolocalización de la generación en España

El objetivo de este capítulo es localizar la generación en España con el propósito de usar esta información para relacionar generación real con restricciones e indisponibilidades. Para ello, se han extraído los datos de las instalaciones de generación de España.

5.1. Extracción de datos geográficos de instalaciones de generación

Los datos de potencia instalada y generada por comunidad y tecnología son públicos y fácilmente accesibles usando la API de REE [33], pero si se quiere mayor granularidad en los datos no hay ninguna alternativa pública.

De la misma forma que con los nudos, se ha buscado si esta información podría estar disponible en algún mapa interactivo *online*.

REE desarrolló un sistema de información llamado Sistema de Información del Operador del Sistema (E-SIOS). Según la legislación vigente, el Operador del Sistema (OS) tiene la obligación de hacer públicos los resultados de los mercados o procesos de operación del sistema, ESIOS se encarga tanto de almacenarlos como de publicarlos, además de funciones adicionales de poca relevancia para el presente estudio [34].

A 24/07/2019, ESIOS publicó una serie de mapas interactivos de las instalaciones de producción de España. Uno de estos mapas contiene la ubicación de

todas las instalaciones de producción de España, con su tecnología y su potencia en el momento de la publicación [35]. Se puede ver una imagen del mapa en la figura 5.1. Cada círculo representa una instalación o agrupación de instalaciones de producción. Cada color representa una tecnología, como se ve en la leyenda y a mayor tamaño del círculo, mayor es la potencia instalada en ese punto. Cuando se pincha en un punto, se indica el municipio en el que se encuentra, su tecnología, sus MW y el número de instalaciones que agrupa.

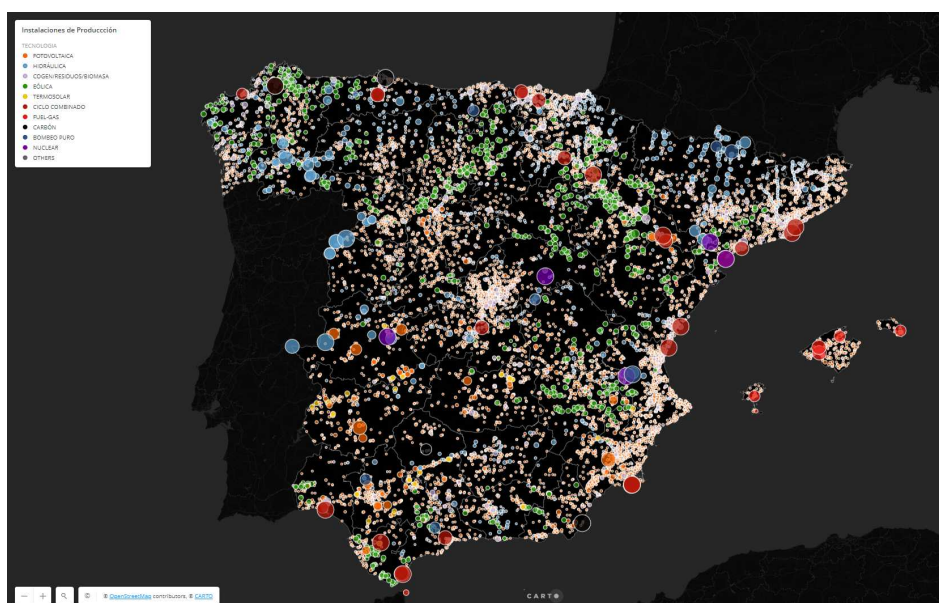


Figura 5.1: Mapa de instalaciones de generacion de Esios

5.1.1. Scraping del mapa de red de ESIOS

Con el objetivo de extraer la información de este mapa se han seguido los mismos pasos que se siguieron con el mapa de ENTSO-E de nudos y líneas. El proveedor del mapa interactivo es *CARTO* [36], que, al igual que *Mapbox*, ofrecen servicios parecidos a los ofrecidos por la API de Google Maps.

En primer lugar, se ha inspeccionado el *html* de la página por si los resultados se mostraban ahí. Se ha visto que el mapa se inserta mediante un *iframe* que incrusta <https://ree.carto.com/u/liquidaciones/builder/1bf...a7ca/embed> en la página web de ESIOS. Analizando la página web del *iframe*, se puede ver que el contenido del mapa no está en el *html*. Sin embargo, al pinchar en un círculo aparece un "bocadillo" que sí aparece en el código *html* bajo la clase *CDB-infowindow-container* y

el contenido de este bocadillo también se encuentra recogido ahí. Aun así, como los círculos no aparecen en el *html*, no se puede automatizar este proceso.

Se ha probado entonces a analizar los recursos de la página en busca de algún archivo que pudiera contener todas las instalaciones y su localización, pero no es el caso.

Por último, se ha analizado el tráfico de la página web, con el objetivo de encontrar *request* a alguna API que le proporcione la información. Entre las diferentes llamadas que se realizan, se han encontrado 2 tipos.

- Vemos por un lado llamadas a `d.basemaps.cartocdn.com/rastertiles/dark_nolabels`. Estas llamadas incluyen al final un `/Z/X/Y@2x.png` en el que se indica el zoom (Z) y la posición que se pide (X,Y). Este sistema de cuadrícula es muy parecido al que usa *Mapbox*. Estas llamadas devuelven imágenes en formato PNG con un trozo del fondo del mapa, se puede ver un ejemplo de una de estas imágenes en la figura 5.2.



Figura 5.2: Foto PNG de un trozo del fondo del mapa de generadores

En la figura 5.3 se puede ver cómo es la cuadrícula que *Carto* usa (esa

cuadrícula es solo para un determinado valor de zoom).

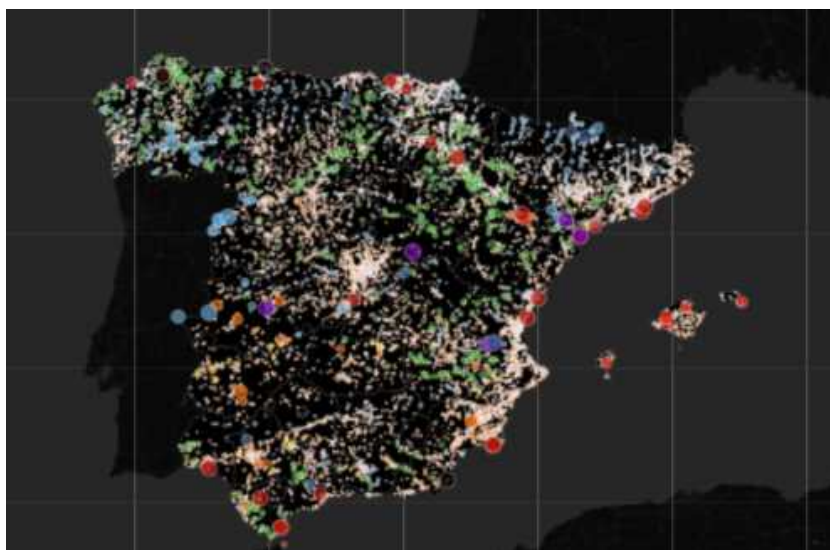


Figura 5.3: Mapa en el que se ven las cuadrículas usadas por Carto para un determinado nivel de zoom

- El otro tipo de llamadas que se hacen son a `cartocdn-gusc-c.global.ssl.fastly.net/liquidaciones/api/v1/map/ liquidaciones0352...77/1,2`. De forma similar a las anteriores llamadas, estas incluyen al final un `/Z/X/Y.png` que indica la zona que se está pidiendo. Como se ve en la figura 5.4 estas llamadas devuelven los círculos de las instalaciones sin ningún fondo, con algunas excepciones dependiendo del nivel de zoom).

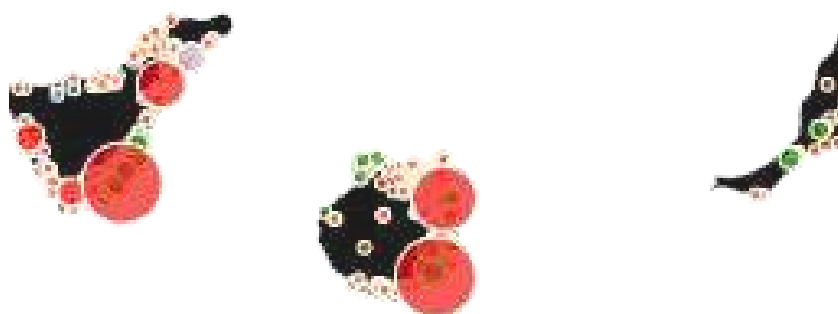


Figura 5.4: Foto PNG de un trozo del mapa de generadores con la representación de las instalaciones

Al superponer las dos respuestas a las llamadas comentadas anteriormente, respetando sus posiciones en la cuadrícula, podemos obtener la imagen que finalmente vemos en el mapa. Esto es poco alentador, ya que al pasar solo imágenes, en ningún momento se está pasando la información en un formato que sea procesable.

Por último, se ha tratado de ver si se hacía algún *request* cuando se pinchaba en algún círculo y así es. Cada vez que se pincha en un círculo se hace una llamada a <https://cartocdn-gusc-b.global.ssl.fastly.net/liquidaciones/api/v1/map/liquidaciones@352...77/2/attributes/> y finalmente se añade un número. La respuesta es un texto estilo diccionario con la información de la instalación. Se puede ver un ejemplo de una de estas respuestas en 5.1.

```
{ "municipio": "NAVARRA_Viana",  
  "tecnologia": "Fotovoltaica",  
  "mw": 7.765,  
  "instalaciones": 369 }
```

Ecuación 5.1: Ejemplo de respuesta de Carto con informacion de una instalación de producción

Después de probar con diferentes círculos se ha visto que el número que se añade al final funciona a modo de ID. Se ha probado entonces a generar nuestras propias llamadas cambiando el ID. Se ha visto que existe una instalación asociada a cada ID entre 1 y 23297. Automatizando estas 23298 llamadas y juntando los resultados se obtienen los datos de todas las instalaciones de producción de España.

Los datos obtenidos no contienen la localización exacta de cada instalación, ya que la información que se tiene sobre su localización es el municipio. Aun así, teniendo en cuenta que España hay más de 8100 municipios es aproximadamente como si España estuviera dividida en una cuadrícula de 90 x 90 y se conociera la localización de cada instalación dentro de esa cuadrícula. Esto se trata tan solo de una aproximación, ya que la densidad de provincias en España no es uniforme. Se puede ver una imagen con los municipios de España en la figura 5.5[37].



Figura 5.5: Mapa de los municipios de España

Es importante comentar que sí hay una forma de extraer la información con

mejor geolocalización que el municipio. Se pueden hacer llamadas que devuelvan los círculos del mapa en formato PNG con el nivel de zoom más alto posible. Después se podría entrenar un reconocedor de círculos que ubique cada uno de los círculos y detecte su tamaño, lo que será importante luego. Una vez finalizado este proceso habría que identificar el color de cada uno de estos círculos y usar tanto la información de la cuadrícula en la que se encuentre la foto PNG, como la posición exacta del círculo dentro de la imagen para obtener la posición geográfica con una exactitud muy elevada. Por último, habría que cruzar esta información con la información de instalaciones por municipio para obtener los datos de cada uno de los círculos. En determinados casos sería necesario el tamaño del círculo, ya que dentro de un mismo municipio puede haber más de una instalación del mismo tipo.

5.2. Análisis de los datos obtenidos

Para facilitar el análisis de estos datos, se han trasladado a una tabla de la que se pueden ver las primeras filas en 5.1. Por cada instalación de generación, se cuenta con 4 variables: municipio, tecnología, MW e instalaciones.

municipio	tecnología	mw	instalaciones
A CORUÑA Ames	Cogen/residuos/biomasa	7.3	2
A CORUÑA Aranga	Cogen/residuos/biomasa	0.5	1
A CORUÑA Aranga	Eólica	36.0	2
A CORUÑA Aranga	Hidráulica	9.0	1
A CORUÑA Aranga	Hidráulica	1.3	1

Tabla 5.1: Datos de instalaciones de generación obtenido de ESIOS (limitado a 5 resultados)

5.2.1. Variables

A continuación, se analiza cada una de estas variables de forma individual.

Tecnología

La variable tecnología contiene la tecnología de la instalación. Como se ve en la tabla 5.2 nuestros datos distinguen entre 11 tipos de tecnologías distintas. Se puede ver que de las 23297 instalaciones, 19823 son de fotovoltaica, eso es un 85 % de las instalaciones. Como se verá más adelante, estas son en su mayoría instalaciones

CAPÍTULO 5. GEOLOCALIZACIÓN DE LA GENERACIÓN EN ESPAÑA

con muy poca potencia instalada. En la tecnología nuclear, los datos no se dan separados por reactor, sino agrupados por central nuclear. Hay 2 instalaciones de tecnología desconocida, aun así estas instalaciones no suponen más que un 0.004 % del total como se detallará más adelante.

Tecnología	Número de instalaciones
Fotovoltaica	19823
Hidráulica	1388
Eólica	1021
Cogen/residuos/biomasa	934
Termosolar	51
Ciclo combinado	30
Fuel-gas	20
Carbón	15
Bombeo puro	8
Nuclear	5
Otras	2

Tabla 5.2: Variable tecnología y número de instalaciones que la tienen

MW

La variable MW informa de la potencia instalada. Mostrar la distribución de MW instalados por central aporta muy poca información. Esto se debe a que la distribución de MW instalados depende mucho de la tecnología. Para solucionarlo, en la tabla 5.3 podemos ver la media, máximo, mínimo, mediana y desviación estándar de los MW instalados en los generadores por tecnología. Por otro lado, en la figura 5.6 se pueden ver los histogramas de la potencia instalada dividida por tecnologías).

Tecnología	Media (MW)	Máximo (MW)	Mínimo (MW)	Mediana (MW)	Desviación (MW)
Bombeo puro	427.53	1511.95	88.85	289.87	452.72
Carbón	637.46	1403.19	299.76	557.20	330.71
Ciclo combinado	831.59	1670.81	385.85	791.18	378.83
Cogen/residuos/biomasa	7.97	197.30	0.01	2.34	16.06
Eólica	25.70	110.02	0.00	25.00	16.49
Fotovoltaica	0.48	499.82	0.00	0.02	6.93
Fuel-gas	231.80	693.34	7.95	163.50	225.99
Hidráulica	12.23	953.32	0.01	1.00	54.67
Nuclear	1423.46	2017.13	1003.41	1063.94	528.98
Termosolar	45.18	50.00	0.10	49.90	12.69

Tabla 5.3: Variable MW por tecnología y sus características (redondeado a 2 decimales)

Las distribuciones varían significativamente entre las distintas tecnologías, aunque casi todas se pueden aproximar con distribuciones unimodales. Muchas de ellas cuentan una desviación estándar muy baja, ya que la mayoría de instalaciones de esa tecnología tiene una potencia instalada muy similar. Esto se ejemplifica en cogen/residuos/biomasa, eólica, fotovoltaica, hidráulica y termosolar. Todas estas tecnologías (excepto termosolar) cuentan con una potencia instalada muy cercana a 0, aunque cuentan con un alto número de instalaciones en algunos casos.

En el caso de la energía termosolar encontramos en casi todas sus instalaciones potencia instalada muy cercana 50 MW, pero nunca superior. Esto es debido a que de acuerdo a lo establecido en el RD 661/2007 25 de mayo [38], la actividad de producción de energía eléctrica de origen renovable tiene la consideración de régimen especial cuando se realiza mediante instalaciones cuya potencia instalada no supera los 50 MW. Las tarifas del régimen especial hacen que sea más rentable tener plantas termosolares juntas pero de 50 MW de potencia máxima cada una, que tener plantas más grandes, como es el caso de la Plataforma Solar Extremadura. Estos resultados hacen dudar de la viabilidad de esta tecnología en un sistema sin tarifas de régimen especial, ya que el 86 % de las instalaciones tienen exactamente 50 MW, lo que parece indicar que es una tecnología cuyo tamaño óptimo está por encima de los 50 MW, al menos con tarifas de régimen especial. Aun así, no es rentable ninguna instalación de más de 50 MW (ya que no existen en España) al no entrar en régimen especial, o al menos su rentabilidad está muy por debajo de la de una central que entra en régimen especial.

Se puede apreciar que hay una cierta acumulación de instalaciones eólicas con 50 MW de potencia instalada, posiblemente por la tarifa de régimen especial. Otras tecnologías como el carbón, fuel-gas o bombeo puro parecen seguir una distribución normal plegada, es decir, sin valores negativos.

Instalaciones

Esta variable indica el número de instalaciones que se encuentran agrupadas, representadas con un número entero. En la figura 5.7 se puede ver la distribución del número de instalaciones agrupadas por tecnología.

La mayoría de las tecnologías parecen seguir una distribución normal de media 1 plegada, lo que ocurre en cogen/residuos/biomasa, fotovoltaica, hidráulica, eólica o carbón. Los ciclos combinados también parecen seguir esta distribución. Fuel-gas sigue una distribución muy diferente al resto, siendo 10 su número de instalaciones más habitual. En las instalaciones termosolares en casi un 98 % de los casos se trata de una única instalación. Por último, en el caso de las instalaciones nucleares se puede ver como de las cinco centrales, tres tienen solo una instalación y dos tienen dos. Realmente el número de instalaciones indica en este caso el número de reactores de la instalación. Las centrales de Asco y Almaraz tienen dos reactores mientras que Trillo, Cofrentes y Vandellós tienen solo uno.

Municipio

Pese a que esta variable se llame *Municipio*, no contiene solo el nombre del municipio, sino que además tiene el nombre en mayúsculas de la provincia y luego el nombre del municipio que comienza con una letra mayúscula. Esto resulta muy útil ya que en España existen municipios con nombres repetidos en 34 ocasiones.

Se pueden encontrar 4617 valores distintos de municipios en la tabla. Sabiendo que en España hay algo más de 8100 municipios, se ve que solo hay instalaciones de generación en algo menos del 60 % de los municipios de España. En la tabla 5.4 podemos ver los 20 valores más repetidos de municipio. Destacan principalmente Murcia y Madrid, aunque, como se verá más adelante, esto no tiene por qué significar que tienen mucha potencia instalada.

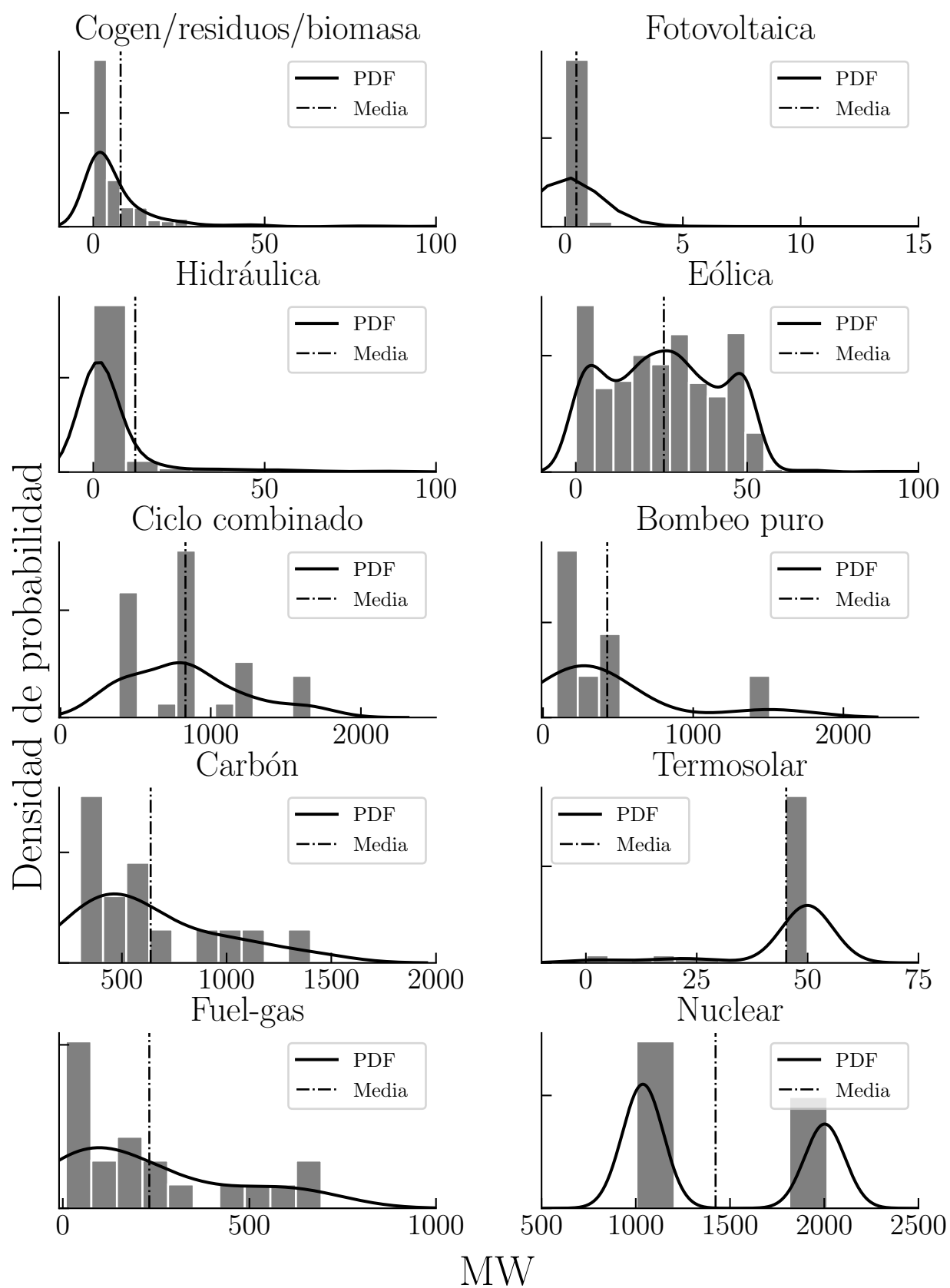


Figura 5.6: Histograma de los MW en instalaciones de generación por tecnología

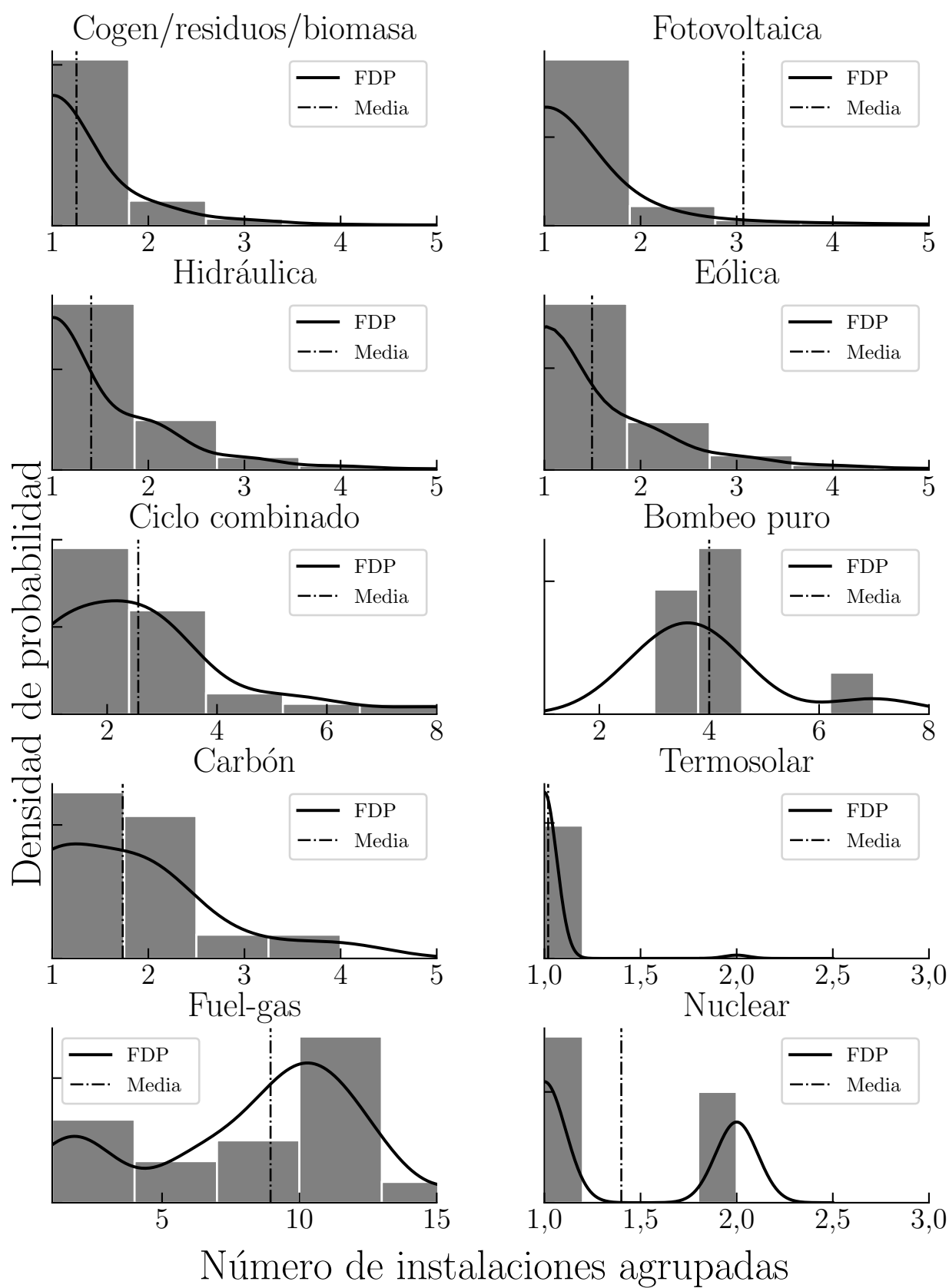


Figura 5.7: Histograma del número de instalaciones de generación agrupadas por tecnología

Municipio	Frecuencia
MURCIA Murcia	257
MADRID Madrid	211
ALBACETE Albacete	178
ZARAGOZA Zaragoza	170
ARABA/ÁLAVA Vitoria-Gasteiz	138
BARCELONA Barcelona	137
VALENCIA/VALÈNCIA Ontinyent	134
MURCIA Lorca	120
VALLADOLID Valladolid	115
PALMAS, LAS Telde	111
PALMAS, LAS Agüimes	104
MÁLAGA Málaga	99
PALMAS, LAS Palmas de Gran Canaria, Las	93
SEVILLA Sevilla	91
PALMAS, LAS Puerto del Rosario	89
NAVARRA Tudela	89
BURGOS Burgos	88
MADRID Rivas-Vaciamadrid	86
GUIPUZKOA Donostia/San Sebastián	83
MURCIA Cartagena	83

Tabla 5.4: Variable municipio y sus 20 valores más repetidos)

En la figura 5.8 podemos apreciar que gran parte de los municipios se repiten menos de 10 veces, aunque se puedan encontrar municipios que aparecen hasta 200 veces. La distribución de probabilidad se asemeja a una Poisson de $\lambda=3$ pero con una cola mayor, lo que explica por qué tiene una media mayor.

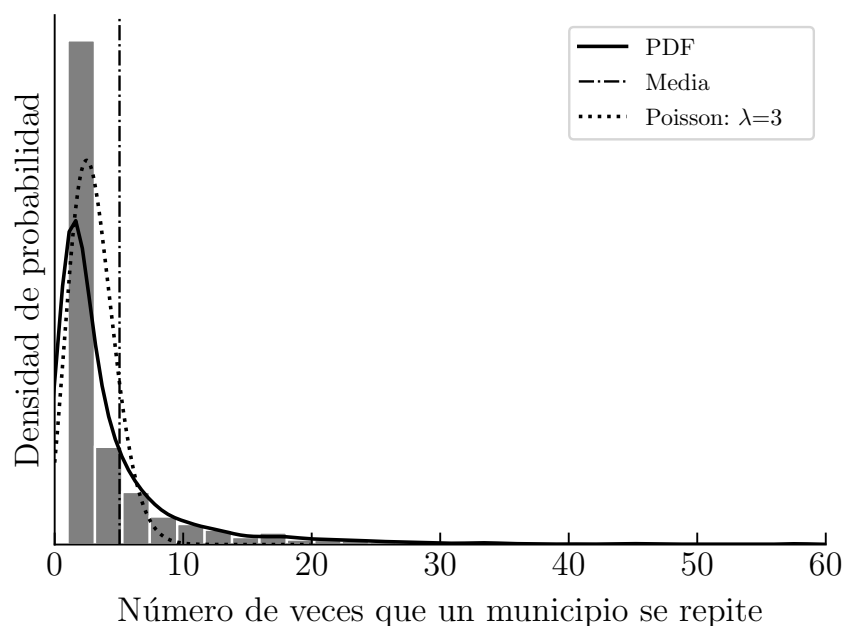


Figura 5.8: Histograma de número de veces que se repite un municipio

El principal uso de esta variable es geolocalizar las instalaciones. Con este objetivo se ha obtenido la latitud y longitud de cada municipio de [39]. Estos datos son de 2010 y aportan la localización de 8114 municipios. Sin embargo, actualmente, existen 8131 municipios en España [INEPadron], 17 más que en nuestros datos, aunque esto no ha supuesto problemas.

Como se puede ver en la tabla 5.5 los datos de [39] aportan, aparte de la latitud y longitud, también la comunidad autónoma, la provincia, la altitud, el número de habitantes y número de hombres y mujeres. Pese a ser todos datos interesantes, los datos de altitud y los referentes a los habitantes y sus géneros no se van a utilizar.

Comunidad	Provincia	Población	Lat	Long	Alt	Hab	Homb	Muj
Andalucía	Almería	Abla	37.141	-2.780	871.168	1504	783	721
Andalucía	Almería	Abrucena	37.133	-2.797	976.938	1341	682	659
Andalucía	Almería	Adra	36.748	-3.0225	10.978	24373	12338	12035
Andalucía	Almería	Albánchez	37.287	-2.181	481.312	815	422	393
Andalucía	Almería	Alboloduy	37.033	-2.621	388.434	674	334	340

Tabla 5.5: Datos de localización de los municipios de España [39]

Juntar los datos de localización de municipios de [39] con los datos obtenidos de ESIOs no ha sido simplemente separar la comunidad autónoma y el municipio de los datos de ESIOs sabiendo que uno está en mayúsculas y el otro no (usando regex) y buscarlos automáticamente en los datos de [39]. Entre ambas tablas había cientos de inconsistencias, muchas de ellas se han podido solucionar automáticamente, al tratarse de espacios, artículos que se usan en una tabla pero la otra no etc. Sin embargo, en otros casos, como municipios o provincias que tienen más de una forma de escribirse, o municipios que han cambiado el nombre se han tenido que solucionar manualmente a través de un diccionario de municipios.

Se ha decidido añadir a los datos de cada una de las instalaciones con las que se cuenta, la latitud, longitud, la comunidad autónoma, la provincia y el municipio al que pertenece. Aunque estas dos últimas variables ya estaban, se encontraban dentro de una única variable *municipio*, que se ha separado. Se puede ver el resultado en la tabla 5.6.

municipio	tecnologia	mw	inst	CCAA	provincia	poblacion	lat	long
A CORUÑA Ames	Cogen/residuos/biomasa	7.3	2	Galicia	A Coruña	Ames	42.900	-8.633
A CORUÑA Aranga	Cogen/residuos/biomasa	0.5	1	Galicia	A Coruña	Aranga	43.233	-8.016
A CORUÑA Aranga	Eólica	36.0	2	Galicia	A Coruña	Aranga	43.233	-8.016
A CORUÑA Aranga	Hidráulica	9.0	1	Galicia	A Coruña	Aranga	43.233	-8.016
A CORUÑA Aranga	Hidráulica	1.3	1	Galicia	A Coruña	Aranga	43.233	-8.016

Tabla 5.6: Datos de instalaciones de generación después de añadir localización (limitado a 5 resultados)

5.2.2. Simplificación de los datos

Con el objetivo de simplificar los datos para su posterior procesamiento y uso se ha decidido agrupar por municipio y tecnología. Al agrupar de esta forma no se pierde información geográfica sobre los generadores, ya que la máxima precisión que se tiene es el municipio. Después de agrupar, se pasa de 23298 generadores distintos a tan solo 6046, que es un 25 % respecto del número original.

Además, se ha añadido un ID a cada uno de estos generadores para facilitar su uso más adelante.

5.2.3. Fecha real de los datos

Una vez se cuenta con los datos geográficos procesados, se pueden agrupar por comunidades, obteniendo la tabla 5.7, donde se pueden ver los MW instalados por comunidad autónoma y tecnología, junto con el porcentaje que representan respecto del total.

Comparar estos valores con los de la API de REE permite confirmar que los datos de las instalaciones con los que se cuentan están temporalmente ubicados entre 2019 y 2020. Se aprecia que se encuentran mucho más cerca de los valores de 2019 que de los de 2020, lo que incita a pensar, que, pese a que los datos se publicaran el 24/07/2019, muy posiblemente datan de unos cuantos meses antes, si asumimos una velocidad de instalación de potencia uniforme a lo largo del año.

