



***ANÁLISIS DE LA DEMANDA ELÉCTRICA EN ESPAÑA MEDIANTE SERIES  
TEMPORALES. DETERMINACIÓN DE VARIABLES EXÓGENAS***

MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS

2021 - 2022

Nombre y Apellidos del alumno

***Bernardo José Llamas Verna***

Nombre y Apellidos del tutor

***Lino González García***

## **Índice de Contenidos**

<b>Resumen y Palabras Clave.</b>	<b>4</b>
<b>Agradecimientos.</b>	<b>5</b>
<b>1. Introducción</b>	<b>7</b>
1.1. Contexto y justificación	7
1.2. Objetivos	7
<b>2. Metodología.</b>	<b>8</b>
<b>3. Estado del arte.</b>	<b>8</b>
3.1. Series temporales	8
3.2. Técnicas de análisis de series temporales	9
3.3. Variables exógenas para mejorar las predicciones	10
<b>4. Extracción y preparación de los datos</b>	<b>10</b>
4.1. Introducción	11
4.2. Conexión con la API de E-SIOS	11
4.3. Extracción de datos de temperatura	12
<b>5. Análisis exploratorio de la demanda eléctrica</b>	<b>13</b>
5.1. Introducción	13
5.2. Evolución y estacionariedad de la demanda eléctrica en el período de estudio	13
5.3. Evolución de la demanda eléctrica	21
5.4. Incorporación de variables exógenas	26
<b>6. Modelos de series temporales</b>	<b>27</b>
6.1. Introducción	27
6.2. Holt- Winters como modelo de partida	28
6.3. SARIMAX	29
6.4. XGBoost	30
6.5. Prophet	30
6.6. Métricas escogidas	31
6.7. Variables exógenas y escenarios escogidos	34

<b>6.8. Comparación entre modelos</b>	<b>35</b>
<b>6.9. Comparación de métricas entre modelos</b>	<b>37</b>
<b>7. Aplicación de los modelos sobre datos a futuro</b>	<b>39</b>
<b>7.1. Comparativa métricas</b>	<b>40</b>
<b>8. Conclusiones</b>	<b>42</b>
<b>9. Siguiendo pasos</b>	<b>44</b>
<b>Bibliografía</b>	<b>46</b>
<b>Anexos</b>	<b>50</b>

### ***Resumen y Palabras Clave.***

El objetivo del presente trabajo es el análisis y modelización de la demanda eléctrica en el Estado Español entre los años 2014-2021. Se realiza un análisis exploratorio con el fin de encontrar patrones en diferentes granularidades temporales: años, meses, semanas y días. Las conclusiones permitirán la selección de un conjunto de variables exógenas con las que enriquecer los modelos planteados en la segunda parte del trabajo.

A la hora de abordar la modelización, se ha partido de un modelo clásico regresivo (Holt-Winters) como modelo baseline a partir del cual poder comparar el resto de modelos, todos basados en técnicas más avanzadas: SARIMAX, XGBoost y Prophet.

Mediante la comparativa de métricas se han determinado cómo afectan las variables exógenas, así como el mejor modelo descriptivo de la serie.

***Palabras Clave:*** Series temporales, forecasting, demanda eléctrica, predicción, variables exógenas.

## ***Agradecimientos.***

2021-2022, un curso complicado, se podría decir en resumen para comenzar estas líneas. El año pasado sobre estas fechas estaba cansado de lo que estaba haciendo con mi vida profesional, afectando a su vez a mi vida personal. La sensación de no encontrar definitivamente mi sitio y una nostalgia cada vez más acentuada de mis tiempos como estudiante de Ciencias Físicas me llevaban a plantearme por qué no seguí un poco más la carrera de investigación, qué sitio podría encontrar donde poder reencontrarme, etc. Obviamente la opción de todo lo relacionado con el mundo del dato estaba ahí, pero me era difícil que en una entrevista de trabajo pudieran ver más allá de mi experiencia como desarrollador de software, más que lo que podía aportar como físico en esa área, y la desesperación crecía.

Por azares de la vida terminé contactando con Lino, que es el tutor del presente trabajo, y me ofreció la posibilidad de hacer un doctorado o hacer este máster. Al final me decanté por el máster viendo que el programa se adaptaba a lo que quería hacer.

Agradezco finalmente el haber realizado el máster a pesar de lo complicado que ha sido trabajar, estudiar e intentar vivir. Finalmente el ansiado cambio de departamento en mi empresa se ha dado y ahora estoy en mi primer proyecto como científico de datos.

Pero también tengo que agradecerle a muchas personas.

A los primeros que toca agradecer es a mis padres por todo lo que han hecho y siguen haciendo por mí. El estar en la vida adulta e intentar realizar un proyecto de vida propio, da una dimensión mucho mayor de todo el sacrificio que han hecho y siguen haciendo por sus hijos y ni de lejos habría conseguido lo que he conseguido hasta ahora sin su apoyo. Aquí mi reconocimiento.

A mis abuelos, ya sea en las montañas de León o en Caracas, por haber sacado adelante a mis padres con todo lo que pudieron.

A mis hermanos, aunque la vida adulta nos lleve por caminos separados, el recuerdo de una bonita infancia no se borrará jamás.

A mis amigos, que este año más de una vez les he tenido que rechazar planes por tener que dedicarle tiempo a este máster. En especial agradecer a Pérez que me ayudase con el curso de DevOps que hice en paralelo a este; las 2 tardes en su casa haciendo prácticas se las agradeceré infinito por siempre.

A Victoria, por animarme a realizar este curso cuando a lo mejor asomaba la sombra de duda y por haber sido paciente cuando no nos podíamos ver o tenía que desaparecer para conseguir realizar una entrega a tiempo.

A mi grupo de amigos de *Data, Blockchain y NFT*, Maddi, Manu, Claudia y Diego por ayudarme a elegir y estructurar este tema de investigación. La extracción de datos de temperaturas y cómo procesarlos se lo debo a Claudia y la selección de las ciudades donde sacar las temperaturas a Manu.

A Diego aquí es al que más hay que agradecer. No solo estructuró el proyecto, cómo abordarlo, etc, sino que también ha estado ahí para resolver cualquier tipo de duda, revisar casi al instante esta tesis, dando ideas, correcciones, etc. Agradecer infinitamente es insuficiente sabiendo el poco tiempo que tiene siendo padre primerizo y corredor semiprofesional. Casi ha sido como volver a nuestras prácticas de laboratorio en la facultad. Muchas gracias amigo.

Antes de terminar quisiera decir que los amigos de verdad son los que se conocen antes de los 30 y, antes de que Diego levante la vista de esta página cual Robin Williams en el Indomable Will Hunting reclamando una frase robada, quisiera añadir un corolario:

*Los amigos de verdad son los que conoces antes de los 30 y los que nunca fallan son aquellos con los que tienes que buscar gasolina en medio de viento, lluvia y nieve en la mitad de la nada en Anatolia.*

Madrid, 16 de julio del año 2022

## **1. Introducción**

### **1.1. Contexto y justificación**

Los mercados de electricidad se diferencian de otros mercados por la necesidad de producir la energía en el mismo instante que es consumida, manteniendo un balance casi perfecto.

El análisis de la demanda de energía eléctrica es importante para comprender no sólo los patrones de consumo de los usuarios y su relación con factores como el económico y climático, sino también para identificar uno de los elementos que afecta la formación del precio final.

En el momento en el que se escribe esta memoria (año 2022) todo lo relacionado con el Sistema Eléctrico español lleva siendo tendencia en los debates a nivel nacional, tanto social como político. Principalmente por el alza de los costes de generación y cómo afecta a nivel económico y social [\[01\]](#), pero también por cómo se tiene que generar la energía en la necesidad de una transición ecológica, sumándose ahora la incertidumbre por los suministros de energía en el nuevo contexto internacional que marcan la guerra en Ucrania y la tensión de las relaciones con Argelia [\[02\]](#), [\[03\]](#).

Es en este contexto que surge el deseo de conocer mediante el análisis de series temporales cómo se comporta la demanda energética en el Estado Español para el período 2014-2021: qué tendencia global sigue, comportamiento estacional, variables que afectan a dicha demanda, etc. Para ello se utilizarán diferentes técnicas para generar modelos y cómo afecta la adición de variables exógenas a la bondad de dichos modelos.

### **1.2. Objetivos**

Los objetivos del presente trabajo son los siguientes:

- Caracterización de la demanda eléctrica en el Estado Español.
- Identificación de las principales variables exógenas que afectan a la demanda eléctrica.
- Modelización
- Identificación del peso de las variables exógenas en la precisión de los modelos.

## **2. Metodología.**

La metodología llevada a cabo es la que se corresponde con la de un estudio de caso en el que se analiza mediante técnicas de análisis de series temporales cómo se comporta la demanda eléctrica en España para el período 2014-2021 y determinar las principales variables exógenas que afectan a la demanda y pueden ayudar a realizar predicciones precisas.

## **3. Estado del arte**

Antes de iniciar el presente trabajo hubo un proceso de recopilación de información para ver qué es lo que se venía haciendo en cuanto a series temporales y determinación de variables exógenas aplicado a la predicción de la demanda eléctrica. Pero antes de comenzar a hablar más en profundidad, cabe repasar algunos conceptos sobre series temporales y cómo abordar su estudio.

### **3.1. Series temporales**

En cuanto a las series temporales, son series cronológicamente ordenadas de datos que pueden estar o no espaciados en intervalos de tiempo iguales. Para predecir los valores a futuro se puede hacer ya sea modelizando la serie solo a partir de datos pasados (opción autoregresiva) o usando otras variables externas [\[04\]](#).

Al principio en el *forecasting* clásico el interés se centraba en el histórico de la variable objeto de estudio para conseguir las predicciones, pero a día de hoy el enfoque moderno es que, junto con el histórico de datos, se consiguen mejores predicciones teniendo en cuenta las variables exógenas y su relación con la variable de interés.

Para tratar una serie temporal hay que saber que se compone de 4 partes:

- **Tendencia:** se define como la evolución a lo largo del tiempo de la serie.
- **Ciclos:** es la información sobre los cambios que pueden repetirse a lo largo del tiempo sin que se deban a un factor estacional.
- **Estacionalidad:** describe los cambios periódicos a lo largo del tiempo.
- **Residuos:** todo lo que no puede ser descrito en los 3 puntos anteriores y que se origina por factores aleatorios.



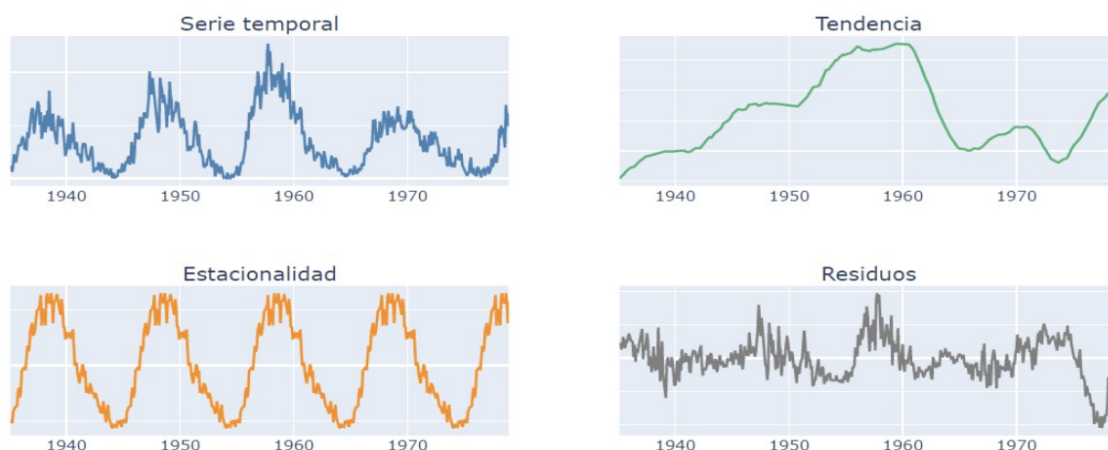


Fig. 01: Ejemplo descomposición de una serie temporal (Fuente: web Keeper [\[05\]](#))

Las series temporales se pueden clasificar en 2 tipos: **multiplicativa** y **aditiva** y se diferencian en cómo se relacionan las componentes de una serie temporal entre sí, siendo en el caso de las aditivas por la suma de sus componentes y en el caso de las multiplicativas por su producto [\[05\]](#).

- Serie temporal = Tendencia + Estacionalidad + Residuos (aditiva)
- Serie temporal = Tendencia \* Estacionalidad \* Residuos (multiplicativa)

### 3.2. Técnicas de análisis de series temporales

Para modelizar series temporales y poder predecir valores a futuro, se han utilizado diversas técnicas que han ido mejorando o siendo abandonadas por otras según aumentaba la capacidad de cálculo.

En la década de 1920-1930, cuando comienza el abordaje estadístico de las series temporales, aparecen los primeros modelos autorregresivos y de media móvil. Después, con la aparición de las primeras máquinas de cálculo en la década de 1950, surgen técnicas de suavizado exponencial (exponencial simple, doble y *Holt-Winters*) que permiten conseguir mejoras en la predicción al tener en cuenta factores como la tendencia y la estacionalidad.

Las principales técnicas para modelizar series temporales han sido hasta ahora las arquitecturas autorregresivas representadas por *ARIMA*, *SARIMA*, *SARIMAX*, etc, que sobrepasaron a las técnicas de modelado exponencial, pero con la irrupción de las redes neuronales convolucionales y arquitecturas basadas en redes neuronales recursivas, como *DeepAR* de AWS, se están consiguiendo modelos con un poder de predicción y precisión nunca vistos. Esto

podría indicar que las técnicas autorregresivas han quedado atrás definitivamente, pero en 2017 Facebook hizo público en un artículo llamado “*Forecasting at Scale*” que habían creado un nuevo modelo llamado *Prophet* [06].

Por último indicar que, junto con las técnicas autorregresivas y las redes neuronales, también se están empleando algoritmos de clasificación como *XGBoost* en series temporales, consiguiendo también buenos resultados.

