



UNIVERSITAT DE
BARCELONA

Facultat de Matemàtiques
i Informàtica

Treball Final de Grau

GRAU DE MATEMÀTIQUES

Modelización de previsiones de consumo de
electricidad

Patricia Kiss

Curso 2023 - 2024

Director: Josep Vives Santa Eulalia

Departamento de Matemática Económica, Financiera y Actuarial

Abstract

This research is focused on enhancing the accuracy of electricity demand forecasts for an energy retailer through the application of mathematical models. The goal is to optimize resources, and to provide valuable insights for strategic decision-making, this way enabling more effective management of electricity demand in the retailer.

This work offers an innovative and practical perspective to improve electricity forecasts within the specific context of an energy retail company.

Resumen

Este trabajo se enfoca en mejorar la precisión de las previsiones de demanda eléctrica para una comercializadora, así como su automatización mediante la aplicación de modelos matemáticos. Se busca optimizar la gestión de recursos y contribuir a una planificación más eficiente de la comercializadora en el sector eléctrico. Los resultados obtenidos aportan información en la toma de decisiones estratégicas, permitiendo de esta manera una gestión más efectiva de la demanda eléctrica en la comercializadora.

Este trabajo aporta una perspectiva práctica para mejorar las previsiones de electricidad para una empresa comercializadora.

Índex

1. Motivación y objetivo	7
1.1. Introducción	7
1.2. El mercado eléctrico	9
1.3. La comercializadora	12
1.4. Previsiones de electricidad	16
2. Metodología actual	18
2.1. Tratamiento de datos	18
2.2. Método actual	22
2.3. Del proceso manual hacia la automatización	25
3. Modelos y metodología	27
3.1. Notación, conceptos básicos y modelos base	27
3.1.1. Proceso estocástico	29
3.1.2. Estacionariedad	30
3.1.3. Modelo autorregresivo de medias móviles ARMA (p, q)	32
3.1.4. Modelo autorregresivo integrado de medias móviles ARIMA (p, q, d)	34
3.2. Modelos propuestos	36
3.2.1. Media del consumo de las últimas 3 semanas	36
3.2.2. SARIMAX (ARIMA)	37
3.3. Prueba retrospectiva	40
3.4. Métricas	41
3.4.1. Cups APE (Error porcentual absoluto)	41
3.4.2. Cartera APE (Error porcentual absoluto)	41
4. Resultados	43
4.1. Tarifas 6.1TD	43
4.1.1. Tarifa 6.1TD (días laborables)	46
4.2. Tarifas 3.0TD	47
4.2.1. Tarifa 3.0TD (días laborables)	50
5. Conclusiones	51
6. Bibliografía	52

1. Motivación y objetivo

1.1. Introducción

El escenario actual del mercado eléctrico está caracterizado por un alto grado de volatilidad, así como una rápida evolución y alta competitividad entre los diferentes agentes de mercado. En particular vemos esta competitividad reflejada en el hecho que las pequeñas comercializadoras se encuentran en la encrucijada de anticipar y satisfacer las demandas de un mercado cada vez más dinámico y además adaptarse a las preferencias de los consumidores.

Una de las tareas de mayor relevancia e importancia en general para una empresa es la planificación. Ésta debe disponer de un plan que permita un correcto funcionamiento de la empresa además de ser una de las claves para conseguir mejores resultados con una eficiencia más alta. La estructura de la planificación no es algo trivial, y una de las tareas que se destaca dentro de las más importantes es la predicción de la demanda de consumo que la empresa tendrá para su cartera de clientes. Esta tarea permitirá llevar a cabo una planificación que visualice las inversiones necesarias a realizar, pero dando la posibilidad de efectuar acciones que minimicen costes y permitan ofrecer un mejor servicio.

Por ello, en este caso en particular, la precisión en las previsiones se ha convertido en el elemento vital que determina el éxito operativo y financiero de estas empresas, ya que la gestión efectiva de la oferta y la demanda es esencial para evitar desvíos costosos y asegurar la continuidad operativa.

Para una pequeña comercializadora, con una cartera reducida de clientes, o que actualmente esté en crecimiento, a la hora de realizar la tarea de elaborar con cierta precisión sus previsiones de consumo, se encuentran con ciertas dificultades:

- **Presupuesto reducido** – se limita la capacidad de implementar soluciones tecnológicas basadas en el análisis de datos; obviamente esto supone una desventaja ante competidores más fuertes que sí pueden aprovechar estos recursos para mejorar sus modelos predictivos
- **Falta de histórico** - igual que el punto anterior, el no tener datos históricos y no poder identificar patrones de consumo para construir un modelo predictivo fiable, supone otra desventaja ante otros agentes de mercado
- **Volatilidad mercado** – existe una necesidad de adaptación a las nuevas normativas, así como a las tendencias del mercado y otro tipo de factores como económicos o estacionales. Para ello, las comercializadores deben adaptarse rápidamente a dichos factores, para no incurrir en pérdidas.

Por tanto, esta necesidad de responder a cambios imprevistos, bien dado por factores exteriores (normativas) o interiores (cambios de consumo en la cartera) se convierte en algo crítico para que estas empresas sobrevivan. Si no existe una previsión precisa, esto puede llevar a sobrecostes innecesarios y hasta pérdidas de clientes.