Comunidad Autónoma	Bomb (MW)	Carb (MW)	Cicl (MW)	Coge (MW)	Eóli (MW)	Foto (MW)	Fuel (MW)	Hidr (MW)	Nucl (MW)	Term (MW)	Total (MW)	Porcentaje (%)
Andalucía	584	1989	5951	1182	3450	1953	-	589	-	1000	16721	14.9 %
Aragón	307	1055	1869	701	3496	997	-	1254	-	-	9691	8.6 %
Asturias	-	2099	854	234	518	-	-	804	-	-	4516	4.0 %
Canarias	-	-	-	41	440	166	2362	1	-	-	3015	2.7 %
Cantabria	360	-	-	318	35	2	-	98	-	-	816	0.7 %
Castilla La Mancha	215	-	758	494	3812	1791	-	652	1003	349	9085	8.1 %
Castilla León	-	2456	-	639	6051	801	-	4401	-	-	14364	12.8 %
Catalunya	439	-	4174	1112	1271	276	-	1921	3032	24	12263	10.9 %
Ceuta y Melilla	-	-	-	2	-	-	166	-	-	-	168	0.2 %
Extremadura	-	-	-	53	39	1574	-	2277	2017	848	6817	6.1 %
Galicia	-	1960	1246	615	3802	18	-	3729	-	-	11383	10.1 %
Islas Baleares	-	-	-	87	3	80	2098	-	-	-	2272	2.0 %
La Rioja	-	-	784	26	448	98	-	52	-	-	1411	1.3 %
Madrid	-	-	-	284	-	63	-	108	-	-	457	0.4 %
Murcia	-	-	3263	339	262	1112	-	34	-	31	5048	4.5 %
Navarra	-	-	1222	194	1196	154	-	237	-	-	3007	2.7 %
País Vasco	-	-	1968	581	153	50	-	174	-	-	2930	2.6 %
Valencia	1511	-	2853	529	1255	364	7	641	1063	50	8285	7.4 %
Total	3420	9561	24947	7439	26237	9508	4635	16980	7117	2304	112157	100 %
Porcentaje (%)	6.1 %	17.1 %	44.5 %	13.3 %	46.8 %	17.0 %	8.3 %	30.3 %	12.7 %	4.1 %	100 %	-

Tabla 5.7: Datos de instalaciones de generación separados por comunidad autónoma y tecnología (valores redondeados)

5.2.4. Visualización de los datos

En la figura 5.9 se puede ver en un mapa de calor la distribución de potencia instalada en España. Se observan lugares con gran concentración de potencia instalada, como Cádiz, Cartagena o Tarragona. Destacan las ubicaciones de las

centrales nucleares, al ser puntos concretos con mucha potencia instalada.

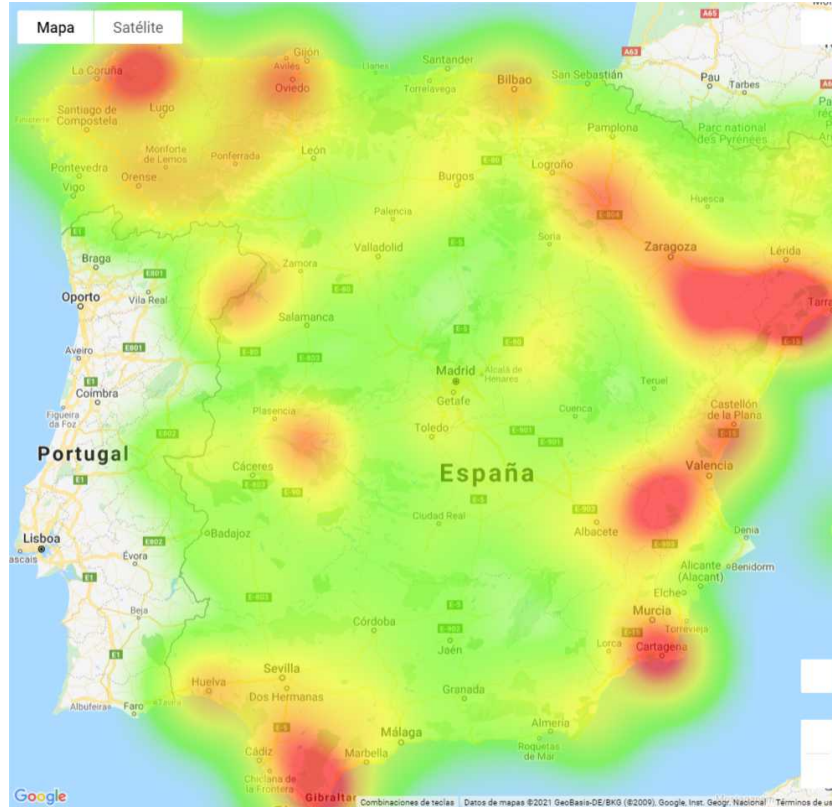


Figura 5.9: Mapa de calor de la potencia instalada en España

5.3. Cálculo de generación con geolocalización

Los datos de generación están disponibles en la pagina web de ESIOS [34], y se pueden encontrar divididos por tecnología. Sin embargo, no se pueden obtener los datos divididos por provincia o por comunidad autónoma, solo se pueden obtener los datos de la generación agregada de toda España.

En este trabajo se ha usado la geolocalización de la potencia instalada y la generación total por tecnología en España de cada día, para obtener los datos de generación con geolocalización cada día.

Se ha creado una tabla con los IDs de los generadores en las columnas y fechas en las filas, donde para cada día, se rellenará la potencia generada por ese generador. Inicialmente, está rellena con ceros y se puede ver en la tabla 5.8

Fecha	ID										
	0	1	2	3	4	...	6041	6042	6043	6044	6045
2020-01-01	0 MWh	0 MWh	0 MWh	0 MWh	0 MWh	...	0 MWh	0 MWh	0 MWh	0 MWh	0 MWh
2020-01-02	0 MWh	0 MWh	0 MWh	0 MWh	0 MWh	...	0 MWh	0 MWh	0 MWh	0 MWh	0 MWh
2020-01-03	0 MWh	0 MWh	0 MWh	0 MWh	0 MWh	...	0 MWh	0 MWh	0 MWh	0 MWh	0 MWh
2020-01-04	0 MWh	0 MWh	0 MWh	0 MWh	0 MWh	...	0 MWh	0 MWh	0 MWh	0 MWh	0 MWh
...	...	...	...	...	...	...	...	...	...	...	...
2021-05-28	0 MWh	0 MWh	0 MWh	0 MWh	0 MWh	...	0 MWh	0 MWh	0 MWh	0 MWh	0 MWh
2021-05-29	0 MWh	0 MWh	0 MWh	0 MWh	0 MWh	...	0 MWh	0 MWh	0 MWh	0 MWh	0 MWh
2021-05-30	0 MWh	0 MWh	0 MWh	0 MWh	0 MWh	...	0 MWh	0 MWh	0 MWh	0 MWh	0 MWh
2021-05-31	0 MWh	0 MWh	0 MWh	0 MWh	0 MWh	...	0 MWh	0 MWh	0 MWh	0 MWh	0 MWh

Tabla 5.8: Tabla de producción de cada generador para cada día

5.3.1. Cálculo de generación de tecnologías independientes del clima

Para las tecnologías independientes del clima el cálculo de generación es muy simple. Se coge la generación total de España de esa tecnología para cada día y se divide entre los generadores proporcionalmente a su potencia instalada. La forma de llevarlo a cabo es la siguiente: para cada día y central se coge la potencia generada de su tecnología, se multiplica por la potencia instalada de la central y se divide por la potencia total de esa tecnología instalada tal como se ve en la formula de 5.2.

$$prod[*dia*, *central*] = \frac{produccion_tecnologia[*dia*] \cdot pot_instalada[*central*]}{pot_total_instalada_tecnologia}$$

Ecuación 5.2: Fórmula para el cálculo de la generación para centrales con tecnologías independientes del clima

En este trabajo solo se ha calculado la generación geolocalizada de la tecnología nuclear, pero se podría aplicar a las demás tecnologías independientes del clima.

Nuclear

Los datos de generación nuclear se han obtenido de *GENERACIÓN MEDIDA NUCLEAR* de ESIOS [40]. Se han descargado los datos desde 2019 a 2021 agregados por día. Se podrían haber usado los datos de *GENERACIÓN T.REAL*

NUCLEAR de ESIOS [41], pero cuando se agregan estos datos por día se devuelve la media, no la generación total del día.

En la figura 5.10 se puede ver la generación nuclear durante el año 2020.

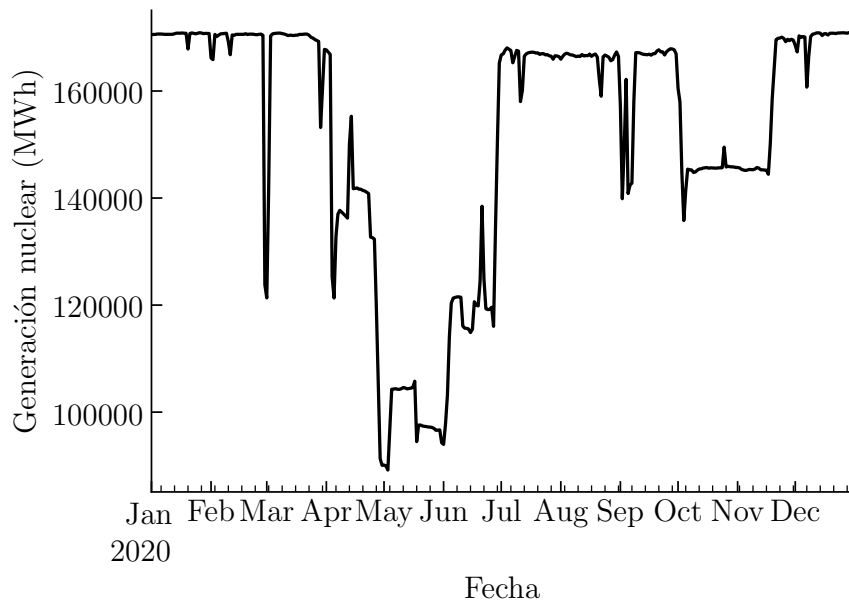


Figura 5.10: Generación nuclear durante el año 2020

La generación de la figura 5.10 se reparte entre las 5 centrales nucleares proporcionalmente a su potencia instalada para cada día y se añaden los valores a la tabla 5.8. Para hacerlo se sigue la formula de 5.2.

Se pueden ver los resultados en la tabla 5.9, donde se muestran solo los datos de las centrales nucleares.

Fecha	ID				
	1952	2237	2792	3235	5023
2020-01-01	25048.92 MWh	24044.86 MWh	25495.35 MWh	48336.78 MWh	47626.75 MWh
2020-01-02	25048.23 MWh	24044.2 MWh	25494.65 MWh	48335.45 MWh	47625.44 MWh
2020-01-03	25052.08 MWh	24047.89 MWh	25498.57 MWh	48342.88 MWh	47632.76 MWh
2020-01-04	25062.15 MWh	24057.56 MWh	25508.82 MWh	48362.32 MWh	47651.92 MWh
2020-01-05	25062.55 MWh	24057.94 MWh	25509.22 MWh	48363.08 MWh	47652.67 MWh
2020-01-06	25063.41 MWh	24058.77 MWh	25510.1 MWh	48364.75 MWh	47654.31 MWh
2020-01-07	25056.2 MWh	24051.85 MWh	25502.76 MWh	48350.83 MWh	47640.6 MWh
2020-01-08	25053.44 MWh	24049.2 MWh	25499.95 MWh	48345.51 MWh	47635.35 MWh

Tabla 5.9: Tabla de producción de centrales nucleares (limitado a 8 resultados)

5.3.2. Cálculo de generación de tecnologías dependientes del clima

Para las tecnologías dependientes del clima, como la eólica o solar (tanto térmica como fotovoltaica) se podría aplicar la misma fórmula que se aplica a las independientes del clima (fórmula 5.2), pero se estaría ignorando la fuerte sensibilidad a las condiciones atmosféricas de estas tecnologías. Es por este motivo por lo que se ha querido tener en cuenta las condiciones atmosféricas en el cálculo de generación de las centrales de tecnologías dependientes del clima. Se puede ver la fórmula para calcular la generación de cualquier central dependiente del clima en 5.3. Para cada central y día se toma la potencia instalada en la central y se multiplica por un factor que depende del clima de la central, este factor es una ecuación que relaciona la variable climatológica con la generación. El resultado se divide por el sumatorio del producto de la potencia de cada central por el factor climatológico del clima de cada central. Esta división permite normalizar los valores para obtener la aportación de cada central individual, que después simplemente se multiplica por la producción total de la tecnología concreta en ese día para obtener la producción de la central en el día. El resultado de $factor_clima(clima_central) \cdot pot_instalada[central]$ no tiene valor en sí y no tiene por qué significar nada físico, ya que luego ese valor se normaliza al dividirlo. La suma de la generación de todas las centrales de una tecnología siempre será igual a la generación en ese día gracias a normalizar los valores.

$$prod[dia, central] = \frac{factor_clima(clima_central) \cdot pot_instalada[central]}{\sum_i^{centrales} factor_clima(clima_i) \cdot pot_instalada[i]} \cdot produccion_tecnologia[dia]$$

Ecuación 5.3: Fórmula para el cálculo de la generación para centrales con tecnologías dependientes del clima

Extracción de datos atmosféricos de AEMET

Los datos atmosféricos que se han usado han sido extraídos de AEMET usando su API, *AEMET OpenData* [42]. Para usar esta API ha sido necesario pedir un API key a AEMET, que ellos entregan de forma gratuita por correo.

Se han hecho *requests* de tipo *GET* a la API para obtener los datos diarios obtenidos por todas las estaciones meteorológicas disponibles entre 2019 y 2021. La tabla donde se han recogido todos los datos se puede ver en las tablas 5.11 y 5.10, se ha dividido en 2 tablas para su visualización, pero los datos de las 2 tablas están alineados horizontalmente.

	fecha	indicativo	nombre	provincia	altitud	tmed	prec	tmin	horatmin
0	2019-01-01	4358X	DON BENITO	BADAJOS	273	8,0	0,0	0,4	08:30
1	2019-01-01	4220X	PUEBLA DE DON RODRIGO	CIUDAD REAL	486	6,9	NaN	-4,2	07:50
2	2019-01-01	C447A	TENERIFE NORTE AEROPUERTO	STA. CRUZ DE TENERIFE	632	15,0	0,0	10,4	07:56
3	2019-01-01	6106X	ANTEQUERA	MALAGA	408	9,8	0,0	1,1	07:00
4	2019-01-01	9698U	TALARN	LLEIDA	807	7,6	0,0	2,4	07:30

Tabla 5.10: Trozo 1 de tabla de datos de AEMET (limitado a 5 resultados)

	tmax	horatmax	dir	velmedia	racha	horaracha	sol	presMax	horaPresMax	presMin	horaPresMin
0	15,7	15:50	15	0,8	3,3	21:00	8,4	1000,3	10	997,8	01
1	18,0	15:50	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	19,5	15:30	16	3,9	11,4	10:11	5,5	956,5	Varias	951,2	01
3	18,6	15:10	15	0,8	7,2	00:20	NaN	982,8	10	980,5	05
4	12,7	14:40	19	0,8	2,8	00:50	7,7	935,8	09	933,5	24

Tabla 5.11: Trozo 2 de tabla de datos de AEMET (limitado a 5 resultados)

Como se puede ver, se proporcionan muchos datos para cada día y estación, pero en este trabajo solo se utilizan los que se pueden ver en la tabla 5.12

CAPÍTULO 5. GEOLOCALIZACIÓN DE LA GENERACIÓN EN ESPAÑA

fecha	indicativo	nombre	provincia	velmedia	sol
2019-01-01	4358X	DON BENITO	BADAJOS	0,8	8,4
2019-01-01	4220X	PUEBLA DE DON RODRIGO	CIUDAD REAL	NaN	NaN
2019-01-01	C447A	TENERIFE NORTE AEROPUERTO	STA. CRUZ DE TENERIFE	3,9	5,5
2019-01-01	6106X	ANTEQUERA	MALAGA	0,8	NaN
2019-01-01	9698U	TALARN	LLEIDA	0,8	7,7

Tabla 5.12: Datos de AEMET que se utilizan (limitado a 5 resultados)

Análisis de las variables de AEMET

- Fecha: Indica la fecha de los datos
- Indicativo y Nombre: Identifican a la estación. El indicativo es más útil para esa función, ya que el nombre puede repetirse para estaciones diferentes. Se cuenta con los datos de 255 estaciones distintas.
- Provincia: AEMET Opendata no proporciona la geolocalización de las estaciones meteorológica, solo su provincia. Además, no hay ninguna forma de obtener una localización más precisa usando la API.

Los datos precisos de geolocalización de las estaciones meteorológica se pueden obtener de la página del gobierno de España bajo el apartado de Cartografía [cartografia]. Los datos que se ofrecen son de cuatro tipos de estaciones (completas, termométricas, pluviométricas y automáticas) y se dan en formato *shapefile*. Después de convertir los archivos a tablas y juntarlos, se obtiene una tabla con la geolocalización de 3282 estaciones meteorológicas. Esta tabla también cuenta con otros datos que no se mencionan ya que no se utilizan en este trabajo. Cruzando los indicativos de estos datos con los de las 255 estaciones que *AEMET Opendata* nos ofrece se ve que se cuenta con la geolocalización de 254 de las 255 estaciones. La estación de la que no se tienen datos geográficos, simplemente no se usará para los cálculos. Se puede ver la tabla resultante en 5.13, esta tabla cuenta con 254 filas y 3 columnas.

INDICATIVO	Lat	Lon
2661	42.588346	-5.651108
8416	39.480566	-0.366386
9262	42.817780	-1.638339
7031	37.788899	-0.803332
1082	43.298060	-2.906390

Tabla 5.13: Tabla de estaciones meteorológicas y su localización (limitado a 5 resultados)

Para facilitar posibles estudios futuros sobre estos datos, se ha modificado la variable *Provincia* para que sus valores se adapten a la tabla de los datos de localización de municipios de España 5.5, donde también aparecen las provincias y las comunidades.

- **Velmedia:** Esta variable es de tipo *float* e indica la velocidad del viento media en km/h para cada día y cada estación medida por un anemómetro. Cuando no se cuentan con valores para un determinado día en su lugar aparece *NaN*. En la figura 5.11 se puede ver un histograma de la velocidad media medida por las estaciones meteorológicas. Se pueden observar que hay desde valores muy cercanos a cero, a 27.2 km/h, que es el valor más alto registrado en Santa Cruz de Tenerife el 16-12-2019.

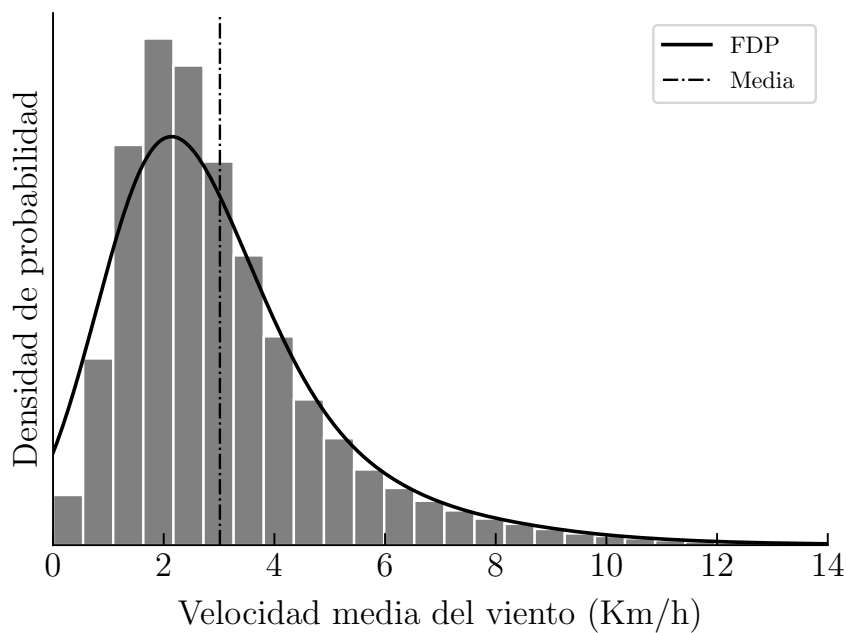


Figura 5.11: Histograma de la velocidad media del viento en las estaciones meteorológicas de España durante 2019, 2020 y parte de 2021

Los valores de velocidad media se pueden agregar fácilmente entre ellos para obtener la media de la provincia, comunidad o toda España. Los valores de cada día se pueden visualizar como una serie temporal, en la figura 5.12 se puede ver la media de la velocidad media en España durante los años 2019, 2020 y parte de 2021. Se aprecia un cierta periodicidad en el comportamiento de la función.

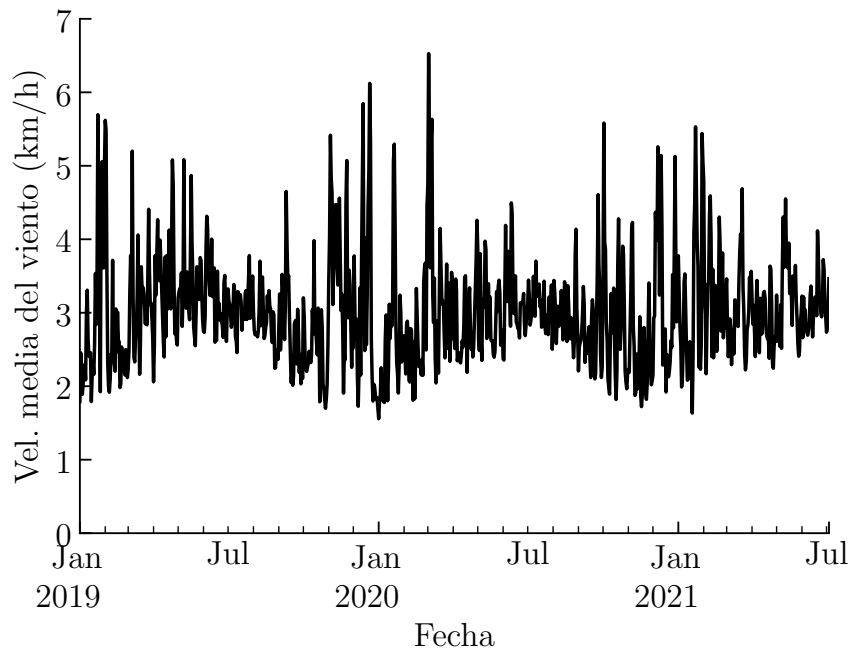


Figura 5.12: Media de la velocidad del viento en España durante 2019, 2020 y parte de 2021

En la figura 5.13 podemos ver un histograma de la media de la velocidad media del viento en España entre 2019 y 2021. Es importante destacar que la diferencia entre esta figura y la figura 5.11 es que 5.13 representa la media de la velocidad media de todas las estaciones juntas (es decir, la media de España) para cada día mientras que 5.11 representa las velocidades medias medidas por cada estación, este es el motivo por el que 5.13 cuenta con una varianza mucho menor (por el teorema del limite central).

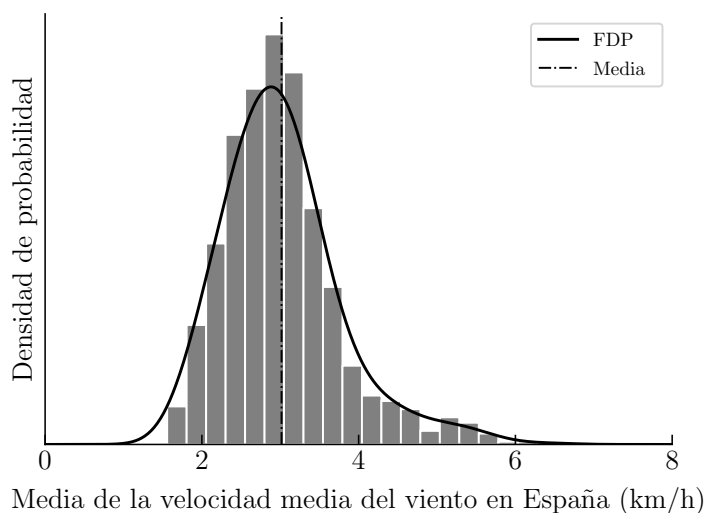


Figura 5.13: Histograma de la media de la velocidad media del viento en España durante 2019, 2020 y parte de 2021

- Sol: igual que *Velmedia*, esta variable es de tipo *float*. Indica el tiempo de insolación en horas para cada día y estación medida con un heliómetro. Es importante destacar que esta variable no indica las horas de sol en un día aunque ambos valores estén íntimamente relacionados. Un 42 % de las estaciones (109 de las 255) no ofrecen datos de *Sol* y tienen esta variable rellena de *NaN*. En la figura 5.14 podemos ver un histograma de la variable sol. Sigue una distribución poco frecuente, con una gran cantidad de valores concentrados cerca de 0.

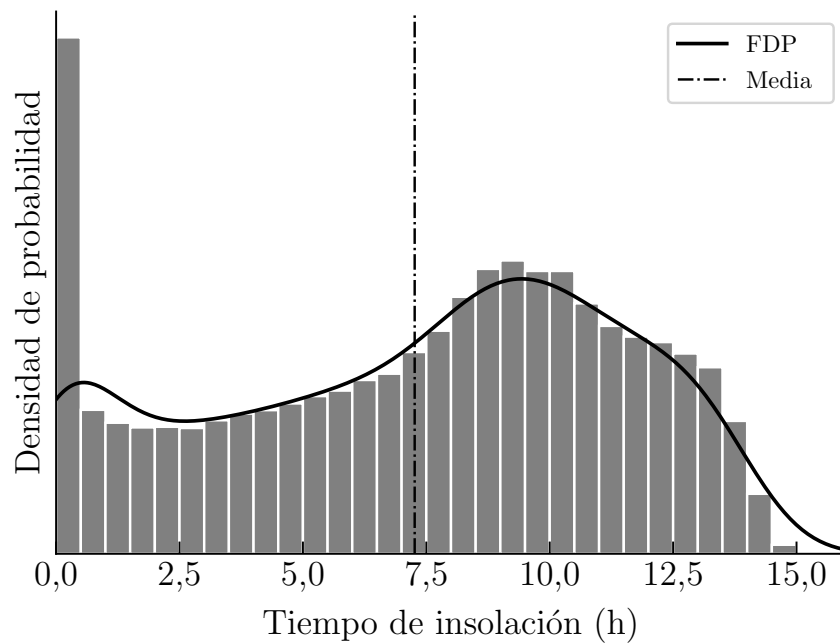


Figura 5.14: Histograma de las horas de insolación en España durante 2019, 2020 y parte de 2021

Estos valores también pueden agregarse para obtener la media de la provincia, comunidad autónoma o España. En la figura 5.15 se puede ver las horas de insolación medias por día en España durante un periodo de 2 años y medio. En este caso, se puede ver un claro comportamiento sinusoidal, con más horas de insolación en los meses de verano, y menos en los meses de invierno.

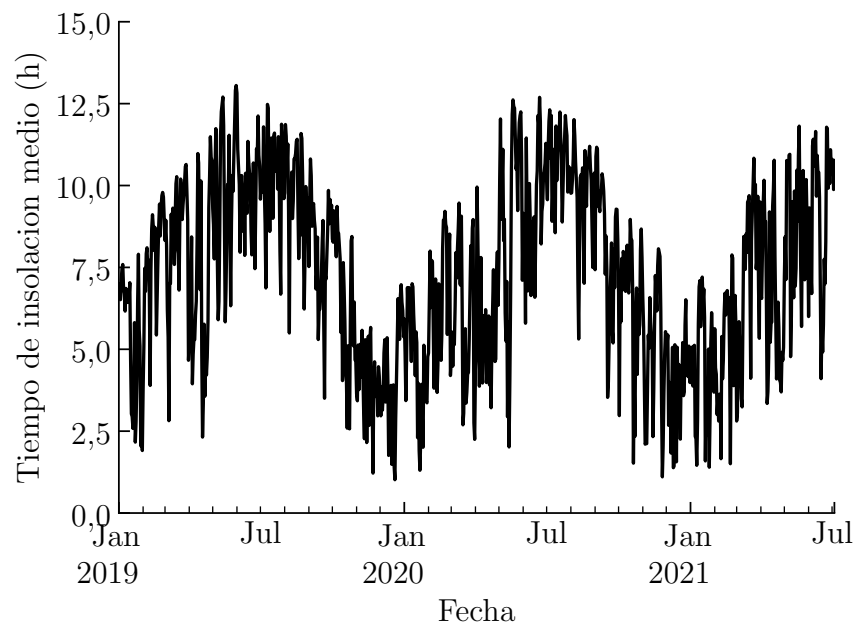


Figura 5.15: Horas de insolación media por día en España durante 2019, 2020 y parte de 2021

Se puede ver un histograma de las horas de insolación medias por día en España en la figura 5.16. Igual que ocurría con la velocidad del viento, al hacer la media de España, la variable tiene menos varianza que en 5.14, de nuevo por el teorema del limite central.

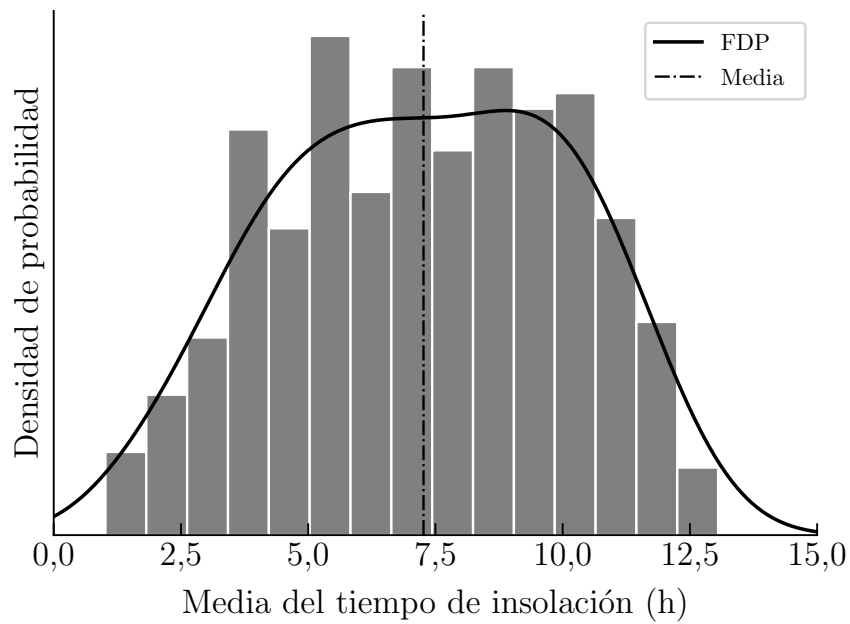


Figura 5.16: Histograma de la media de horas de insolación en España durante 2019, 2020 y parte de 2021

Eólica

Igual que con la generación nuclear, los datos de generación eólica diarios se han obtenido de ESIOS, más concretamente de *GENERACIÓN MEDIDA EOLICA* [43].

En la figura 5.17 se puede ver la generación eólica diaria de España durante 2020 y parte de 2021

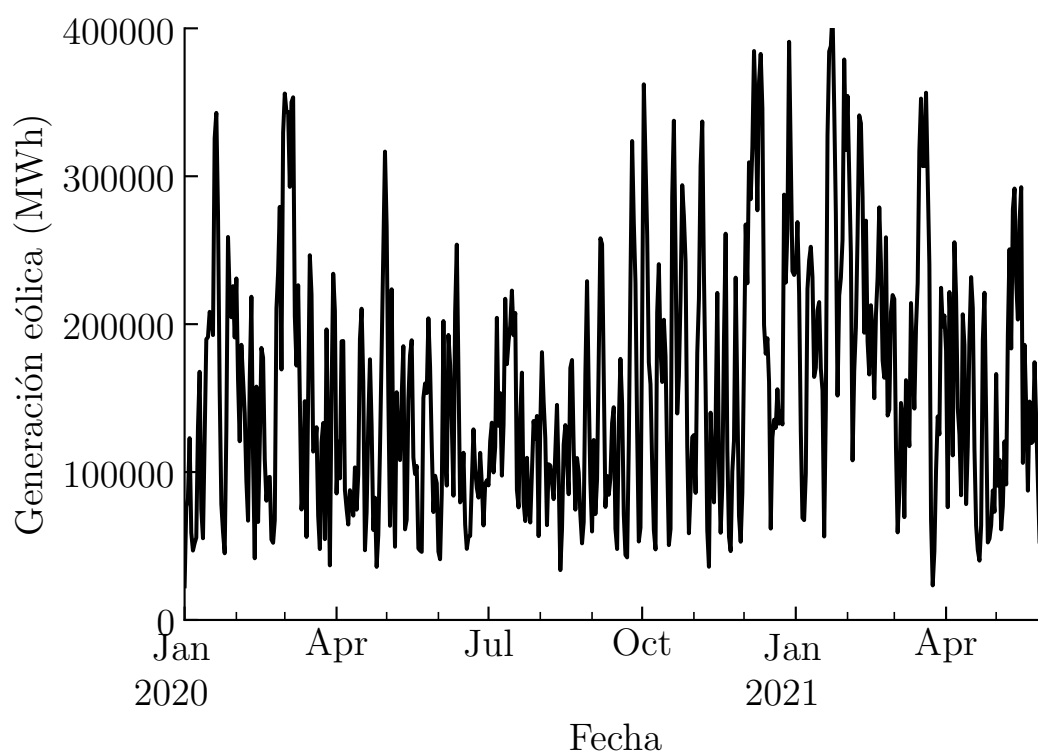


Figura 5.17: Generación eólica durante 2020 y parte de 2021

Al unir los datos de generación eólica diarios de la tabla 5.17 con los datos de velocidad media diarios de la tabla 5.12, se puede obtener la relación entre ambas. El diagrama de dispersión resultante se puede ver en la figura 5.18 en la que cada punto representa un día, y el viento medio que hacía en España ese día frente a cuánta generación eólica hubo. Se pueden ver por separado los datos de 2020 y los de 2021, y se puede apreciar que no hay cambios significativos entre el comportamiento de ambos años.