### **3.3. Variables exógenas para mejorar las predicciones**

Cuando se intenta analizar una serie temporal, hay una variable objeto de estudio que es la que interesa conocer cómo irá evolucionando a lo largo del tiempo. Muchas variables objeto de estudio evolucionan a lo largo del tiempo mostrando tendencias al alza o baja, estacionalidad, etc debido a la interacción con una o varias variables, algunas fáciles de identificar, otras no tanto. Dichas variables pueden ser endógenas o exógenas.

Una variable exógena es aquella cuyo valor se determina fuera del modelo y se impone en dicho modelo sin ser afectada por él [07].

Estudios como el de Ramón Christen et al. (2020) [08] exponen que se pueden llegar a identificar 50 variables exógenas diferentes agrupadas en las siguientes categorías:

- **Variables climáticas:** humedad, precipitación, temperatura, viento, etc.
- **Calendario:** fechas, eventos señalados, vacaciones, Navidad, etc.
- **Información del día:** si el día es antes/después de un festivo, día laboral/festivo, día de la semana, etc.
- **Socioeconómico:** tendencias de la economía, PIB, tasa de empleo, etc.
- **Información demográfica:** tasa de nacimientos, número de habitantes, número de viviendas, etc
- **Otros:** sensores, aparatos conectados a la red, etc.

Para obtener los mejores grados de precisión posible Ramón Christen et al. (2020) propone establecer pesos que den más o menos importancia a cada una de las variables. Estos pesos puede variar según la granularidad del estudio (años, meses, días, horas, etc) y también del sitio donde se realice la investigación; por ejemplo, la variable climática puede ser más importante en

Polonia que en España, la variable demográfica más importante en una ciudad del mismo país que en otra, etc.

Para identificar estas variables exógenas, además de la observación de los patrones que afectan a la variable objeto de estudio, como se realiza en el presente trabajo, Janusz Sowinski (2021) [09] propone emplear el algoritmo Hellwig, utilizado en econometría.

Como se ha indicado, en el presente trabajo la variable objeto de estudio es la demanda eléctrica en el Estado Español. Del estudio de su evolución a lo largo del tiempo, patrones de estacionalidad, etc es como se obtendrán las variables exógenas.

## **4. Extracción y preparación de los datos**

### **4.1. Introducción**

Los datos que se analizan en el presente trabajo se han extraído de 2 fuentes:

- **Demanda eléctrica:** los datos se han extraído de la API que ofrece el operador del sistema eléctrico en España, Red Eléctrica, y que se conoce como *e.sios*.
- **Temperatura:** de la web de *Copernicus Climate Data Store*.

### **4.2. Conexión con la API de E-SIOS**

Para poder extraer los datos de la API de *e.sios*, lo primero que se tiene que hacer es pedir un *token* de acceso que tiene que acompañar todas las consultas que se hagan. El *token* se solicita a través de correo electrónico a la dirección [consultasios@ree.es](mailto:consultasios@ree.es). Una vez conseguido el token, las consultas se han realizado usando la librería *requests* de Python que permite realizar la gestión de la conexión *http* y la descarga de sus datos [10], [11], [12].

La API de *e.sios* ofrece información de una amplia gama de indicadores. Los indicadores se extraen con la siguiente consulta:

<https://api.esios.ree.es/indicators>

Con la respuesta se identifica el indicador de demanda eléctrica que tiene como índice 1293. Por lo tanto, para extraer los datos de demanda eléctrica realizamos una consulta a la API de e.sios usando la siguiente url.

[https://api.esios.ree.es/indicators/1293/?start\\_date='+start\\_date+'&end\\_date='+end\\_date+'&time\\_trunc=\[VARIABLE\\_TEMPORAL\]](https://api.esios.ree.es/indicators/1293/?start_date='+start_date+'&end_date='+end_date+'&time_trunc=[VARIABLE_TEMPORAL])

Donde *start\_date* y *end\_date* son campos de fechas con los que fijar el intervalo de tiempo para el cual se solicita información de la demanda eléctrica. Por otro lado, el campo [VARIABLE\_TEMPORAL] es un campo que puede tomar los valores *hour*, *day*, *month*, *year*, etc para fijar en qué régimen de frecuencia queremos los datos (horario, diario, mensual, anual, etc).

Para más información de cómo se han extraído los datos de demanda eléctrica, consultar [Anexo 01](#).

#### **4.3. Extracción de datos de temperatura**

Los datos de temperatura se extrajeron de la web de *Copernicus Climate Data Store*. Copernicus es una iniciativa de la Comisión Europea y de la Agencia Espacial Europea para construir un sistema autónomo de observación de la Tierra que permita la observación del medio ambiente y cómo le afectan los cambios ambientales, el origen de estos cambios y la influencia en la vida de las personas. Copernicus tiene como objetivo entonces ser parte fundamental en la protección del medio ambiente, la salud y seguridad de los ciudadanos de la Unión Europea [13].

La solicitud de datos se realiza en el siguiente enlace:

<https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=form>

Para poder solicitar datos es necesario hacerse una cuenta de usuario.

En el presente trabajo se han extraído datos de temperaturas para el período 2014-2021 de las 5 ciudades más importantes de España: Madrid, Barcelona, Bilbao, Valencia y Sevilla. Además de ser las ciudades más importantes de España, están bastante bien repartidas por su geografía, lo que

permite captar con relativamente pocos datos la diversidad de temperaturas en todo el territorio en el mismo instante.

Para solicitar los datos de cada ciudad han sido necesarios indicar su ubicación geográfica, el período de tiempo (2014-2021) y la frecuencia de toma de datos, que ha sido horaria.

Una vez obtenidos los 5 datasets de temperaturas se han juntado junto con los de demanda para formar un único dataset.

Para más información sobre el tratamiento de los archivos de temperaturas, consultar [Anexo 02](#).

## ***5. Análisis exploratorio de la demanda eléctrica***

### ***5.1. Introducción***

En el presente apartado se va a realizar un análisis pormenorizado de la serie temporal. Se empezará desde un análisis general de toda la serie, viendo qué patrones presenta la demanda eléctrica en el período 2014-2021. Después se pasará a un análisis más granular de cómo se distribuye en promedio la demanda por meses y por días, buscando otros patrones y variables que afectan a nuestra variable de estudio.

### ***5.2. Evolución y estacionariedad de la demanda eléctrica en el período de estudio***

#### ***5.2.1. Análisis de la evolución de la demanda eléctrica***

Una serie temporal de demanda eléctrica es esperable que cumpla una serie de patrones estacionales. Para ello hemos partido de una visualización general de la demanda eléctrica para los años 2014-2021, tal y como se observa en la siguiente figura:

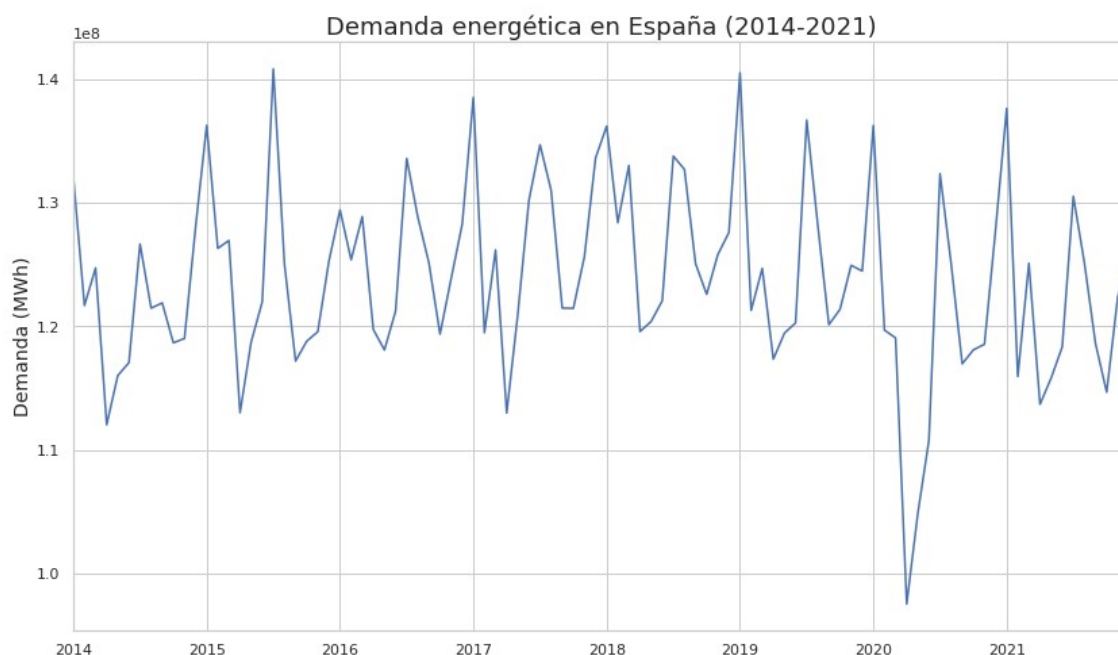


Fig. 02: Demanda energética en España (2014-2021)

La demanda de energía oscila de forma periódica a lo largo del tiempo con unos máximos y mínimos dentro de un margen bien definido entre  $(1.1-1.4) \cdot 10^8$  MWh, excepto en 2020 donde vemos una caída abrupta que coincide con la irrupción del COVID-19 en España y el inicio del confinamiento estricto que duró casi 3 meses. Después la demanda de energía remonta coincidiendo con la vuelta a cierta normalidad.

Se intuye que el pico máximo cada año se produce a principios de año y en los meses de verano debido a la mayor necesidad de energía en los hogares para calefacción o aire acondicionado, respectivamente. Por otra parte, los meses de menor consumo son los correspondientes a los meses de primavera y otoño. Esto sería un primer indicativo de cómo afecta la temperatura a la demanda eléctrica: los valores extremos implican un mayor consumo de energía. Más adelante se verá qué rangos de temperaturas provocan estas mayores demandas.

Ahora se centrará la atención en cómo se distribuye la demanda de energía por año y mes en el intervalo de tiempo tratado. La demanda anual permite ver una tendencia a lo largo del tiempo, mientras que la demanda mensual permite ver si hay estacionalidad.

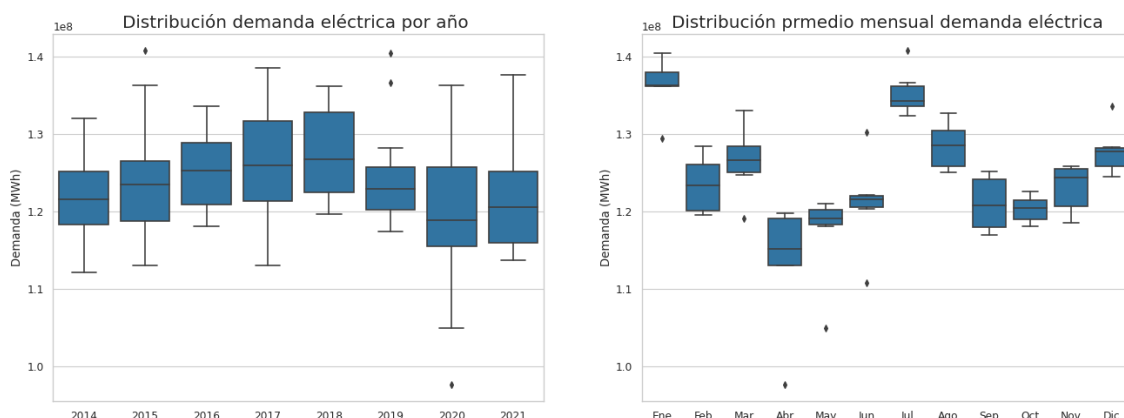


Fig. 03: Distribución promedio de la demanda energética por año (izda). Distribución promedio de la demanda energética por mes (dcha).

En el primer gráfico, se puede apreciar cómo la media de la demanda de energía cada año sube hasta llegar al año 2018, a partir del cual se produce un descenso bastante notable en 2019 para caer aún más en 2020 y volver a remontar en 2021, aunque todavía bastante por debajo de 2018 y los años anteriores.

En el segundo gráfico se confirma lo indicado anteriormente: los mayores consumos se producen en los meses de invierno (diciembre-marzo) y verano (julio-agosto), mientras que los de menor consumo se corresponden con los meses de primavera (abril-junio) y otoño (septiembre-noviembre).

Para apreciar mejor los patrones de demanda anuales se puede realizar una superposición de la demanda energética por año de forma superpuesta.

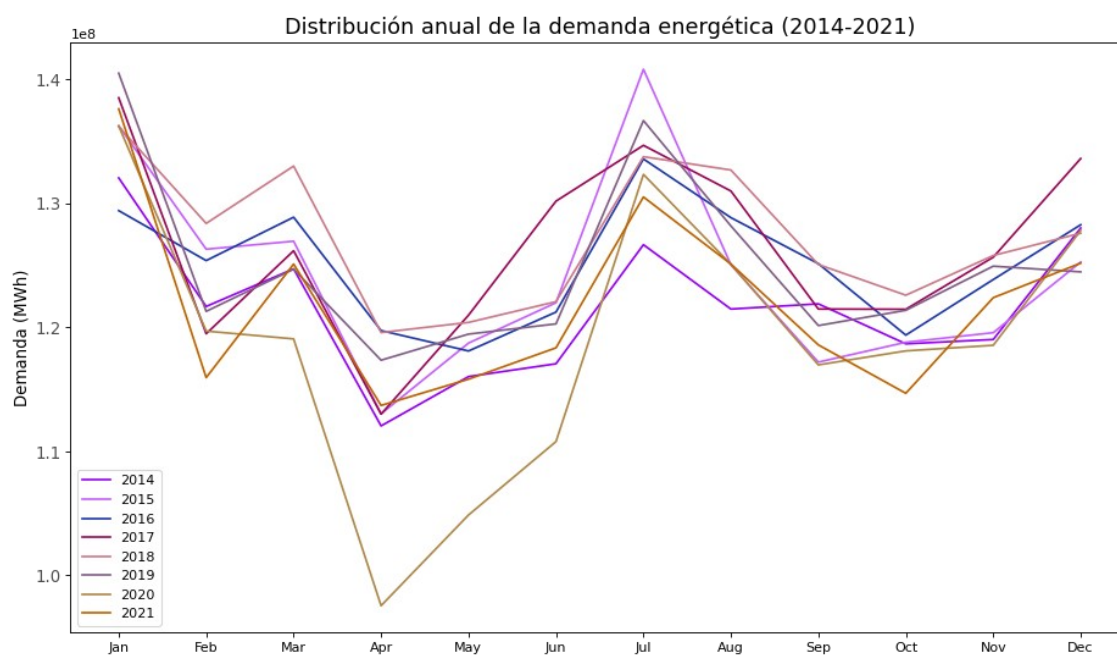


Fig. 04: Distribución superpuesta de la demanda energética período (2014-2021)

Se puede apreciar bastante bien cómo cada año se repiten los patrones de demanda de energía en los mismos meses.