Para ello, muchas comercializadoras, sobre todo en sus inicios, buscan estrategias para mejorar su capacidad predictiva, siendo una de las más recurrentes la colaboración con consultoras especializadas en ofrecer soluciones para que las comercializadoras puedan sobrellevar las continuas complejidades del mercado y a la vez ofrecer un buen servicio a sus clientes.

Este trabajo se enfoca a estudiar detalladamente la manera de buscar un modelo de previsión lo más preciso, adaptable y escalable a diferentes tipologías de carteras de comercializadoras. La solución a este problema que se plantea, además de tener un resultado inmediato, puede sentar las bases para el continuo crecimiento de dichas comercializadoras a largo plazo.

¿Optimización de las previsiones → Estabilidad y éxito?

Vamos a intentar comprender los motivos que pueden llevar al éxito el uso de un método de predicción proporcionado por una consultora con cierta visión y experiencia:

1. Integrar datos externos

El uso de datos históricos ya experimentados, como por ejemplo tendencias del mercado, ciertas situaciones sociales, o hasta problemas globales, tipo geopolíticas por ejemplo (guerras, pandemias), pueden facilitar la comprensión del escenario actual y anticiparse en la toma de decisiones.

2. Aumento de la precisión

El poder incorporar tecnologías avanzadas y usar por ejemplo la inteligencia artificial para analizar un conjunto de datos grande de manera más eficaz permite identificar patrones que quizás con métodos más básicos pasarían desapercibidos.

3. Flexibilidad y adaptación

El uso de tecnologías avanzadas permite una mejor capacidad de adaptación a las nuevas demandas del mercado; para una mayor competitividad se requiere ser flexible y tener una respuesta rápida ante los cambios que se producen.

4. Automatización y eficiencia

Obviamente la parte de automatización permite liberar recursos internos, prevenir el error humano y ganar tiempo en la operativa, permitiendo así que la empresa se centre en buscar estrategias de mejora y no en tareas rutinarias.

5. Análisis en tiempo real

El poder analizar datos al instante permite una toma de decisiones inmediata, de manera efectiva, y ajustar así estrategias en función de las demandas existentes.

En resumen, vemos que la combinación de experiencia y tecnologías avanzadas puede ser un buen camino para mejorar las previsiones, aumentar la eficiencia y dar cierta estabilidad y el triunfo para pequeñas empresas comerciales. Sin embargo, se deben implementar estas estrategias de manera cuidadosa y considerada, adaptándolas a las necesidades específicas de cada empresa. Es este aspecto, la modelización de previsiones de electricidad para una comercializadora implica un enfoque matemático y estadístico significativo.

1.2. El mercado eléctrico

Antes de empezar a detallar la estructura del mercado eléctrico, vamos a hablar de cómo se estructura el sistema eléctrico, es decir de cómo se regula, se genera, se transporta, se distribuye y comercializa la electricidad.



Ilustración 1. Agentes del mercado eléctrico español

- **Generadores**

Los agentes generadores de energía eléctrica tienen un rol fundamental en el mercado eléctrico al proporcionar la oferta de energía necesaria para satisfacer la demanda. Su papel es transformar diversas fuentes de energía (renovables o no), como la térmica, hidroeléctrica, eólica, solar, nuclear, entre otras, en electricidad para el uso final de los consumidores. Estos deben cumplir la normativa y las diferentes regulaciones

ambientales, de seguridad y de calidad de servicio, incluyendo la reducción de emisiones, la gestión de residuos nucleares y la participación en programas de respuesta a la demanda.

En España, así como en muchos otros sistemas eléctricos, la electricidad se mueve dentro de un mercado mayorista. Es decir, los generadores participan en subastas o transacciones bilaterales para vender la electricidad que producen a las empresas distribuidoras o directamente a otros sujetos del mercado.

Los precios que fijan estos generadores varían según la disponibilidad de recursos, la demanda en tiempo real y costes de producción. Por ejemplo, la electricidad producida mediante una fuente natural de energía será más barata, que otra que necesita materias primas para su producción.

La oferta de electricidad se debe gestionar de manera eficiente para satisfacer la demanda variable a lo largo del tiempo, siendo fundamental disponer de capacidad de respuesta y flexibilidad para equilibrar la oferta y la demanda y mantener la estabilidad del sistema eléctrico.

- **Reguladores y operadores**

Los agentes reguladores son los encargados de establecer las normas y asegurarse que todos los agentes de mercado las cumplan. Se trata en este caso de la Administración del Estado y la CNMC.

Red Eléctrica de España, operador del sistema y OMIE, operador del mercado, son los responsables del correcto funcionamiento del sistema y del mercado eléctrico español; casan y balancean la oferta y la demanda de energía, y se encargan de establecer costes para los diferentes agentes de mercado, así como liquidárselos.

- **Transportistas**

Son los responsables de llevar la electricidad desde las plantas de generación hasta la red de distribución para el consumo. La empresa encargada de gestionar las infraestructuras de transporte de energía eléctrica y que tiene monopolio es Red Eléctrica de España (REE).