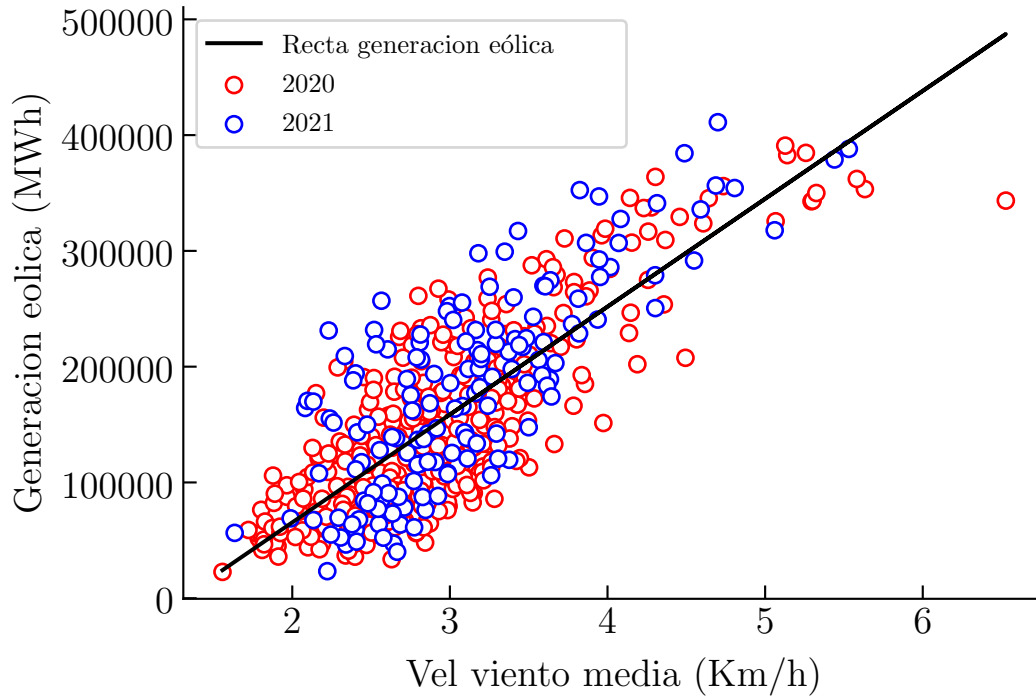


Figura 5.18: Diagrama de dispersión de la generación eólica en función de la velocidad media del viento en España durante 2020 y parte de 2021

La ecuación de la recta de generación eólica es

$$Recta_{gen\ eolica}(MWh) = 93170,42\left(\frac{MWh}{km/h}\right) \cdot vel_viento(km/h) - 120846,51(MWh)$$

Ecuación 5.4: Ecuación de la recta de generación eólica

Se aprecia una clara relación lineal entre la generación eólica y la velocidad media del viento. Se ha calculado el coeficiente de correlación de Pearson y su significancia.

$$Coeficiente\ de\ Pearson = 0,8028 \quad p - value \approx 0$$

El coeficiente de Pearson de 0.8028 indica una alta correlación entre ambas variables y con un p valor aproximado a 0. Esto nos permite afirmar que la velocidad

media del viento medida por las estaciones meteorológicas se puede utilizar para predecir la generación eólica de las instalaciones de generación cercanas usando la ecuación de la recta de generación eólica. Al hacer esto, estamos asumiendo que el comportamiento entendido como la relación entre la pendiente y la constante de la recta de cada generador eólico individual de España, respecto de la velocidad del viento es igual que el que se ve en la figura 5.18. Esto es equivalente a afirmar que cada generador se comporta como la media de los generadores. El principal problema de esta aproximación es que solo podemos afirmar que hay un comportamiento lineal entre velocidades de viento medio de 1 km/h a 6 km/h, pero como se ve en 5.11 las estaciones pueden medir velocidades medias más pequeñas y cercanas a cero, y otras mucho mayores de 6. Para estas velocidades, no podemos confirmar un comportamiento lineal, pero al no poder modelar ese comportamiento asumimos linealidad independientemente de la velocidad media del viento.

Si se observa la ecuación de la recta de generación eólica de 5.4 se puede ver que para una velocidad de viento de 0 km/h, la generación eólica es cercana a -121 GWh, y que para velocidades de viento menores de 1.3 km/h la generación eólica es negativa. Físicamente esto no tiene sentido ya que significaría que en vez de generar energía, el generador la está consumiendo. Como no sabemos realmente como se comportan los generadores por debajo de velocidades de viento de 2 km/h, se va a asumir linealidad hasta el corte con el eje X, a partir del cual el valor es 0. Se puede ver esta ecuación de generación eólica por tramos en 5.5

$$Gen_{eolica} (MWh) = \begin{cases} 0 & si \quad vel \leq 1,3km/h \\ 93170,42 * vel_{viento} (km/h) - 120846,51 & si \quad vel > 1,3km/h \end{cases}$$

Ecuación 5.5: Ecuación de la recta de generación eólica por tramos

Geneolica en este caso representa el término *factor_clima* de la fórmula 5.3 para cálculo de generación en centrales con tecnología dependiente del clima. Para obtener la velocidad del viento de cada central eólica para cada día se ha cogido la media de la velocidad del viento de las 3 estaciones meteorológicas más cercanas.

Al aplicar la fórmula 5.3 con la ecuación 5.5 podemos obtener la generación de cada central eólica para cada día.

En la figura 5.19 podemos ver un mapa de calor con la distribución geográfica de la generación eólica durante el año 2020 calculada. En la figura se ven princi-

palmente los picos, que coinciden también con los principales lugares de potencia eólica instalada en el país.

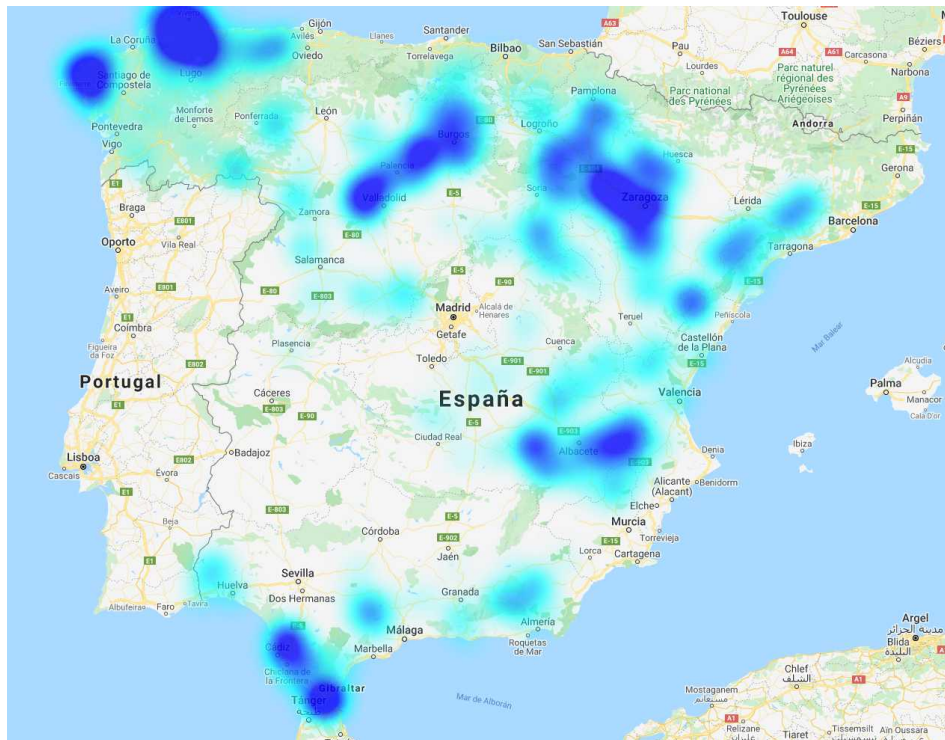


Figura 5.19: Mapa de calor de la generación eólica durante 2020

Solar Fotovoltaica

Los datos de generación solar fotovoltaica se han obtenido de ESIOS agrupados por día, del apartado de *GENERACIÓN MEDIDA SOLAR FOTOVOLTAICA* [44]. En la figura 5.20 se puede ver la generación fotovoltaica desde 2019 a 2021. Se aprecia la clara dependencia climática de la generación fotovoltaica. También se observa como la generación tiene una tendencia ascendente, debido a la gran cantidad de potencia solar fotovoltaica instalada en estos últimos 2 años.

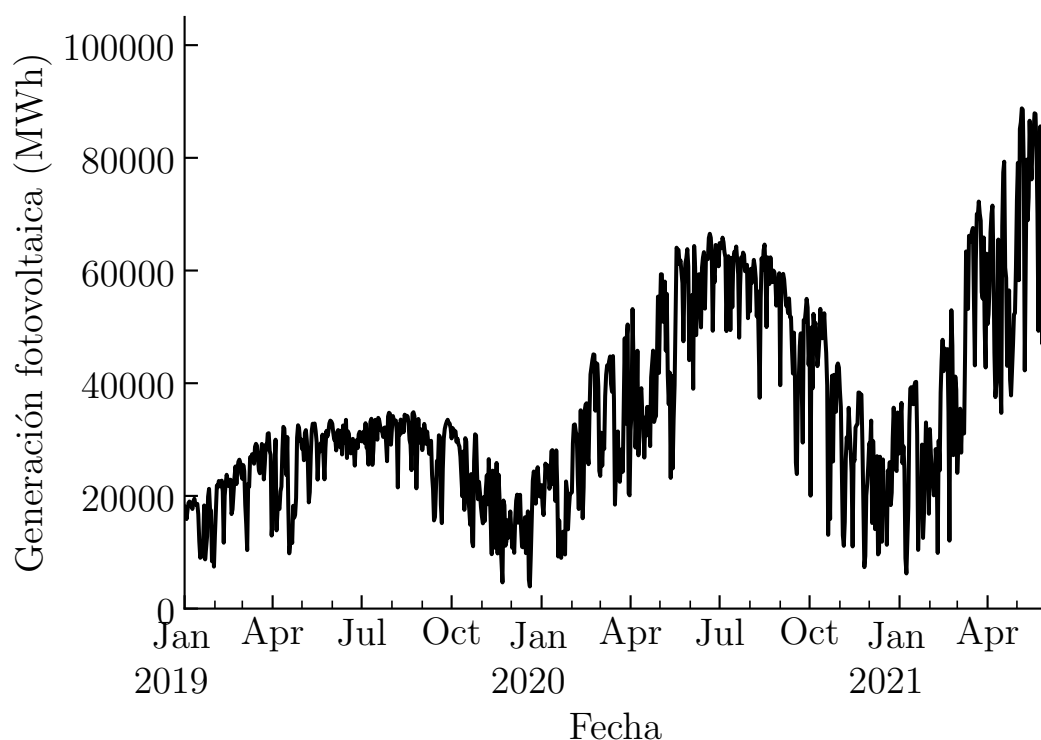


Figura 5.20: Generación solar fotovoltaica durante 2019, 2020 y parte de 2021

Si se juntan los datos de generación solar fotovoltaica de 5.20 con los datos de la media de horas de insolación en España para cada día, se consigue el diagrama de dispersión de la figura 5.21. En la figura cada punto representa un día y la generación fotovoltaica frente a las horas de insolación medias que hubo. Se ven por separado los datos de 2019, 2020 y 2021. Las diferencias en su pendiente se deben a las diferencias entre la potencia instalada de cada año, pero si nos fijamos en la segunda derivada vemos que también es diferente entre los distintos años. En 2019 el comportamiento no es lineal y la segunda derivada tiene un alto valor negativo. Esto físicamente es como decir que en 2019 el rendimiento se reduce al aumentar las horas de insolación. Este efecto es menor en 2020 (quizás se instaló una tecnología más eficiente durante este año) aunque sigue habiendo un cierto comportamiento no lineal, con una segunda derivada ligeramente negativa, como se ve en la ecuación 5.6. No se cuenta con suficientes datos de 2021 como para interpretarlos, aunque si se ve claramente una mayor potencia instalada que en 2019.

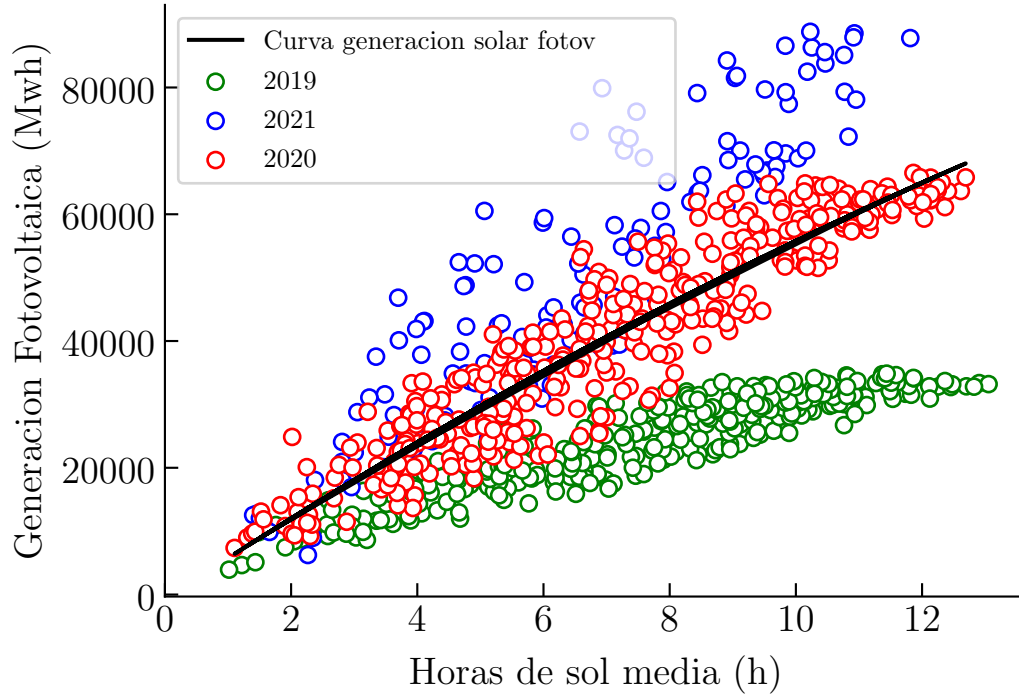


Figura 5.21: Diagrama de dispersión de la generación solar fotovoltaica en función de la horas de insolación medias en España durante 2019, 2020 y parte de 2021

La ecuación de la curva de generación fotovoltaica para el año 2020 es

$$Curva\ gen\ fotov.(MWh) = -86,79\left(\frac{MWh}{h^2}\right) \cdot sol^2(h^2) + 6512,24\left(\frac{MWh}{h}\right) \cdot sol(h) + -654,79(MWh)$$

Ecuación 5.6: Ecuación de la recta de generación fotovoltaica

La ecuación 5.6 se utilizará para predecir la generación fotovoltaica para cada día de 2020 de las centrales de esta tecnología en función de las horas de insolación que haya tenido ese día. Igual que con la eólica, al hacer esto se asume que cada central individual se comporta como la "central gigante" que representan todas las centrales juntas. El principal problema es que al haber cogido la media de España, no se sabe el comportamiento para valores de insolación diarios de menos de 1 hora o más de 13 horas y podría no obedecer a la ecuación de 5.6.

Al mirar la ecuación de 5.6 se puede ver que para valores de insolación menores

a 0.1 horas (puede ocurrir como se ve en la figura 5.14), la generación se vuelve negativa. Físicamente esto no tiene sentido así que la generación se limita a valores positivos, obteniendo la ecuación de generación fotovoltaica a tramos que se ve en 5.7

$$Gen\ fotov\ (MWh) = \begin{cases} 0 & si\ sol \leq 0,1h \\ -86,79 \cdot sol^2(h^2) + 6512,24 \cdot sol(h) + -654,79 & si\ sol > 0,1/h \end{cases}$$

Ecuación 5.7: Ecuación de la recta de generación fotovoltaica por tramos

$Gen\ fotov$ representa el término $factor_clima$ de la fórmula 5.3 para cálculo de generación en centrales con tecnología dependiente del clima. Para obtener las horas de insolación de cada central fotovoltaica para cada día se ha cogido la media de las horas de insolación de las 2 estaciones meteorológicas más cercanas (excluyendo las 109 que no son capaces de obtener esos datos).

Al aplicar la fórmula 5.3 con la ecuación 5.7 obtenemos la generación de cada central fotovoltaica para cada día.

En la figura 5.22 podemos ver un mapa de calor con la distribución geográfica de la generación fotovoltaica durante el año 2020. En la figura se ven principalmente los picos, que coinciden también con los principales lugares de potencia fotovoltaica instalada en el país.

5.3. Cálculo de generación con geolocalización

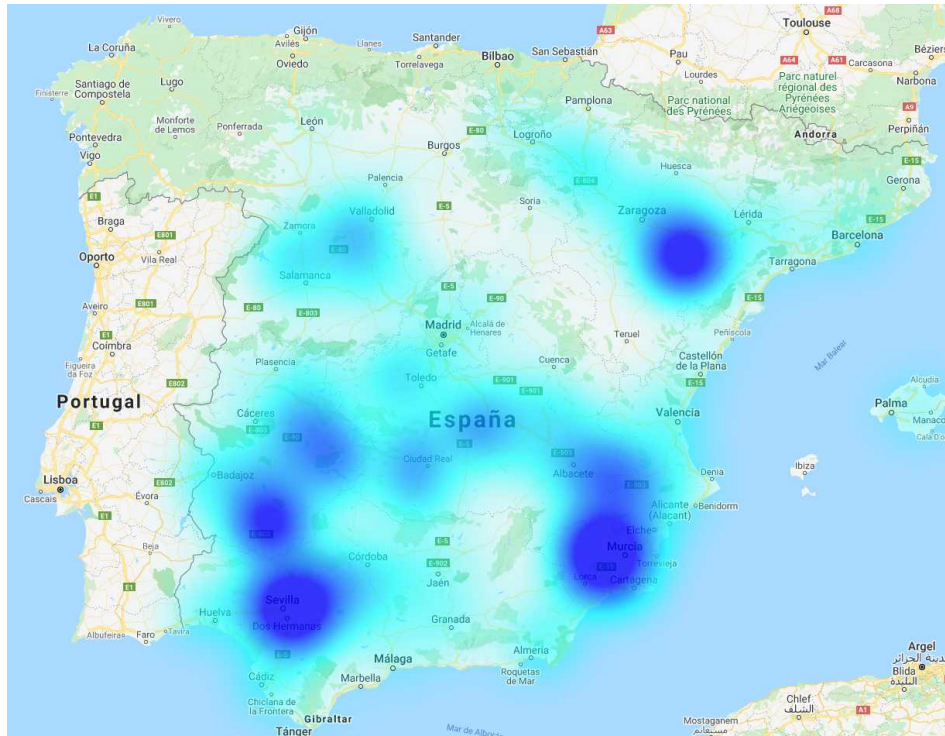


Figura 5.22: Mapa de calor de la generación fotovoltaica durante 2020

Solar térmica

Los datos de generación solar térmica ha sido obtenida de ESIOS, del apartado de *GENERACIÓN MEDIDA SOLAR TÉRMICA* [ESIOSsolarterm]. Se pueden ver estos datos en la figura 5.23. Igual que con la fotovoltaica se ve una dependencia climática.

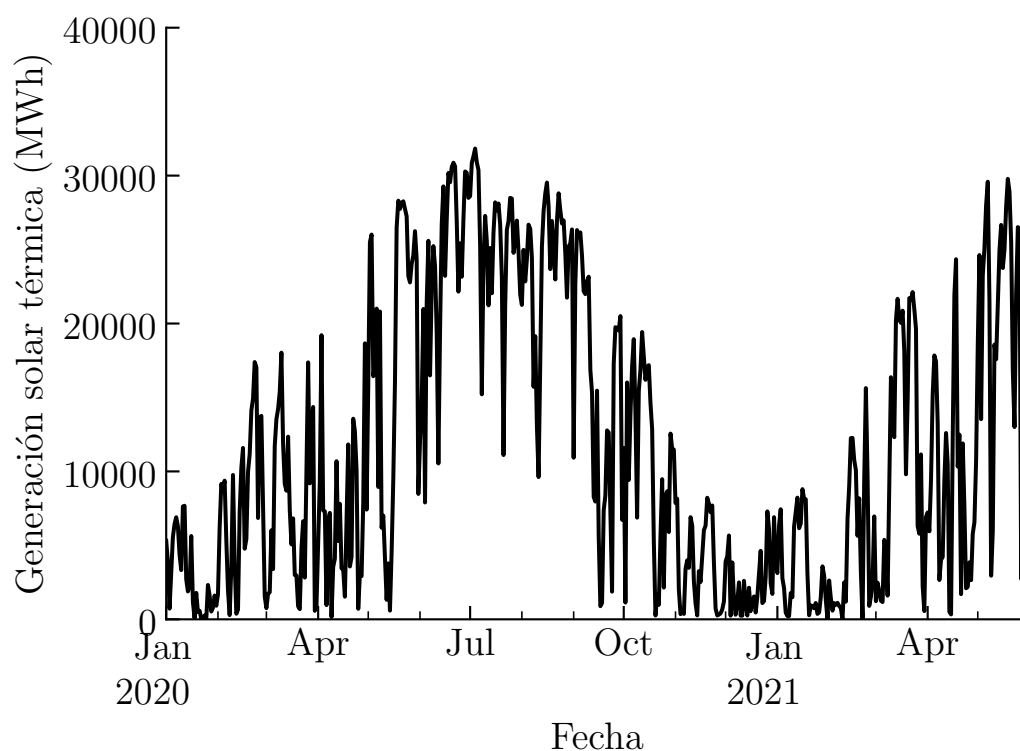


Figura 5.23: Generación solar térmica durante 2020 y parte de 2021

Al juntar los datos de generación solar térmica de 5.23 con los datos de la media de horas de insolación de España se obtiene el diagrama de dispersión de la figura 5.24. Se aprecia una clara relación entre las dos variables, pese a que no sea tan alta como la de la fotovoltaica. No se aprecia ningún cambio entre los años 2019 y 2020.

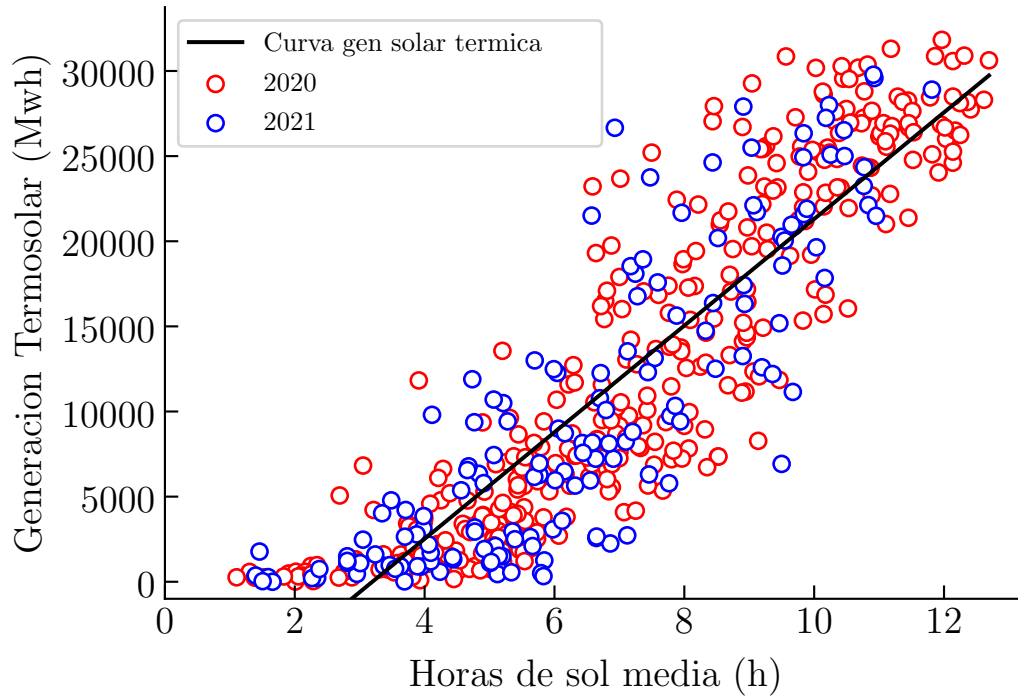


Figura 5.24: Diagrama de dispersión de la generación solar térmica en función de la horas de insolación medias en España durante 2020 y parte de 2021

La ecuación de la recta de generación solar térmica es

$$Recta\ gen\ termica(MWh) = 3131,08\left(\frac{MWh}{h}\right) \cdot sol(h) - 10004,28(MWh)$$

Ecuación 5.8: Ecuación de la recta de generación eólica

Como se comentaba, se aprecia una clara relación lineal entre la generación térmica y las horas de insolación diarias así que se ha calculado el coeficiente de correlación de Pearson y su significancia.

$$Coeficiente\ de\ Pearson = 0,89 \quad p - value \approx 0$$

Estos resultados nos permiten afirmar que se puede usar las horas de insolación

medidas por las estaciones meteorológicas para predecir la generación térmica de las centrales térmicas cercanas.

Si se observa el gráfico de dispersión y la recta de generación térmica se puede ver que parece que hay un mínimo de horas de sol para que se empiece a producir solar térmica, en torno a 3.2 horas. Para evitar valores negativos de generación (no tiene sentido físicamente) se modifica la ecuación de la generación térmica, creando una nueva ecuación a tramos, como se ve en 5.9

$$Gen_{termica}(MWh) = \begin{cases} 0 & si \quad sol \leq 3,2h \\ 3131,08 \cdot sol(h) - 10004,28 & si \quad sol > 3,2/h \end{cases}$$

Ecuación 5.9: Ecuación de la recta de generación fotovoltaica por tramos

Gen_{termica} es el término *factor_clima* de la fórmula 5.3 para cálculo de generación en centrales con tecnología dependiente del clima. Para obtener las horas de insolación de cada central térmica se ha hecho lo mismo que con la fotovoltaica y para cada día se ha cogido la media de las horas de insolación de las 2 estaciones meteorológicas más cercanas (excluyendo las 109 que no son capaces de obtener esos datos).

Al aplicar la fórmula 5.3 con la ecuación 5.9 obtenemos la generación de cada central térmica para cada día.

En la figura 5.25 podemos ver un mapa de calor con la distribución geográfica de la generación térmica durante el año 2020.

5.3. Cálculo de generación con geolocalización

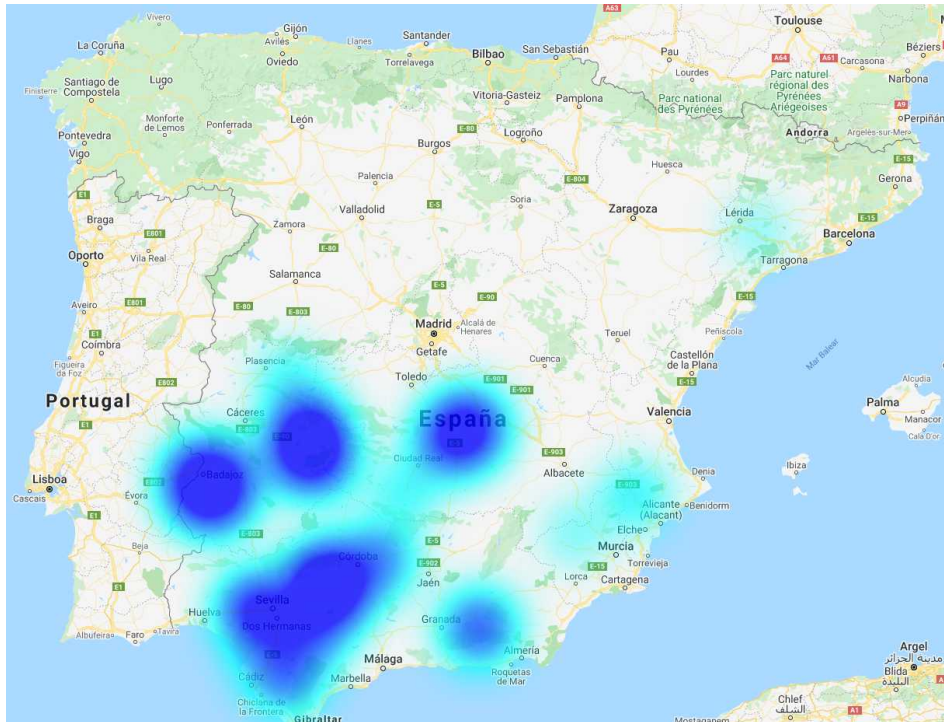


Figura 5.25: Mapa de calor de la generación térmica durante 2020

Capítulo 6

Nudos con alta indisponibilidad en 2020

Con el objetivo de saber qué nudos concretos son los más problemáticos de España se ha hecho un estudio sobre la tabla de indisponibilidades con los nudos asociados 4.8, obtenida en la sección 4.3.

Para ello, se ha creado una nueva tabla con el nombre de cada nudo que ha aparecido en alguna indisponibilidad durante 2020 y una variable llamada *Tiempo total*, donde se irá rellenando el tiempo que ese nudo tiene indisponibilidades. Para rellenar esta tabla, que se encuentra inicialmente rellena con ceros, se ha recorrido la tabla de indisponibilidades 3.2 y para los datos de cada día se ha hecho lo siguiente: Se han cogido todos los nudos que aparecen ese día (variables *Nudo 1* y *Nudo 2*) y, para cada uno de esos nudos se han tomado todos sus variables *INICIO* y *FINAL*. Con el propósito de no caer en repeticiones con el tiempo en caso de que un nudo aparezca en más de una indisponibilidad al mismo tiempo se pasan los *INICIO* y *FINAL* a una función que devuelve una lista con los *INICIO* y *FINAL*, pero eliminando las superposiciones en el tiempo. Estos tiempos se suman para obtener el tiempo total que cada nudo tiene indisponibilidades ese día. Por último, este tiempo se suma a la variable *Tiempo total* de la tabla de nudos para el nudo concreto, que muestra el tiempo en segundos que dura la indisponibilidad.

La tabla resultante se puede ver en la tabla 6.1.

Nudo	Tiempo total (s)
A. Zinc	725400.0
Abadiano	1555200.0
Abanto	2104320.0
Abegondo	0.0
Aboño	0.0
Acebizk	572400.0
Aceca	8200800.0
Aceriasa	1452600.0
Acerinox	0.0
Adrall	1011600.0

Tabla 6.1: Nudos y su tiempo total con indisponibilidades durante el año 2020 (limitado a 10 resultados)

Dividiendo el *Tiempo total* entre el número de segundos en un año (31536000) podemos obtener el porcentaje del año, que un nudo ha tenido indisponibilidades, que es un valor más fácil de interpretar. Podemos ver los nudos con más porcentaje del año con indisponibilidades asociadas en la tabla 6.2.

Nudos	Tiempo tot (s)	Porcentaje del año (%)
Tabiella	30153600.0	95.616438
Soto de Ribera	30153600.0	95.616438
Guadalquivir Medio	30132000.0	95.547945
Balboa	29789220.0	94.460997
Sider. Balboa	29789220.0	94.460997
Galapagar	29300340.0	92.910769
Valdeconejos	25898400.0	82.123288
S. Costera	25833600.0	81.917808
Magallón	25372800.0	80.456621
Talavera	23306400.0	73.904110

Tabla 6.2: Nudos con más tiempo con indisponibilidades (limitado a 10 resultados)

Es importante destacar que, los nudos concretos no tienen por qué haber aparecido tanto porcentaje del año, ya que al no contar con los datos de todos los

nudos, algunos se asociaban al nudo más cercano. Este es por ejemplo, el caso de Tabiella, donde no se contaba con información de Trasona así que sus indisponibilidades se asocian a su nudo más cercano, en este caso Tabiella. Por lo tanto Tabiella, representa tanto a Tabiella como a Trasona.

En la figura 6.1 podemos ver el histograma del porcentaje del tiempo de 2020 que un nudo tiene indisponibilidades.

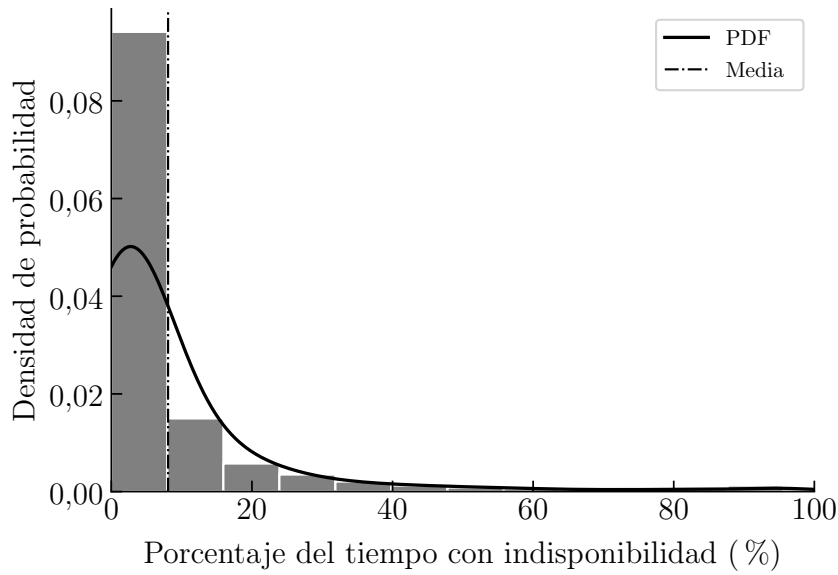


Figura 6.1: Histograma del porcentaje del tiempo que un nudo tiene indisponibilidades asociadas

6.1. Programas de Indisponibilidades de Red

A continuación se establecerán distintos programas, en función del porcentaje del tiempo que los nudos tienen indisponibilidades. Se han creado 3 programas, P25, P50 y P90.

6.1.1. P25

El programa P25 incluye los nudos que tienen indisponibilidades al menos un 25 % del tiempo. Se puede ver una tabla con los 46 nombres de estos nudos en el anexo 11.4.