Destaca 2015 con el máximo valor registrado en la serie en el mes de julio, mientras que el valor mínimo aparece en abril de 2020. Este valor mínimo está por debajo de los  $10^8$  MWh, cuando el resto de valores mínimos se encuentran entre  $(1.1-1.2) \cdot 10^8$  MWh y se debe no solo por ser abril el mes habitualmente de menor consumo, sino también por la irrupción del COVID y la adopción del confinamiento total como medida preventiva.

En 2017, una vez alcanzado el mínimo en abril, la demanda de energía crece de forma rápida manteniendo valores muy cercanos en los meses de junio, julio y agosto, lo que puede indicar que fue un año bastante caluroso y no solo en julio, como viene siendo habitual en el resto de años. En ese aspecto, también destaca 2018 por tener un consumo muy parecido entre julio y agosto.

### **5.2.2. Descomposición de la serie temporal**

De la tendencia y estacionalidad que hemos analizado antes podemos obtener conclusiones parecidas descomponiendo la señal en sus 3 componentes: tendencia, estacionalidad y ruido.



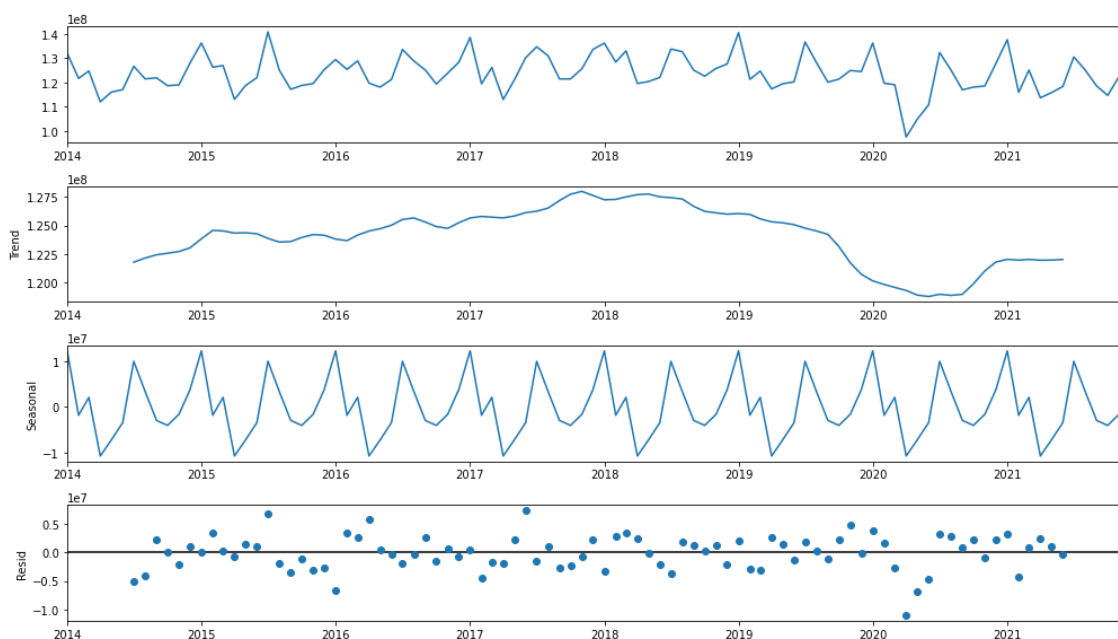


Fig. 05: Descomposición de la demanda energética (2014-2021) en sus componentes (tendencia, estacionalidad y residual)

Se puede apreciar en la primera gráfica el perfil de la demanda eléctrica para el período de estudio. La siguiente gráfica es la tendencia de la demanda donde se puede apreciar un aumento sostenido año a año hasta llegar al año 2018, desde el cual la demanda empieza a caer hasta alcanzar un mínimo en los primeros meses de 2020 y volver a remontar en 2021.

Ya se explicó que la caída en la demanda de energía en 2020 se debe a la aparición del COVID y el inicio del confinamiento, pero lo indicado en el 2018 como el año de mayor consumo podría ser indicativo de que este año marca el inicio de un período de contracción económica.

La tercera gráfica describe el carácter estacional de la demanda eléctrica donde se pueden apreciar los máximos de consumo a principios de año (meses de invierno) y mediados de año (meses de verano), siendo los mínimos los correspondientes a los meses de primavera y otoño.

### **5.2.3. Estacionariedad de la demanda eléctrica**

Antes de realizar cualquier modelo de series temporales, es importante saber si es estacionaria o no, es decir, si su media y varianza no están en función del tiempo. La estacionariedad es importante porque la mayoría de métodos estadísticos para realizar análisis y predicciones se basan en la suposición de

que las propiedades estadísticas de la serie (media, varianza, correlación, etc.) son constantes en el tiempo. Para comprobarlo se va a emplear el test de Dickey-Fuller que establece como hipótesis nula ( $H_0$ ) que la serie es no estacionaria [14],[15].

Aplicando la función definida para obtener el Test de Dickey-Fuller sobre la serie obtenemos el siguiente resultado:

**ADF Test Statistic:** -1.34347257117945

**p-value:** 0.6090564874005809

**Lags Used:** 12

**Number of Observations Used:** 83

**Conclusion:** weak evidence against null hypothesis, time series has a unit root, indicating it is non-stationary

Como el p-valor es mayor que el nivel de significación (0.05), no se puede rechazar la hipótesis nula ( $H_0$ ) y la serie se tiene que considerar como no estacionaria.

#### **5.2.4. Funciones de autocorrelación (ACF) y autocorrelación parcial (PACF)**

Antes de construir los modelos se necesita saber cómo se correlacionan los datos entre sí mediante las funciones de autocorrelación (ACF) y las funciones de autocorrelación parcial (PACF). Los análisis de autocorrelación ayudan a detectar patrones ocultos y estacionalidad, así como buscar aleatoriedad [16].

ACF mide y dibuja la correlación promedio entre los datos de una serie temporal y sus valores previos para diferentes valores de retardos (lags). También mide el grado de asociación lineal que existe entre dos variables del mismo proceso estocástico.

Una PACF es similar a una ACF midiendo la correlación entre dos variables separadas por  $k$  periodos cuando no se considera la dependencia creada por los retardos intermedios existentes entre ambas. Es decir, mide la autocorrelación que existe entre dos variables separadas  $k$  períodos descontando los posibles efectos debidos a variables intermedias [17].

Gracias a ACF se pueden responder las siguientes preguntas:

- ¿Es la serie temporal ruido blanco o random?
- ¿Una observación está relacionada con su adyacente?...¿y la siguiente? ¿y la siguiente?
- ¿La serie temporal se puede describir con un modelo MA? ¿de qué orden?

Mientras que PACF:

- ¿La serie temporal se puede modelar con un modelo AR? ¿de qué orden?

Por lo tanto, el valor de un ACF y PACF en el primer retraso es igual porque ambos miden la correlación entre datos en un tiempo  $t$  con puntos atrasados un tiempo  $t-1$ . Sin embargo, en el segundo retraso, la ACF mide la correlación entre los puntos en un instante  $t$  con los puntos en un instante  $t-2$ , mientras que PACF mide la misma correlación, pero después de controlar la correlación entre los puntos en un instante  $t$  con los que están en  $t-1$ .

Gracias a ACF y PACF podemos entender la dinámica temporal de una serie temporal individual.

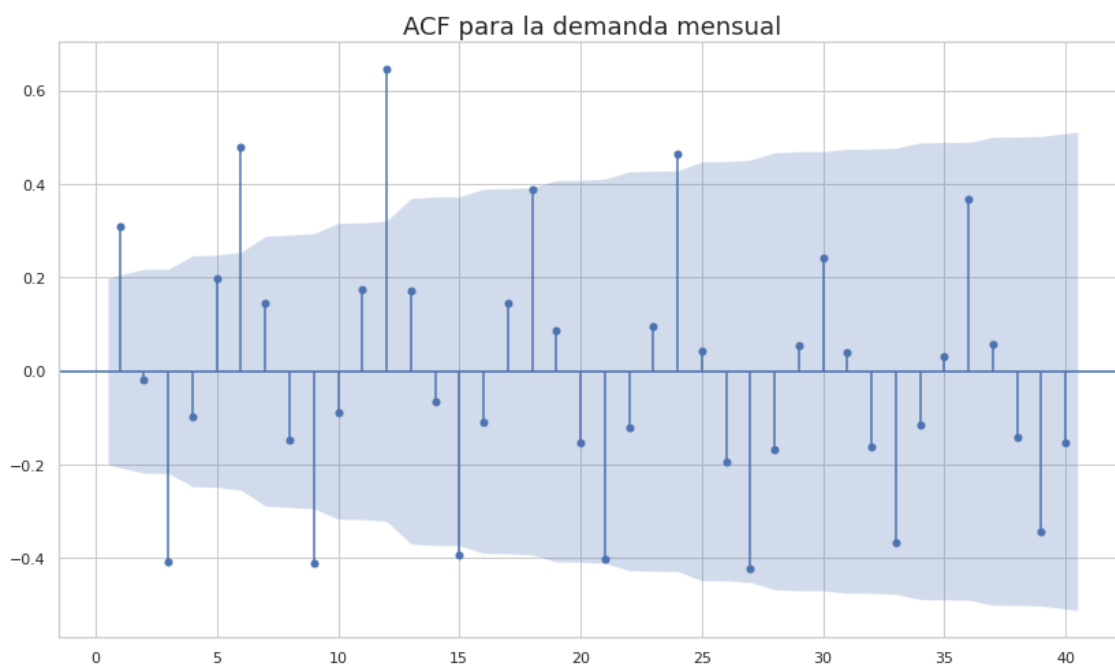


Fig. 06: Autocorrelación para la distribución mensual de la demanda energética (2014-2021)

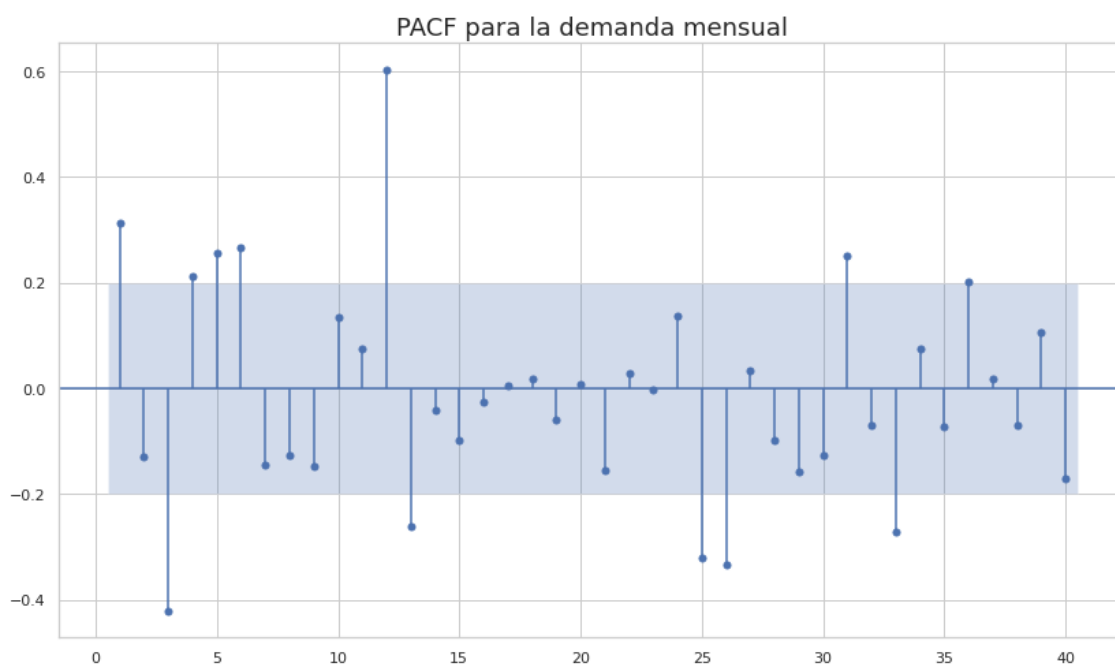


Fig. 07: Autocorrelación parcial para la distribución mensual de la demanda energética (2014-2021)

Los valores dentro del área azul en ambas gráficas indican que son estadísticamente cercanos a cero, mientras que sí que es estadísticamente significativo todo lo que hay fuera.

En ACF se puede apreciar una fuerte correlación positiva con el primer y tercer valor. Más adelante se pueden ver fuertes correlaciones negativas y positivas cada 3 valores (3, 6, 9, etc) destacando 12 y 24.

En PACF, por otro lado, hay una fuerte correlación positiva con el valor 1, negativa en el valor 3, en los valores 4, 5 y 6 y de nuevo destaca el valor 12. La fuerte correlación con 12, sus múltiplos y divisores, se puede explicar por el marcado período de 12 meses que tiene la serie.

En resumen, tanto para ACF como PACF, la fuerte correlación de un valor con los anteriores es un indicativo de que un modelo autorregresivo puede explicar bien la serie.