- **Distribuidoras**

Se encargan de llevar la energía hasta el punto de consumo del cliente final, y se encargan del correcto funcionamiento de la instalación de suministro del mismo. Las empresas distribuidoras se reparten por zonas geográficas, y a cada punto de suministro le corresponde una en particular, sin opción a cambio.

Actualmente hay más de 300 distribuidoras, aunque destacamos 5 grandes de ellas que tienen la mayoría de los puntos de suministro.

MAPA DE DISTRIBUIDORES



Ilustración 2. Mapa distribuidoras España

- **Comercializadores**

Las empresas comercializadoras se encargan de vender a los clientes finales la energía que producen y distribuyen las distribuidoras según la normativa vigente en cada momento.

Estas compran la energía en el mercado, bien mediante subasta, o bien mediante ciertos acuerdos bilaterales a largo o cortos plazos adquiriéndola directamente de los productores.

Esta actividad está liberalizada y existen una gran cantidad de empresas que realizan la actividad en libre competencia. Actualmente existen más de 400 comercializadoras de electricidad, y aunque hay ciertas comercializadoras grandes conocidas, como Endesa, Iberdrola, Gas Natural, también se están dando a conocer otras como Factor Energía, Bonpreu Energía o Axpo por ejemplo.

1.3. La comercializadora

Para poder entender la importancia del objetivo de este trabajo, vamos a intentar exponer cómo pueden afectar los desvíos en una comercializadora, los desvíos siendo la penalización aplicada sobre la diferencia entre la previsión y compra realizada en el mercado y el consumo real final de los clientes de la cartera de la comercializadora.

Previamente vamos a ver ciertas definiciones para poder tener una visión completa de los temas que trataremos:

- **CUPS:** Código Universal de Punto de Suministro, una clave de 20 dígitos alfanuméricos, que comienza por ES, seguida por cuatro dígitos que identifica la distribuidora y doce números restantes son asignados por la propia distribuidora para reconocer el punto de suministro final.
- **Curvas de carga:** es la curva que se genera a partir de las lecturas reales registradas cada cuarto de hora por el contador digital instalado en el punto de suministro. Esta puede ser o bien cuarto-horaria, o bien horaria agregada.
- **Tarifas de acceso:** Se asigna una tarifa en función de la potencia contratada en cada CUPS y de la tensión de la línea de distribución. Las tarifas 2.0TD son de baja tensión, las 3.0TD media tensión y la resta alta tensión

Tensión	Nivel de tensión	Potencia contratada	Tarifa	Períodos	
				Potencia	Energía
Menos de 1 kV	NT0	P < 15 kW	2.0TD	2	3
		P > 15 kW P _{n+1} ≥ P _n	3.0TD	6	6
Entre 1 y 30 kV	NT1	P _{n+1} ≥ P _n	6.1TD	6	6
Entre 30 y 72,5kV	NT2	P _{n+1} ≥ P _n	6.2TD	6	6
Entre 72,5 y 145 kV	NT3	P _{n+1} ≥ P _n	6.3TD	6	6
Mayor de 145 kV	NT4	P _{n+1} ≥ P _n	6.4TD	6	6

Ilustración 3. Tarifas de acceso TD

- **Periodos:** Llamamos periodos a los distintos grupos de discriminación horaria que tiene cada tarifa. Tal y como se indica en la Ilustración 3, para la tarifa 2.0TD tendremos 2 periodos de potencia y 3 de energía (punta P1, llano P2, valle P3), y para la resta de tarifas tenemos 6 periodos tanto de energía como de potencia.

Nos centramos en los de energía, estos van desde la P1 a la P6, siendo los primeros periodos los de demanda alta, y por tanto más caros, y van hasta los de menos demanda y más baratos.