CAPÍTULO 6. NUDOS CON ALTA INDISPONIBILIDAD EN 2020

En la figura 6.2 se puede ver un mapa de calor con los nudos del P25. En este mapa, todos los nudos aportan la misma cantidad de "calor" independientemente de que algunos tengan indisponibilidades solo el 25 % del tiempo y otros mucho más.

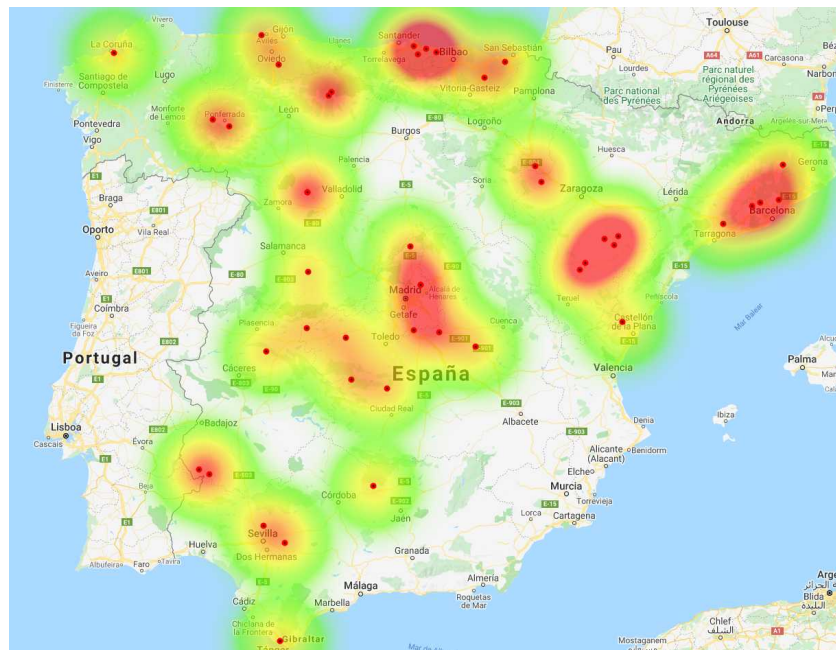


Figura 6.2: Mapa de calor de los nudos del P25

6.1.2. P50

El programa P50 incluye los nudos que tienen indisponibilidades al menos un 50 % del tiempo. En este caso solo hay 18 nudos y se pueden ver en la tabla 6.3.

Nudo	Porcentaje del año (%)
Tabiella	0.956164
Soto de Ribera	0.956164
La Lomba	0.591895
Magallón	0.804566
Almaraz	0.581849
Sider. Balboa	0.944610
Balboa	0.944610
Alcores	0.676598
Galapagar	0.929108
Talavera	0.739041
P. Ingenieros	0.527310
Villaverde	0.552397
Guadalquivir Medio	0.955479
Morata	0.509587
S. Costera	0.819178
Valdeconejos	0.821233
Rubi	0.685443
Can Jordi	0.586357

Tabla 6.3: Nudos de P50

En la figura 6.3 se puede ver un mapa de calor con los nudos del P50, igual que en el anterior mapa de calor, todos los nudos aportan la misma cantidad de "calor".

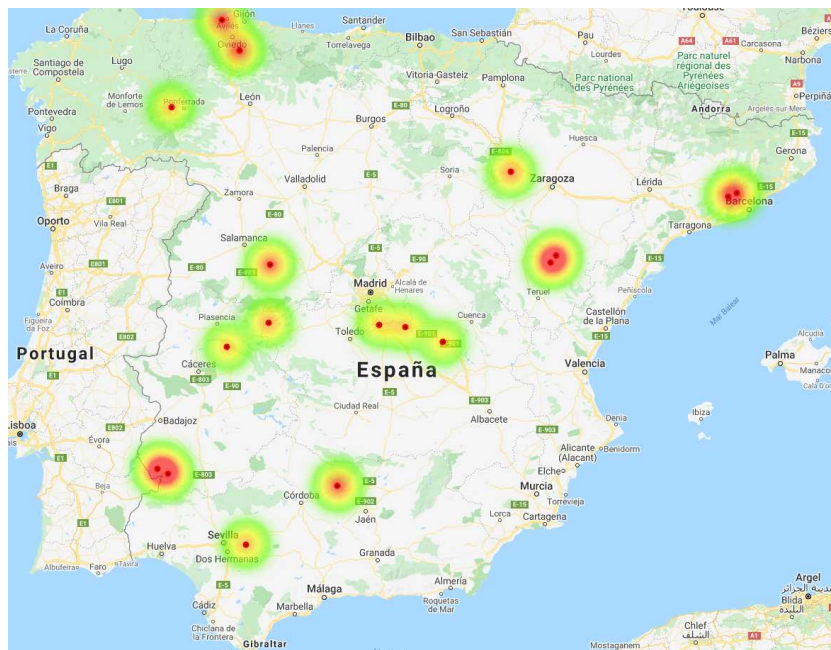


Figura 6.3: Mapa de calor de los nudos del P50

6.1.3. P90

El programa P90 incluye los nudos que tienen indisponibilidades al menos un 90 % del tiempo.

Los 6 nudos de P90 se pueden ver en la tabla 6.4. Parece que hay 2 pares de nudos relacionados entre ellos al tener el mismo porcentaje del año. Esto se debe a 2 indisponibilidades que han estado activas gran parte del año. Estas indisponibilidades son *400 kV BALBOA SIDERURGICA BALBOA* y *220 kV SOTO DE RIBERA TRASONA*.

Nudo	Porcentaje del año (%)
Tabiella	0.956164
Soto de Ribera	0.956164
Sider. Balboa	0.944610
Balboa	0.944610
Galapagar	0.929108
Guadalquivir Medio	0.955479

Tabla 6.4: Nudos de P90

6.2. Cambios en los nudos con Indisponibilidades de Red bajo condiciones de generación extremas

En la figura 6.4 se puede ver un mapa de calor con los nudos del P90.

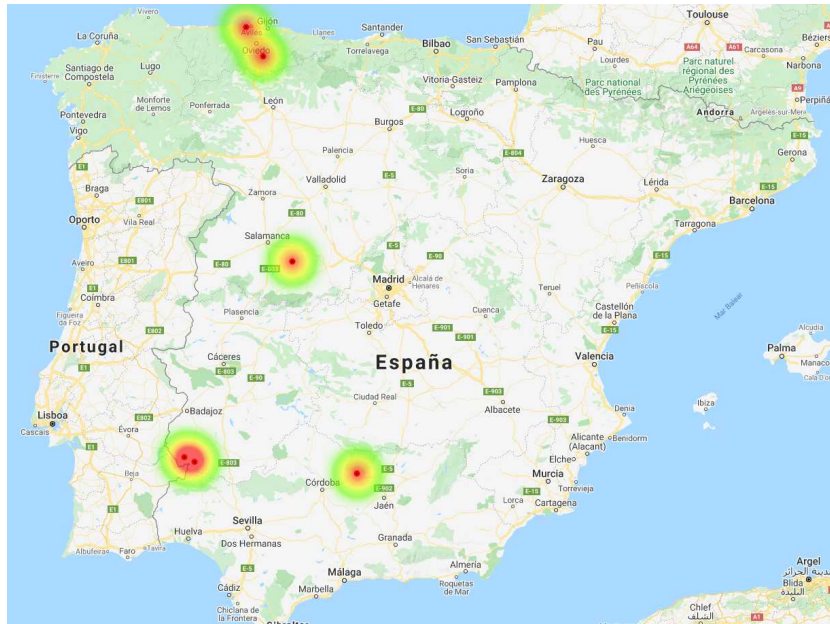


Figura 6.4: Mapa de calor de los nudos del P90

6.2. Cambios en los nudos con Indisponibilidades de Red bajo condiciones de generación extremas

Las tecnologías de generación con más variabilidad e importancia del mix de generación son la eólica y la fotovoltaica. Es por esto por lo que se ha querido hacer un estudio sobre la variabilidad de los nudos en los que hay indisponibilidades cuando nos encontramos en los máximos y mínimos de generación de estas dos tecnologías. Las fuentes de datos de generación son las empleadas en la sección 5.3.

6.2.1. Escenarios de generación eólica

Generación eólica en máximos

Para este caso, se han analizado únicamente los días de 2020 en los que la generación eólica estaba en el top 20 % de su generación anual. Se pueden ver estos días en la figura 6.5. Estos días están no están distribuidos uniformemente a

lo largo del año, como se ve en 6.6 lo que disminuye la confianza en los resultados. Si todos están concentrados en una franja del año, se puede fácilmente confundir la influencia de la generación eólica con cualquier otro factor, como por ejemplo, que en esa franja del año ha habido una indisponibilidad por reparación de una línea.

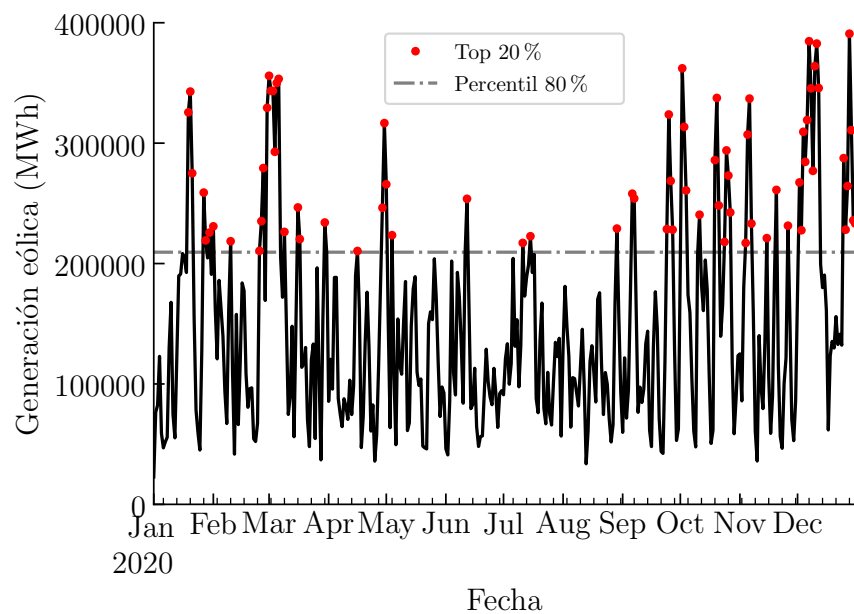


Figura 6.5: Top 20 % días de generación eólica durante 2020

6.2. Cambios en los nudos con Indisponibilidades de Red bajo condiciones de generación extremas

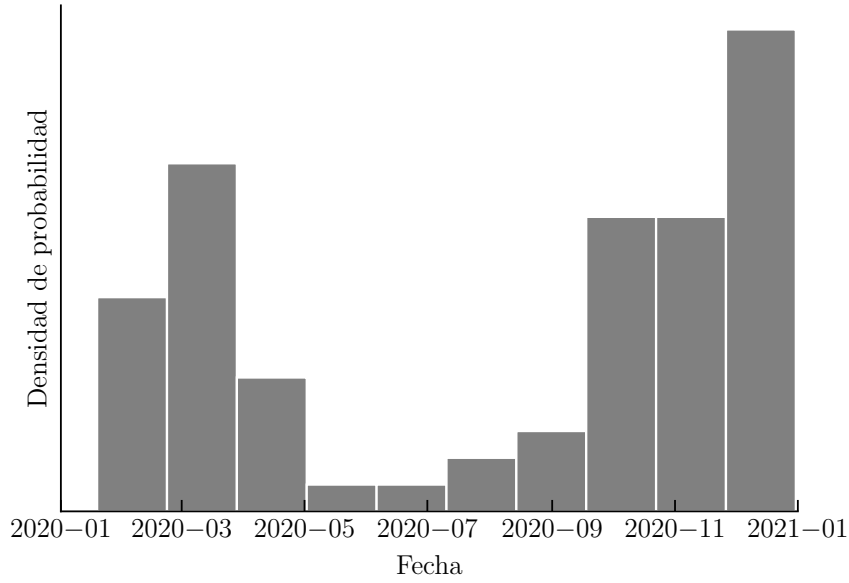


Figura 6.6: Distribución en el año del top 20 % días de generación eólica durante 2020

Para estos días se ha calculado el porcentaje del tiempo que cada nudo tiene indisponibilidades de la misma forma que se comentó al inicio de este capítulo. En este caso, para calcular el porcentaje no se ha dividido entre el número de segundos del año sino el número de segundos en 73 días (20 % de 365 días). El porcentaje de cada nudo se ha comparado con el porcentaje de ese nudo durante todo 2020 para calcular el incremento de la siguiente forma, que se usará para todos los demás escenarios. Los valores positivos significan que el porcentaje de tiempo con indisponibilidades aumenta y los valores negativos significan que disminuyen. La variable *Incremento* se utilizará para detectar las variaciones de nudos más extremas.

$$\text{Incremento}(\%) = \frac{\text{Porcentaje nuevo} - \text{Porcentaje 2020}}{\text{Porcentaje 2020}}$$

Ecuación 6.1: Cálculo del incremento del porcentaje de tiempo con indisponibilidad en cada nudo

También se ha calculado la *diferencia* entre el porcentaje de 2020 y el porcentaje en escenario de generación eólica top 20 % de la manera mostrada en 6.2. De la misma forma que en la variable *Incremento*, los valores positivos indican que el porcentaje de indisponibilidades en el nudo ha aumentado. Esta variable se usará para estudiar la tendencia general calculando la media, ya que no se puede usar la variable *Incremento* con este objetivo dado que la variable *Incremento* y su media son muy sensibles a porcentajes pequeños que aumenten mucho en proporción pero poco en valores absolutos. De esta forma, pasar de 1 % a 3 % es un *Incremento* de 200 %, que posiblemente responda más a la casualidad por ser pocos datos, mientras que su *diferencia* es tan solo 2 % que, al hacer la media entre todas las .aleatorias” deberían de compensarse entre ellas.

$$Diferencia(\%) = Porcentaje\ nuevo - Porcentaje\ 2020$$

Ecuación 6.2: Cálculo de la diferencia del porcentaje de tiempo con indisponibilidad en cada nudo

En la tabla 6.5 se pueden ver los nudos con las mayores diferencias entre el porcentaje de 2020 y el porcentaje en circunstancias de generación eólica top 20 %. Con el objetivo de no introducir nudos que hayan variado mucho por casualidad, solo se muestran los nudos que tengan más de un 30 % del tiempo indisponibilidades (ya sea en el porcentaje de 2020 o en el porcentaje en circunstancias de generación eólica de top 20 %). Se pueden observar cambios muy significativos en estos nudos; el nudo de Cofrentes tiene un incremento de casi el 90 %, siendo el más extremo. Por otro lado, el nudo de Montearenas es el que más disminuye, ya que lo hace en un 37 %. La media de *diferencia* es +0.17 %, sin la condición del 30 % comentada antes, ya que la variable *diferencia* no es sensible a variaciones absolutas pequeñas. Esto implica que, en circunstancias de máximos de generación eólica los nudos de media tienen prácticamente las mismas indisponibilidades.

6.2. Cambios en los nudos con Indisponibilidades de Red bajo condiciones de generación extremas

Nudos	Porcentaje 2020 (%)	Porcentaje top 20 % eólica (%)	Diferencia (%)	Incremento (%)
Confrentes	17.066020	32.420091	15.354072	89.968673
Arañuelo	22.203196	37.157534	14.954338	67.352185
Gramanet	28.892694	43.378995	14.486301	50.138285
Teruel	39.908676	57.534247	17.625571	44.164760
Itxaso	36.175799	51.484018	15.308219	42.316188
Aceca	26.004566	36.073059	10.068493	38.718174
Puerto de la Cruz	27.647451	38.013699	10.366248	37.494409
...	...	...	...	...
Talavera	73.904110	66.780822	-7.123288	-9.638554
Cicero	39.474886	35.616438	-3.858447	-9.774436
Begues	46.832192	42.237443	-4.594749	-9.811091
Torrijos	34.520548	29.794521	-4.726027	-13.690476
Magallón	80.456621	68.721461	-11.735160	-14.585698
Alcores	67.659817	49.143836	-18.515982	-27.366290
Montearenas	31.055936	19.691781	-11.364155	-36.592538

Tabla 6.5: Nudos con mayores cambios en escenario de generación eólica máxima

Generación eólica en mínimos

Para este caso se han analizado únicamente los días de 2020 en los que la generación eólica estaba en el mínimo 20 % de su generación durante el año. Se pueden ver estos días en la figura 6.7. Estos días están extremadamente uniformemente distribuidos (mucho más que en los días de generación eólica en máximos) a lo largo del año, como se ve en 6.8 lo que aumenta mucho la confianza en los resultados.

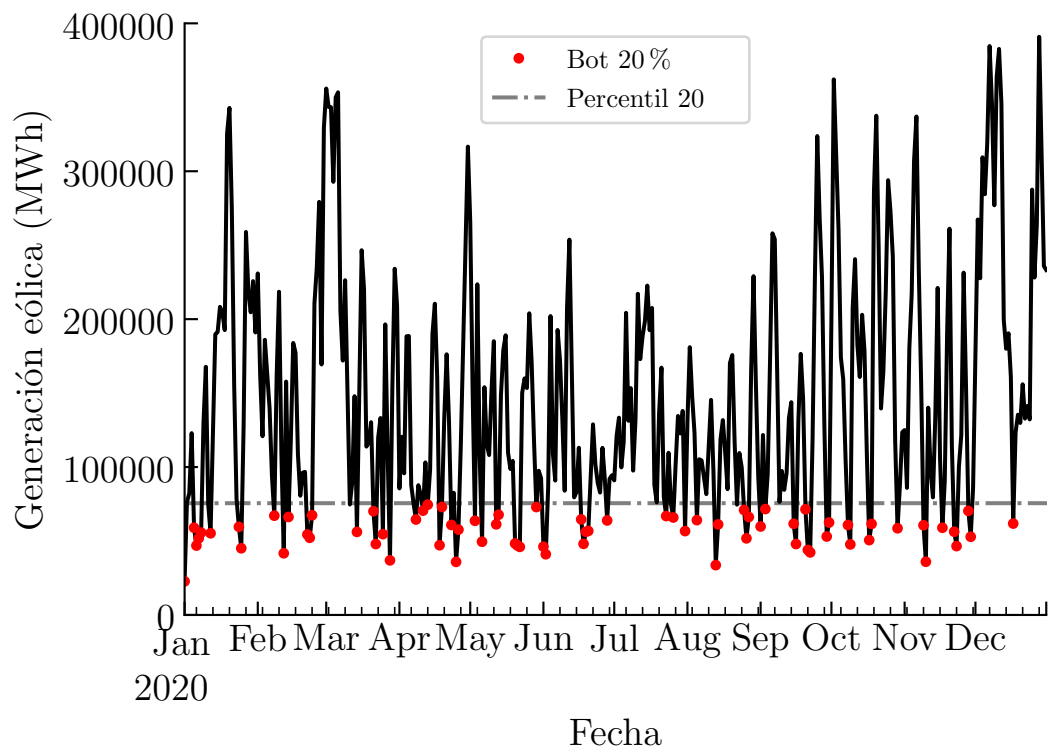


Figura 6.7: Bot 20 % días de generación eólica durante 2020

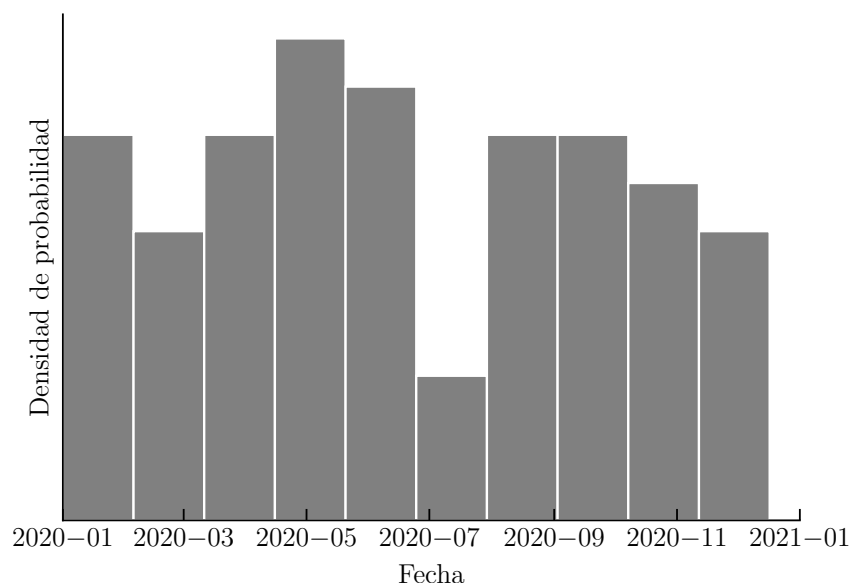


Figura 6.8: Distribución en el año del bot 20 % días de generación eólica durante 2020

Para estos días se ha calculado el porcentaje del tiempo que cada nudo tiene indisponibilidades de la misma forma que en el escenario de mínimos de generación eólica. El cálculo del *incremento* se hace con la formula de 6.1 y el cálculo de *diferencia* con 6.2.

En la tabla 6.6 se pueden ver los nudos con las mayores diferencias entre el porcentaje de 2020 y el porcentaje en circunstancias de generación eólica bot 20 %. Igual que con los máximos de generación eólica, solo se muestran los nudos que tengan más de un 30 % del tiempo indisponibilidades (ya sea en el porcentaje de 2020 o en el porcentaje en circunstancias de generación eólica de bot 20 %). Los cambios en este escenario son significativamente menores que en el escenario de generación eólica en máximos. La media de *diferencia* es -0.0224 % lo que significa que en circunstancias de mínimos de generación eólica los nudos tienen de media el mismo porcentaje del tiempo indisponibilidades.

Nudos	Porcentaje 2020 (%)	Porcentaje bot 20 % eólica (%)	Diferencia (%)	Incremento (%)
Montearenas	31.055936	38.755708	7.699772	24.793237
Santurce	25.273782	31.506849	6.233067	24.662185
Cicero	39.474886	46.575342	7.100457	17.987276
Zierbana	27.899543	32.648402	4.748858	17.021277
Solorzano	43.732877	49.315068	5.582192	12.764291
Coslada	33.105023	35.730594	2.625571	7.931034
Guillena	45.793379	49.115297	3.321918	7.254144
...	...	...	...	...
Villaverde	55.239726	47.260274	-7.979452	-14.445133
La Lomba	59.189498	48.972603	-10.216895	-17.261331
Can Jordi	58.635654	47.602740	-11.032915	-18.816051
Almaraz	58.184932	45.776256	-12.408676	-21.326270
Velilla	47.123288	36.986301	-10.136986	-21.511628
Guardo	47.853881	36.986301	-10.867580	-22.709924
Villaviciosa	32.317352	24.200913	-8.116438	-25.114800

Tabla 6.6: Nudos con mayores cambios en escenario de generación eólica mínima

6.2.2. Escenarios de generación solar fotovoltaica

Generación fotovoltaica en máximos

Para este caso se han analizado únicamente los días de 2020 en los que la generación solar fotovoltaica estaba en el top 20 % de su generación durante el año. Se pueden ver estos días en la figura 6.9. Se puede ver que estos días están todos agrupados en solo 4-5 meses del año en la figura 6.10, esto hace que los resultados se obtengan en este apartado no tengan tanta confianza, ya que la probabilidad de que un incremento sea casualidad es más alta.

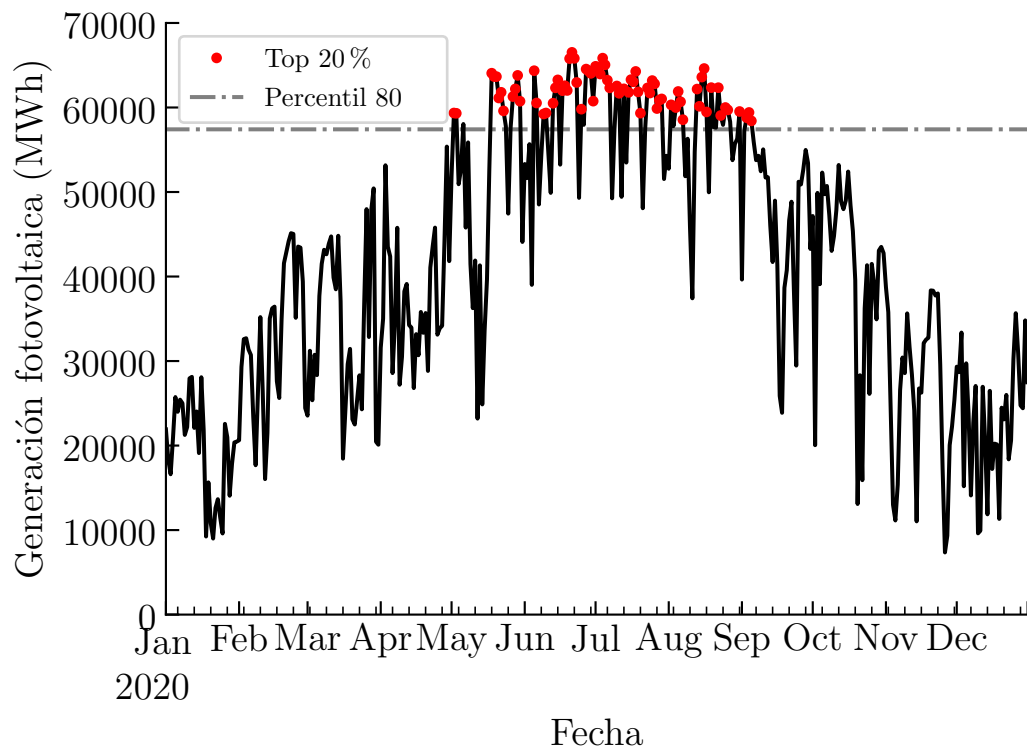


Figura 6.9: Top 20 % días de generación fotovoltaica durante 2020

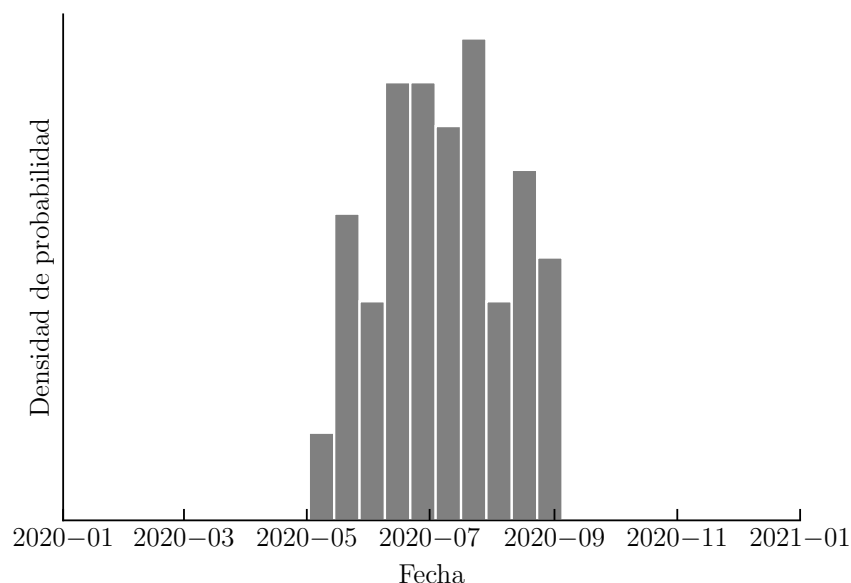


Figura 6.10: Distribución en el año del top 20 % días de generación solar fotovoltaica durante 2020

Para estos días se ha calculado el porcentaje del tiempo que cada nudo tiene indisponibilidades de la misma forma que en los escenarios de generación eólica. El cálculo del incremento se hace con la formula de 6.1 y el cálculo de *diferencia* con 6.2.

En la tabla 6.7 se pueden ver los nudos con las mayores diferencias entre el porcentaje de 2020 y el porcentaje en circunstancias de generación fotovoltaica top 20 %. Igual que con los escenarios de generación eólica, solo se muestran los nudos que tengan más de un 30 % del tiempo indisponibilidades (ya sea en el porcentaje de 2020 o en el porcentaje en circunstancias de generación fotovoltaica de top 20 %). Los incrementos son mucho más extremos que en los escenarios de eólica, muchos de estos incrementos pueden ser debidos a otros factores distintos a los factores de generación fotovoltaica debido a la poca distribución a lo largo del año de los días de generación fotovoltaica top 20 %. La media de *diferencia* es +1.39 %, esto indica que los nudos tienen ligeramente más indisponibilidades en escenario de generación fotovoltaica máxima. Aun así esto muy probablemente no es causa de una mayor generación fotovoltaica sino que son los meses de verano y con el calor hay mas indisponibilidades.

6.2. Cambios en los nudos con Indisponibilidades de Red bajo condiciones de generación extremas

Nudos	Porcentaje 2020 (%)	Porcentaje top 20 % fotovoltaica (%)	Diferencia (%)	Incremento (%)
Telleo	17.442922	51.141553	33.698630	193.193717
Boimente	14.315068	41.666667	27.351598	191.068581
La Robla	14.794140	41.609589	26.815449	181.257234
Aguayo	22.814307	62.557078	39.742770	174.201081
La Pereda	11.426941	30.878995	19.452055	170.229770
V. Gasteiz	14.942922	39.611872	24.668950	165.087853
A. de la Front. (Sur)	20.296804	51.655251	31.358447	154.499438
...	...	...	...	...
Begues	46.832192	36.101598	-10.730594	-22.912858
Mudejar	42.328767	28.082192	-14.246575	-33.656958
Teruel	39.908676	24.657534	-15.251142	-38.215103
Solorzano	43.732877	18.378995	-25.353881	-57.974419
Cicero	39.474886	15.068493	-24.406393	-61.827646
Guillena	45.793379	14.155251	-31.638128	-69.088870
Puerto	39.348364	8.447489	-30.900875	-78.531538

Tabla 6.7: Nudos con mayores cambios en escenario de generación fotovoltaica máxima

Generación fotovoltaica en mínimos

Para este caso se han analizado únicamente los días de 2020 en los que la generación solar fotovoltaica estaba en el bot 20 % de su generación durante el año. Se pueden ver estos días en la figura 6.11. Estos días están más distribuidos a lo largo del año que los de generación fotovoltaica máxima y además están en 2 periodos distintos como se ve en la figura 6.12. Esto permite otorgar a estos resultados de más confianza que en el caso de generación fotovoltaica máxima aunque no es tan alta que si la distribución fuera uniforme a lo largo del año.

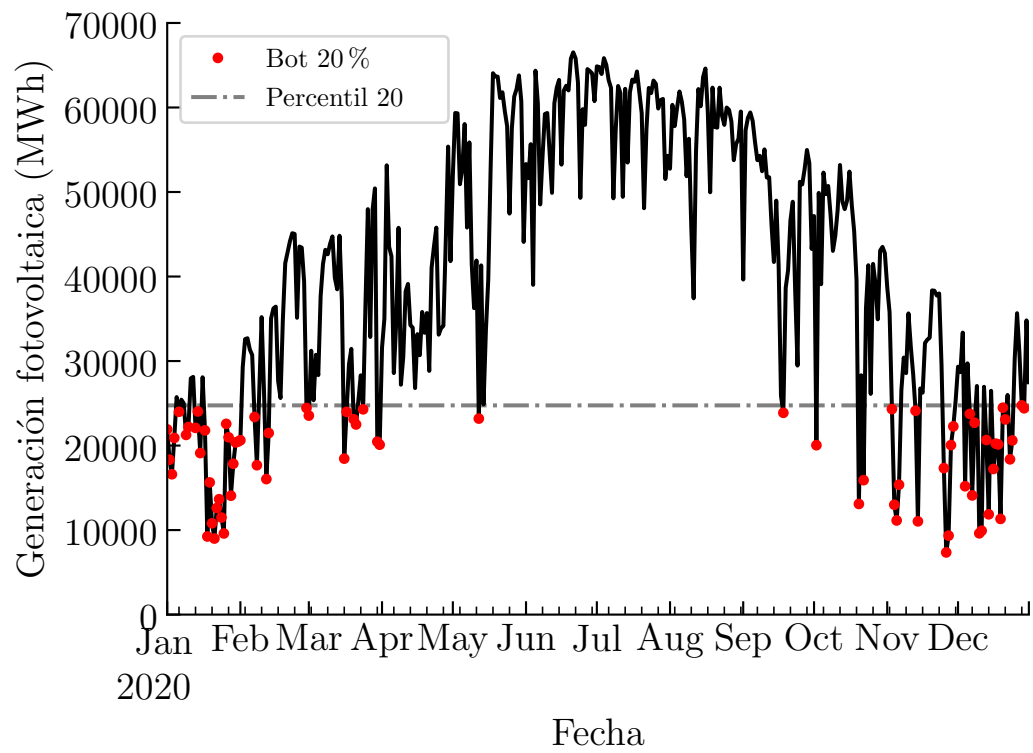


Figura 6.11: Bot 20 % días de generación fotovoltaica durante 2020

6.2. Cambios en los nudos con Indisponibilidades de Red bajo condiciones de generación extremas

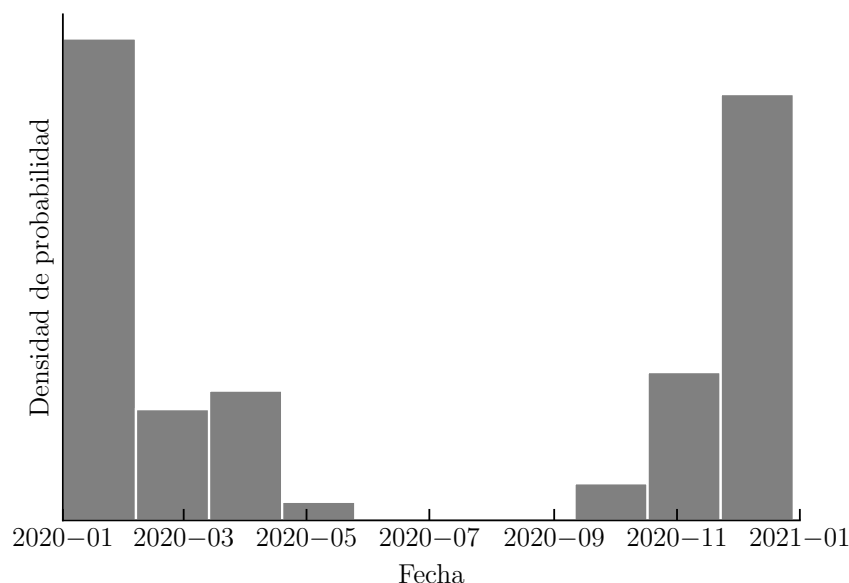


Figura 6.12: Distribución en el año del bot 20 % días de generación fotovoltaica durante 2020

Para estos días se ha calculado el porcentaje del tiempo que cada nudo tiene indisponibilidades de la misma forma que en los escenarios de generación eólica. El cálculo del incremento se hace con la fórmula de 6.1 y el cálculo de *diferencia* con 6.2.