### **5.3. Evolución de la demanda eléctrica**

#### **5.3.1. Evolución de la demanda eléctrica diaria**

Para ver cómo evoluciona y se distribuye la demanda eléctrica a lo largo de los días, se seleccionan 2 semanas aleatorias (14-01-2019 - 28-01-2019) para ver si hay diferencias en la demanda energética según sea el día laboral de la semana.

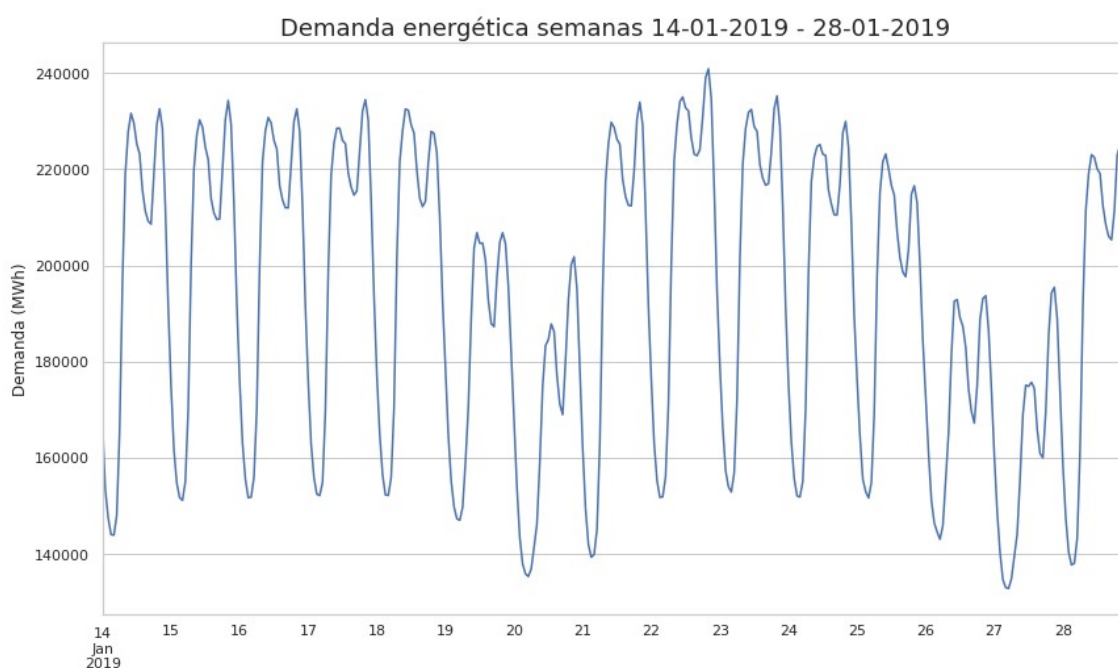


Fig. 08: Distribución diaria de la demanda energética (semanas 14-01-2019 - 28-01-2014)

Se puede observar claramente cómo los patrones de demanda de energía se repiten día a día y a lo largo de la semana.

De lunes a viernes (días laborables) el patrón de demanda y sus valores son casi los mismos, para luego decaer en sábado y domingo (días no laborables), siendo el domingo el día de menor demanda energética.

Para ver si de verdad afectan o no los festivos a la demanda energética semanal, el estudio de la Semana Santa es bastante ilustrativo, ya que siempre hay dos días laborables (jueves y viernes) que se vuelven festivos.

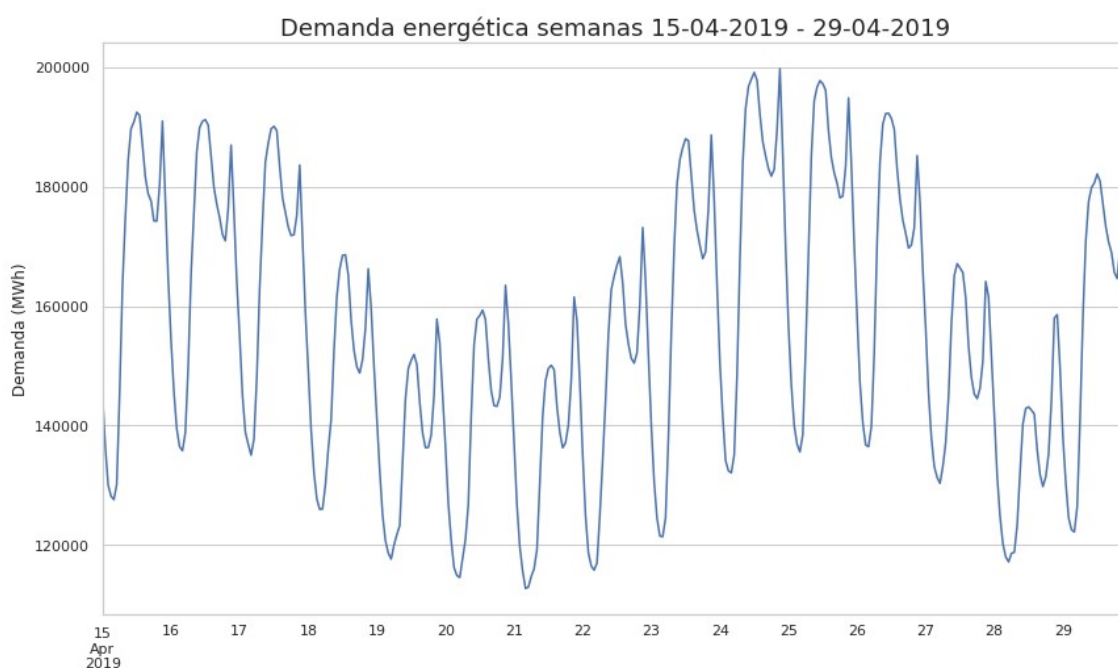


Fig. 09: Distribución diaria de la demanda energética (semanas 15-04-2019 - 29-04-2019)

De lunes a miércoles el consumo es casi el mismo, para luego decaer durante 4 días (jueves, viernes, sábado y domingo) por ser festivos, para luego volver a aumentar durante la semana siguiente a patrones habituales: 5 días de demanda elevada, 2 días de demanda reducida.

A nivel global el cálculo de la demanda media energética según el día sea laboral o festivo permite ver si de verdad en los días festivos la demanda energética es menor o no.

<b><i>Tipo de día</i></b>	<b><i>Demanda media (MWh)</i></b>	<b><i>Porcentaje (%)</i></b>
Festivo	145177.96	46
Laborable	170217.51	54

Tabla 1: Demanda energética media para días laborables y festivos

Por lo tanto, se puede apreciar cómo la demanda promedio un día laborable es un 8% mayor que en un día festivo.

Ahora se pasa a ver a nivel global cómo es la distribución de la demanda energética según el día de la semana.

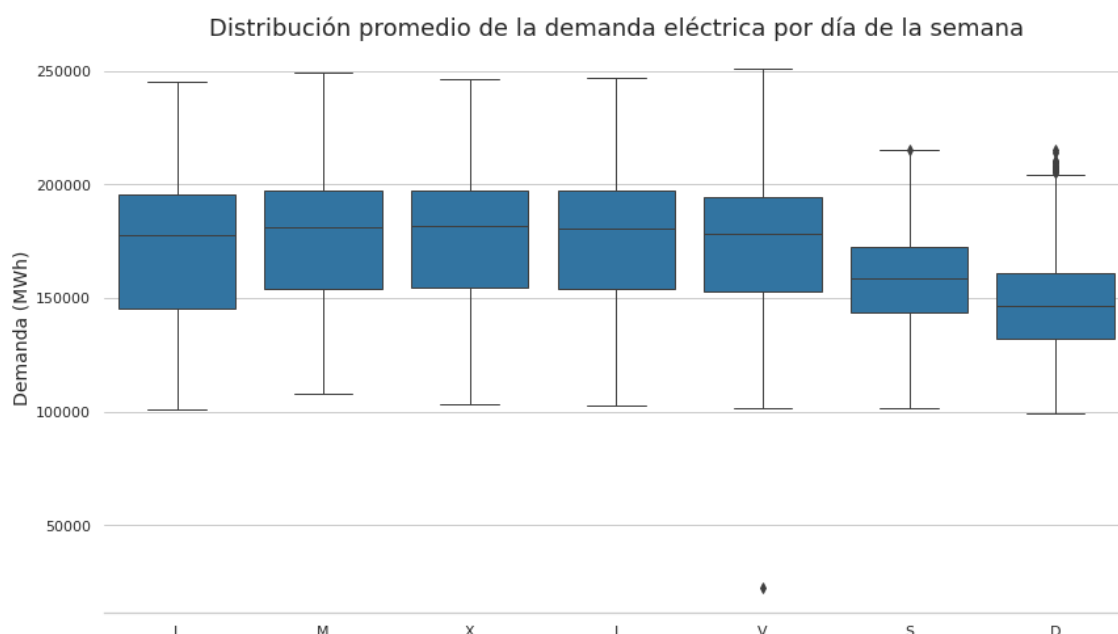


Fig. 10: Distribución promedio de la demanda energética por día de la semana (2014-2021)

Se puede ver cómo los días usualmente laborables (de lunes a viernes) concentran la mayor demanda energética, mientras que ésta decae los fines de semana (sábado y domingo), siendo el domingo el día de menor demanda. Martes, miércoles y jueves son los que tienen una mayor media y cabe destacar el viernes como el día de mayor variabilidad teniendo los picos más extremos de mayor y menor demanda.

### **5.3.2. Evolución de la demanda eléctrica mensual**

Centramos de nuevo la atención en la distribución promedio de la demanda eléctrica a lo largo de cada mes.

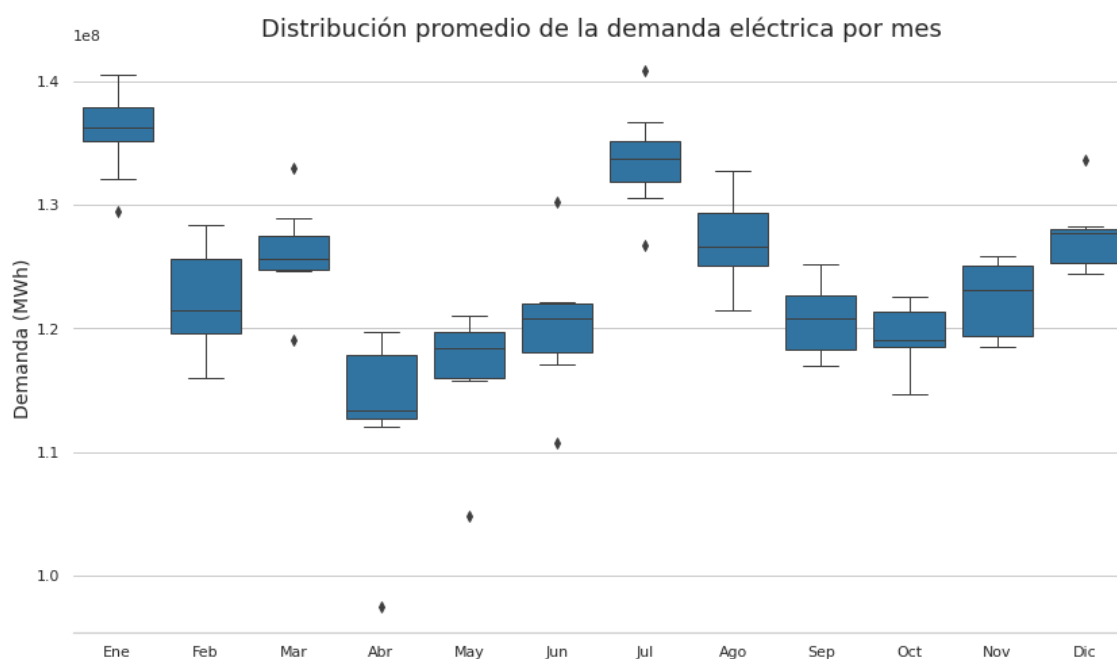


Fig. 11: Distribución promedio mensual de la demanda energética período (2014-2021)

Como se explicó en el apartado *5.2.1.- Análisis de la evolución de la demanda eléctrica*, los meses de mayor consumo promedio de energía son los meses de invierno (diciembre-marzo) y verano (julio-agosto), mientras que los de menor consumo se corresponden con los meses de primavera (abril-junio) y otoño (septiembre-noviembre). En concreto son los meses de enero y julio los de mayor consumo promedio, mientras que abril, mayo y octubre los de menor consumo.

Esto refleja la influencia que tiene la temperatura en la demanda para valores extremos. En el siguiente apartado vamos a ver cómo se materializa esa influencia.

### **5.3.3. Influencia de la temperatura en la demanda eléctrica**

Para entender la influencia de la temperatura en la demanda eléctrica, se van a comparar los gráficos demanda y temperatura media durante el intervalo de un año.