Periodos horarios para energías

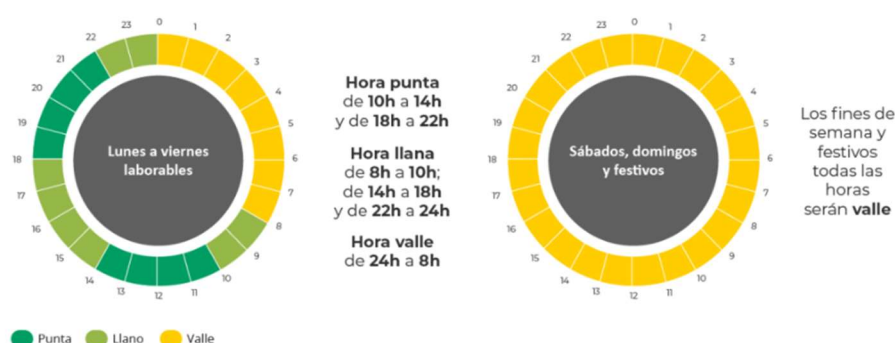


Ilustración 4. Periodos tarifas 2.OTD

TARIFA 3.0 y 6.X	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Enero	6	6	6	6	6	6	6	6	2	1	1	1	1	1	2	2	2	2	1	1	1	1	2	2
Febrero	6	6	6	6	6	6	6	6	2	1	1	1	1	1	2	2	2	2	1	1	1	1	2	2
Marzo	6	6	6	6	6	6	6	6	3	2	2	2	2	2	3	3	3	3	2	2	2	2	3	3
Abril	6	6	6	6	6	6	6	6	5	4	4	4	4	4	5	5	5	5	4	4	4	4	5	5
Mayo	6	6	6	6	6	6	6	6	5	4	4	4	4	4	5	5	5	5	4	4	4	4	5	5
Junio	6	6	6	6	6	6	6	6	4	3	3	3	3	3	4	4	4	4	3	3	3	3	4	4
Julio	6	6	6	6	6	6	6	6	2	1	1	1	1	1	2	2	2	2	1	1	1	1	2	2
Agosto	6	6	6	6	6	6	6	6	4	3	3	3	3	3	4	4	4	4	3	3	3	3	4	4
Septiembre	6	6	6	6	6	6	6	6	4	3	3	3	3	3	4	4	4	4	3	3	3	3	4	4
Octubre	6	6	6	6	6	6	6	6	5	4	4	4	4	4	5	5	5	5	4	4	4	4	5	5
Noviembre	6	6	6	6	6	6	6	6	3	2	2	2	2	2	3	3	3	3	2	2	2	2	3	3
Diciembre	6	6	6	6	6	6	6	6	2	1	1	1	1	1	2	2	2	2	1	1	1	1	2	2
Sábados, domingos y festivos	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6

Ilustración 5. Periodos tarifas 3.0TD y 6.XTD

1.3.1. El precio de la electricidad

El precio final aplicado al cliente por una comercializadora se compone de costes regulados, los que se corresponden directamente con el coste de producción de la energía y por costes de mercado, siendo estos el coste de la electricidad propiamente dicho. En la *Ilustración 8.* podemos ver como afecta cada componente al cálculo final del precio por periodo de una tarifa 3.0TD.

- **Materia prima**

El coste de la electricidad se fija diariamente tras una subasta realizada por OMIE (Operador del Mercado Ibérico Español) cada día de 12:00 a 12:45. En esta subasta participan de un lado los agentes compradores (comercializadoras) y los agentes vendedores (generadoras).

Mediante el llamado algoritmo Euphemia, los agentes vendedores van a poner sus ofertas de venta a diferentes precios, en función del coste de producción de la energía, desde las más caras, como por ejemplo las de ciclo combinado, hasta las mas baratas, tipo eólicas. A la vez, los agentes compradores van a poner sus ofertas de compra al precio más alto dispuestos a pagar, siempre que sus garantías en el mercado se lo permitan.

El punto medio en el cual se vende toda la energía necesaria para satisfacer la demanda será el punto que determinará el precio resultado de la casación diaria.