En la tabla 6.8 se pueden ver los nudos con las mayores diferencias entre el porcentaje de 2020 y el porcentaje en circunstancias de generación fotovoltaica bot 20 %. Igual que con los escenarios de generación eólica, solo se muestran los nudos que tengan más de un 30 % del tiempo indisponibilidades (ya sea en el porcentaje de 2020 o en el porcentaje en circunstancias de generación fotovoltaica de bot 20 %). Los incrementos son mucho más extremos que en los escenarios de eólica, muchos de estos incrementos pueden ser debidos a otros factores distintos a los factores de generación fotovoltaica debido a la poca distribución a lo largo del año de los días de generación fotovoltaica top 20 %. La media de *diferencia* es -2.15 %, esto indica que los nudos tienen ligeramente menos indisponibilidades en escenario de generación fotovoltaica mínima. Aun así esto muy probablemente no es causa de una menor generación fotovoltaica sino que son los meses de invierno y con el frío hay menos indisponibilidades.

CAPÍTULO 6. NUDOS CON ALTA INDISPONIBILIDAD EN 2020

Nudos	Porcentaje 2020 (%)	Porcentaje bot 20 % fotovoltaica (%)	Diferencia (%)	Incremento (%)
Monforte del Cid	13.858257	34.245624	20.387367	147.113497
Carmona	14.600457	32.305936	17.705479	121.266615
FARDIOUA	24.999049	49.315068	24.316020	97.267780
Tarifa	24.999049	49.315068	24.316020	97.267780
Amorebieta	20.810122	38.525495	17.715373	85.128636
Puerto de la Cruz	27.647451	50.342466	22.695015	82.087190
Puerto	39.348364	67.465753	28.117390	71.457583
...	...	...	...	...
Begues	46.832192	26.369863	-20.462329	-43.692870
Alcores	67.659817	37.671233	-29.988584	-44.322592
La Mudarra	37.254566	20.490868	-16.763699	-44.997702
Itxaso	36.175799	17.979452	-18.196347	-50.299779
Morata	50.958714	24.657534	-26.301180	-51.612723
Montearenas	31.055936	12.214612	-18.841324	-60.668995
Torrijos	34.520548	9.246575	-25.273973	-73.214286

Tabla 6.8: Nudos con mayores cambios en escenario de generación fotovoltaica mínima

Capítulo 7

Herramienta de visualización de indisponibilidades de red y generación

En este trabajo se ha decidido crear una herramienta que permitiera la visualización interactiva de los datos de geolocalización de indisponibilidades de red del capítulo 4 y los datos de geolocalización de la generación del capítulo 5. En la figura 7.1 se puede ver como es la herramienta.

El objetivo de esta herramienta es mostrar la evolución de indisponibilidades y generación a lo largo del tiempo o de días concretos. La herramienta también se puede usar para buscar picos no habituales de generación para ver su efecto en las indisponibilidades. También se podría usar para la toma de decisiones sobre la localización de una instalación de generación con el objetivo de que no se perjudique por indisponibilidades.

CAPÍTULO 7. HERRAMIENTA DE VISUALIZACIÓN DE INDISPONIBILIDADES DE RED Y GENERACIÓN

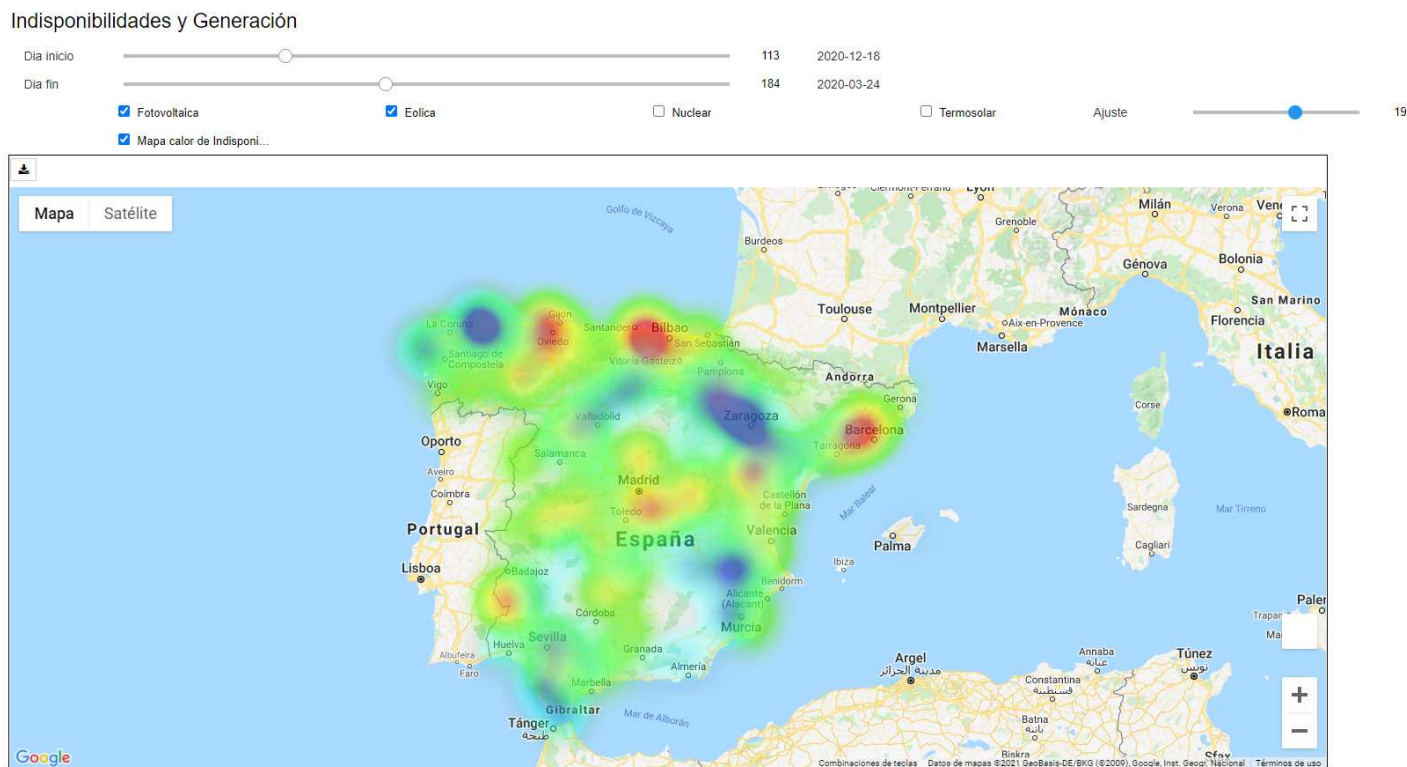


Figura 7.1: Herramienta de visualización de indisponibilidades de red y generación

Como se puede ver en la figure 7.1 la herramienta permite seleccionar con una barra deslizable una fecha de inicio y una fecha de fin, que delimitan el periodo que se ve en el mapa. También se pueden seleccionar qué tecnologías de generación se quieren ver en la producción.

La herramienta muestra en el mapa de calor rojo/amarillo/verde las indisponibilidades de red durante el periodo elegido y en el mapa de calor azul la generación durante el periodo. Si se desactivan todas las tecnologías el mapa de calor de generación desaparece, y si se desactiva la opción de *Mapa calor Indisponibilidades* se desactiva el de Indisponibilidades de red.

Por último la herramienta cuenta con una barra deslizable que se llama *Ajuste*, esta barra permite limitar el máximo de "calor" de cada punto del mapa de calor de generación. Esto es necesario ya que los mapas de calor son muy sensibles a los picos, y al mostrar la generación de tecnologías distintas estos picos son muy

comunes. El ajuste es exponencial debido a la gran diferencia entre los valores de generación entre tecnologías.

Capítulo 8

Modelos de predicción de Restricciones Técnicas

El objetivo de este capítulo es crear un modelo que pueda predecir con precisión la energía diaria por restricciones en Fase I. Los modelos usados son modelos de regresión. Para hacerlo se usaran los informes diarios de Restricciones Técnicas del PDBF comentados en el capítulo 3 junto con otras variables. Es importante destacar que, al contrario de los anteriores capítulos, estos modelos son de predicción y por tanto no pueden utilizar información que no se conociera en el momento de la predicción.

Estos modelos se entrenan con el objetivo de predecir cada día el valor de la energía de restricciones en fase I a subir al día siguiente.

8.1. Variable de salida

La variable de salida es la energía por restricciones en Fase I (a subir) agrupada de forma diaria, y es extraída de *ENERGÍA ASIGNADA RESTRICCIONES TÉCNICAS DIARIO FASE I* de ESIOS [45]. Se han descargado los datos desde el 01/01/2020 hasta el 21/03/2021 ya que solo contamos con los informes de Restricciones Técnicas comprendidos entre esas fechas. Se puede ver la serie temporal en la figura 8.1. Se aprecian valores negativos en la serie temporal, estos ocurren cuando la energía por restricciones en Fase I es a bajar. Como se puede ver en la figura esto es algo raro (solo el 2.5 % de los días de nuestra serie temporal) y es la Fase II la que tiene que bajar energía.

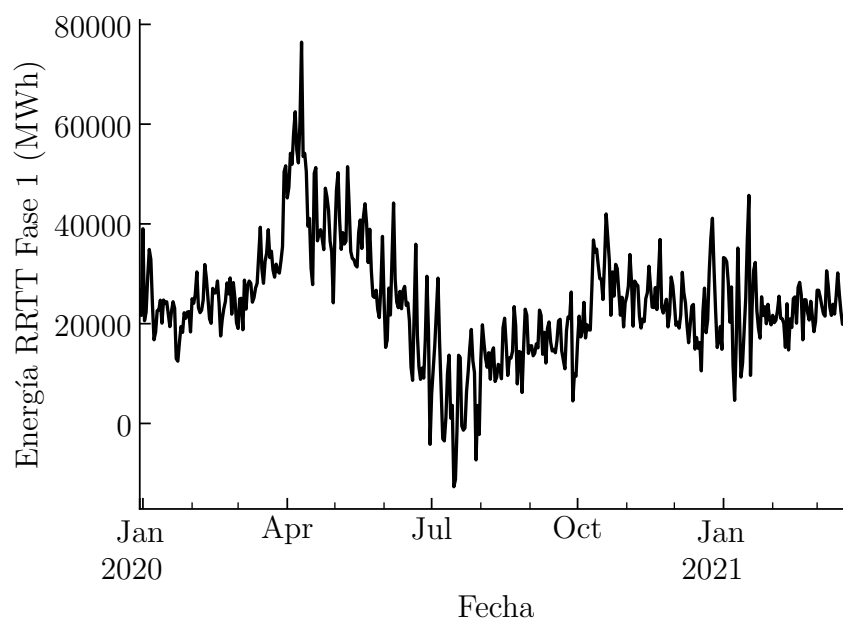


Figura 8.1: Energía restricciones técnicas Fase I (a subir) durante 2020 y parte de 2021

Si observamos la figura 8.2 se puede ver la serie temporal desde 2017. Parece claro que la pandemia de 2020 ha tenido un gran impacto sobre la Energía asignada a Restricciones Técnicas y sin duda entrenar nuestros modelos con estos datos anómalos impacta negativamente en los resultados de los mismos.

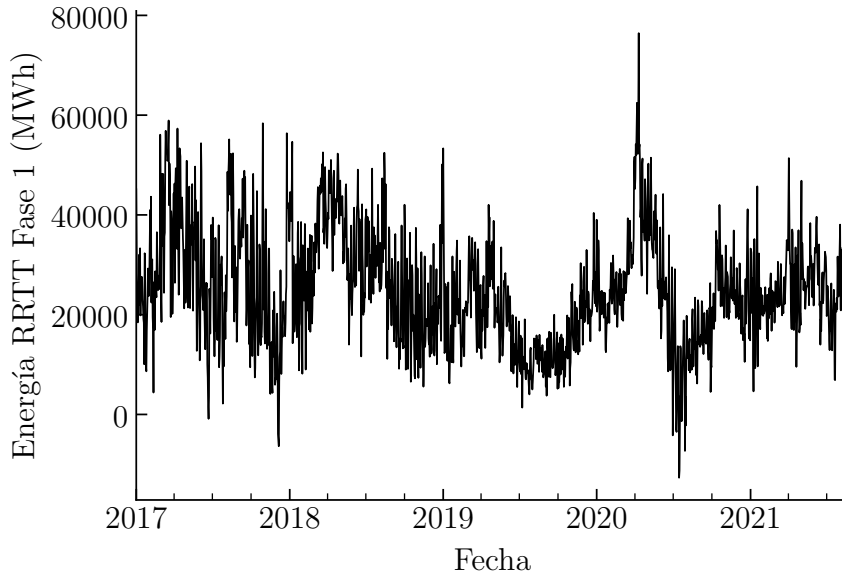


Figura 8.2: Energía restricciones técnicas Fase I (a subir) de 2017 a 2021

8.2. Variables de entrada

Como variables de entrada se han utilizado variables íntimamente relacionadas con las Restricciones Técnica, como los Informes de Restricciones Técnicas y otras no directamente relacionadas, pero sí que ayudaran al modelo a entender el estado de la red eléctrica en ese momento, como diferentes previsiones de generación, en particular renovable por su gran variabilidad.

8.2.1. Informes de Restricciones Técnicas

Como ya se ha comentado en la introducción de este capítulo los informes de Restricciones Técnicas son una de las variables de entrada, en concreto la tabla de Restricciones Técnicas de Seguridad. Se puede ver esta tabla en 3.1 aunque también se usaran los cambios introducidos en esta tabla durante el capítulo 3. El número de restricciones que aparece por día es muy variable, tal como se veía en 3.1. Esto supone un problema ya que los modelos tradicionalmente esperan una única fila de variables por cada output. Si el número de restricciones por día no fuera variable esto no sería un problema grave ya que las filas de un mismo día simplemente se concatenarían horizontalmente, sin embargo esto no se puede hacer en este caso ya que habría que adaptarlo para el máximo de restricciones por día que tenemos registrado (15) y esto provocaría que para gran parte de los días la

CAPÍTULO 8. MODELOS DE PREDICCIÓN DE RESTRICCIONES TÉCNICAS

mitad de las columnas estarían vacías. Para solucionar este problema y agrupar un número de filas variables en una sola se han probado 2 alternativas.

Semáforo de zonas

La primera forma que se ha probado ha sido crear unas variables que indiquen cuantas restricciones hay en cada zona a modo de semáforo. Para hacerlo primero se rellena el número de horas que cada zona tiene Restricciones Técnicas (usando la variable *Escenario*) para cada día. Esto nos permite obtener la tabla de 8.1

	Centro	Este	Levante	Noroeste	Norte	Sur
Fecha						
2020-01-01	18.0	24.0	28.0	24.0	18.0	24.0
2020-01-02	7.0	7.0	26.0	36.0	7.0	0.0
2020-01-03	8.0	24.0	30.0	36.0	0.0	0.0
2020-01-04	9.0	24.0	24.0	40.0	18.0	9.0
2020-01-05	18.0	24.0	29.0	39.0	18.0	24.0

Tabla 8.1: Tabla del número de horas con Restricciones por Zona

En la figura 8.3 se puede ver la distribución de estas horas en cada día por zona.

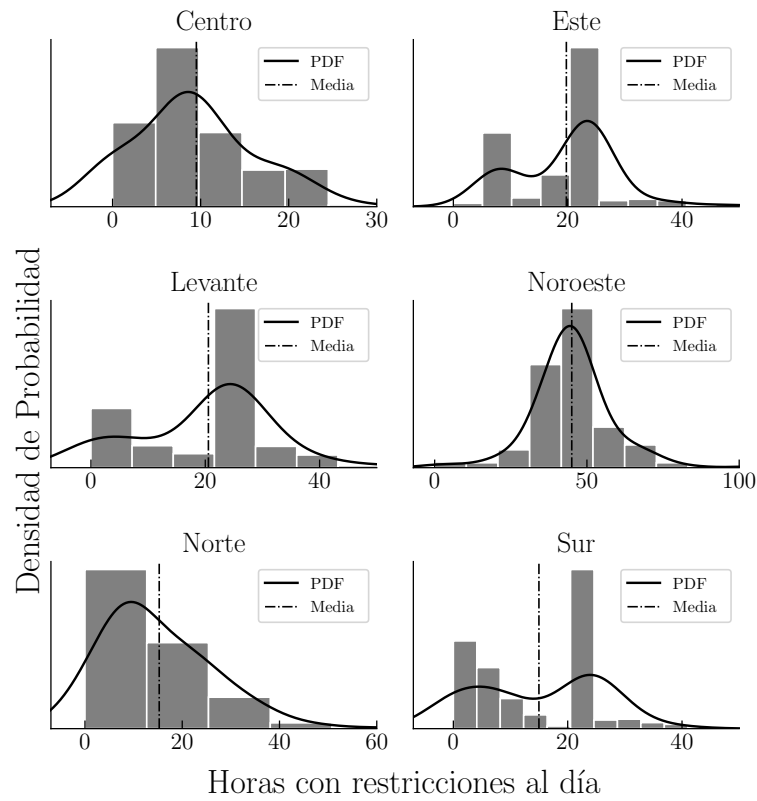


Figura 8.3: Distribución de las horas con restricción al día en diferentes zonas

Como se puede ver existe mucha variación entre la distribución de las diferentes zonas. Para solucionar esto se ha decidido sustituir el valor de las horas por un valor que represente el percentil en el que se encuentren (respecto de los valores de esa zona) siguiendo la tabla 8.2. Esta transformación es como una transformación de cuartil (que obtiene una salida uniforme) pero simplificada al usar valores discretos.

Percentil	Valor del Semáforo
0 % - 10 %	1
10 % - 20 %	2
20 % - 30 %	3
30 % - 40 %	4
40 % - 50 %	5
50 % - 60 %	6
60 % - 70 %	7
70 % - 80 %	8
80 % - 90 %	9
90 % - 100 %	10

Tabla 8.2: Valor del semáforo en función del percentil

Se puede ver el resultado final en la tabla 8.3. Los valores son muy intuitivos y fácilmente interpretables.

	Centro	Este	Levante	Noroeste	Norte	Sur
Fecha						
2020-01-01	9	5	9	1	6	6
2020-01-02	3	1	9	2	2	1
2020-01-03	4	5	9	2	1	1
2020-01-04	5	5	4	3	6	4
2020-01-05	9	5	9	3	6	6

Tabla 8.3: Tabla del semáforo de Restricciones por Zonas

Agrupar por dos variables

La otra forma consiste en agrupar dos variables, esto se realiza mediante la creación de una nueva tabla, donde las columnas son cada combinación de las dos variables escogidas y se tiene una fila por día (que es lo que necesitamos). El valor con el que se rellena es con el número de horas que estuvo esa Restricción (el número de horas se calculaba usando la variable *Escenario*). En función de las dos variables que se usen para agrupar se obtienen más o menos columnas.

Si se agrupa por *Zona* y *Tipo* se obtienen 36 columnas, ya que *Zona* puede adoptar 6 valores distintos y *Tipo* otros 6. Se puede ver un trozo de la tabla resultante en 8.4. Se ve como las columnas son valores de la variable *Zona* unidos

con valores de la variable *Tipo*. Esto resulta en muchas celdas con valor de 0. Con el objetivo de quitar columnas que no aportan casi información se mantienen únicamente las columnas cuya suma de horas totales a lo largo del año sea mayor de 10 horas. Este es un valor muy pequeño y por tanto solo se eliminan las columnas que sean combinaciones de *Zona* y *Tipo* que no se han repetido casi en los 15 meses que utilizamos. Después de hacer esto se pasa de 36 columnas a solo 19.

Fecha	Centro_Control Tension	Centro_Sobrecargas	Centro_seguridad en la red de distribucion	Este_Control Tension	Este_Sobrecargas
2020-01-01	18.0	0.0	0.0	24.0	0.0
2020-01-02	7.0	0.0	0.0	7.0	0.0
2020-01-03	8.0	0.0	0.0	24.0	0.0
2020-01-04	9.0	0.0	0.0	24.0	0.0
2020-01-05	18.0	0.0	0.0	24.0	0.0

Tabla 8.4: Restricciones Técnicas agrupadas por Zona y Tipo (limitado a 5 columnas)

También se ha probado a agrupar por *Zona* y *Localizacion*, por la gran variedad de valores de *Localizacion* (aunque tiene inconsistencias). Al agrupar por estas dos variables se obtienen 1272 columnas, que al aplicar la regla de las 20 horas comentada en el párrafo anterior resultan en únicamente 39. Se puede ver un trozo de esta tabla en 8.5.

	Centro_Zona Centro	Centro_L-400 kV Moraleja - T Galapagar	Centro_C-220 kV Arganda - Loeches	Centro_L-220 kV Aceca - Anover	Centro_L-220 kV Aceca - Pradillos
2020-01-01	18.0	0.0	0.0	0.0	0.0
2020-01-02	7.0	0.0	0.0	0.0	0.0
2020-01-03	8.0	0.0	0.0	0.0	0.0
2020-01-04	9.0	0.0	0.0	0.0	0.0
2020-01-05	18.0	0.0	0.0	0.0	0.0

Tabla 8.5: Restricciones Técnicas agrupadas por Zona y Localización (limitado a 5 columnas)

8.2.2. Desplazamiento de los valores de un día

Como ya se ha comentado, los modelos no pueden usar valores que no se puedan obtener al momento de realizar la predicción (un día antes). Los informes se publican en el día y por tanto no se cuenta con sus valores en el momento de la predicción, es por esto que hay que desplazar los valores 1 día y usar para la predicción los valores del día anterior.

8.2.3. Días laborales

Como los modelos usados no comprenden series temporales, a cada día se le añade una nueva variable que informa si se trata de un día laborable o un fin de semana, en caso de que sea día laborable esta variable adopta el valor de 1 y en caso de que sea fin de semana adopta un 0. No se tienen en cuenta fiestas ni vacaciones.

	Valor
Día laborable	1
Fin de semana	0

Tabla 8.6: Variable laboral

8.2.4. Ciclos combinados acoplados

Se ha añadido como variable de entrada la potencia acoplada de los ciclos combinados de toda España pero agrupados en 10 zonas distintas. Estos datos representan la generación programada de los ciclos en la última hora del día anterior. Los datos se pueden ver en la tabla 8.7.

Fecha	Zona 1 (MW)	Zona 2 (MW)	Zona 3 (MW)	Zona 4 (MW)	Zona 5 (MW)	Zona 6 (MW)	Zona 7 (MW)	Zona 8 (MW)	Zona 9 (MW)	Zona 10 (MW)
2020-01-02	664	0	677	0	610	0	0	401	303	0
2020-01-03	418	250	371	160	594	0	0	532	318	5
2020-01-04	360	380	383	180	176	0	381	376	311	50
2020-01-05	370	325	274	400	166	0	0	259	260	0
2020-01-06	383	370	360	685	241	0	0	360	334	320

Tabla 8.7: Potencia acoplada de los ciclos combinados dividido en 10 zonas

Las zonas tienen cada una las unidades de programación asociadas que se ven en la tabla 8.8.

8.2.5. Previsiones

También se han añadido como variables a cada día las siguientes previsiones.

- Eólica: Se han extraído y añadido la previsión de generación eólica para $D+1$ de *PREVISIÓN DIARIA $D+1$ EÓLICA* de ESIOS [46] agrupada por días. Al ser estos valores una predicción no es necesario moverlos al día anterior.

8.2. Variables de entrada

Zona 1	Zona 2	Zona 3	Zona 4	Zona 5	Zona 6	Zona 7	Zona 8	Zona 9	Zona 10
STC4	ARRU1R	ESCC1	ARCOS1	BES4	PALOS1	PTVENT2	CTN3	PGR1	ACE3
BAHIAB	ARRU2R	ESCC2	ARCOS2	BES3	PALOS2	PTVENT1	CTN4	PGR2	ACE4
AMBIETA	CTJON1	ESCC3	ARCOS3	BES5	PALOS3	TAPOWER	SAGU1	PGR3	
	CTJON3	CTGN1	ALG3	PBCN1	COL4		SAGU2	PGR4	
		CTGN2		PBCN2			SAGU3	SBO3	
		CTGN3							

Tabla 8.8: División en 10 zonas de las unidades de programación

Estos valores se han separado por comunidad autónoma usando la tabla 5.7(Excluyendo las que tienen un porcentaje ínfimo). Se puede ver un trozo de la tabla resultante en 8.9.

Fecha	Eolica	Eolica.Castilla y Leon	Eolica.Castilla La Mancha	Eolica.Galicia	Eolica.Andalucia
2020-01-02	76392	17570.16	10847.664	10694.88	9701.784
2020-01-03	84897	19526.31	12055.374	11885.58	10781.919
2020-01-04	116823	26869.29	16588.866	16355.22	14836.521
2020-01-05	57087	13130.01	8106.354	7992.18	7250.049
2020-01-06	40363	9283.49	5731.546	5650.82	5126.101

Tabla 8.9: Previsión eólica dividida por Comunidad Autónoma (limitado a 5 columnas)

En la figura 8.4 se puede ver la previsión de generación eólica frente a la generación eólica real durante la mitad de 2020. El error absoluto medio (MAE) es de tan solo 9070 MWh y el coeficiente de determinación (R^2) de 0.97.

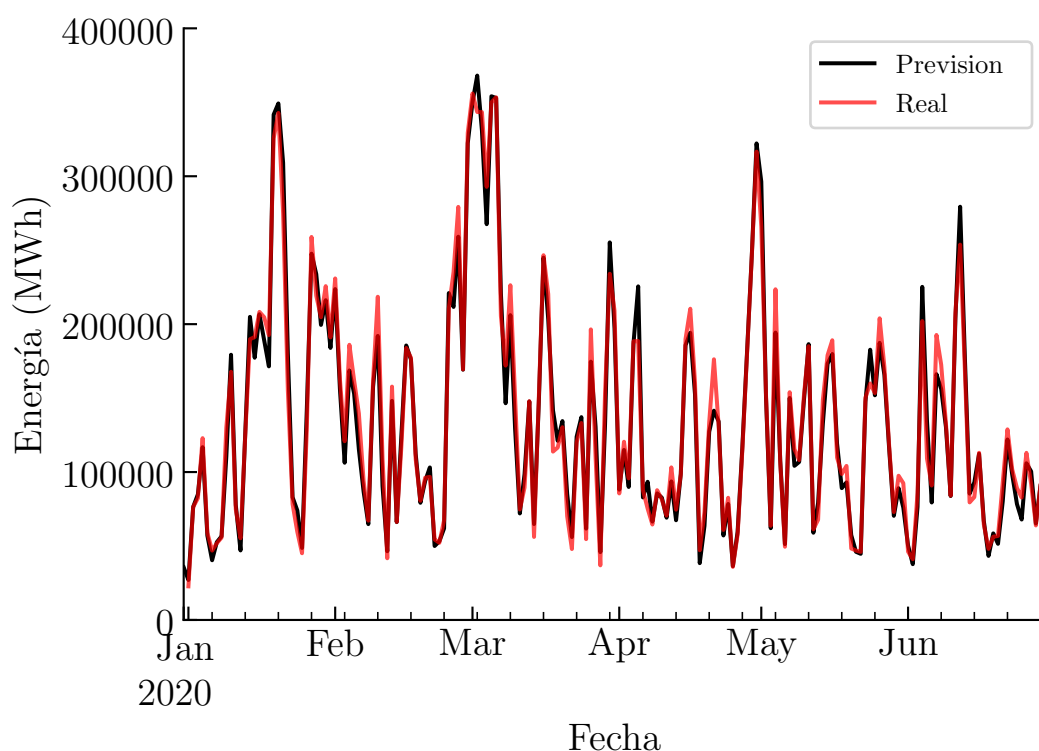


Figura 8.4: Previsión de generación eólica frente a generación real

- Fotovoltaica: Igual que con la eólica se ha extraído la previsión de generación fotovoltaica para D+1 de *PREVISIÓN DIARIA D+1 FOTOVOLTAICA* de ESIOS [47] agrupada por días. De nuevo los valores se agrupan por comunidad autónoma usando la tabla 5.7. Se puede ver un trozo de la tabla resultante en 8.10.

Fecha	Solar	Solar.Castilla y Leon	Solar.Castilla La Mancha	Solar.Galicia	Solar.Andalucia
2020-01-02	23290.0	2096.100	4564.8400	46.5800	4658.00
2020-01-03	18281.3	1645.317	3583.1348	36.5626	3656.26
2020-01-04	24351.4	2191.626	4772.8744	48.7028	4870.28
2020-01-05	26347.4	2371.266	5164.0904	52.6948	5269.48
2020-01-06	26885.7	2419.713	5269.5972	53.7714	5377.14

Tabla 8.10: Previsión solar fotovoltaica dividida por Comunidad Autónoma (limitado a 5 columnas)

En la figura 8.5 se puede ver la previsión de generación eólica frente a la generación eólica real (extraída de ESIOS en el capítulo 5) durante la mitad de 2020. El MAE es de tan solo 2720 MWh y el R^2 de 0.94. Estos resultados son ligeramente peores que los de la eólica pero aun así son muy buenos.

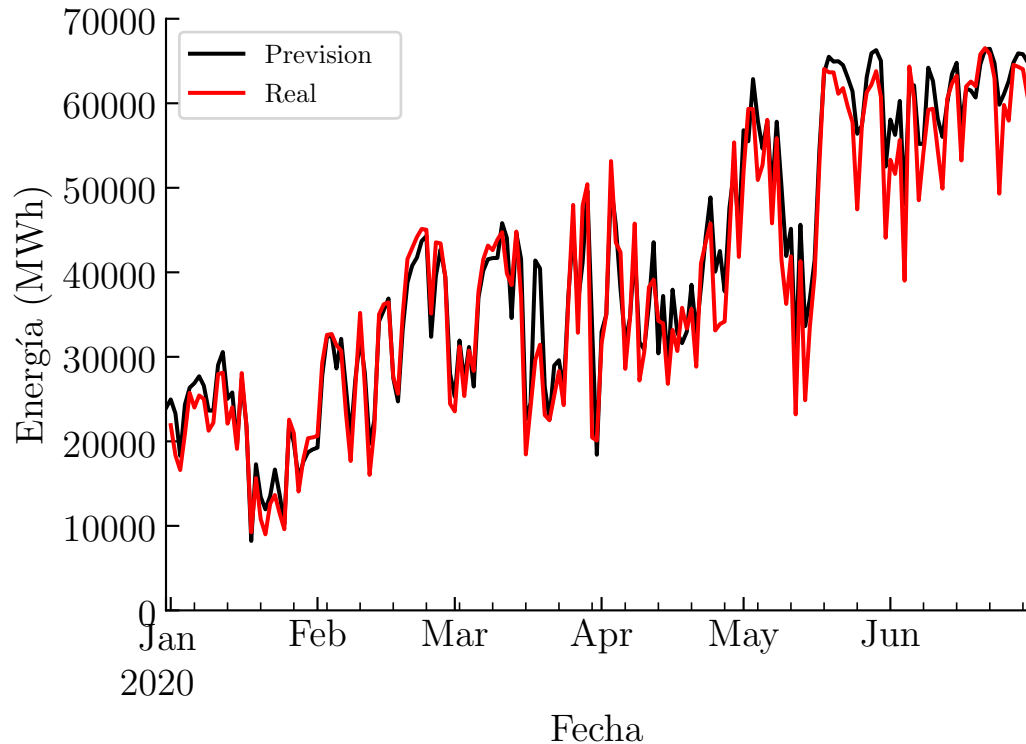


Figura 8.5: Previsión de generación fotovoltaica frente a generación real

- Nuclear: En el caso de la nuclear no se obtiene la previsión sino la generación programada para D+1. La información se extrae de *GENERACIÓN PROGRAMADA PBF NUCLEAR* de ESIOS [41]. En la figura 8.6 se puede ver la generación nuclear programada frente a la generación nuclear real (extraída de ESIOS en el capítulo 5) durante la mitad de 2020. Se aprecia como en determinados días no coincide exactamente pese a estar programada. Aun así tiene tan solo un MAE de 1805 MWh y un R^2 de 0.98.