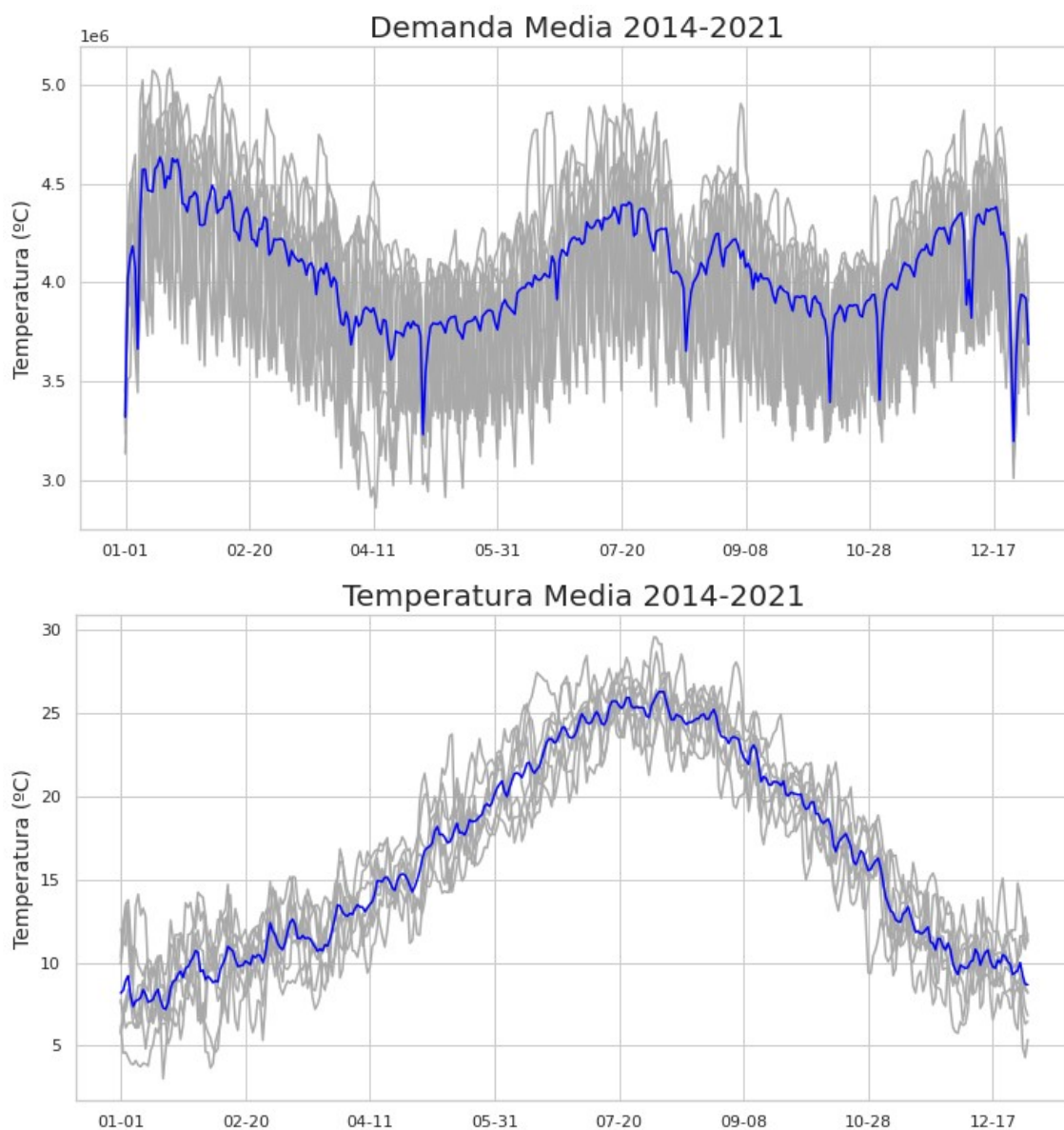


Fig. 12: Comparativa entre demanda energética media y temperaturas medias (2014-2021)

Se puede confirmar la influencia que ejerce la temperatura en la demanda energética: para bajas temperaturas (meses de invierno) y para meses de altas temperaturas (meses de verano) la demanda crece considerablemente, mientras que en los meses de otoño y primavera la demanda se reduce.

En los períodos extremos de invierno y verano podemos ver que se producen las mayores demandas, pero a la par grandes caídas. Esto se debe a que ambos períodos coinciden con festivos o con la costumbre generalizada de tomar vacaciones en esos espacios de tiempo.

Por último, lo importante de esta comparativa es el poder establecer que los rangos de temperatura que provocan un aumento de la demanda están por debajo de los 15°C y por encima de los 20°C.

Por lo tanto, para reflejar mejor en el dataset esta influencia de la temperatura para valores extremos, se van a añadir 2 columnas al dataset que notifiquen si el día y mes de cada registro está en verano o en invierno.

#### ***5.4. Incorporación de variables exógenas***

Después de lo comentado en el análisis exploratorio sobre los factores que afectan a la demanda eléctrica, se han escogido para los modelos las siguientes variables exógenas que deben ayudar a realizar mejores predicciones:

- Día de la semana (0-6 lunes-domingo)
- Días laborables (1) / Festivos (0)
- Temperatura
- Estacionalidad (invierno / verano)

De la extracción de la temperatura se ha hablado en el apartado anterior. Los días de la semana se han obtenido a partir de las fechas de los registros de la base de datos, mientras que los días festivos en el Estado Español se han recopilado para el período 2014-2021 y después se han contrastado con las fechas que disponemos. Por último, con la estacionalidad, se ha procedido de forma similar a los anteriores: si las fechas caen o no en invierno se marcan como 1 o 0 en la columna *is\_winter* y lo mismo para la columna *is\_summer* en caso del verano.

	value	weekday	working_day	madrid_temp	barcelona_temp	bilbao_temp	sevilla_temp	valencia_temp	is_winter	is_summer
date										
2014-01-01	3315035.0	2.0	0.0	6.785193	8.006851	11.418970	12.184413	13.189499	1	0
2014-01-02	4059590.0	3.0	1.0	9.109185	10.764683	12.935602	14.524325	15.136449	1	0
2014-01-03	4098986.0	4.0	1.0	9.776889	11.935958	13.029657	15.653018	16.565861	1	0
2014-01-04	3867911.0	5.0	1.0	6.703811	11.307489	9.814500	14.106683	14.699295	1	0
2014-01-05	3648038.0	6.0	1.0	5.285333	8.682916	9.971085	10.653680	10.970076	1	0
...	...	...	...	...	...	...	...	...	...	...
2021-12-27	3686647.0	0.0	1.0	11.606002	13.531011	14.285517	17.009720	17.443636	1	0
2021-12-28	3729169.0	1.0	1.0	9.792129	14.261092	14.133620	16.015375	16.007968	1	0
2021-12-29	3692921.0	2.0	1.0	8.401157	13.250848	12.936356	13.320164	15.173263	1	0
2021-12-30	3689439.0	3.0	1.0	7.203344	11.992122	12.347234	12.666560	12.678245	1	0
2021-12-31	3331056.0	4.0	1.0	8.081628	11.068301	12.862481	13.293043	11.608819	1	0

2922 rows × 10 columns

Fig. 13: Base de datos resumen

Para más información de cómo se han incorporado las variables exógenas al dataset de demanda eléctrica, consultar [Anexo 01](#).

## 6. Modelos de series temporales

### 6.1. Introducción

Una vez obtenidos y analizados los datos de demanda eléctrica y los factores que pueden influir sobre ella, se va a modelizar a través de diferentes técnicas para ver cómo afectan cada uno de estos factores sobre la precisión de los modelos.

Las técnicas de modelización empleadas han sido las siguientes:

- *Holt-Winters*
- *SARIMAX*
- *XGBoost*
- *Prophet*

El motivo de escoger estas técnicas ha sido la de probar si afectan por igual los factores exógenos a la hora de hacer predicciones entre algoritmos más simples y avanzados .

*Holt-Winters* se ha empleado como modelo de partida sobre el que se ha aplicado el dataset de la forma más sencilla posible: usando sólo datos de demanda a lo largo del tiempo y de ahí comparar con el resto de modelos y cómo cambian sus métricas con la adición de factores exógenos.

*Holt-Winters*, *SARIMAX* y *Prophet* son algoritmos específicos para series temporales, mientras que *XGBoost* es una técnica de clasificación que se ha visto que también tiene aplicaciones bastante buenas en predicción de series temporales. La selección de *Holt-Winters* y *SARIMAX* se debe a que son técnicas más sencillas e interesa saber si los resultados que ofrecen difieren mucho o no de algoritmos más avanzados como *Prophet* y *XGBoost*.

## **6.2. Holt- Winters como modelo de partida**

El método de *Holt-Winters* pertenece al conjunto de técnicas de suavizado exponencial y es una ampliación perfeccionada del método de suavización exponencial de Holt que considera solo dos exponentes suavizantes. *Holt-Winters* considera tres exponentes suavizantes (nivel, tendencia y estacional) de una serie temporal dada. A diferencia de otras técnicas de suavizado exponencial, *Holt-Winters* puede adaptarse fácilmente a cambios y tendencias, así como a patrones de estacionalidad [\[18\]](#), [\[19\]](#), [\[20\]](#), [\[21\]](#).

Según el tipo de estacionalidad tenemos dos modelos:

- **Modelo Multiplicativo estacional:** este modelo supone que a medida que aumentan los datos, también aumenta el patrón estacional en la serie. La gran mayoría de series temporales muestran este patrón. En este modelo la tendencia y los componentes de estación se multiplican y luego se suman al componente de error.
- **Modelo Aditivo Estacional:** en este modelo los efectos de los factores individuales se diferencian y se agrupan para modelar los datos. Este método es el mejor para datos con tendencia y estacionalidad que no aumentan a lo largo del tiempo. Su resultado es una previsión de curva que muestra los cambios estacionales en los datos.

En [Anexo 03](#) se puede consultar en detalle cómo se ha creado el modelo, proceso de entrenamiento y test en el apartado *Holt-Winters (Baseline)*.

### 6.3. SARIMAX

*SARIMAX* es el acrónimo de *Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors* que se traduciría como: Media móvil integrada autorregresiva estacional con factores exógenos.

*SARIMAX* supone una versión mejorada de los modelos *ARIMA* y *SARIMA* que incorpora tanto estacionalidad como factores exógenos a la elaboración del modelo. Otra ventaja es que a diferencia de *ARIMA*, *SARIMAX* se puede aplicar sobre series temporales no estacionarias, como ocurre en nuestro caso [\[22\]](#).

Los parámetros de entrada del modelo serían los siguientes:

```
model_sarimax = SARIMAX(  
    train / test data,  
    exo train/ test,  
    order=(3, 2, 3),  
    seasonal_order=(3, 0, 3, 4)  
)
```

El primer parámetro serían los datos de entrenamiento/test, el segundo parámetro los datos exógenos; el parámetro *order* = (*p*, *d*, *q*) está formado por tres variables: *p* es el orden del componente AR no estacional, *d* el orden diferenciado no estacional y *q* el orden del componente MA no estacional; el parámetro *seasonal\_order* = (*P*, *D*, *Q*, *s*) está formado por 4 variables: *P* es el orden del componente RA estacional, *D* es el orden de integración de estacionalidad de las series temporales, *Q* es el orden del componente estacional MA y *s* es la duración de la estacionalidad [\[23\]](#).

Debido al amplio rango de valores y combinaciones de los parámetros, para obtener el mejor modelo posible que se ajuste a la serie temporal se ha definido un modelo iterativo que genera modelos *SARIMAX*, los compara entre sí por el valor de AIC [\[24\]](#) siendo el mejor modelo el que tiene menor valor de AIC.

En [Anexo 04](#) se puede consultar cómo se ha realizado el proceso de optimización de parámetros según cada escenario.

#### **6.4. XGBoost**

*XGBoost* es el acrónimo de *eXtreme Gradient Boosting* (Aumento del Gradiente Extremo) es un algoritmo de machine learning de tipo supervisado más popular a día de hoy [\[25\]](#).

Su popularidad se debe al poco esfuerzo relativo que necesita para obtener buenos resultados de predicción compitiendo muy bien frente a modelos complejos computacionalmente, sobre todo en lo relativo a tratar con datos heterogéneos.

El principio de *boosting* que emplea se basa en la generación de múltiples modelos de predicción que se pueden calificar como débiles secuencialmente, pero con la particularidad de que cada uno de esos modelos tome los resultados del anterior generando un modelo cada vez más fuerte en cuanto a mejor poder predictivo y resultados más estables. Los modelos débiles de partida son árboles de decisión de diferentes tipos que se pueden emplear en tareas de clasificación y regresión.

El algoritmo de optimización entre modelos es el del descenso del gradiente [\[26\]](#). El proceso para cuando se alcanza un número máximo de iteraciones o cuando la diferencia entre modelos es insignificante.

Para más detalles de la construcción del modelo, entrenamiento y testeo se pueden ver en el [Anexo 05](#) y en el apartado *XGBoost* del [Anexo 03](#).

#### **6.5. Prophet**

Prophet es una herramienta de código abierto creada por Facebook para el tratamiento y modelado de series temporales que se basa en un modelo aditivo descomponible donde las tendencias no lineales se ajustan a la estacionalidad, además de a eventos como las vacaciones y sus efectos en una serie [\[27\]](#).

Resulta muy útil para series que abarquen períodos de tiempo amplios con frecuencias de hora, día, semana, mes, etc; también para series que puedan llegar a tener varias estacionalidades muy marcadas, eventos irregulares, que les falten datos, así como series con tendencias de crecimiento no lineal que se aproximen a un límite [\[28\]](#).

La expresión en la que se basa el modelo es la siguiente:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

Donde  $g(t)$  es la función que modeliza la tendencia del modelo, es decir, el comportamiento a largo plazo;  $s(t)$  modeliza la estacionalidad del modelo mediante series de Fourier;  $h(t)$  modela el efecto que tienen eventos como las vacaciones, ofertas, etc; y  $\epsilon_t$  representa el término del error irreducible [29].

Para más detalles de la construcción del modelo, entrenamiento y testeo se pueden ver en el [Anexo 06](#) y en el apartado *Prophet* del [Anexo 03](#).

### **6.6. Métricas escogidas**

A la hora de comparar modelos entre sí aplicados a una misma serie temporal es importante saber que la selección puede afectar no solo a la optimización de los modelos, sino también al juicio e interpretación de los mismos.

Cada vez son más las métricas de errores que se están proponiendo para el estudio de series temporales. Hay que tener en cuenta que no hay métrica mejor que otra, ya que cada una condensa información sobre una gran cantidad de datos en un solo valor, de forma que solo da una proyección de los errores de los modelos haciendo énfasis en un aspecto determinado de los errores de sus características [30].

Hay 3 grandes grupos de métricas de errores para series temporales:

- **Dependientes de la escala:** métricas que vienen dadas en las mismas unidades que los datos que se usan. Su principal ventaja es que son fáciles de calcular y de interpretar, pero no se pueden usar para comparar diferentes series por la dependencia que tienen de las unidades. En este grupo entrarían métricas como el Error Absoluto Medio (MAE por sus siglas en inglés), el Error Cuadrático Medio (MSE) y la Raíz del Error Cuadrático Medio (RMSE).
- **Errores porcentuales:** dado que los errores dependientes de la escala no permiten la comparación entre diferentes series temporales, los errores porcentuales sí que permiten resolver este problema. Sin embargo, tienen el punto débil de que no admiten valores iguales a cero, ya que

convierten a este tipo de errores en infinito o indefinido, lo que no permite interpretarlos. En este grupo de errores entrarían el Error Porcentual Absoluto (MAPE) y el Error Porcentual Absoluto Simétrico (sMAPE).