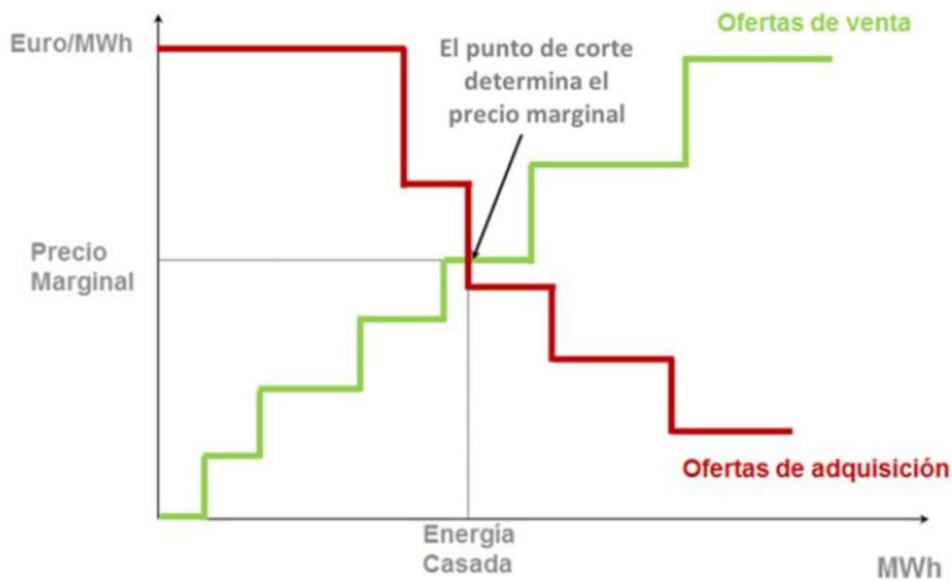


Ilustración 6. Algoritmo Euphemia aplicado en la subasta diaria

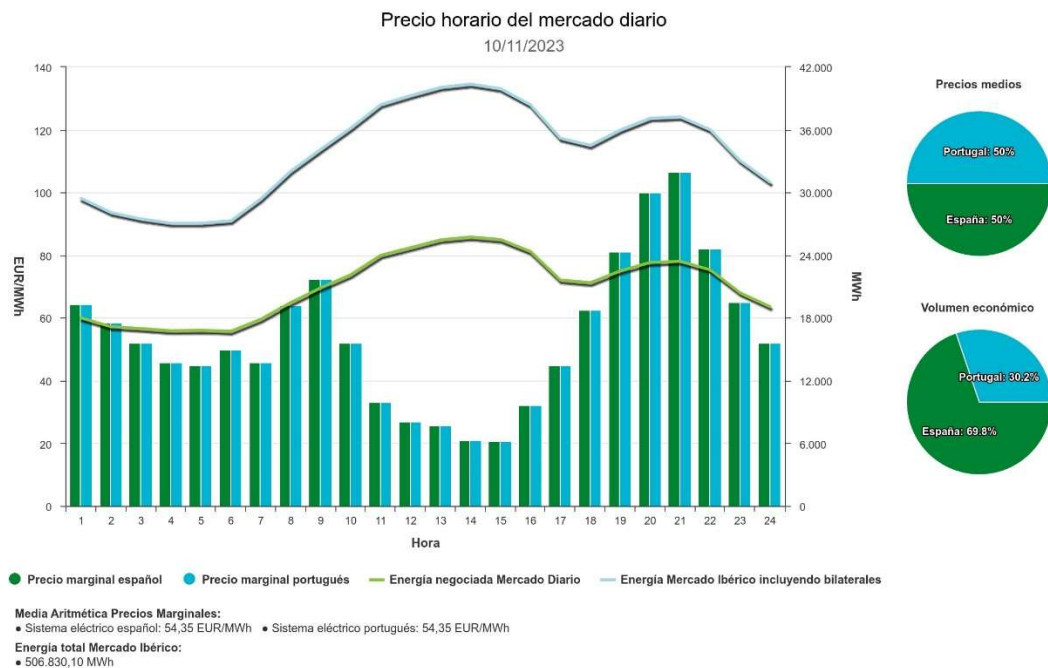


Ilustración 7. Resultado casación diaria 10/11/2023

- **Costes regulados**

- **Pagos por capacidad:** son ayudas a las centrales de energía que actúan como respaldo del sistema
- **Servicios complementarios y restricciones técnicas:** costes asociados al aumento o disminución de la generación para adaptarse a la demanda real necesaria en el sistema.
- **Remunerador del operador de mercado y operador de sistema:** coste aplicado para financiar tanto la actividad de REE y la de OMIE.
- **Fondo nacional de eficiencia energética:** ayudas para mejorar la eficiencia energética de las empresas y los particulares.
- **Pérdidas:** el porcentaje de energía que se pierde desde la generación de la energía hasta el momento de consumirla; por ese motivo se le añade el % adicional a los costes aplicados a la energía consumida
- **Acceso de terceros a la red:** coste de las distribuidoras y se les abonan por realizar la función de encargarse de hacer llegar la energía de forma adecuada hasta el contador de los clientes.
- **Coste de desvíos:** penalización que se aplica a la comercializadora por la diferencia de ajuste entre la compra realizada y en consumo final de sus clientes. Lo hay de 2 tipos, a subir, o a bajar, si se ha comprado de más energía, o de menos respecto el consumo.

Sin embargo, no todas las horas tienen penalización. Por ejemplo, si compramos de menos en una hora que se ha generado menos energía que la demanda que había, habremos favorecido el sistema y por tanto no se aplica penalización.

Actualmente los desvíos medios rondan los 20-25 €/MWh.

3.0TD



	P1	P2	P3	P4	P5	P6	MP OMIP	*perd+perfil
Perfil Materia Prima	1,14	1,08	1,09	1,01	0,91	0,95	75,0	
Precio Materia prima OMIP	85,57	81,04	81,35	75,78	68,56	71,33		
Pagos por Capacidad	1,443	0,667	0,444	0,333	0,333	0,000		
Perfil SSCC+RRTT	1,05	1,05	1,05	1,00	1,02	1,08	14	
Precio SSCC+RRTT	14,66	14,63	14,71	14,04	14,22	15,18		
Remun. Op Mercado y Sistema	0,212	0,212	0,212	0,212	0,212	0,212		
Fondo Nacional Eficiencia Energética	0,97	0,97	0,97	0,97	0,97	0,97		
Coste Desvíos	3	3	3	3	3	3	3	
Garantías de Origen	0,0	0	0	0	0	0	0,0	
Pérdidas	17,4%	18,4%	17,3%	17,3%	14,5%	18,9%		
Costes Mercado	124,14	118,82	117,97	110,51	99,81	107,65		
Bono Social	0	0	0	0	0	0	0	
Margen Comercializadora	2,00	2,00	2,00	2,00	2,00	2,00	2,00	
Tasa Municipal	1,892	1,812	1,8	1,688	1,527	1,645		
Coste Financiero	0%	0%	0%	0%	0%	0%	0%	
Peajes Transporte y Distribución	23,974	12,820	7,573	5,495	0,424	0,234		
Cargos del Sistema	24,469	18,118	9,788	4,894	3,137	1,958		
PRECIO SIN IEE (€/MWh)	176,473	153,566	139,132	124,591	106,895	113,487	131,45	
Precio con margen 0 €/MWh	174,443	151,536	137,102	122,561	104,865	111,457		

Ilustración 8. Pricing de electricidad

1.4. Previsiones de electricidad

¿Como de importante es tener una previsión de consumo precisa?