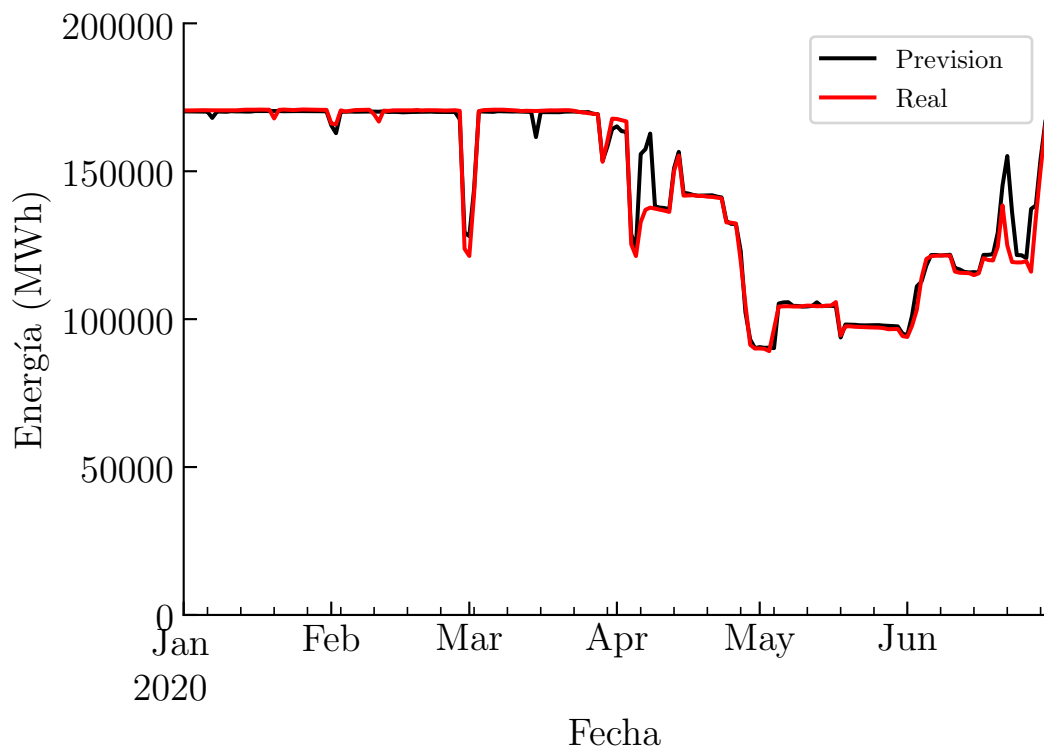


Figura 8.6: Generación programada nuclear frente a generación real

- Demanda: La previsión de demanda diaria D+1 se ha obtenido de *PREVISIÓN DIARIA D+1 DEMANDA* de ESIOS [48] agregada por días. En la figura 8.7 se puede ver la previsión de demanda frente a la demanda real (obtenida de *DEMANDA DEL SISTEMA A EFECTOS DEL CÁLCULO DE PERFILES* de ESIOS [49]). La precisión de la previsión es altísima, con un MAE de tan solo 9455 y un R^2 de 0.96.

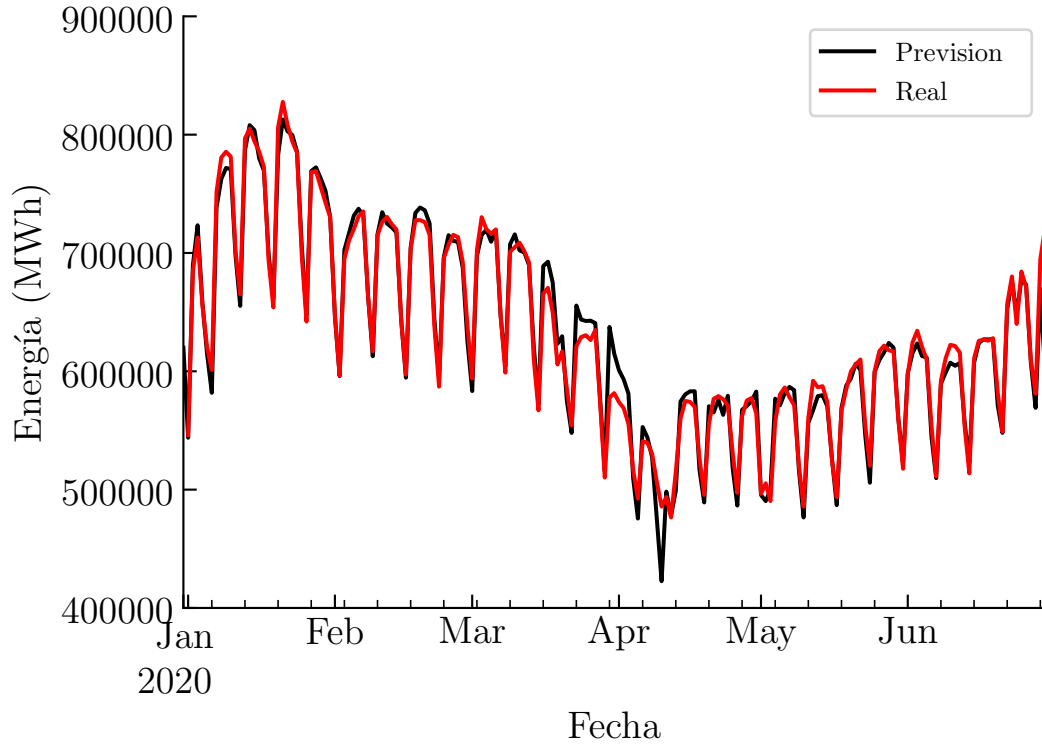


Figura 8.7: Previsión de demanda frente a demanda real

- Hueco Térmico: El hueco térmico es la parte de la demanda que no queda cubierta por tecnología nuclear o renovable y que ha de ser cubierta con la generación de electricidad a partir de térmicas convencionales y ciclos combinados. Por lo tanto la previsión del hueco térmico se puede calcular de la siguiente forma. *Prev_Solar* incluye tanto fotovoltaica como solar térmica pero lamentablemente ESIOS no ofrece una previsión diaria D+1 de la solar térmica así que se va a aproximar *Prev_Solar* a *Prev_Solar_fotovoltaica*, dado que la generación solar térmica es mucho más pequeña que la generación fotovoltaica no se introduce mucho error. Una alternativa sería utilizar la media de generación solar térmica del año o el valor del día anterior, pero la gran estacionalidad y la altísima variación de un día para otro (se pueden apreciar ambas en la figura 5.23) hacen que estas alternativas introduzcan un error más impredecible.

$$Prev_Hueco_Termi = Prev_Demand - Prog_nucle - Prev_Eoli - Prev_Solar - Prev_Hidrau$$

Ecuación 8.1: Cálculo del Hueco Térmico

Los R^2 de todas las previsiones de ESIOs son muy altas, lo que hace que nuestros modelos no acumulen casi error al utilizar estas previsiones y cuenten con información muy precisa del estado de la red.

8.2.6. Valor anterior

Como ya se ha comentado, los modelos usados no comprenden series temporales y de nuevo con el objetivo de transmitir parte de la componente temporal, al modelo se le introduce el valor de la energía por restricciones del día anterior.

8.3. Modelos

A continuación se comentan de forma breve los distintos modelos de *Machine Learning* (ML) que se han probado.

8.3.1. Regresión Lineal

Los modelos de regresión lineal son algoritmos de aprendizaje supervisado que modelan una variable dependiente como la suma de transformaciones lineales de las variables independientes más una constante. Estos modelos son muy rápidos de entrenar y ofrecen unos resultados muy fáciles de interpretar.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

Ecuación 8.2: Ecuación del modelo de regresión lineal

Este algoritmo cuenta solo con un parámetro:

- Constante: Se decide si se va a usar constante o se quiere que la regresión corte en 0.

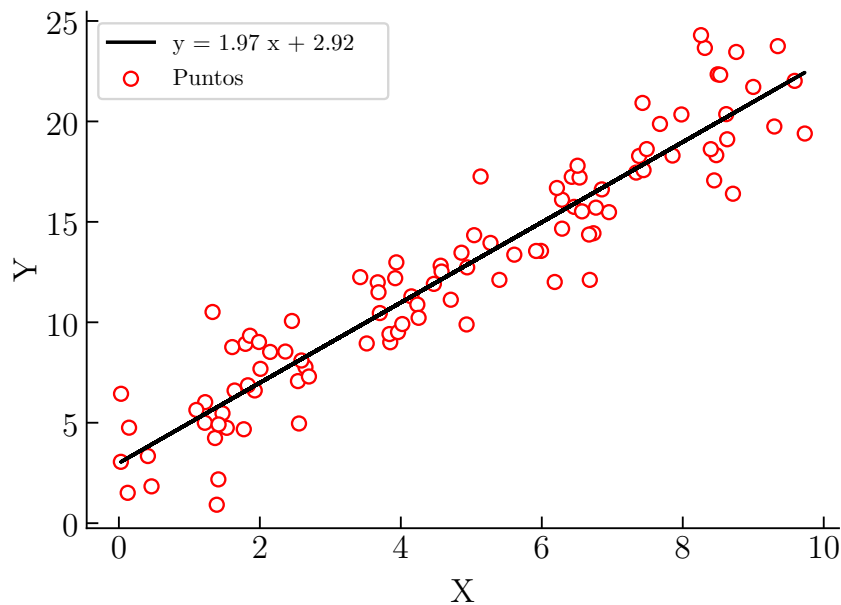


Figura 8.8: Ejemplo de regresión lineal

8.3.2. K-Nearest-Neighbors

Este algoritmo de aprendizaje determina la salida, en función de las k muestras etiquetadas que estén más cerca de la muestra a etiquetar. Si por ejemplo k es igual a 3, se cogerán las 3 muestras más cercanas y la salida será la media de ellas.

Normalmente se usa la distancia euclídea para encontrar muestras cercanas, pero se pueden usar otros criterios.

Este algoritmo cuenta con un hiperparámetro clave:

- Número de vecinos: Este parámetro indica cuantas muestras cercanas se miran para clasificar la nueva muestra.

8.3.3. Random Forrest

Random Forrest es un algoritmo de aprendizaje supervisado, utiliza árboles de decisión. Los árboles de decisión tienen diferentes nodos conectados y en cada nodo se toma una decisión en función de los valores de la muestra en diferentes dimensiones.

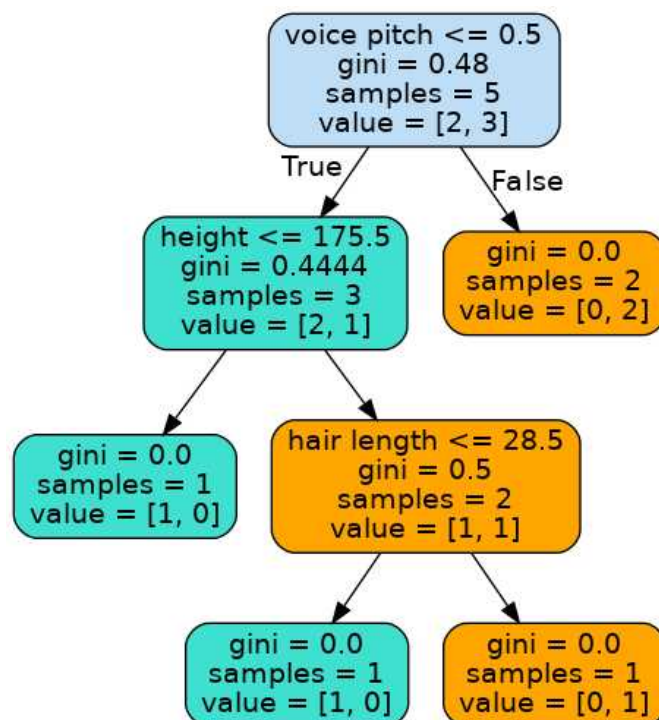


Figura 8.9: Árbol de decisión [50]

Un Random Forrest está compuesto por más de un árbol de decisión, cada árbol tiene un cierto componente aleatorio para que los árboles sean distintos entre ellos.

Tienen algunos hiperparámetros para ajustar el modelo, los principales son:

- Número de árboles: Determina el número de árboles de decisión de los que el Random Forrest está compuesto.
- Número de hojas máximo: Determina el número de nodos máximo que puede tener cada árbol.

8.3.4. Perceptrón multicapa

El perceptrón multicapa (MLP) es el único modelo de *Deep Learning* usado en este trabajo. Este modelo conecta las entradas con la salida pasando por una o más capas intermedias de perceptrones simples.

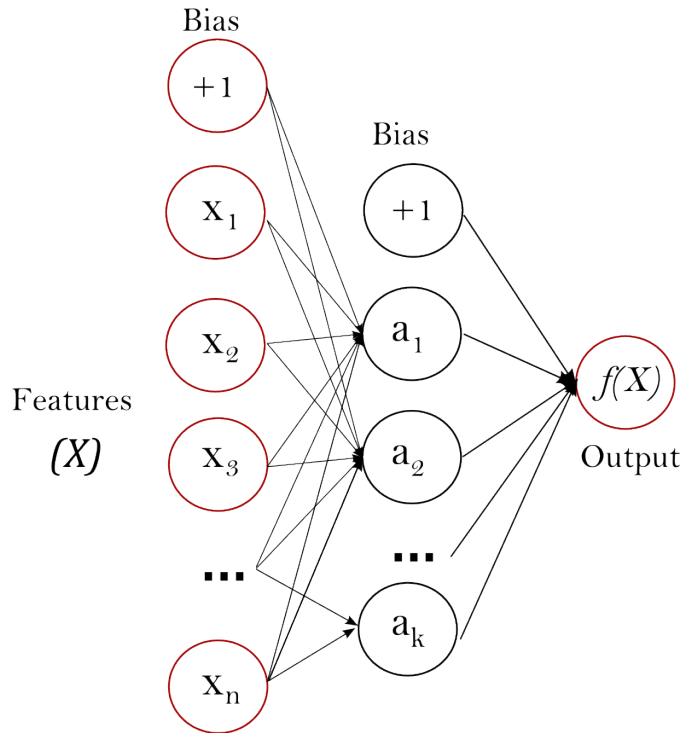


Figura 8.10: Perceptrón multicapa
[51]

Los MLP cuentan con muchos hiperparámetros, los principales son:

- Número de capas ocultas: El número de capas de perceptrones simples que hay entre la entrada y la salida.
- Función de activación: Determina la función de dentro de cada perceptrón. Para modelos lineales se usa la función identidad pero cuando se quiere que el modelo se adapte a relaciones no lineales se usan funciones como la función umbral, la función sigmoide o la función Relu.
- Alpha: Este parámetro impone una penalización a la complejidad del MLP. Permite combatir el *overfitting* al conjunto de entrenamiento.

8.4. Preprocesamiento

Antes de entrenar algunos de los modelos se han introducido algunas transformaciones al conjunto de datos con el objetivo de facilitar el entrenamiento y mejorar los resultados.

8.4.1. Normalización/Estandarización

Dado que las variables de nuestro conjunto de datos tienen unidades distintas y ordenes de magnitud muy diferentes (Algunas valores de 0,1 como la variable *Laboral* y otras valores de miles de MWh como la variable de salida) se ha probado a normalizar o estandarizar los valores.

- Normalización: La normalización fuerza todos los valores de cada variable en un rango. Este rango puede variar en función del tipo de normalización, si se aplica una normalización de tipo *MinMaxScaler* se fuerzan todos los valores entre 0 y 1 (se puede cambiar si se quisiera). Esta normalización es muy sensible a outliers, al utilizar estos como máximo y mínimo al realizar la transformación. Nuestros datos no tienen mucha presencia de outliers y por tanto esta transformación no debería de presentar problemas.
- Estandarización: La estandarización más típica transforma los datos para que no tengan media y su varianza sea 1 (y por tanto también la desviación típica). De nuevo es sensible a outliers aunque menos que la normalización *MinMaxScaler*. Esta transformación asume una distribución normal en las variables, lo cual no es cierto para gran parte de las variables (todas las que hablan de horas con restricciones). Aun así se ha decidido probar en algunos modelos que se benefician mucho de la estandarización de las variables.

8.4.2. Valores incrementales

Gracias a trabajar con una serie temporal y con el objetivo de que la información que no esta en nuestras variables pero varíe poco de un día a otro pueda ser asimilada por el modelo, todas las variables se han transformado en incrementales. De esta forma el modelo ya no predice el valor de energía por restricciones sino la variación respecto del día anterior.

En cierta forma es algo parecido a lo que se buscaba al introducir el valor del día anterior pero a una escala mayor.

8.5. División del conjunto de entrenamiento y de test

Al tratarse de una serie temporal, no se puede coger el conjunto de datos y de forma aleatoria separarlos en un 80 %/20 %. Si se hiciera esto, se usarían aveces días del futuro para predecir algunos días del conjunto de test. Por tanto si se va a hacer una división 80 %/20 % esta tiene que respetar el orden temporal.

Aun así, con un escenario eléctrico tan cambiante y con el objetivo de obtener los mejores resultados se ha decidido aplicar *Online Learning*. *Online Learning* va re-entrenando el modelo según pasa el tiempo y se puede usar cuando el tamaño del conjunto de datos aumenta secuencialmente con el tiempo, como en nuestro caso que cada día añadimos una fila a nuestro conjunto. De esta forma nuestros modelos para cada día predicen el valor de energía de restricciones de un día entrenándose con todos los días anteriores y para la predicción del día siguiente el modelo añade los datos del día anterior al conjunto de entrenamiento, se re-entrena y se hace la predicción. Por lo tanto no hay un conjunto de entrenamiento y de test fijo, sino que estos van cambiando según pasan los días, el conjunto de entrenamiento se hace cada día una fila más grande y el conjunto de test es únicamente una fila (la de día de la predicción) que cambia cada día. Se va a usar *Online Learning* para predecir un 35 % del conjunto de datos.

8.6. Resultados

Las métricas que se han usado para evaluar los resultados han sido MAE, RMSE y R^2 . Como se verá más adelante la principal métrica utilizada ha sido R^2 .

Se han probado diferentes marcas estándar de comparación, como el *naive approach* [7], que consiste en coger un valor similar, es decir para predecir un viernes coger el valor del día anterior. Los resultados de esta predicción para el conjunto de test se pueden ver en la tabla 8.11 y si se pone la predicción frente al valor real se obtiene el gráfico de dispersión de la figura 8.11. Si la predicción fuera perfecta todos los puntos se encontrarían en la línea discontinua. Se puede ver que el R^2 es muy negativo, esto puede parecer raro pero la métrica R^2 no es en realidad una variable R^2 al cuadrado. Se obtienen valores negativos cuando la predicción es peor que aproximar por la media del conjunto de test, como en este caso.

	MAE	RMSE	R^2
Test	4949.7540	6855.9694	-0.3121

Tabla 8.11: Resultados del *naive approach*

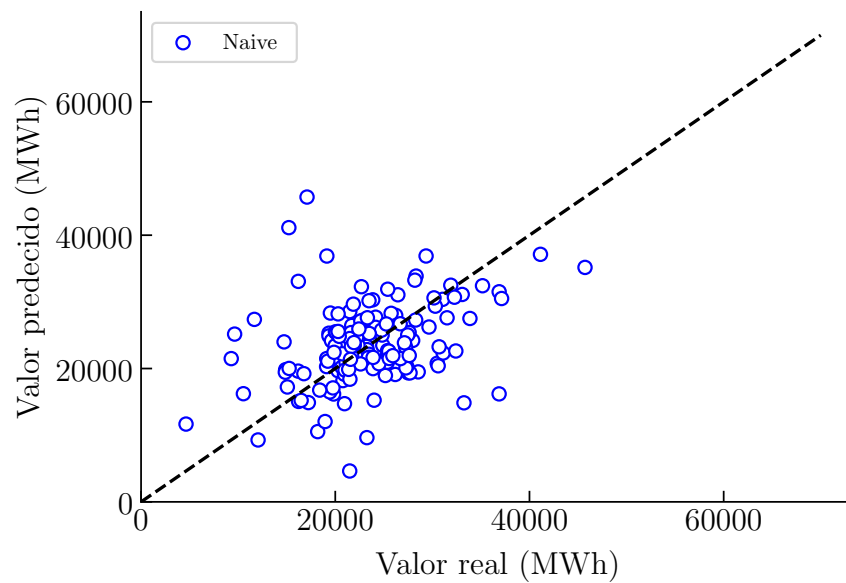


Figura 8.11: Resultados del *naive approach*

Otro test que se ha probado ha sido simplemente aproximar por la media, los resultados de esta predicción para el conjunto de test se pueden ver en la tabla 8.12 y si se pone la predicción frente al valor real se obtiene el gráfico de dispersión de la figura 8.12. En este caso el valor sigue siendo negativo pero es casi igual que 0, ya que la media del conjunto de train es muy parecida a la media del conjunto de test. Se han probado otras marcas estándar de comparación pero ninguna ha podido superar a esta y por tanto esta es la que se usará. A partir de ahora denominaremos a este estándar como *mean pred.*

	MAE	RMSE	R^2
Test	4565.2592	6038.6189	-0.0179

Tabla 8.12: Resultados de aproximar por la media del conjunto de train

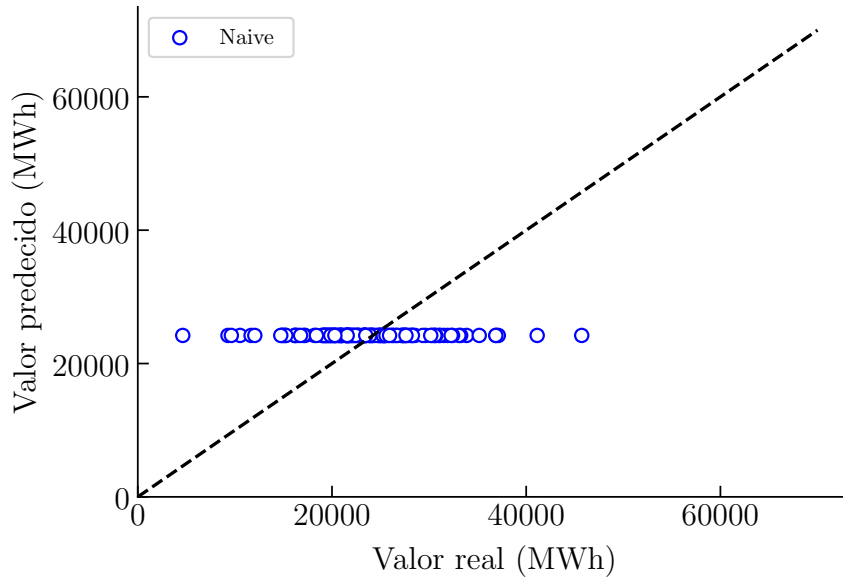


Figura 8.12: Resultados de aproximar por la media del conjunto de train

Para obtener una medida de cuanto *overfitting* hay de los modelos al conjunto de datos, en cada modelo se calculan las métricas para el conjunto de entrenamiento considerando este como el 65 % de los datos. Como ya se ha comentado antes esto en realidad no es así ya que se está usando *online learning* y el conjunto de entrenamiento se hace más grande según se avanza pero aun así puede ofrecer una buena medida de cuanto *overfitting* hace el modelo a los datos.

Existen más de 100 combinaciones entre formas de agrupar los informes, tipos de transformaciones de datos y modelos, sin contar ni siquiera modelos con distintos parámetros y pequeñas variaciones de las transformaciones. No es posible mostrar y comentar todas las que se han probado así que solo se presentarán una pequeña proporción, aquellas con los mejores resultados de cada modelo y las que aportan alguna conclusión interesante. En algunos de los casos también se comentará que una alternativa es mejor que otra sin detallar todos los resultados que lo sustentan, como en los casos de optimización de parámetros de los modelos.

8.6.1. Regresión Lineal

Agrupación de zona y tipo

Al entrenar el modelo de regresión lineal agrupando las variables zona y tipo se obtienen los resultados de la tabla 8.13 y al ver la predicción frente al valor real se obtiene el gráfico de dispersión de la figura 8.13. Pese a que el MAE y el RMSE no

es mucho mayor en el Test el R^2 es muy diferente, en el conjunto de entrenamiento (importante destacar de nuevo que no es realmente el conjunto de entrenamiento ya que este cambia, es solamente el resultado de entrenar el modelo con el 65 % de los datos y ver el resultado) es de 0.78 y en el de test es -0.73. Parte del problema se ve a simple vista, el conjunto de test tiene una varianza mucho menor que la del conjunto de entrenamiento y por tanto al modelo le cuesta adaptarse a circunstancias normales (final del año 2020 y inicio del 2021) cuando ha sido entrenado en gran medida con datos de circunstancias excepcionales (año 2020 afectado por la pandemia). Por tanto comparar el MAE y RMSE entre el conjunto de entrenamiento y de test no tiene sentido y no se hará de aquí en adelante, solo se comparan usando el R^2 , aunque el MAE y RMSE del test se seguirán usando para comparar modelos.

Este modelo no supera el *mean pred.*

	MAE	RMSE	R^2
Test	5984.098082	7864.235103	-0.726438
Train	4900.104529	6299.887180	0.785028

Tabla 8.13: Resultados de modelo de regresión lineal usando variables zona y tipo agrupadas

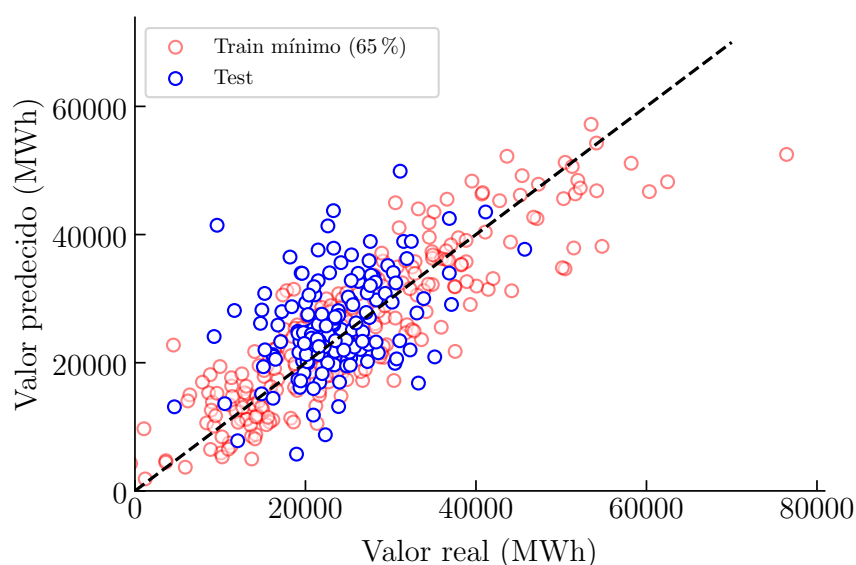


Figura 8.13: Resultados de modelo de regresión lineal usando variables zona y tipo agrupadas

Agrupación de zona y Localización

Al entrenar el modelo de regresión lineal agrupando las variables zona y localización se obtienen los resultados de la tabla 8.14 y al ver la predicción frente al valor real se obtiene el gráfico de dispersión de la figura 8.13. Los resultados son mucho mejores que al agrupar por zona y tipo, pese a que la variable localización sea una variable con muchas inconsistencias. Con esta forma de agrupación el R^2 pasa de -1.04 a -0.37 y el MAE se reduce en un 9 %. Aun así este modelo no supera el *mean pred.*

	MAE	RMSE	R^2
Test	5467.106220	7011.122711	-0.372186
Train	4283.476049	5758.979504	0.820358

Tabla 8.14: Resultados de modelo de regresión lineal usando variables zona y localización agrupadas

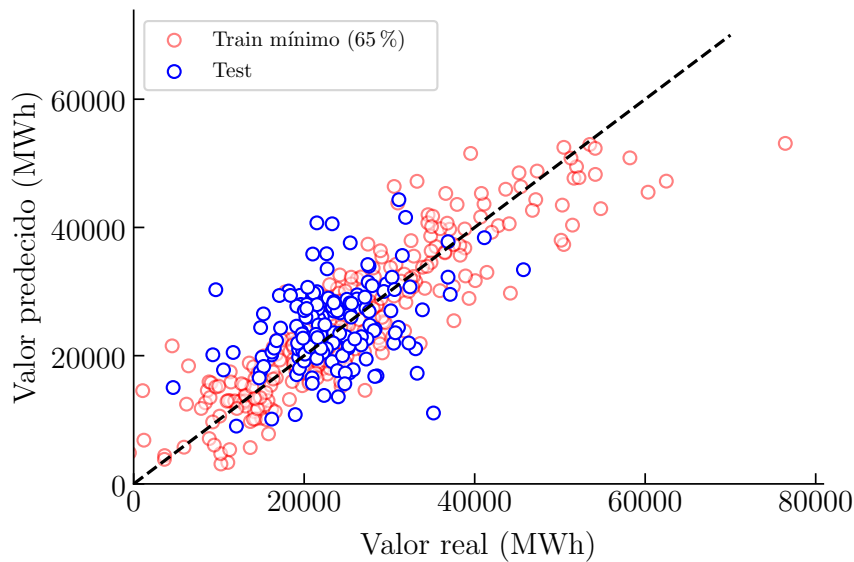


Figura 8.14: Resultados de modelo de regresión lineal usando variables zona y localización agrupadas

Agrupación de semáforo y estandarización de variables

Al entrenar el modelo de regresión lineal usando el semáforo se obtienen los resultados de la tabla 8.15 y al ver la predicción frente al valor real se obtiene el gráfico de dispersión de la figura 8.15. Los resultados son mucho mejores que en los anteriores, en parte gracias a la estandarización. Con esta forma de agrupación el R^2 pasa de -0.37 a -0.087 y el MAE se reduce en un 11 %. Este modelo sigue sin superar el *mean pred.*

	MAE	RMSE	R^2
Test	4854.854129	6240.097521	-0.086978
Train	4888.901599	6449.250054	0.775518

Tabla 8.15: Resultados de modelo de regresión lineal usando el semáforo de restricciones

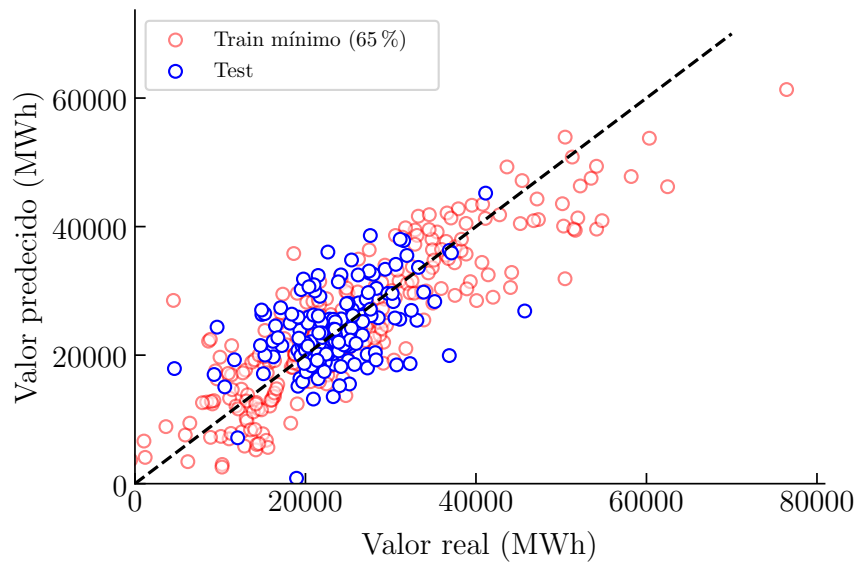


Figura 8.15: Resultados de modelo de regresión lineal usando el semáforo de restricciones

Prueba con valores incrementales y valores estandarizados

Por último se ha decidido probar el mejor modelo de los anteriores, pero con las variables incrementales obteniéndose los resultados de la tabla 8.16 y al ver la predicción frente al valor real se obtiene el gráfico de dispersión de la figura

8.16. Los resultados son ligeramente mejores que en el modelo del semáforo de restricciones, con un R^2 que ha subido en 0.04 y un MAE que se ha reducido un 3.6 %.

	MAE	RMSE	R^2
Test	4678.405070	6109.391656	-0.041919
Train	4240.370048	5741.643662	0.822640

Tabla 8.16: Resultados de modelo de regresión lineal usando el semáforo de restricciones y variables incrementales

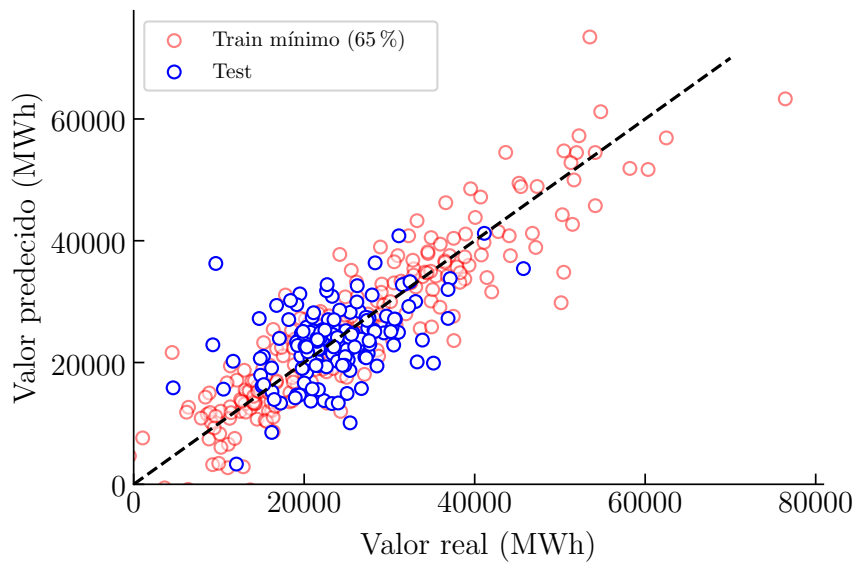


Figura 8.16: Resultados de modelo de regresión lineal usando el semáforo de restricciones y variables incrementales

8.6.2. K-Nearest-Neighbors

Agrupación de zona y tipo

Al entrenar el modelo KNN con 15 vecinos (se ha visto que los mejores resultados se obtienen con 15 vecinos) agrupando las variables zona y tipo se obtienen los resultados de la tabla 8.17 y al ver la predicción frente al valor real se obtiene el gráfico de dispersión de la figura 8.17. El R^2 es negativo pero mayor que -1, aun así no supera el *mean pred.* Visualmente parece que el modelo tiene tendencia a predecir un valor menor que el que realmente es.