- **Errores libres de escala:** este tipo de errores viene a subsanar el problema de los errores dependientes de escala. En este grupo se encuentra el Error Absoluto Medio Escalado (MASE) que viene a mejorar el error MAE.

En el presente trabajo se han seleccionado las siguientes métricas:

- **MAE (Mean Absolute Error):** se calcula calculando la media de las diferencias absolutas entre los valores reales de la serie y los predichos por el modelo. Este cálculo tan simple es su principal ventaja, pero, además de no permitir comparar series temporales diferentes, no se debe usar si se quieren penalizar los valores atípicos.
- **MSE (Mean Squared Error):** esta métrica nos sirve si se quiere poner atención en los valores atípicos. Se calcula como la media del cuadrado del MAE. Debido a que se eleva al cuadrado, da más peso a los errores grandes que a los pequeños, lo que puede ser una desventaja en algunas situaciones. Por lo tanto, el MSE es apropiado para situaciones en las que quieres poner el foco en errores grandes. Indicar que al estar elevado al cuadrado hace que pierda sus unidades.

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

- **RMSE (Root Mean Squared Error):** para evitar la pérdida de unidades del MSE, podemos tomar su raíz cuadrada. Tiene las mismas ventajas que el MAE y el MSE, sin embargo, al igual que el MSE, es sensible a los valores atípicos.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$



- **MAPE (Mean Absolute Percentage Error):** es una de las métricas más populares en series temporales. Se calcula como la media de la diferencia absoluta entre los valores reales y predichos dividido por los valores reales. La principal ventaja de MAPE es su independencia de las escalas de las unidades, así como su fácil interpretabilidad, lo que permite usarlo para comparar modelos de diferentes series. Sin embargo, MAPE tiene algunas desventajas como generar infinitos o valores indefinidos para valores iguales a cero o cercanos a cero. También penaliza fuertemente sobre errores negativos más que sobre errores positivos, lo que genera una asimetría. Por último, MAPE no se puede usar cuando el uso de porcentajes no tiene sentido, como en el caso de series de valores de temperaturas, en las que puede haber valores negativos y nulos.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| * 100$$

- **sMAPE (Symmetric Mean Absolute Percentage Error):** esta métrica surge para evitar la asimetría de MAPE. Hay un debate tanto en qué expresión matemática debería definir este error, como si de verdad es simétrico, tal y como indica su nombre. El sMAPE es la media a lo largo de todas las predicciones hechas hasta un determinado límite. Tiene como ventajas que no tiene los mismos problemas que MAPE cuando los valores de la serie son cercanos a cero.

$$sMAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|)}$$

- **MASE (Mean Absolute Scaled Error):** se calcula tomando el MAE y dividiéndolo por el MAE de una submuestra. Los valores de MASE por encima de 1 indican que las predicciones son peores, en promedio, que en la submuestra escogida. Dado que MASE no depende de escalas, se puede usar para comparar modelos de series temporales diferentes. Sin embargo, hay controversia con esta métrica porque algunos estudios apuntan a que introduce un sesgo por ser sensible a valores atípicos.

$$MASE = \frac{MAE}{MAE_{in-sample, naive}}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

### **6.7. Variables exógenas y escenarios escogidos**

En el apartado 3.4 se plantearon las siguientes variables exógenas y cómo se traducen en el dataset:

- **Día de la semana (weekday):** se refleja como una columna tipo *int* con valores que van del 0 (lunes) al 6 (domingo).
- **Días laborables y festivos (working\_day):** se refleja como una columna de tipo *boolean* en el que un día laboral se refleja como verdadero (1) y un festivo como falso (0).
- **Temperatura de cada ciudad (ciudad\_temp):** 5 columnas de tipo *float* donde se refleja la temperatura media por día de cada ciudad considerada.
- **Estacionalidad (is\_winter / is\_summer):** 2 columnas de tipo *boolean* que según el día del dataset esté o no en invierno/verano se marca como verdadero/falso (1/0).

Para probar cómo las variables exógenas afectan a los modelos se han creado 4 escenarios diferentes:

- **A:** sin variables exógenas
- **B:** día de la semana y día laboral/festivo.
- **C:** día de la semana, laboral/festivo y temperatura.
- **D:** día de la semana, laboral/festivo, temperatura y estacionalidad (invierno, verano).

### 6.8. Comparación entre modelos

Comenzamos con las métricas de Holt-Winters que es el modelo que vamos a utilizar como modelo base (baseline) sobre el que comparar con los demás casos.

<b>Métrica</b>	<b>Holt-Winters</b>
<b>MAE</b>	4.36e+6
<b>MSE</b>	2.61e+11
<b>RMSE</b>	5.11e+6
<b>MAPE (%)</b>	11.39
<b>sMAPE</b>	10.56
<b>MASE</b>	1.79

Tabla 2: Resumen de métricas del modelo Holt-Winters

#### 6.8.1. SARIMAX

<b>SARIMAX</b>				
<b>Métrica</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
<b>MAE</b>	3.2e+6	2.37e+6	2.49e+6	2.5e+6
<b>MSE</b>	1.35e+13	1.03e+11	1.19e+11	1.20e+11
<b>RMSE</b>	3.67e+6	3.21e+6	3.44e+6	3.46e+6
<b>MAPE (%)</b>	81.59	5.86	6.12	6.16
<b>sMAPE</b>	126.33	5.94	6.24	6.28
<b>MASE</b>	13.37	0.97	1.02	1.03

Tabla 3: Resumen de métricas del modelo SARIMAX

Podemos ver cómo el escenario A para *SARIMAX* es pésimo comparado con las métricas de los otros escenarios. Destacan especialmente las métricas MSE, sMAPE y MASE que tienen valores enormes. Esto es comprensible ya que *SARIMAX* está pensado para tener en cuenta variables exógenas. Todas las métricas, excepto MSE, son mejores que en *Holt-Winters*.

En cuanto a los escenarios B, C y D podemos ver cómo mejoran las métricas bastante con respecto al escenario A. Podemos ver cómo el escenario B (día de la semana y día laboral/festivo) es el que mejores métricas tiene de los cuatro escenarios, seguido de los escenarios C y D. Por lo tanto, podemos decir que la adición de variables exógenas más allá del día de la semana y festivos, penaliza negativamente las métricas del modelo.

### 6.8.2. XGBoost

<b>XGBoost</b>				
<b>Métrica</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
<b>MAE</b>	2.8e+6	2.8e+6	2.8e+6	2.8e+6
<b>MSE</b>	1.12e+11	1.12e+11	1.13e+11	1.13e+11
<b>RMSE</b>	3.35e+6	3.35e+6	3.37e+6	3.37e+6
<b>MAPE (%)</b>	7.12	7.12	7.21	7.21
<b>sMAPE</b>	6.99	6.99	7.06	7.06
<b>MASE</b>	0.61	0.61	0.61	0.61

Tabla 4: Resumen de métricas del modelo *XGBoost*

En todos los escenarios las métricas son mejores que las del modelo base *Holt-Winters*.

Podemos ver en este caso unas métricas muy parecidas entre los diferentes escenarios, siendo los mejores los escenarios A y B, lo que indica que *XGBoost* funciona bien con pocas o ninguna variable exógena. Esto es comprensible por lo comentado de cómo funciona *XGBoost* para encontrar el mejor modelo posible para unos datos dados.

Por último decir, que *XGBoost*, como en *SARIMAX*, responde bien a la adición de días de la semana y festivos y laborables en el conjunto de datos. Por otro lado, temperatura y estacionalidad penalizan negativamente, pero a diferencia de *SARIMAX*, esta penalización es mínima por cómo *XGBoost* busca el mejor modelo posible.

### 6.8.3. Prophet

<b>PROPHET</b>				
<b>Métrica</b>	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
<b>MAE</b>	2.56e+6	2.45e+6	2.33e+6	2.42e+6
<b>MSE</b>	8.93e+10	8.28e+10	7.62e+10	8.24e+10
<b>RMSE</b>	2.99e+6	2.88e+6	2.76e+6	2.87e+6
<b>MAPE (%)</b>	6.37	6.08	5.77	5.99
<b>sMAPE</b>	6.61	6.33	6.00	6.24
<b>MASE</b>	1.05	1.01	0.96	1.00

Tabla 5: Resumen de métricas del modelo *PROPHET*

En este caso, como los anteriores modelos, tenemos mejores métricas que en el caso de *Holt-Winters*.

Podemos ver cómo la adición de variables exógenas ayuda a mejorar las métricas en todos los escenarios, menos en el D, donde la adición de la estacionalidad lejos de ayudar, empeora ligeramente el modelo. Por lo tanto, para Prophet, a diferencia del resto de modelos, la adición de la temperatura junto con los días de la semana, festivos y laborables, ayuda a la mejora de la precisión del modelo (escenario C). Sin embargo, penaliza la adición de la estacionalidad en todas las métricas (escenario D).

## 6.9. Comparación de métricas entre modelos

A continuación se resumen los valores de cada métrica en cada escenario para los diferentes modelos

### 6.9.1. MAE

<b><u>Escenario</u></b>	<b><i>Holt-Winters</i></b>	<b><i>SARIMAX</i></b>	<b><i>XGBoost</i></b>	<b><i>Prophet</i></b>
<b>A</b>	4.36e+6	3.2e+6	2.8e+6	2.56e+6
<b>B</b>	-	2.37e+6	2.8e+6	2.45e+6
<b>C</b>	-	2.49e+6	2.8e+6	2.33e+6
<b>D</b>	-	2.5e+6	2.8e+6	2.42e+6

Tabla 6: Resumen valores de MAE en cada modelo y escenario considerados.

Todos los modelos tienen un valor de MAE mejor que el del modelo base *Holt-Winters*. *XGBoost* tiene todos los valores iguales en cada escenario y nunca es mejor que los de *SARIMAX* y *Prophet*.

*SARIMAX* obtiene el mejor valor de MAE en el escenario B, mientras que *Prophet* en los escenarios A, C y D.

### 6.9.2. MSE

<b><u>Escenario</u></b>	<b><i>Holt-Winters</i></b>	<b><i>SARIMAX</i></b>	<b><i>XGBoost</i></b>	<b><i>Prophet</i></b>
<b>A</b>	2.61e+11	1.35e+13	1.12e+11	8.93e+10
<b>B</b>	-	1.03e+11	1.12e+11	8.28e+10
<b>C</b>	-	1.19e+11	1.13e+11	7.62e+10
<b>D</b>	-	1.20e+11	1.13e+11	8.24e+10

Tabla 7: Resumen valores de MSE en cada modelo y escenario considerados.

En el caso del MSE, todos los modelos obtienen un mejor valor que el del modelo base, excepto para *SARIMAX* en el escenario A.

En todos los escenarios, *Prophet* obtiene el mejor valor (menor valor) de la métrica. Entre *SARIMAX* y *XGBoost*, *SARIMAX* consigue el mejor valor solo en el escenario B.

### 6.9.3. RMSE

<b><i>Escenario</i></b>	<b><i>Holt-Winters</i></b>	<b><i>SARIMAX</i></b>	<b><i>XGBoost</i></b>	<b><i>Prophet</i></b>
<b>A</b>	5.11e+6	3.67e+6	3.35e+6	2.99e+6
<b>B</b>	-	3.21e+6	3.35e+6	2.88e+6
<b>C</b>	-	3.44e+6	3.37e+6	2.76e+6
<b>D</b>	-	3.46e+6	3.37e+6	2.87e+6

Tabla 8: Resumen valores de RMSE en cada modelo y escenario considerados.

También en este caso todos los modelos tienen mejores métricas que *Holt-Winters*. En todos los casos *Prophet* es la que obtiene mejores valores. Entre *SARIMAX* y *XGBoost*, *SARIMAX* solo consigue mejor valor en el escenario B.

### 6.9.4. MAPE

<b><i>Escenario</i></b>	<b><i>Holt-Winters</i></b>	<b><i>SARIMAX</i></b>	<b><i>XGBoost</i></b>	<b><i>Prophet</i></b>
<b>A</b>	11.39	81.59	7.12	6.37
<b>B</b>	-	5.86	7.12	6.08
<b>C</b>	-	6.11	7.21	5.77
<b>D</b>	-	6.16	7.21	5.99

Tabla 9: Resumen valores de MAPE en% en cada modelo y escenario considerados.

Obtenemos valores mejores que el modelo base, excepto en *SARIMAX* en el escenario A.

*SARIMAX* obtiene la mejor métrica en el escenario B, mientras que *Prophet* en los escenarios A, C y D. Los valores de *XGBoost* son bastante parecidos entre sí y obtienen los peores valores en los escenarios B, C y D.

### 6.9.5. sMAPE

<b><u>Escenario</u></b>	<b><i>Holt-Winters</i></b>	<b><i>SARIMAX</i></b>	<b><i>XGBoost</i></b>	<b><i>Prophet</i></b>
<b>A</b>	10.56	126.33	6.99	6.61
<b>B</b>	-	5.94	6.99	6.33
<b>C</b>	-	6.24	7.06	6.00
<b>D</b>	-	6.28	7.06	6.24

Tabla 10: Resumen valores de sMAPE en cada modelo y escenario considerados.

Excepto para *SARIMAX* en el escenario A, todos los modelos obtienen mejores métricas que el modelo base de *Holt-Winters*.

*SARIMAX* obtiene la mejor métrica en el escenario B, mientras que en los escenarios A, C y D es *Prophet* quien obtiene las mejores métricas. *XGBoost* vuelve a obtener valores muy parecidos entre sí, y obtiene las peores métricas excepto en el escenario A.