La previsión del consumo energético tiene un papel fundamental en mejorar la asignación de recursos y con la misma, obtener un buen ahorro a la vez que asegurar un suministro de energía de confianza. Sobre estas 3 áreas el impacto que tiene es el siguiente:

Asignación de Recursos:

- **Previsión de generación optima:** Una previsión de consumo precisa permite a los diferentes generadores del sistema hacer una previsión de la energía que necesitan generar y así distribuir sus recursos de manera más eficiente. Así, sabiendo cuando y cuanta energía se necesita horariamente (actualmente nos estamos trasladando a un sistema cuarto-horario), los distribuidores pueden planificar la generación de diferentes fuentes para igualar la demanda existente. Esto evita tanto el exceso de generación, como que haya deficiencia en la generación.
- **Gestión de red:** Ayuda a la red de transporte gestionar el transporte y distribución de la electricidad. Esto asegura que la electricidad se transmita de manera eficaz y que el sistema de distribución pueda soportar la carga esperada. Esto hace que haya menos pérdidas y más descongestión de la red.
- **Balance de carga:** Tener un equilibrio de carga en la red es fundamental para prevenir apagones. Una previsión precisa de consumo permite a los operadores de red tener un balance en su sistema ajustando la generación. Por ejemplo, activando las plantas de energía de reserva, durante los periodos de pico de demanda.

Ahorro de costes:

- **Reducción de uso combustibles y coste operativo:** Producir la cantidad justa de energía cuando se necesita conlleva ahorro de dinero al evitar el uso de plantas caras en momentos de alta demanda, reducción de costos de combustible al utilizar diferentes fuentes de energía de manera más eficiente, así como mayor durabilidad del equipo al no sobrecargarlo en momentos de pico de demanda.
- **Reducción de costes de emisión:** Una planificación de generación más eficiente conlleva una reducción de las emisiones de gases de efecto invernadero, ya que permite priorizar

fuentes de energía más limpias y eficientes. Esto no solo reduce el impacto ambiental, sino también los costes relacionados con las emisiones.

- **Evitar costes de penalización:** Una previsión precisa ayuda a los comercializadoras a evitar incurrir o minimizar las sanciones económicas relacionadas con no ajustar su compra al consumo de sus clientes. La liquidación de costes de REE incluye el coste por desvíos que se calcula horariamente en base a coste global que le ha supuesto al sistema tener un exceso o una deficiencia de energía.

Suministro de energía de confianza:

- **Estabilidad en la red:** Una previsión precisa contribuye a que haya estabilidad en la red. Los operadores en la red pueden ajustar de manera proactiva el balance generación-demanda, de esta manera evitar ciertos riesgos de tensión inestable, desviaciones de frecuencia o cortes de energía.
- **Situación de emergencia:** El sistema de distribución puede estar mejor preparado para casos extremos al tener una idea clara de la demanda. Esto es especialmente importante en regiones propensas a desastres relacionados con el clima, donde un suministro de energía fiable es esencial para los servicios de emergencia y la seguridad pública,
- **Clientes satisfechos:** Cuanto más seguro sea el suministro eléctrico y menos apagones o cortes haya, más satisfechos van a estar los usuarios de la red. Cuando los Usuarios que necesiten un suministro constante (fábricas, por ejemplo) se benefician de un suministro de calidad, sin cortes, esto conlleva menos pérdidas para las empresas.

En resumen, la previsión del consumo de energía es una herramienta fundamental para que los proveedores de energía gestionen sus recursos de manera efectiva, reduzcan los costos operativos y garanticen un suministro de energía seguro, fiable. Contribuye a una asignación más eficiente de recursos, ahorros financieros y un mayor nivel de fiabilidad del servicio, beneficiando en última instancia tanto a los proveedores de energía como a los consumidores.

2. Metodología actual

Para empezar a comentar el método actual que se plantea mejorar, antes vamos a exponer los datos que vamos a tratar, comenzando por exponer que tipologías de clientes nos encontramos, cual es la proporción de consumo que tienen en referencia a la cartera total, y qué diversas situaciones y problemas nos encontramos a la hora de manipular los datos.

Las tarifas de electricidad se clasifican según la tensión a la que están conectados a la red eléctrica. La tensión se mide en voltios (V), y las categorías incluyen la baja tensión (BT), la media tensión (MT) y la alta tensión (AT). A continuación, veremos qué incluyen estas categorías y cómo se tratan para hacer previsiones:

2.1. Tratamiento de datos

1. **Baja tensión:** suelen incluir voltajes inferiores a 1 kV (kilovoltio), puede variar ligeramente en función de la región y la normativa aplicable. Estos peajes se aplican a los negocios que requieren una potencia inferior o igual a los 15kW, con requerimientos energéticos similares a una vivienda, incluyendo la gran mayoría de puntos de suministro domésticos, pequeñas empresas, despachos, oficinas, etc. Son los que tienen asignados tarifa 2.0TD.

¿Como se tratan los datos en este caso?

- **Se analizan patrones de consumo residencial y comercial.**
- **Se consideran estacionalidades y eventos que afecten a estas categorías.**
- **Se evalúan cambios en las tarifas de baja tensión que puedan influir.**

Se dispone de curvas de carga de consumo para un 80-90 % de este tipo de clientes, aunque la mayoría no suelen estar publicadas por distribuidoras hasta al cabo de ciertas semanas. Consideraremos que tenemos unas curvas completas al cabo de un mes para unificar datos.