	MAE	RMSE	R^2
Test	6085.546577	7415.268304	-0.534941
Train	6482.404404	8328.263291	0.624314

Tabla 8.17: Resultados de modelo KNN usando variables zona y tipo agrupadas

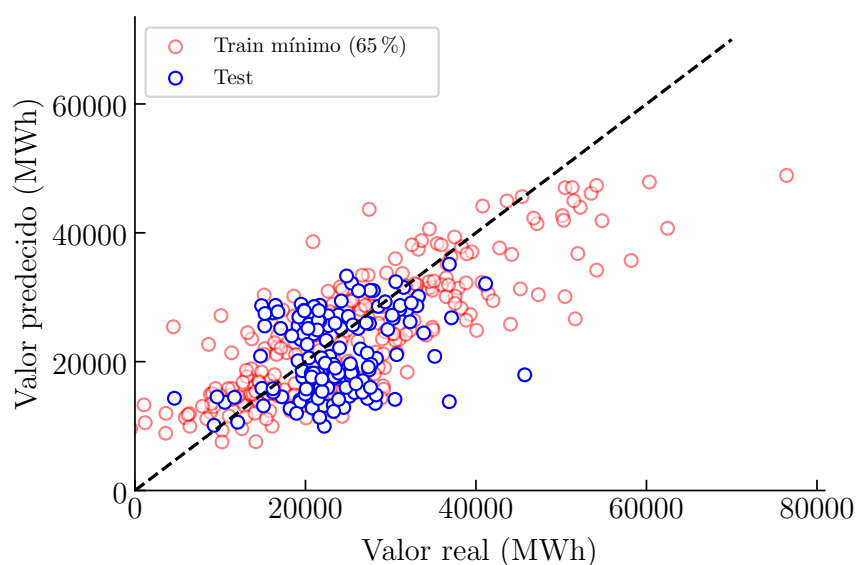


Figura 8.17: Resultados de modelo KNN usando variables zona y tipo agrupadas

Agrupación de zona y Localización con normalización

Al entrenar el modelo KNN con 20 vecinos y aplicar una normalización a las variables de entrada agrupando las variables zona y localización se obtienen los resultados de la tabla 8.18 y al ver la predicción frente al valor real se obtiene el gráfico de dispersión de la figura 8.18. KNN es muy sensible a la magnitud de sus variables y por esto este modelo es mucho mejor que el anterior. Se puede ver como el R^2 es casi 0, mejorando el *mean pred*, con un MAE mayor y un RMSE peor (mas error medio que el *mean pred* pero menos desviaciones grandes). Si se observa el gráfico de dispersión se puede ver como lo que esta haciendo este modelo es conseguir aproximar a la media, es por eso que su R^2 es casi igual a 0 (se obtiene un R^2 de 0 cuando se aproxima una distribución por su media).

	MAE	RMSE	R^2
Test	4726.304189	5951.616504	0.011201
Train	4724.737546	6286.067830	0.786735

Tabla 8.18: Resultados de modelo KNN usando variables zona y localización agrupadas

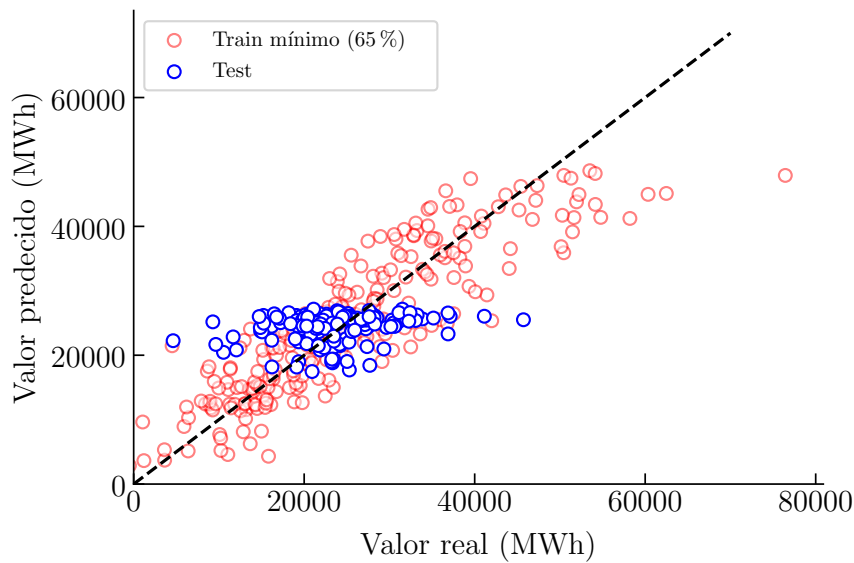


Figura 8.18: Resultados de modelo KNN usando variables zona y localización agrupadas

Agrupación de semáforo con normalización

Al entrenar el modelo KNN con 7 vecinos y aplicar una normalización a las variables de entrada usando el semáforo de restricciones se obtienen los resultados de la tabla 8.19 y al ver la predicción frente al valor real se obtiene el gráfico de dispersión de la figura 8.19. El resultado en este caso es mucho peor que al agrupar por zona y localización, con un R^2 que en este caso ya no supera el *mean pred.* En este caso parece que intenta aproximarse a la media pero falla, prediciendo siempre valores más bajos de lo real.

	MAE	RMSE	R^2
Test	6103.050579	7633.139448	-0.626464
Train	4124.596180	5535.361017	0.834631

Tabla 8.19: Resultados de modelo KNN usando el semáforo de restricciones

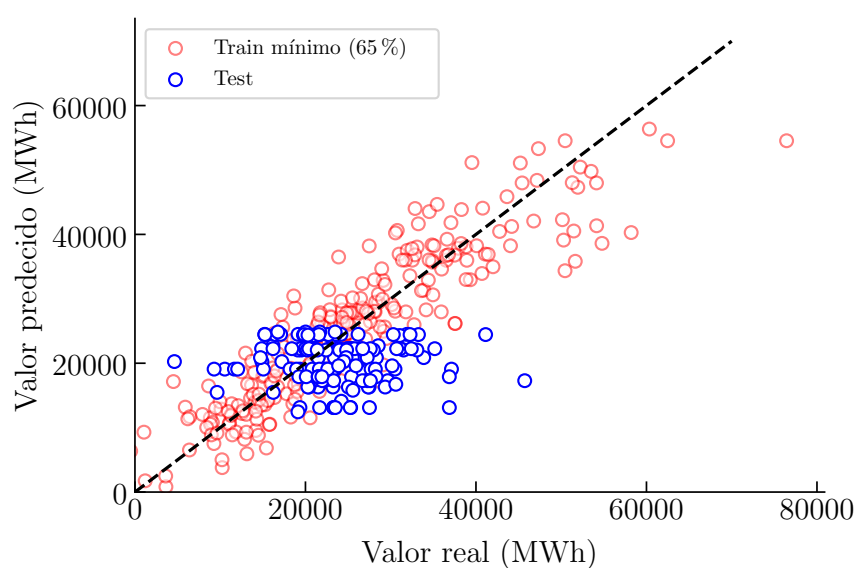


Figura 8.19: Resultados de modelo KNN usando el semáforo de restricciones

Prueba con valores incrementales

Por último se ha entrenado el mejor modelo (Agrupación de zona y Localización con normalización) con variables incrementales obteniéndose los resultados de la tabla 8.19 y al ver la predicción frente al valor real se obtiene el gráfico de dispersión de la figura 8.19. El resultado es peor ya que en este caso no es capaz de aproximar a la media como hacía.

	MAE	RMSE	R^2
Test	4818.067207	6409.092278	-0.146650
Train	4583.980159	6137.733742	0.797326

Tabla 8.20: Resultados de modelo KNN usando variables incrementales

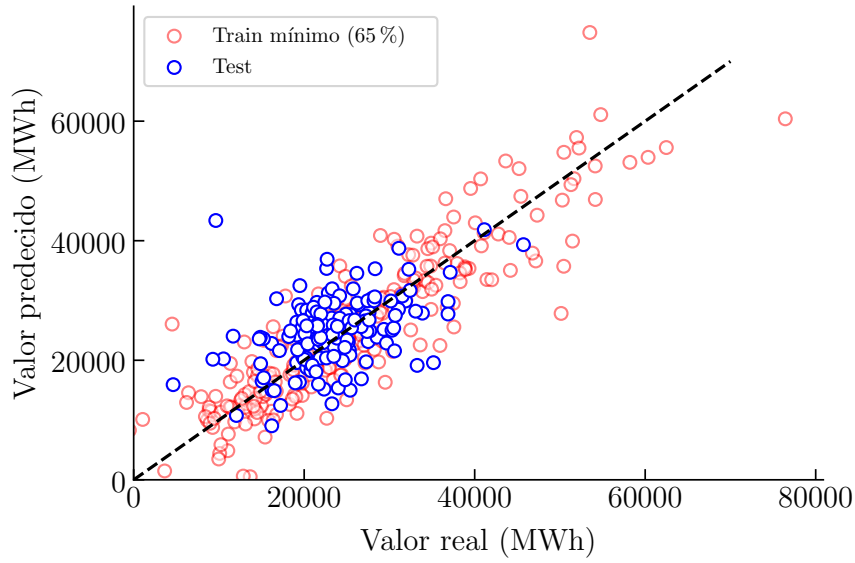


Figura 8.20: Resultados de modelo KNN usando variables incrementales

8.6.3. Random Forest Regressor

Agrupación de zona y tipo

Al entrenar el modelo RFR (con 300 árboles y 5 nodos máximo cada uno) agrupando las variables zona y tipo se obtienen los resultados de la tabla 8.21 y al ver la predicción frente al valor real se obtiene el gráfico de dispersión de la figura 8.21. Este modelo hace cierto *overfitting* al conjunto de train (o quizás es solo la gran diferencia entre los datos de train y de test), con un R^2 de 0.91. El R^2 en el test es positivo pero casi igual a 0, batiendo por tanto ligeramente al *mean pred.*

	MAE	RMSE	R^2
Test	4580.071682	5906.529918	0.026126
Train	2978.501334	3974.582194	0.914434

Tabla 8.21: Resultados de modelo RFR usando variables zona y localización agrupadas

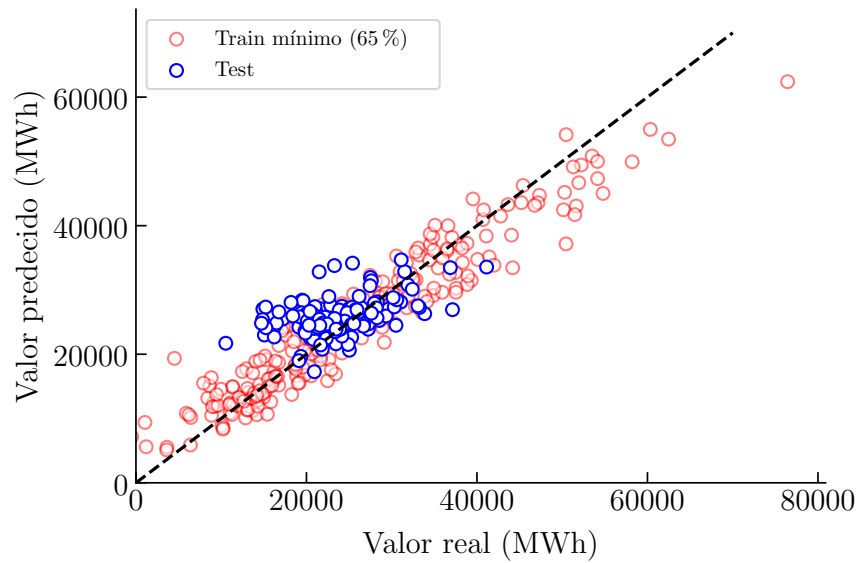


Figura 8.21: Resultados de modelo RFR usando variables zona y tipo agrupadas

Agrupación de zona y Localización

Al entrenar el modelo RFR (con 100 arboles y 8 nodos máximo cada uno) agrupando las variables zona y localización se obtienen los resultados de la tabla 8.22 y al ver la predicción frente al valor real se obtiene el gráfico de dispersión de la figura 8.22. Este modelo hace cierto *overfitting* al conjunto de train (o quizás es solo la gran diferencia entre los datos de train y de test), con un R^2 de 0.86. El R^2 en el test es muy positivo, con un valor de 0.2 batiendo por mucho al *mean pred.*

	MAE	RMSE	R^2
Test	4118.659761	5342.133502	0.203350
Train	1921.985704	2518.429306	0.965769

Tabla 8.22: Resultados de modelo RFR usando variables zona y localización agrupadas

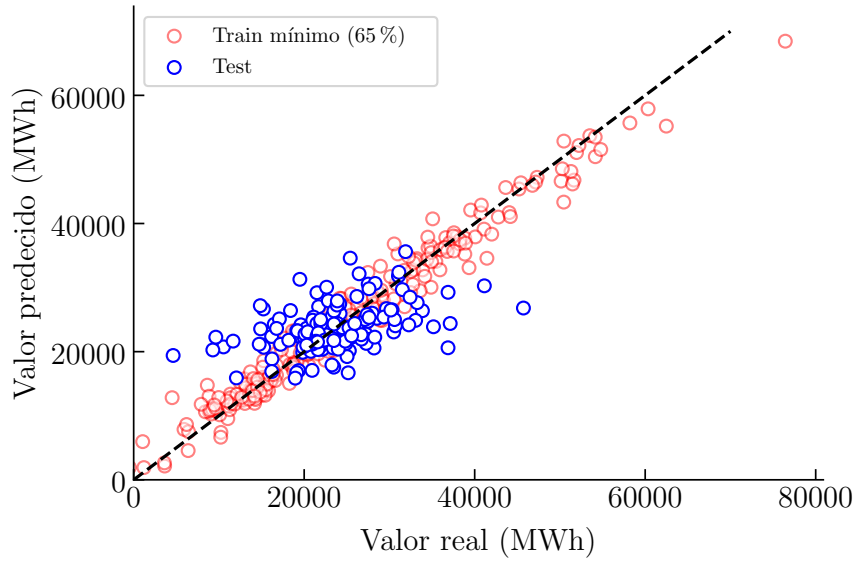


Figura 8.22: Resultados de modelo RFR usando variables zona y localización agrupadas

Agrupación de semáforo

Al entrenar el modelo RFR (con 100 árboles y 5 nodos máximo cada uno) usando el semáforo de restricciones se obtienen los resultados de la tabla 8.23 y al ver la predicción frente al valor real se obtiene el gráfico de dispersión de la figura 8.23. El R^2 en el test es casi 0, aun así batiendo aun así al *mean pred* pero peor que RFR agrupando por zona y localización.

	MAE	RMSE	R^2
Test	4523.196737	5865.534337	0.039598
Train	1886.745106	2612.038021	0.963045

Tabla 8.23: Resultados de modelo RFR usando el semáforo de restricciones

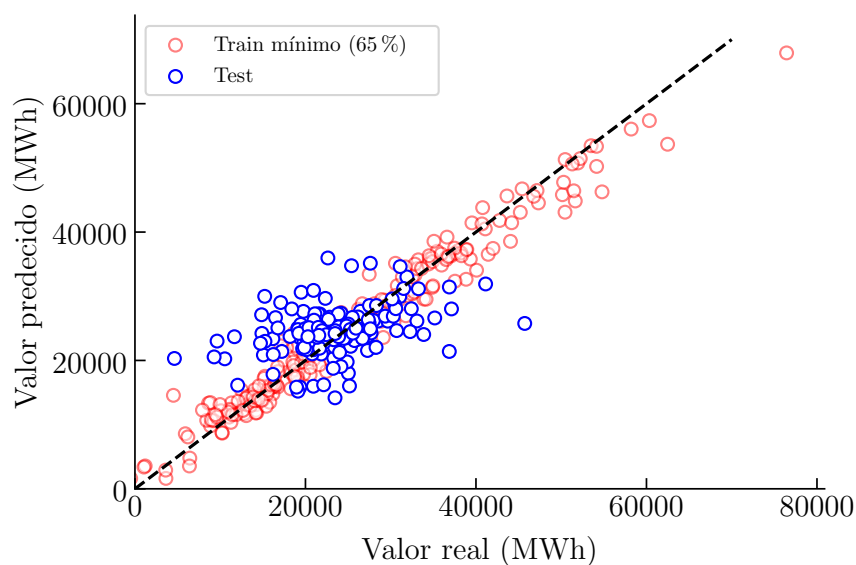


Figura 8.23: Resultados de modelo RFR usando el semáforo de restricciones

Prueba con valores incrementales

Por ultimo se ha entrenado el mejor de los modelos anteriores (Agrupación por zona y localización) con variables incrementales obteniendo los resultados de la tabla 8.24 y al ver la predicción frente al valor real se obtiene el gráfico de dispersión de la figura 8.24. El R^2 en el test es bastante peor, con un valor de casi 0, aunque el MAE no empeora tanto, aumentandolo un 6.1 %.

	MAE	RMSE	R^2
Test	4373.238470	5905.768355	0.026377
Train	2685.086549	3462.140126	0.935308

Tabla 8.24: Resultados de modelo RFR usando variables incrementales

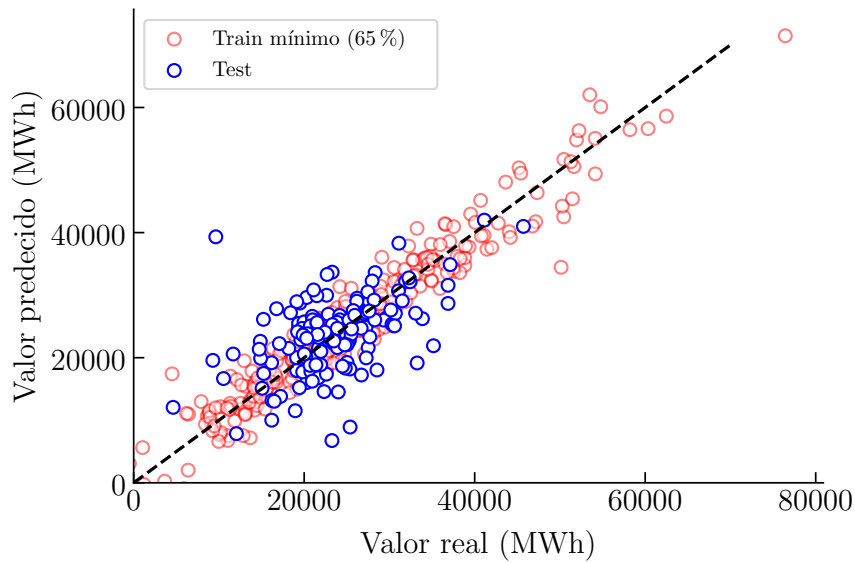


Figura 8.24: Resultados de modelo RFR usando variables incrementales

8.6.4. Perceptrón multicapa

El perceptrón multicapa es muy sensible a la escala de las variables de entrada, por tanto todos los modelos entrenados de MLP tendrán estandarización, se ha probado también con normalización pero la estandarización funciona mucho mejor en este modelo.

Agrupación de zona y tipo con estandarización

Al entrenar el modelo MLP (con 2 capas de 100 perceptrones cada una y activación de ambas capas relu) y entradas estandarizadas, agrupando las variables zona y tipo se obtienen los resultados de la tabla 8.25 y al ver la predicción frente al valor real se obtiene el gráfico de dispersión de la figura 8.25. Este modelo hace cierto *overfitting* al conjunto de train, con un R^2 de 0.91. El R^2 en el test es negativo, y peor que el *mean pred.*

	MAE	RMSE	R^2
Test	5002.404091	6460.603245	-0.165156
Train	2911.749650	3934.355507	0.916457

Tabla 8.25: Resultados de modelo MLP usando variables zona y tipo agrupadas

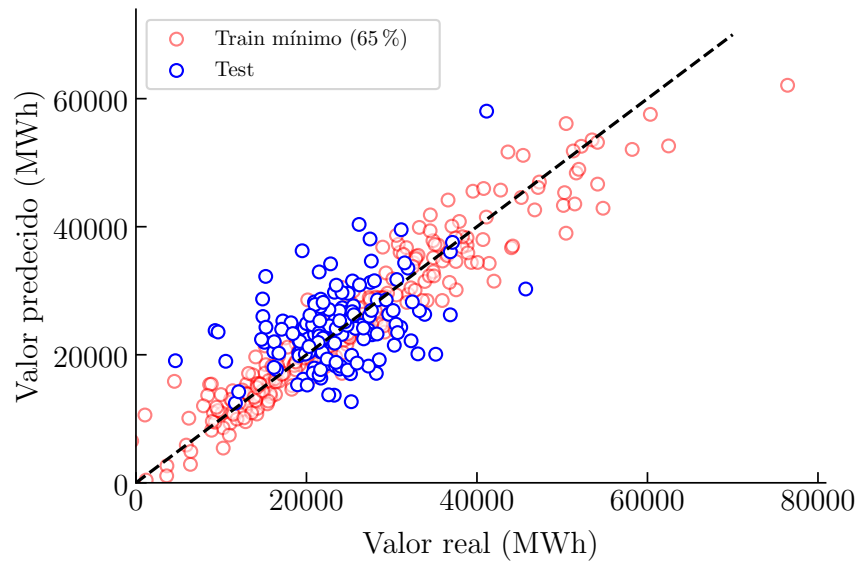


Figura 8.25: Resultados de modelo MLP usando variables zona y tipo agrupadas

Agrupación de zona y Localización con estandarizacion

Al entrenar el modelo MLP (con 5 capas de 200 perceptrones cada una y activación de ambas capas relu) y entradas estandarizadas, agrupando las variables zona y tipo se obtienen los resultados de la tabla 8.26 y al ver la predicción frente al valor real se obtiene el gráfico de dispersión de la figura 8.26. Este modelo hace bastante *overfitting* al conjunto de train, con un R^2 de 0.96. El R^2 en el test es positivo y de 0.07, superando ligeramente al *mean pred.*

	MAE	RMSE	R^2
Test	4645.171596	5765.620114	0.072038
Train	1820.125075	2576.074701	0.964184

Tabla 8.26: Resultados de modelo MLP usando variables zona y localización agrupadas

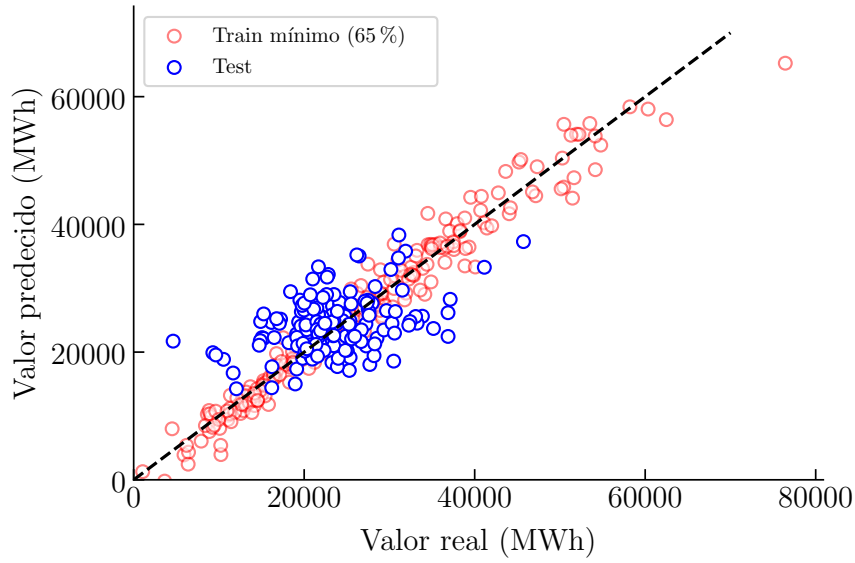


Figura 8.26: Resultados de modelo MLP usando variables zona y localizacion agrupadas

Agrupación de semáforo

Al entrenar el modelo MLP (con 3 capas de 100 perceptrones cada una y activación de las capas relu) y entradas estandarizadas, agrupando las variables zona y tipo se obtienen los resultados de la tabla 8.27 y al ver la predicción frente al valor real se obtiene el gráfico de dispersión de la figura 8.27. El R^2 en el test es mayor que en las otras agrupaciones con un valor de 0.13, superando por mucho al *mean pred.*

	MAE	RMSE	R^2
Test	4392.257055	5577.396938	0.131637
Train	2940.678352	3893.319949	0.918191

Tabla 8.27: Resultados de modelo MLP usando el semáforo de restricciones

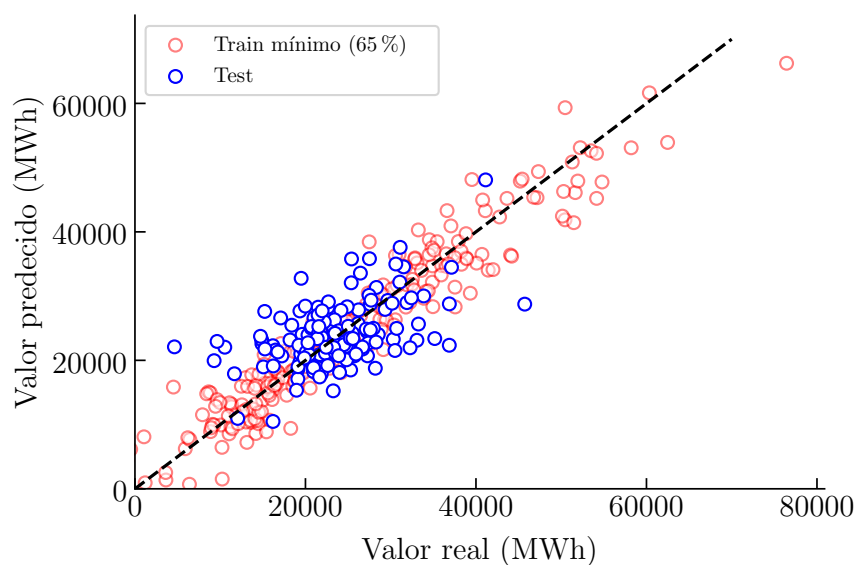


Figura 8.27: Resultados de modelo MLP usando el semáforo de restricciones

Prueba con valores incrementales

Por ultimo se ha probado el mejor modelo anterior con variables incrementales y se ha entrado un MLP con 5 capas de 200 perceptrones. Se obtienen los resultados de la tabla 8.28 y al ver la predicción frente al valor real se obtiene el gráfico de dispersión de la figura 8.28. Los resultados empeoran mucho con un R^2 que ni supera el *mean pred*

	MAE	RMSE	R^2
Test	5297.890942	6827.845434	-0.301384
Train	2073.339232	2870.364952	0.955674

Tabla 8.28: Resultados de modelo MLP usando el semáforo de restricciones y variables incrementales

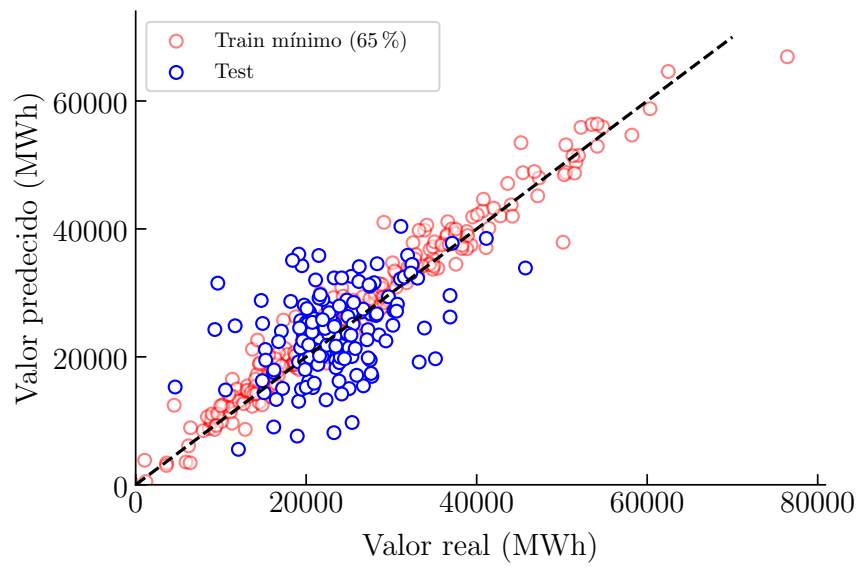


Figura 8.28: Resultados de modelo MLP usando el semáforo de restricciones y variables incrementales

8.6.5. Tabla de Resultados

Modelo		MAE	RMSE	R^2
Benchmarks	<i>Naive approach</i>	4949.7540	6855.9694	-0.3121
	<i>Mean pred</i>	4565.2592	6038.6189	-0.0179
LR	Agrupación Zona y Tipo	5984.09	7864.23	-0.726
	Agrupación Zona y Localización	5467.10	7011.12	-0.372
	Agrupación Semaforo	4854.85	6240.09	-0.086
	Mejor modelo con variables Incrementales	4678.40	6109.39	-0.041
KNN	Agrupación Zona y Tipo	6085.54	7415.26	-0.534
	Agrupación Zona y Localización	4726.30	5951.61	0.011
	Agrupación Semaforo	6103.05	7633.13	-0.626
	Mejor modelo con variables Incrementales	4818.06	6409.09	-0.146
RF	Agrupación Zona y Tipo	4580.07	5906.52	0.026
	Agrupación Zona y Localización	4118.65	5342.13	0.203
	Agrupación Semaforo	4523.19	5865.53	0.039
	Mejor modelo con variables Incrementales	4373.23	5905.76	0.026
MLP	Agrupación Zona y Tipo	5002.40	6460.60	-0.165
	Agrupación Zona y Localización	4645.17	5765.62	0.072
	Agrupación Semaforo	4392.25	5577.39	0.131
	Mejor modelo con variables Incrementales	5297.89	6827.84	-0.301

Tabla 8.29: Resultados de los modelos

8.6.6. Conclusiones

Los modelos de regresión lineal no han sido capaces de conseguir un R^2 positivo (aunque han estado cerca), lo que significa que no han conseguido superar la predicción de la media. Estos modelos se han visto muy favorecidos por la estandarización de las variables de entrada lo cual es algo raro en regresión lineal. Curiosamente es el único modelo que se ha beneficiado del uso de variables incrementales.

Los modelos de KNN son modelos demasiado simples para este problema y se ha visto muy claramente en los resultados. Aun así uno de los modelos ha sido capaz aproximar la media con bastante exactitud, consiguiendo un R^2 positivo aunque casi 0. Estos modelos son muy sensibles a la escala de sus variables y se han visto beneficiados de la normalización de las mismas.

Random Forrest ha conseguido los mejores resultados con mucha diferencia consiguiendo en el modelo de agrupación de zona y localización un R^2 de 0.2. Estos modelos son los únicos que no se han beneficiado de normalización/estandarización lo cual es normal en RF. El mejor modelo tiene un MAE de 4118 MWh, lo que supone tan solo un error del 16.5 % respecto del valor medio.

El perceptrón multicapa ha sido el segundo mejor de los modelos, obteniendo en el modelo del semáforo de restricciones un R^2 de 0.13. Estos modelos son los que más se han beneficiado de la estandarización de las variables. Muy posiblemente modelos más sofisticados o más ajustados de MLP podrían llegar a superar al mejor modelo conseguido de RF.

Todos los modelos han conseguido resultados que se pueden considerar pobres, ya que el máximo R^2 conseguido es 0.2 (solo se ha conseguido explicar un 20 % de la varianza de la variable de salida), el cual es muy bajo. Sin embargo, estos resultados son mucho mejores que los del *mean pred* o el *naive approach*.

Es importante destacar que el uso de *online learning* ha sido clave para mejorar los resultados.

Los resultados obtenidos han estado muy condicionados por las grandes diferencias entre el entrenamiento y test de la energía por restricciones (causadas por la pandemia de 2020). Los modelos han sido entrenados para predecir una variable cuya varianza es mucho mayor que la varianza del conjunto de test y sin duda esto es el motivo del pobre resultado de los modelos. Confiamos que si este modelo se sigue utilizando en el tiempo, y re-entrenando cada día (y eventualmente eliminando la parte anómala del conjunto de entrenamiento) los resultados probaran ser mucho mejores. Aun así, los modelos entrenados son muy robustos frente a nuevas situaciones, durante el periodo de la borrasca Filomena los resultados se mantienen muy consistentes, mientras que *mean pred* o el *naive approach* empeoran mucho.

Capítulo 9

Conclusiones

Este trabajo tiene como objetivo la predicción de Restricciones Técnicas usando los Informes de Restricciones Técnicas de seguridad, no obstante los trabajos desarrollados han sido mas amplios.

En primer lugar se ha hecho un análisis de estos informes, donde se ha visto que la información recogida en ellos es muy reducida y presenta con muchas inconsistencias. Con un esfuerzo de procesamiento de los datos la información resultante mejora mucho de calidad.

También se ha conseguido extraer información, no directamente pública, sobre la localización de los nudos y líneas de la red. Esta información ha permitido localizar las Indisponibilidades de Red y realizar un estudio básico sobre ellas. Se ha realizado un procesamiento de esta información para presentarla forma gráfica. Sería interesante completar este estudio con el objetivo de buscar patrones geográficos más allá de la representación gráfica.

Los datos no públicos de la localización de la potencia instalada de España también se han extraído y analizado. Combinando y procesando esta información con datos climatológicos de AEMET se ha conseguido estimar la localización de la generación en España.