### 6.9.6. MASE

<b><u>Escenario</u></b>	<b><i>Holt-Winters</i></b>	<b><i>SARIMAX</i></b>	<b><i>XGBoost</i></b>	<b><i>Prophet</i></b>
<b>A</b>	1.79	13.37	0.61	1.05
<b>B</b>	-	0.97	0.61	1.01
<b>C</b>	-	1.02	0.61	0.96
<b>D</b>	-	1.03	0.61	1.00

Tabla 11: Resumen valores de MASE en cada modelo y escenario considerados.

Una vez más, excepto en el escenario A de *SARIMAX*, todos los modelos obtienen mejores métricas que el modelo base. En este caso, los mejores valores se obtienen siempre con *XGBoost*.

Entre *SARIMAX* y *Prophet*, este último obtiene las mejores métricas en el escenario A, C y D.

## 7. Aplicación de los modelos sobre datos a futuro

Al momento de escribir el presente trabajo, los datos disponibles de temperatura que se podían extraer de *Copernicus* llegaban hasta el 28 de febrero del 2022. Una vez contruidos los modelos, entrenados y testeados sobre el período 2014-2021, se ha considerado importante aplicarlos sobre datos a futuro para probar su capacidad de predicción y volverlos a comparar.

Dado que el alcance del presente trabajo no era el de definir un modelo predictivo de temperaturas que ayudase junto con el resto de variables exógenas, a predecir la demanda eléctrica a futuro, se ha considerado el predecir la demanda eléctrica sin tener en cuenta la temperatura.

Una vez formado el dataset se han entrenado los modelos para el período 2014-2022 (hasta el 28 de febrero del 2022) sin considerar la temperatura y acto seguido se han hecho las predicciones para el mes de marzo.

La comparativa de las métricas para cada modelo es la siguiente:

### 7.1. Comparativa métricas

<b><i>Escenario</i></b>	<b><i>Holt-Winters</i></b>	<b><i>SARIMAX</i></b>	<b><i>XGBoost</i></b>	<b><i>Prophet</i></b>
<b><i>MAE</i></b>	8.25e+5	3.55e+5	2.62e+5	2.69e+5
<b><i>MSE</i></b>	8.97e+11	1.61e+11	1.31e+11	1.04e+11
<b><i>RMSE</i></b>	9.47e+5	4.02e+5	3.62e+5	3.22e+5
<b><i>MAPE (%)</i></b>	20.90	8.86	7.13	7.06
<b><i>sMAPE</i></b>	24.23	9.37	6.76	6.79
<b><i>MASE</i></b>	82.24	35.45	25.94	26.85

Tabla 12: Comparativa de métricas en predicción para los 10 primeros días de marzo.

Se pueden apreciar patrones parecidos a los que se han visto a lo largo del estudio. Para empezar, ningún modelo tiene peores métricas que el modelo base Holt-Winters. Las mejores métricas están en los modelos *XGBoost* y *Prophet*, siendo mejores en *Prophet* las métricas MSE, RMSE y MAPE.



La representación de los modelos sería así:

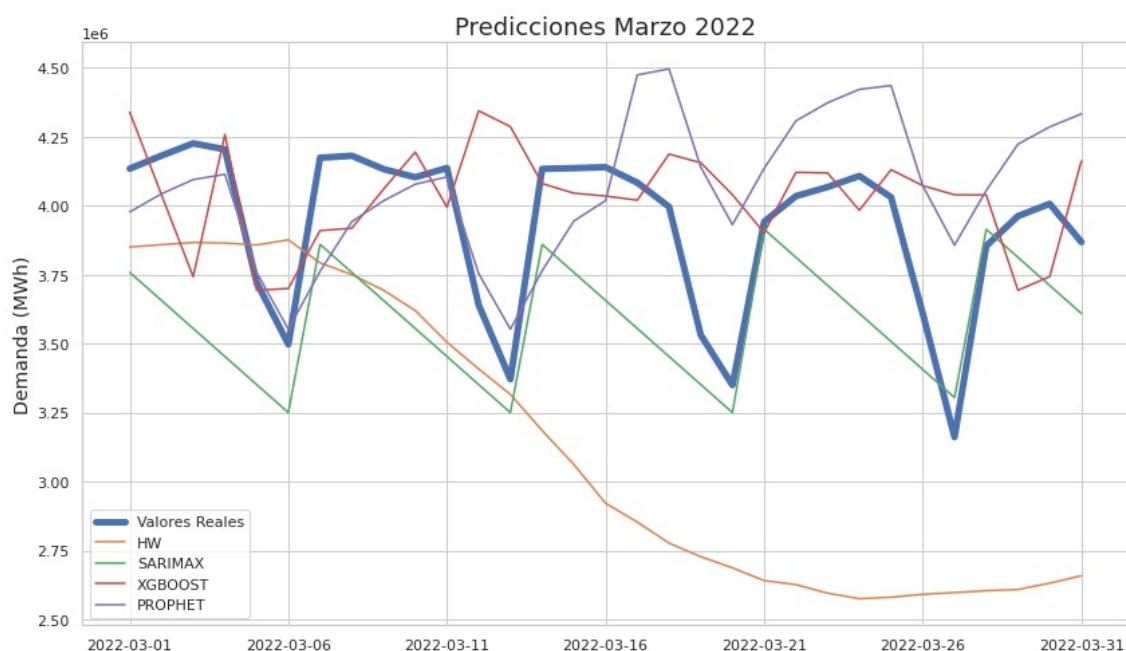


Fig. 13: Representación de los modelos para el mes de marzo 2022

Podemos apreciar cómo al principio todos los modelos, menos *Holt Winters*, siguen bastante bien la tendencia de los valores reales, pero a medida que pasa el tiempo se empiezan a desacoplar. *SARIMAX* consigue captar la estacionalidad diaria con una forma muy acusada de “sierra” que le impide captar bien otros matices de la tendencia del consumo diario. *XGBoost* y *Prophet* consiguen captar bastante bien los mínimos diarios, pero a medida que pasa el tiempo sus predicciones se alejan bastante de esos mínimos, en especial *XGBoost*. Por último, sobre *Prophet*, cabe mencionar que mantiene mejor que *XGBoost* un perfil similar al de la demanda real.

Dado que *XGBoost* y *Prophet* son los que obtienen mejores métricas es oportuno calcular los errores de sus predicciones con respecto al valor real de demanda y representarlos para ver qué modelo sería el mejor.

Se han calculado como el porcentaje obtenido de la diferencia en valor absoluto entre los valores reales y predichos, divididos por el valor real.

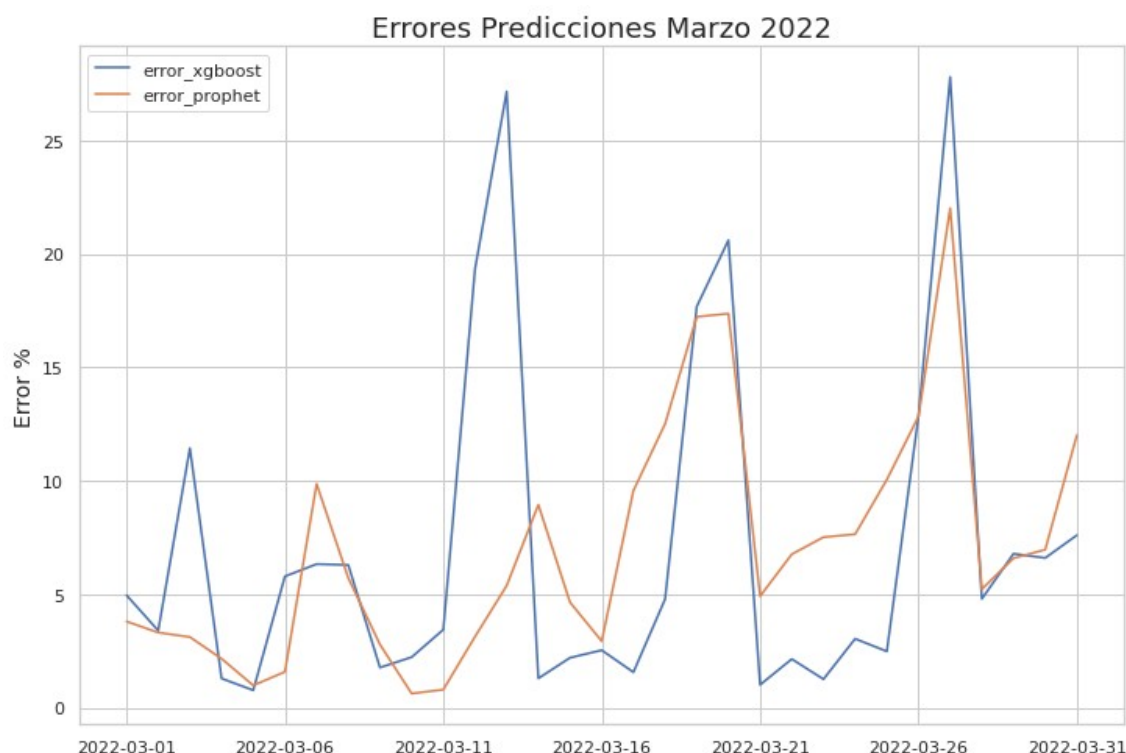


Fig. 14: Representación de los errores

Se puede apreciar cómo los errores de *Prophet* en promedio son menores que los de *XGBoost* al principio de la serie, pero a medida que pasa el tiempo van creciendo hasta obtener valores muy parecidos. Cabe destacar cómo *XGBoost* desde muy pronto tiene unos márgenes de error bastante grandes cuando hay picos de demanda. Esto se podría usar para escoger definitivamente *Prophet* como modelo, ya que permite obtener una predicción de demanda más cercana al valor real cuando hay picos de consumo.

## 8. Conclusiones

Después de haber comparado cada uno de los modelos en los escenarios planteados, podemos decir que, lejos de lo que se podría pensar a priori de que cuanto mayor número de variables exógenas, mayor precisión de los modelos, parece ser que con identificar las principales variables que tienen influencia sobre nuestra variable de interés es suficiente.

De los modelos estudiados y sus métricas, podemos decir que *Prophet* es el mejor modelo en la mayoría de los escenarios para casi todas las métricas. Por lo tanto *Prophet* es el modelo que mejor sabe usar la información de las variables exógenas consideradas en este estudio. Cabe destacar cómo

*SARIMAX* en tres casos (MAE, MAPE y sMAPE), y siempre en el escenario B (días festivos y laborables), es el mejor modelo.

*XGBoost* solo es el mejor modelo para la métrica MASE, pero cabe destacar cómo los valores de las métricas en cada escenario suelen ser muy parecidos entre sí. Esto tiene que ver por cómo *XGBoost* obtiene el mejor modelo entre todos los posibles para los datos que se le pasan.

Para el caso de este estudio nuestra variable de interés ha sido la demanda eléctrica en el Estado Español y, aunque se ha demostrado en la visualización de los datos que la temperatura tiene influencia sobre la demanda, en concreto para valores extremos (por debajo de 15°C y por encima de 25°C), podemos decir que no tiene un peso decisivo a la hora de mejorar los modelos a diferencia de los días laborables y festivos.

Esto se debe a la dificultad de determinar cómo influye la temperatura para un mismo día en diferentes partes de España teniendo, por su diversidad geológica, una gran variedad de microclimas en un territorio relativamente pequeño.

Por otro lado, que las variables exógenas que tengan más peso sobre la demanda sean los días festivos y laborables, nos indica que la demanda eléctrica viene dada principalmente por patrones de consumo regulares que marca la industria.

Con relación a lo indicado en el último párrafo podemos ver que, aunque en la mayoría de las métricas y escenarios el mejor modelo siempre es Prophet, el hecho de que *SARIMAX* tenga el mejor valor en dos de las métricas más empleadas (MAE y MAPE) en el escenario B, nos da una idea del peso importante que tiene el conocer los patrones de demanda y consumo en los días laborables y festivos. Esto es bastante importante ya que la temperatura es una variable que no podemos conocer a futuro con precisión más allá de 3-5 días y, por encima de ese valor, el ruido en nuestras predicciones aumenta considerablemente.

## **9. Siguiendo pasos**

Una vez hemos conseguido un cierto conocimiento de cómo funciona la demanda eléctrica en el caso del Estado Español, cabe preguntarse si el presente estudio podría mejorarse de cara al futuro pudiendo ser en los siguientes ejes:

- **Temperatura:** investigar si el considerar la temperatura media de cada día en cada una de las cinco ciudades consideradas es acertado, teniendo en cuenta que hay épocas del año con variaciones de temperatura bastante amplias, siendo estas variaciones más acusadas según la ciudad.

Ver si conseguir otro tratamiento de lo que llamamos “estacionalidad” etiquetando los meses como invierno o verano podría ayudar que estas etiquetas pudieran ayudar a mejorar las métricas en vez de empeorarlas.

Otra línea en cuanto a la temperatura podría ser el tener en cuenta la temperatura por hora considerando, por tanto, esa misma frecuencia horaria en nuestros datos de demanda. Quizás en este sentido la temperatura cobre más peso como variable exógena que ayude a aumentar la precisión de los modelos en lugar de penalizar, como hemos visto para el caso diario.

Por último, el considerar otra ciudad en el noroeste atlántico, como Coruña, podría ayudar a tener una información más completa sobre la distribución de temperaturas y su influencia en el conjunto de la demanda eléctrica nacional.

- **Población:** podría ser interesante que junto con la temperatura se pudiera tener en cuenta la población de las ciudades en consideración. Que determinada densidad de población se vea afectada por las mismas condiciones climáticas y, por lo tanto, comparta patrones de consumo similares, podría ayudar a una mejora de las métricas. Haría falta investigar qué se ha hecho hasta ahora en este sentido.
- **Días laborables y festivos:** realizar una investigación más granular sobre la demanda energética de los viernes podría ser de ayuda ya que presenta mayor variabilidad que el resto de días laborales. Lo mismo

ocurre con los sábados, que tienen un promedio de demanda mayor que los domingos. Por lo tanto, el considerar indicadores (tipo booleano o de otro tipo) que nos indiquen si el día es viernes, sábado o domingo, podría ayudar a mejorar aún más las métricas del modelo.