Veremos en el siguiente gráfico, en la *Ilustración 9.*, los consumos acumulados por tarifa 2.0TD. Por ejemplo, veremos un consumo bastante estable y lineal en la semana S2 de mayo, inferior al que podemos observar en la semana S1 de febrero, pudiendo ser esto resultado del uso de la calefacción en pequeñas empresas. Notamos sin embargo cierta similitud en los consumos acumulados durante los fines de semana de estas semanas.

Si observamos la semana escogida de agosto en la *Ilustración 9.*, veremos que el consumo es elevado, pero estable a lo largo de la semana. Dado que es una semana de cierre (mayoría de oficinas y despachos cerrados), podemos suponer que el aumento y estabilidad de consumo puede ser consecuencia del uso de aires acondicionados por el aumento de temperatura.

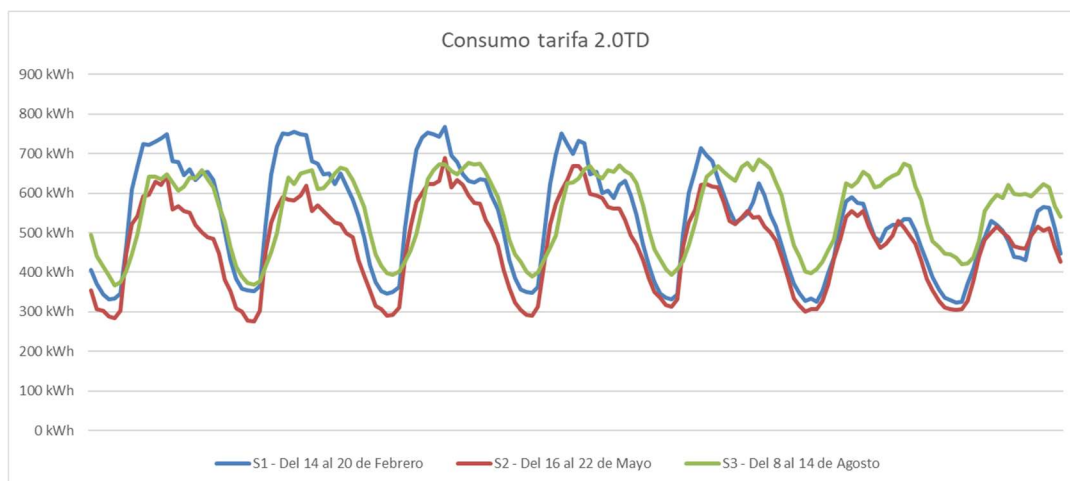


Ilustración 9. Consumo 2022 tarifa 2.0TD

2. **Media tensión:** suelen incluir voltajes inferiores a 1 kV (kilovoltio), y son los que tienen como asignada la tarifa 3.0TD. Estos peajes se aplican a los negocios que requieren una potencia superior a los 15kW, como por ejemplo grandes oficinas, empresas tipo pymes, o pequeñas naves industriales.

¿Como se tratan los datos en este caso?

- Se evalúa la demanda industrial y comercial
- Se consideran cambios en la producción, o bien paradas de consumo.
- Se estudia la tipología de clientes

Se dispone de curvas de carga de consumo para un 90 % de este tipo de clientes, y dependiendo de la distribuidora se suelen publicar la mayoría al día siguiente o al cabo de dos días. Para la resta de clientes utilizaremos en este orden, un perfilado mensual o anual. Priorizaremos el mensual, ya que puede haber clientes estacionales, y por último el consumo anual.

Veamos algunos ejemplos:

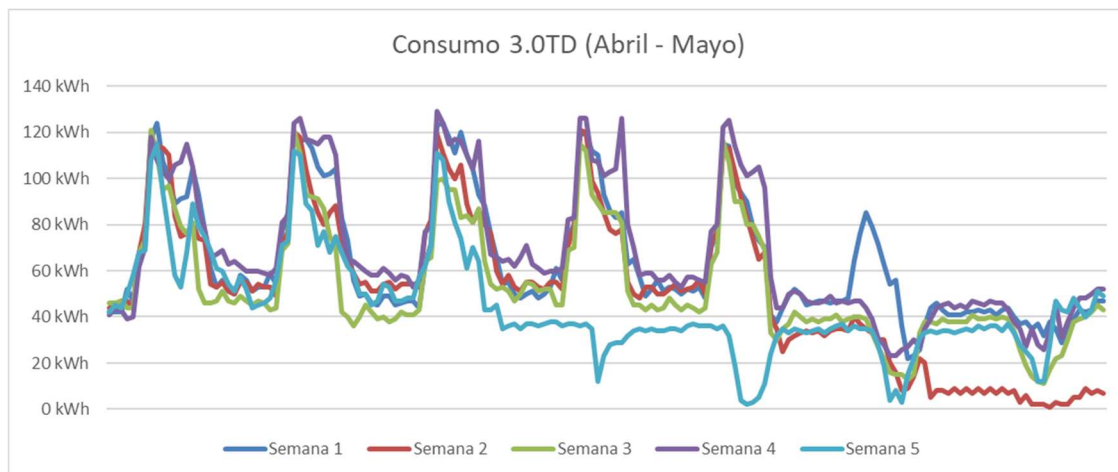


Ilustración 10. Consumo 2022 tarifa 3.0TD (Abril – Mayo)

El CUPS cuyo consumo podemos observar en la *Ilustración 10*. sería el tipo de cliente con un patrón típico de consumo. Un negocio que funciona de lunes a viernes, con cierto consumo residual el fin de semana. Se cogen aleatoriamente unas semanas de consumo de abril-mayo.