Por último, se ha dado respuesta al objetivo inicial de la tesis de máster y se han entrenado modelos que predicen el valor de la energía de restricciones técnicas en fase I a subir usando los informes de restricciones y un conjunto de variables adicionales seleccionadas. El mejor de estos modelos supera con creces el *benchmark* establecido, aunque ha quedado claro que la predicción del valor de la energía de Restricciones Técnicas es una tarea muy compleja, sobre todo al entrenar con datos de 2020, que ha sido tan condicionado por una pandemia. El siguiente paso sería

CAPÍTULO 9. CONCLUSIONES

añadir a estos modelos la información de localización de las Indisponibilidades de Red y la previsión de la localización de la generación, con estas nuevas variables los modelos deberían de mejorar sus predicciones enormemente.

Capítulo 10

Alineación con los Objetivos de Desarrollo Sostenible (ODS)

El Objetivo de Desarrollo Sostenible de la Agenda 2030 más relacionado con el trabajo desarrollado es el ODS 7 “Energía Asequible y no contaminante”. En 2019 el 19,6 % de los hogares españoles sufrían un gasto energético desproporcionado (gasto en energía superior a la mediana de sus ingresos) [52].

Además, se estima que en España hay 19,1 millones de personas (41 % de la población española), que cae en alguna de las condiciones asociadas a la pobreza energética: incapacidad para mantener el hogar a una temperatura adecuada, personas con uno o varios recibos sin pagar, gasto energético desproporcionado y la pobreza energética escondida, constituida por aquellas casas que generan muy poco gasto energético por no poder permitirse más pero no aparecen en las estadísticas [53]. Conociendo estos datos, se comprende la necesidad a nivel nacional y global de que la energía sea asequible para todos, ya que se estima que en España mueren unas 7000 personas al año debido a estas condiciones. En ese sentido, el arbitraje eléctrico (entre el mercado diario e intradiario) es positivo ya que aplana el precio de la electricidad, lo que permite reducir el precio medio para el consumidor final.

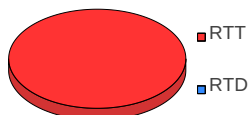
Capítulo 11

Anexos

11.1. Anexo I: Informe de restricciones

INFORME DIARIO DE RESTRICCIONES TÉCNICAS DEL PDBF: Miércoles 01 de Enero de 2020
INFORMACIÓN GENERAL
1. Energía diaria programada: Valores totales y por tipo de limitación (MWh)

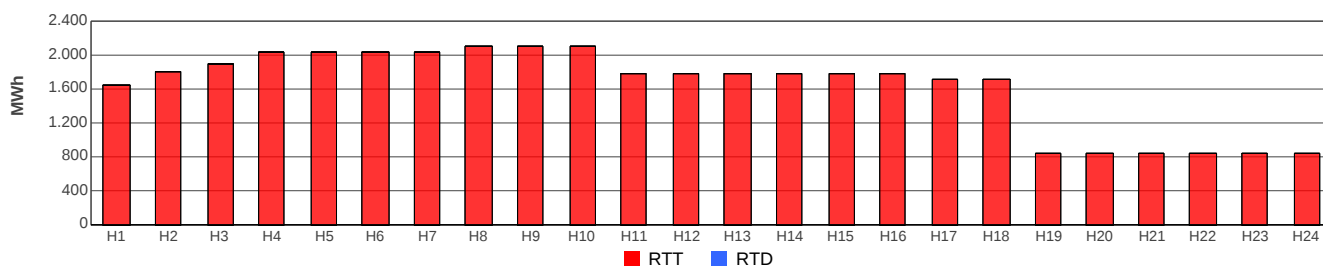
Fase I	Subir	39.029	Total(*)
	Bajar	-	39.029
Fase II	Subir	-	Total(*)
	Bajar	39.029	39.029



TIPO	Energía (*)	%
RTT	39.029	100
RTD	0	0

(*) Suma del valor absoluto de las energías programadas a subir y a bajar

(*) Energía gestionada

2. Energía horaria programada (FASE I) (MWh)


Tipo redespacho (*)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
RTT	1.638	1.811	1.893	2.043	2.043	2.043	2.043	2.108	2.108	2.108	1.778	1.778	1.778	1.778	1.778	1.778	1.713	1.713	851	851	851	851	851	851
RTD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	1.638	1.811	1.893	2.043	2.043	2.043	2.043	2.108	2.108	2.108	1.778	1.778	1.778	1.778	1.778	1.778	1.713	1.713	851	851	851	851	851	851

(*) Valores de energía neta

3. Energía limitada por seguridad (MWh)

		H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15	H16	H17	H18	H19	H20	H21	H22	H23	H24
A Subir	RTT	0	0	0	0	90	90	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	RSI	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5.370	5.408	5.370	5.408	5.370	0
A Bajar	RTT	2.478	2.478	2.478	2.478	2.478	2.478	2.478	2.478	2.478	2.478	2.148	2.148	2.148	2.148	2.148	2.148	2.148	2.148	851	851	851	851	851	851
	RTD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

4. Costes de la resolución de restricciones técnicas (*)

	€	€/MWh Programado	% s/PMHMD (**)
Resolución de Restricciones (FASE I)	2.884.979	73,92	207,99%
Reequilibrio Generación - Demanda (FASE II)	-1.303.174	33,39	93,95%
Sobrecoste Total (FASE I + FASE II)	1.581.805		

(*) En estos campos los signos "+" y "-" representan costes e ingresos para el sistema eléctrico, respectivamente.

(**) % Sobre valor medio diario del precio marginal area española del mercado MIBEL

5. Horarios de publicación

Pdbc	Pdbf	Pdvp
12:44	13:10	14:39

CONTINGENCIAS ANALIZADAS	ETIQUETAS
FALLO SIMPLE DE LÍNEAS Y TRANSFORMADORES	RTT Restricción Técnica en RdT
FALLO DE LÍNEAS DE DOBLE CIRCUITO (Compartiendo apoyos durante >= 30 km.)	RTD Restricción Técnica en RdD
FALLO DE GRUPOS GENERADORES	MEX Medidas Excepcionales
FALLO CONSECUTIVO DE GRUPO GENERADOR Y LÍNEA DE INTERCONEXIÓN DE ZONA	

6. Observaciones

RESTRICCIONES POR SEGURIDAD EN LA RED DE DISTRIBUCIÓN

No existe información en Base de Datos del SIOSbi para el informe solicitado.

RESTRICCIONES TECNICAS DE SEGURIDAD

Zona	Escenario	Contingencias	Tipo	Nivel	Localización
Centro	H1 - H18	Regular Tensión y Seguridad en la Zona	Control Tensión	220 y 400 kV	Zona Centro
Este	H1 - H24	Regular Tensión y Seguridad en la Zona	Control Tensión	220 y 400 kV	Zona Cataluña
Levante	H1 - H24	Regular Tensión y Seguridad en la Zona	Control Tensión	220 y 400 kV	Zona Levante
	H21 - H24	L-400 kV Eliana - Plana	Sobrecargas	> 115 %	L-400 kV Gaussa - Plana
Noroeste	H1 - H24	Regular Tensión y Seguridad en la Zona	Control Tensión	220 y 400 kV	Zona Galicia
Norte	H1 - H18	Regular Tensión y Seguridad en la Zona	Control Tensión	220 y 400 kV	Zona País Vasco
Sur	H1 - H24	Regular Tensión y Seguridad en la Zona	Control Tensión	220 y 400 kV	Zona Andaluza

OTRAS OBSERVACIONES

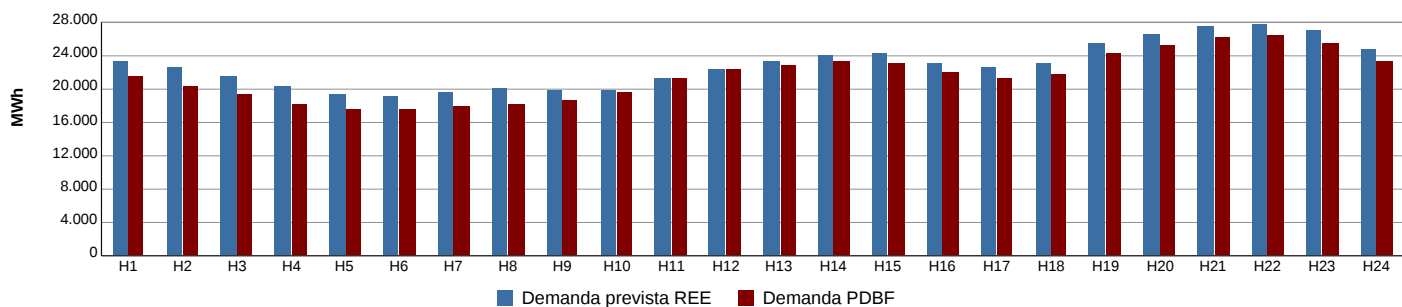
Destacar que de acuerdo con lo establecido en el procedimiento de operación P.O. 1.2, durante el periodo que va desde el día 1 de noviembre hasta el 31 de marzo los análisis de seguridad de la red de transporte se harán con los parámetros correspondientes a los límites térmicos estacionales de invierno.

7. Limitaciones zonales por seguridad (MWh)

No existe información en Base de Datos del SIOSbi para el informe solicitado.

ANEXO: INFORMACIÓN CON INFLUENCIA EN LOS ANALISIS DE SEGURIDAD

1. Contrastación demanda prevista REE - demanda programada en PDBF (MWh)

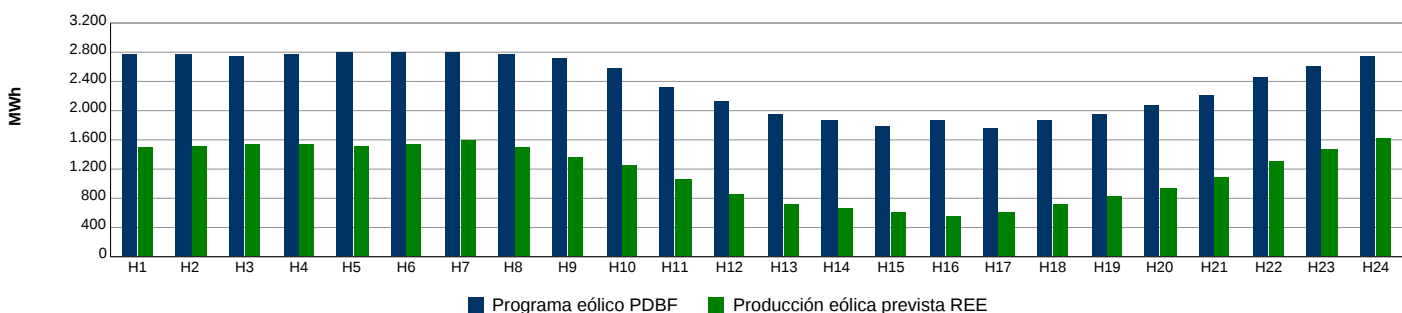


DEMANDA	Hora	PREVISTA REE	PDBF (*)	DIFERENCIA (**)	%PDBF/PREV.REE
Mínimo Horario	H6	19.225	17.522	1.703	91,14%
Máximo Horario	H22	27.830	26.407	1.423	94,89%
TOTAL DIARIO		549.335	518.502	30.833	94,39%

(*) Total programas PDBF venta energía (produccion + saldo importador programas int.int.) - programas PDBF consumo bombeo

(**) Diferencia entre demanda prevista REE y demanda PDBF

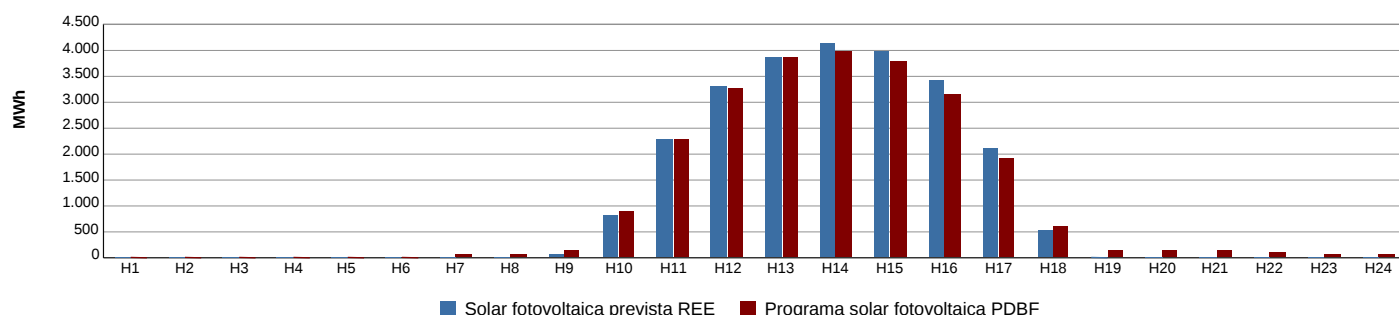
2. Contrastación producción eólica prevista REE - programa eólico PDBF (MWh)



EÓLICA	Hora	PREVISTA REE	PDBF	DIFERENCIA (*)	%PDBF/PREV.REE
Mínimo Horario	H16	558	1.874	1.316	335,86%
Máximo Horario	H24	1.623	2.751	1.128	169,51%
TOTAL DIARIO		27.978	57.107	29.129	204,12%

(*) Diferencia entre programa eólico PDBF y producción eólica prevista por REE

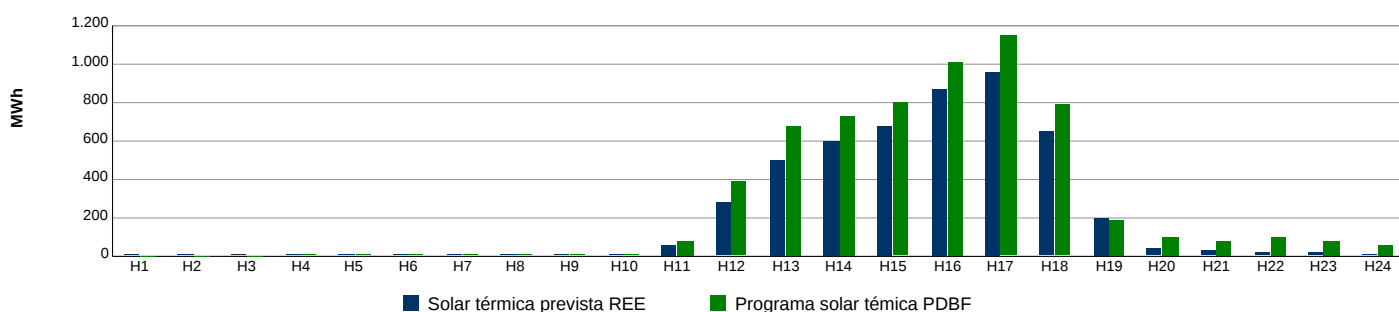
3. Contrastación de solar fotovoltaica prevista REE - programa solar fotovoltaica PDBF (MWh)



SOLAR FOTOVOLTAICA	Hora	PREVISTA REE	PDBF	DIFERENCIA *	% PDBF/PREV. REE
Mínimo Horario	H1	0	8	8	
Máximo Horario	H14	4.124	3.974	150	96,36%
TOTAL DIARIO		24.502	24.936	434	101,77%

(*) Diferencia entre producción solar fotovoltaica prevista por REE y programa solar fotovoltaica PDBF

4. Contrastación de solar térmica prevista REE - programa solar térmica PDBF (MWh)



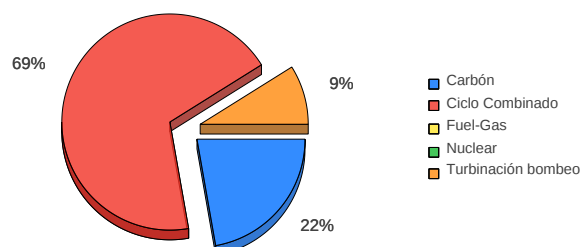
SOLAR TÉRMICA	Hora	PREVISTA REE	PDBF	DIFERENCIA *	% PDBF/PREV. REE
Mínimo Horario	H10	7	11	4	157,14%
Máximo Horario	H17	960	1.154	194	120,21%
TOTAL DIARIO		5.042	6.322	1.280	125,39%

(*) Diferencia entre producción solar térmica prevista por REE y programa solar térmica PDBF

5. Indisponibilidades de unidades de producción

Tecnología	P.Ind.(MW) (*)	P.Inst. (MW)	% Indisp.
Carbón	1.168,6	9.213,9	12,68%
Ciclo Combinado	3.601,9	24.559,6	14,67%
Fuel-Gas	0,0	7,9	0,00%
Nuclear	0,0	7.117,2	0,00%
Turbinación bombeo	460,2	3.417,5	13,47%
Total	5.230,7	44.316,1	11,80%

Potencia indisponible por tecnología respecto a total indisponible (MW y %)

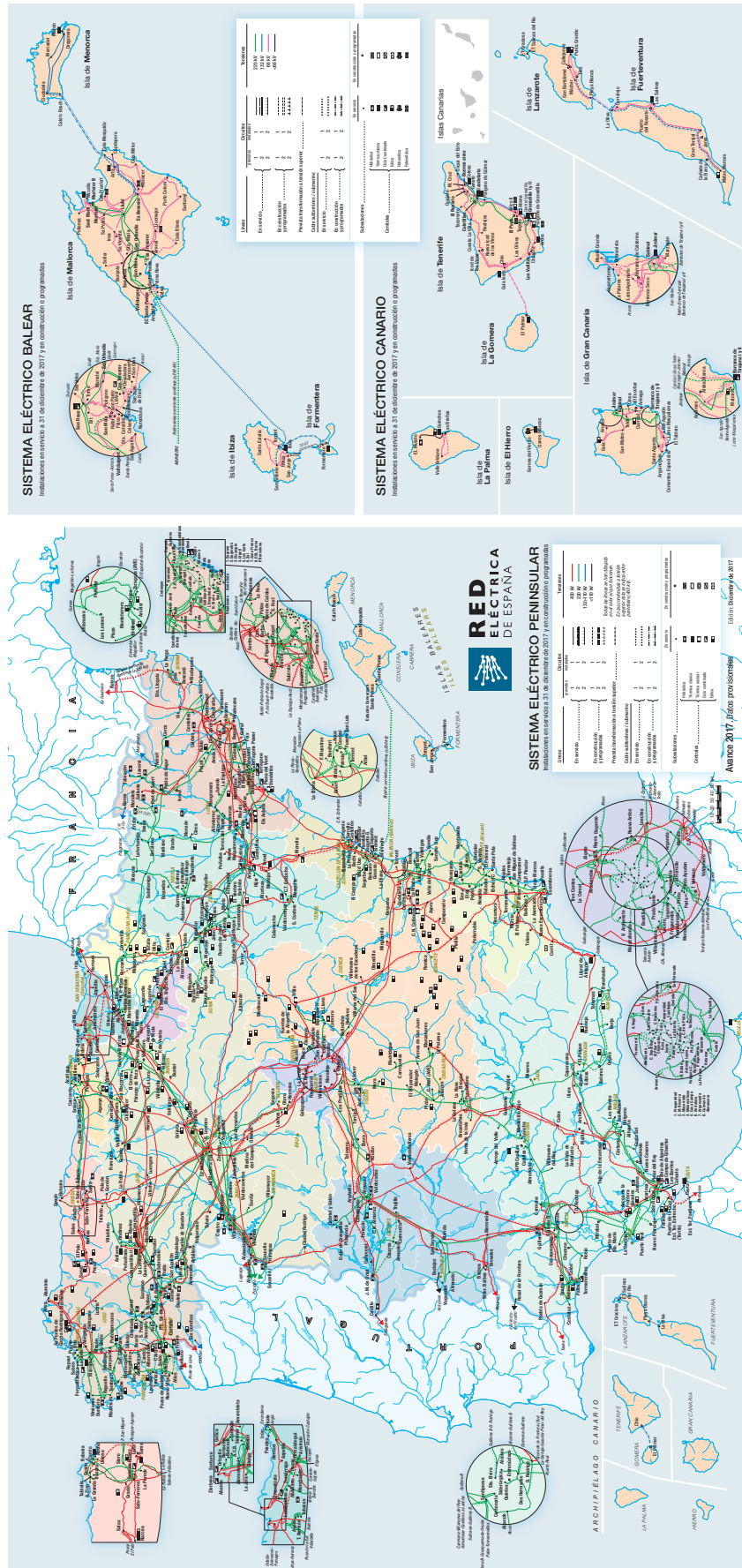


(*) Valores de potencia indisponible considerando periodos horarios completos para todas las indisponibilidades.

6. Indisponibilidades de red

INDISPONIBILIDADES DE RED CON INFLUENCIA EN LOS ANÁLISIS DE RESTRICCIONES	INICIO	FINAL	TIPO	REPOSICIÓN DIARIA	REPOSICIÓN DE EMERGENCIA	OBSERVACIONES
220 kV ACERIASA CACICEDO	19/12/2019 09:29	04/01/2020 18:00	C.P.	NO	NO	
220 kV ADIF MONFORTE DEL CID MONFORTE DEL CID 1	01/08/2019 09:25	13/01/2020 23:59	P.S.	NO	NO	
220 kV CICERO SOLORZANO 1	29/09/2016 11:02	31/12/2021 23:59	P.S.	NO	NO	
220 kV CICERO SOLORZANO 2	29/09/2016 11:02	31/12/2021 23:59	P.S.	NO	NO	
220 kV SOTO DE RIBERA TRASONA	09/10/2015 12:49	28/02/2030 23:59	P.S.	NO	NO	
400 kV BALBOA SIDERURGICA BALBOA	01/03/2018 17:13	21/12/2020 12:00	P.A.	NO	NO	
400 kV PUERTO DE LA CRUZ TARIFA 2	14/11/2019 10:59	13/09/2020 18:00	P.A.	NO	NO	Modifica la capacidad de intercambio con Marruecos (Ver SIOS).
400 kV TARIFA FARDIOUA 2	14/11/2019 10:59	13/09/2020 18:00	P.A.	NO	NO	Modifica la capacidad de intercambio con Marruecos (Ver SIOS).
BANCO DE CONDENSADORES - GALAPAGA GAL400-BCO1	28/01/2019 09:57	12/02/2020 18:00	P.S.	NO	NO	
JUEGO DE BARRAS - AMOREBIE AMO400-BP1	02/12/2019 10:43	10/01/2020 23:59	P.A.	NO	08:00h	
JUEGO DE BARRAS - CHANTADA CHA220-BP2	29/11/2019 11:21	31/01/2020 18:00	P.S.	NO	NO	Se autoriza a partir del 29/11 condicionado a la obtención de la correspondiente APES, para lo cual quedan requisitos pendientes de cumplimentar (al menos el alta del punto frontera en SIMEL).
JUEGO DE BARRAS - TALAVERA TAL220-BP1A	30/09/2019 12:01	11/05/2020 18:00	P.A.	NO	NO	
JUEGO DE BARRAS - TALAVERA TAL220-BP2B	04/11/2019 11:22	11/05/2020 18:00	P.S.	NO	08:00h	
REACTANCIA - GUADAME GMD33-RAT1	01/05/2019 10:51	31/12/2020 18:00	P.A.	NO	NO	
REACTANCIA - MAGALLON MAG400-RPT1	30/07/2019 17:26	31/03/2020 18:00	C.P.	NO	NO	

11.2. Anexo II: Mapa de nudos y líneas



11.3. Anexo III: Nudos transformados y excluidos de las transformación de nudos de Madrid

Lista de nudos transformados: [La Cereal, Tres Cantos, Alcobendas, San Sebastian de los Reyes, Algete, Daganzo, Meco, Galapagar, Majadahonda, E.T. Majadahonda, V. Arcipestre, E.T. Fuencarral, Mirasierra, Boadilla, Aravaca, Villaviciosa, Casa campo, V. Batan, Lucero, Poligono C, Prado Santo, Fuencarral, El Pilar, C. Deportiva, Sanchinarro, Azca, Arroyo de la Vega, Hortaleza, Prosperidad, Coto, Norte, Mazarredo, Palafox, Melancolicos, Aguacate, La Fortuna, Arganzuela, Mediodia, M.Becerra, La Estrella, Pradolongo, Pte Princesa, Retamar, C. de la Plata, P. Ingenieros, Aena, E.T., Coslada, Simancas, E.T., Canillejas, Coslada, Vallengas, Pinto Ayuden, Hornillo, Villaverde, Fuenlabrada, Leganes, C. Fregacedos, Buenavista, Moraleja, Anover, Getafe, El Emperador, Parla, Pinto, Valdemoro, Madrudejos, P. de S. Fernando, Nueva Ardoz, La Torrecilla, Anchuelo, Loeches, Arganda, Morata, C.V.]

Lista de nudos excluidos: [Mora, Lastras del Pozo, Segovia, Otero de Herreros, Torrijos, Los Pradillos, Aceca, Carroyuelas, Arenas de San Juan, Trillo, Fuentes de la Alcarria, J. Cabrera, Huelves]

11.4. Anexo IV: Nudos problematicos

Lista de nudos del P25: Puerto, Tabiella, Soto de Ribera, La Lomba, Montearenas, Cicero, Solorzano, Zierbana, Santurce, Guardo, Velilla, La Mudarra, La Mudarra, San Sebastian de los Reyes, Hernani, Itxaso, La Serna, Magallón, Mudjar, Teruel, Aragon, Almaraz, Sider. Balboa, Balboa, Guillena, Alcores, Puerto de la Cruz, Galapagar, Talavera, Villaviciosa, Torrijos, P. Ingenieros, Coslada, Coslada, Villaverde, Aceca, Guadalquivir Medio, Morata, S. Costera, Valdeconejos, La Plana, Begues, Rubi, Can Jordi, Vic, Gramanet

Bibliografía

- [1] Gobierno de España. *Boletín Oficial del Estado*. URL: [https://www.boe.es/eli/es/res/2019/10/10/\(1\)/dof/spa/pdf](https://www.boe.es/eli/es/res/2019/10/10/(1)/dof/spa/pdf).
- [2] OMIE. *Mercado de electricidad*. URL: <https://www.omie.es/es/mercado-de-electricidad>.
- [3] R. Weron. “Electricity price forecasting: A review of the state-of-the-art with a look into the future”. En: 2014. Cap. 30, págs. 1030-1081.
- [4] A. Ramos, M. Ventosa y M. Rivier. “Modeling competition in electric energy markets by equilibrium constraints”. En: vol. 7. 4. 1999, págs. 233-242.
- [5] T. A Johnsen. “Demand, generation and price in the Norwegian market for electric power”. En: vol. 22. 3. 2001, págs. 227-251.
- [6] R. Becker, S. Hurn y Pavlov V. “Modelling spikes in electricity prices”. En: vol. 263. 86. 2007, págs. 371-382.
- [7] A. J. Conejo y col. “Forecasting electricity prices for a day-ahead pool-based electric energy market”. En: vol. 21. 3. 2005, págs. 435-462.
- [8] J. W. Taylor. “Triple seasonal methods for short-term electricity demand forecasting”. En: vol. 204. 1. 2010, págs. 139-152.
- [9] A. Cruz y col. “The effect of wind generation and weekday on Spanish electricity spot price forecasting”. En: vol. 81. 10. 2011, págs. 1924-1935.
- [10] R. Weron y A. Misiorek. “Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models”. En: vol. 24. 4. 2008, págs. 744-763.
- [11] R. Pino y col. “Forecasting next-day price of electricity in the Spanish energy market using artificial neural networks”. En: vol. 21. 1. 2008, págs. 53-62.
- [12] C. Wan y col. “A hybrid approach for probabilistic forecasting of electricity price”. En: vol. 5. 1. 2013, págs. 463-470.
- [13] A. Ramos. *Análisis de restricciones técnicas en el Sistema Eléctrico español (Trabajo de fin de grado)*. 2016.

- [14] S. Hagemann. “Price determinants in the German intraday market for electricity: an empirical analysis”. En:
- [15] *Proyecto de Ley de Cambio Climático y Transición Energética*. 2019.
- [16] J. R. Andrade y col. “Probabilistic price forecasting for day-ahead and intraday markets: Beyond the statistical model”. En: vol. 9. 11. 1990.
- [17] J. P. Chaves-Ávila, R. A. Hakvoort y A. Ramos. “Short-term strategies for Dutch wind power producers to reduce imbalance costs”. En: vol. 52. 2013, págs. 573-582.
- [18] C. Monteiro y col. “Short-term price forecasting models based on artificial neural networks for intraday sessions in the Iberian electricity market”. En: vol. 9. 9. 2016, pág. 721.
- [19] R. Aïd, P. Gruet y H. Pham. “An optimal trading problem in intraday electricity markets”. En: vol. 10. 1. 2016, págs. 49-85.
- [20] *SimplePDF*. URL: simplepdf.com.
- [21] *Camelot*. URL: <https://pypi.org/project/camelot-py/>.
- [22] *Tabula*. URL: <https://pypi.org/project/tabula-py/>.
- [23] *REE*. URL: <https://www.ree.es/es/clientes/generador/acceso-conexion/conoce-la-capacidad-de-acceso>.
- [24] *REE2*. URL: https://www.ree.es/sites/default/files/01%5C_ACTIVIDADES/Documentos/Mapas-de-red/mapa%5C_transporte%5C_iberico%5C_2018.pdf.
- [25] *ENTSO*. URL: <https://www.entsoe.eu/>.
- [26] *ENTSO2*. URL: <https://www.entsoe.eu/data/map/>.
- [27] *Mapbox*. URL: <https://www.mapbox.com/>.
- [28] *Protocol Buffers*. URL: <https://developers.google.com/protocol-buffers>.
- [29] *Vector Tile Mapbox*. URL: <https://github.com/mapbox/vector-tile-py>.
- [30] *GeoJSON*. URL: <https://geojson.org/>.
- [31] *ISO 3166*. URL: <https://www.iso.org/iso-3166-country-codes.html>.
- [32] *Jupyter GMaps*. URL: <https://jupyter-gmaps.readthedocs.io/>.
- [33] URL: <https://www.ree.es/es/apidatos>.
- [34] *ESIOS REE*. URL: <https://www.esios.ree.es/>.
- [35] *ESIOS Mapa de instalaciones de producción*. URL: <https://www.esios.ree.es/es/mapas-de-interes/mapa-instalaciones-produccion>.

- [36] *Carto*. URL: <https://carto.com/>.
- [37] *Municipios de España*. URL: https://upload.wikimedia.org/wikipedia/commons/e/e0/Spain%5C_municipalities.png.
- [38] *BOE 2007*. URL: <https://www.boe.es/buscar/act.php?id=BOE-A-2007-10556>.
- [39] *Longitud y latitud*. URL: <https://www.businessintelligence.info/varios/longitud-latitud-pueblos-espana.html>.
- [40] *Análisis ESIOS*. URL: <https://www.esios.ree.es/es/analisis/1153>.
- [41] *Análisis ESIOS*. URL: <https://www.esios.ree.es/es/analisis/549?>.
- [42] *Open data AEMET*. URL: <https://opendata.aemet.es/>.
- [43] *ESIOS generación eólica*. URL: <https://www.esios.ree.es/es/analisis/10037>.
- [44] *ESIOS generación fotovoltaica*. URL: <https://www.esios.ree.es/es/analisis/1161>.
- [45] *ESIOS energía asociada a restricciones*. URL: <https://www.esios.ree.es/es/analisis/10051>.
- [46] *ESIOS prevision diaria eólica*. URL: <https://www.esios.ree.es/es/analisis/1777>.
- [47] *ESIOS prevision diaria fotovoltaica*. URL: <https://www.esios.ree.es/es/analisis/1779>.
- [48] *ESIOS previsión demanda*. URL: <https://www.esios.ree.es/es/analisis/1775>.
- [49] *ESIOS real*. URL: <https://www.esios.ree.es/es/analisis/1335>.
- [50] *Árbol de Decisión*. URL: <https://pythonprogramminglanguage.com/what-are-the-advantages-of-using-a-decision-tree-for-classification>.
- [51] *Multilayer Perceptron*. URL: https://scikit-learn.org/stable/%5C_images/multilayerperceptron%5C_network.png.
- [52] *An optimal trading problem in*. URL: [https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/transicion-ecologica/Paginas/2020/121120-indicadores.aspx#:~:text=MITECO%5C%20presenta%5C%20los%5C%20indicadores%5C%20de%5C%20pobreza%5C%20energ%C3%5C%A9tica%5C%20de%5C%202019,-Jueves%5C%2012%5C%20de&text=En%5C%202019%5C%2C%5C%20un%5C%2016%5C%2C7,%5C%2C9%5C%25%5C%20del%5C%20a%5C%C3%5C%B1o%5C%20anterior.\)%5C%20\(https://elpais.com/elpais/2019/02/20/3500_millones/1550664003_825237.html](https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/transicion-ecologica/Paginas/2020/121120-indicadores.aspx#:~:text=MITECO%5C%20presenta%5C%20los%5C%20indicadores%5C%20de%5C%20pobreza%5C%20energ%C3%5C%A9tica%5C%20de%5C%202019,-Jueves%5C%2012%5C%20de&text=En%5C%202019%5C%2C%5C%20un%5C%2016%5C%2C7,%5C%2C9%5C%25%5C%20del%5C%20a%5C%C3%5C%B1o%5C%20anterior.)%5C%20(https://elpais.com/elpais/2019/02/20/3500_millones/1550664003_825237.html).

BIBLIOGRAFÍA

- [53] G. L Morales. *Más del 40 % de la población española sufre pobreza energética*. URL: https://elpais.com/elpais/2019/02/20/3500%5C_millones/1550664003_825237.html.
- [54] *Mapa nudos y líneas*. URL: https://www.ree.es/sites/default/files/downloadable/mapa%5C_lineas%5C_2017%5C_v1.pdf.