Por otro lado, sería interesante analizar qué comportamientos de demanda se producen según el festivo nacional caiga en un día de la semana u otro, permitiendo que haya puentes o no, etc.

Por último, un análisis del comportamiento de la demanda comparada entre los días laborables post-festivos y los no post-festivos podría ser interesante para tenerlo en cuenta.

- **Análisis pormenorizado por día:** encontrar los días en los que los contrastes entre los valores reales y las predicciones sean más grandes y analizar el por qué puede estar ocurriendo esto y si puede deberse a factores que no estamos teniendo en cuenta.
- **Descomposición de la demanda:** conocer qué porcentaje por día/hora de la demanda se corresponde con industria, servicios, hogar, etc. Quizá con esa descomposición variables exógenas como la temperatura puedan tener mayor relevancia en según qué tipo de demanda, como puede ser el consumo de los hogares.
- **Empleo de otras técnicas de modelado:** otras técnicas como Redes Neuronales Convolucionales, Redes Neuronales Recurrentes, etc si además de tener interés en determinar variables exógenas y su peso sobre la demanda, queremos encontrar el mejor modelo posible, teniendo en cuenta el mayor coste de entrenamiento y optimización usando RNRs.

## ***Bibliografía***

[01]:

*La guerra en Ucrania dispara El Precio de los Suministros Energéticos.*  
www.efe.com. (2022, June 3). Retrieved July 16, 2022, from  
<https://www.efe.com/efe/espana/economia/la-guerra-en-ucrania-dispara-el-precio-de-los-suministros-energeticos/10003-4821572#:~:text=La%20media%20registrada%20antes%20de,y%20un%2017%20%25%2C%20respectivamente.>

[02]:

*La guerra en Ucrania dispara El Precio de los Suministros Energéticos.*  
www.efe.com. (2022, June 3). Retrieved July 16, 2022, from  
<https://www.efe.com/efe/espana/economia/la-guerra-en-ucrania-dispara-el-precio-de-los-suministros-energeticos/10003-4821572#:~:text=La%20media%20registrada%20antes%20de,y%20un%2017%20%25%2C%20respectivamente.>

[03]:

Debate, R. E. (2022, July 3). *Argelia confirma Una Subida del Precio del Gas Para España.* El Debate. Retrieved July 16, 2022, from  
<https://www.eldebate.com/economia/20220703/argelia-confirma-subida-precio-gas-espana.html>

[04]:

*Forecasting electricity demand with python.* py29-forecasting-electricity-power-demand-python. (n.d.). Retrieved July 16, 2022, from  
<https://www.cienciadedatos.net/documentos/py29-forecasting-electricity-power-demand-python-0.3.html>

[05]:

Castro, C. (2021, May 11). *Cloud Data Driven Partner: Evolución del forecasting.* Keeper. Retrieved July 16, 2022, from  
<https://keeper.io/es/2021/04/evolucion-del-forecasting/>

[06]:

Taylor, S. J., & Letham, B. (2017). *Forecasting at scale.*  
<https://doi.org/10.7287/peerj.preprints.3190v2>

- [07]:  
Arzamendia, O. (2021, July 12). *Predicción con series De tiempo - una guía inicial*. Medium. Retrieved July 16, 2022, from <https://medium.com/datos-y-ciencia/predicci%C3%B3n-con-series-de-tiempo-una-gu%C3%ADa-inicial-2bd62d55675a#:~:text=Una%20variable%20ex%C3%B3gena%20es%20aquella,sin%20ser%20afectadas%20por%20%C3%A9l>
- [08]:  
Christen, R., Mazzola, L., Denzler, A., & Portmann, E. (2020). Exogenous data for Load Forecasting: A Review. *Proceedings of the 12th International Joint Conference on Computational Intelligence*.  
<https://doi.org/10.5220/0010213204890500>
- [09]:  
Sowinski, J. (2021). The impact of the selection of exogenous variables in the ANFIS model on the results of the daily load forecast in the Power Company. *Energies*, 14(2), 345. <https://doi.org/10.3390/en14020345>
- [10]:  
*Sign in*. RPubS. (n.d.). Retrieved July 16, 2022, from <https://rpubs.com/AlbertoFuente/408243>
- [11]:  
Darco. (2020, November 3). *Conocer Las Tarifas de Luz cada hora, script en python*. iPad Books. Retrieved July 16, 2022, from <https://ipadbooks.contrataweb.com/2020/11/03/conocer-las-tarifas-de-luz-cada-hora-script-en-python/>
- [12]:  
*Cómo conseguir datos de ESIOS con su api*. Cómo conseguir datos de ESIOS con su API · Energía Chi-cuadrado. (n.d.). Retrieved July 17, 2022, from <https://www.energychisquared.com/post/c%C3%B3mo-conseguir-datos-de-esios-con-su-api>
- [13]:  
*Copernicus Climate Data Store*. Copernicus Climate Data Store I. (n.d.). Retrieved July 17, 2022, from <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=form>
- [14]:  
Zach. (2021, May 25). *Augmented dickey-fuller test in Python (with example)*. Statology. Retrieved July 17, 2022, from <https://www.statology.org/dickey-fuller-test-python/>

- [15]: David U. Ordiz Todas las entradas de este autor. (2020, August 2). *Analizando series Temporales con python (parte 2)*. QuantSpace. Retrieved July 17, 2022, from <https://quantspace.es/2020/08/01/analisis-de-series-temporales-con-python-parte-2/>
- [16]: Rodriguez, R. J. T., Alvarez, J., Sergio, & López, C. M. (2020, October 23). *series Temporales Con R → Autocorrelación*. Finanzas Zone. Retrieved July 17, 2022, from [https://finanzaszone.com/analisis-y-prediccion-de-series-temporales-con-r-iii-autocorrelacion/#:~:text=La%20funci%C3%B3n%20de%20autocorrelaci%C3%B3n%20parcial%20\(PACF\)&text=La%20funci%C3%B3n%20ACF%20es%20usada,del%20proceso%20autoregresivo%20\(AR\).](https://finanzaszone.com/analisis-y-prediccion-de-series-temporales-con-r-iii-autocorrelacion/#:~:text=La%20funci%C3%B3n%20de%20autocorrelaci%C3%B3n%20parcial%20(PACF)&text=La%20funci%C3%B3n%20ACF%20es%20usada,del%20proceso%20autoregresivo%20(AR).)
- [17]: *Interpretar la Función de Autocorrelación Parcial (PACF)*. Minitab. (n.d.). Retrieved July 17, 2022, from <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/modeling-statistics/time-series/how-to/partial-autocorrelation/interpret-the-results/partial-autocorrelation-function-pacf/>
- [18]: *Trabajar Con Planificación Predictiva en smart view*. Moved. (2021, May 11). Retrieved July 17, 2022, from [https://docs.oracle.com/cloud/help/es/pbcs\\_common/CSPPU/holt-winters\\_additive.htm](https://docs.oracle.com/cloud/help/es/pbcs_common/CSPPU/holt-winters_additive.htm)
- [19]: Martínez, J., & Ciencia UANL. (n.d.). *Aplicación de Holt-Winters Para pronósticos de inventarios*. Ciencia UANL. Retrieved July 17, 2022, from <https://cienciauanl.uanl.mx/?p=7948#:~:text=MODELO%20HOLT%20DWINTERS&text=El%20m%C3%A9todo%20Holt%20Winters%20es,de%20pron%C3%B3sticos%20a%20corto%20plazo.>
- [20]: Villaverde, N. R. (n.d.). *Holt-Winters*. RPubS. Retrieved July 17, 2022, from <https://rpubs.com/nanrosvil/283121>
- [21]: Wikimedia Foundation. (2020, October 15). *Suavizamiento exponencial*. Wikipedia. Retrieved July 17, 2022, from [https://es.wikipedia.org/wiki/Suavizamiento\\_exponencial#:~:text=El%20modelo%20Holt%20Winters%20incorpora,as%C3%AD%20como%20a%20patrones%20estacionales.](https://es.wikipedia.org/wiki/Suavizamiento_exponencial#:~:text=El%20modelo%20Holt%20Winters%20incorpora,as%C3%AD%20como%20a%20patrones%20estacionales.)



- [22]:  
Verma, Y. (2021, July 29). *Complete guide to SARIMAX in Python for time series modeling*. Analytics India Magazine. Retrieved July 17, 2022, from <https://analyticsindiamag.com/complete-guide-to-sarimax-in-python-for-time-series-modeling/>
- [23]:  
*SARIMA\_FIT - sarima model fitted values – help center*. (n.d.). Retrieved July 17, 2022, from <https://support.numxl.com/hc/en-us/articles/214286406-SARIMA-FIT-SARIMA-Model-Fitted-Values>
- [24]:  
Wikimedia Foundation. (2022, March 19). *Criterio de Información de Akaike*. Wikipedia. Retrieved July 17, 2022, from [https://es.wikipedia.org/wiki/Criterio\\_de\\_informaci%C3%B3n\\_de\\_Akaike#:~:text=El%20criterio%20de%20informaci%C3%B3n%20de,y%20la%20complejidad%20del%20modelo.](https://es.wikipedia.org/wiki/Criterio_de_informaci%C3%B3n_de_Akaike#:~:text=El%20criterio%20de%20informaci%C3%B3n%20de,y%20la%20complejidad%20del%20modelo.)
- [25]:  
Vega, J. B. M. (2022, January 24). *Tutorial: Xgboost en python*. Medium. Retrieved July 17, 2022, from <https://medium.com/@jboscomendoza/tutorial-xgboost-en-python-53e48fc58f73>
- [26]:  
Khan Academy. (n.d.). *Descenso de gradiente (artículo)*. Khan Academy. Retrieved July 17, 2022, from <https://es.khanacademy.org/math/multivariable-calculus/applications-of-multivariable-derivatives/optimizing-multivariable-functions/a/what-is-gradient-descent#:~:text=El%20descenso%20de%20gradiente%20es%20un%20algoritmo%20que%20estima%20num%C3%A9ricamente,0%20como%20hemos%20visto%20antes.>
- [27]:  
RC, &A. (2020, April 26). *Forecasting Con Prophet*. Medium. Retrieved July 17, 2022, from <https://medium.com/@angel.r.chicote/forecasting-con-prophet-7fb36b25eb4b>
- [28]:  
*Algoritmo Prophet - Amazon forecast - AWS documentation*. (n.d.). Retrieved July 17, 2022, from [https://docs.aws.amazon.com/es\\_es/forecast/latest/dg/aws-forecast-recipe-prophet.html](https://docs.aws.amazon.com/es_es/forecast/latest/dg/aws-forecast-recipe-prophet.html)

- [29]:  
*Acervo Lima*. Análisis de series de tiempo usando Facebook Prophet –  
*Acervo Lima*. (n.d.). Retrieved July 17, 2022, from <https://es.acervolima.com/analisis-de-series-de-tiempo-usando-facebook-prophet/>
- [30]:  
Rink, K. (2021, November 3). *Time Series Forecast Error Metrics you should know*. Medium. Retrieved July 17, 2022, from  
<https://towardsdatascience.com/time-series-forecast-error-metrics-you-should-know-cc88b8c67f27>

## **Anexos:**

- [01]:  
Llamas Verna, B. J. (2022). *Extracción Datos Demanda Eléctrica*. Bernardo José Llamas Verna.  
[https://github.com/BernardoJoseLlamasVerna/tfm\\_data\\_science/blob/master/Demanda\\_Electrica/Extraccion\\_Datos\\_Demanda%20Electrica.ipynb](https://github.com/BernardoJoseLlamasVerna/tfm_data_science/blob/master/Demanda_Electrica/Extraccion_Datos_Demanda%20Electrica.ipynb)
- [02]:  
Llamas Verna, B. J. (2022b). *Extracción y Tratamiento Datos de Temperatura*. Bernardo José Llamas Verna.  
[https://github.com/BernardoJoseLlamasVerna/tfm\\_data\\_science/blob/master/Datasets\\_Temperaturas.ipynb](https://github.com/BernardoJoseLlamasVerna/tfm_data_science/blob/master/Datasets_Temperaturas.ipynb)
- [03]:  
Llamas Verna, B. J. (2022a). *Análisis de la Demanda Eléctrica en España 2014–2021*. Bernardo José Llamas Verna.  
[https://github.com/BernardoJoseLlamasVerna/tfm\\_data\\_science/blob/master/Demanda\\_Electrica/An%C3%A1lisis%20Demanda%20El%C3%A9ctrica%20Espa%C3%B1a.ipynb](https://github.com/BernardoJoseLlamasVerna/tfm_data_science/blob/master/Demanda_Electrica/An%C3%A1lisis%20Demanda%20El%C3%A9ctrica%20Espa%C3%B1a.ipynb)
- [04]:  
Llamas Verna, B. J. (2022d). *Modelado SARIMAX*. Bernardo José Llamas Verna.  
[https://github.com/BernardoJoseLlamasVerna/tfm\\_data\\_science/blob/master/Demanda\\_Electrica/SARIMAX\\_TEST.ipynb](https://github.com/BernardoJoseLlamasVerna/tfm_data_science/blob/master/Demanda_Electrica/SARIMAX_TEST.ipynb)
- [05]:  
Llamas Verna, B. J. (2022e). *Modelado XGBoost*. Bernardo José Llamas Verna.  
[https://github.com/BernardoJoseLlamasVerna/tfm\\_data\\_science/blob/master/Demanda\\_Electrica/XGBoost\\_Test.ipynb](https://github.com/BernardoJoseLlamasVerna/tfm_data_science/blob/master/Demanda_Electrica/XGBoost_Test.ipynb)

[06]:

Llamas Verna, B. J. (2022d). *Modelado Prophet*. Bernardo José Llamas Verna.

[https://github.com/BernardoJoseLlamasVerna/tfm\\_data\\_science/blob/master/Demanda\\_Electrica/PROPHET\\_TEST.ipynb](https://github.com/BernardoJoseLlamasVerna/tfm_data_science/blob/master/Demanda_Electrica/PROPHET_TEST.ipynb)