Para la semana 5 del mismo CUPS vemos cuál es la tendencia de consumo a lo largo de la semana del 5 al 11 de diciembre en la *Ilustración 11*. Observamos que aun siendo festivo el día 6, se decide parar la producción solo el 8 y hacer el puente.

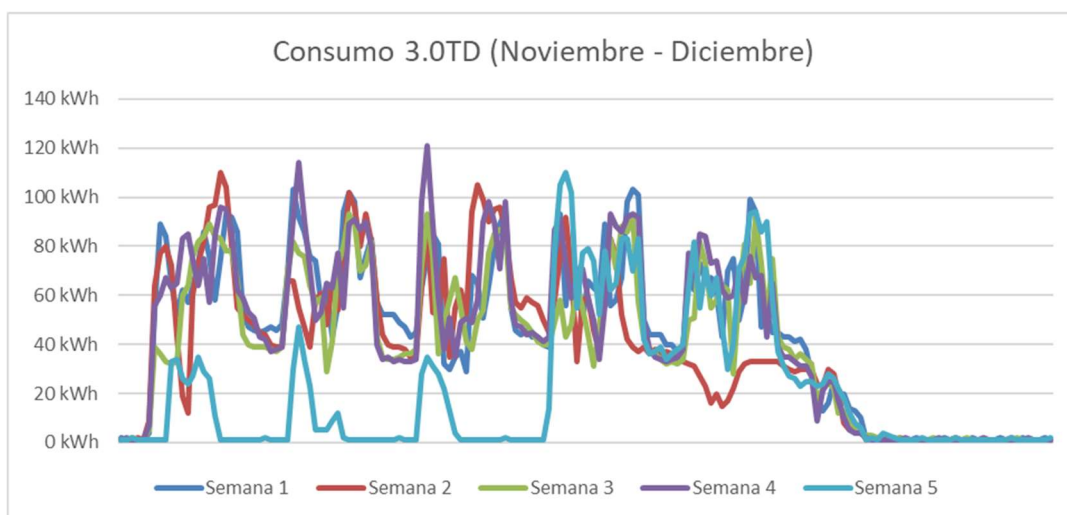


Ilustración 11. Consumo 2022 tarifa 3.0TD (Noviembre – Diciembre)

Este segundo CUPS de la *Ilustración 11*. por ejemplo, vemos que tiene cierta regularidad, aun con algunos picos de consumo que son imprevisibles. Sin embargo, vemos que lo que sería la semana del 5 al 11 de diciembre, como en el anterior caso, este negocia opta por bajar producción durante los primeros 3 días de la semana, totalmente contrario al anterior CUPS.

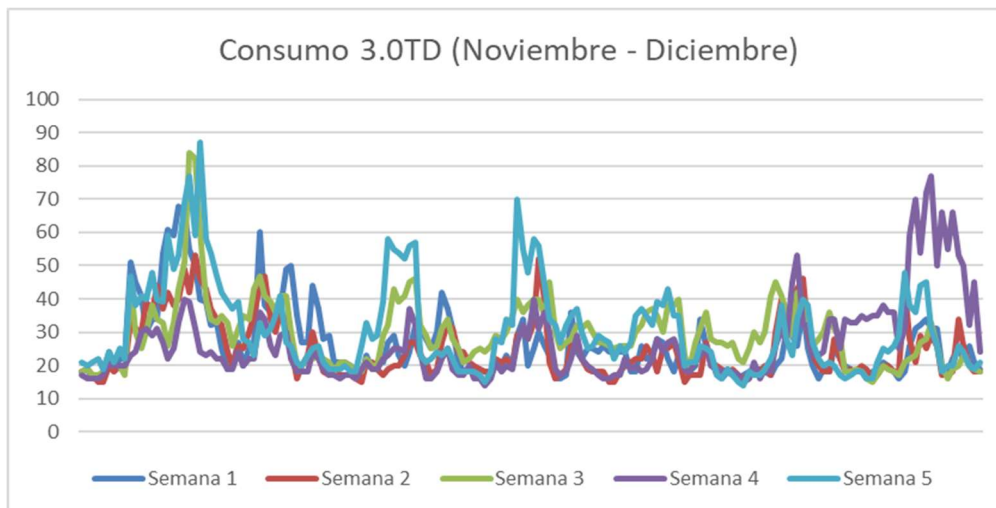


Ilustración 12. Consumo 2022 tarifa 3.0TD (Noviembre – Diciembre)

Y, por último, en la *Ilustración 12*, tenemos un tercer punto de suministro con un consumo totalmente aleatorio, que por lo que podemos observar, no atiende ningún patrón diferente durante los días festivos mencionados anteriormente.

3. **Alta tensión:** son las tarifas que engloban los puntos de suministros con voltajes superiores a 1 kV (kilovoltio). Las tarifas que se incluyen en esta categoría son los 6.1TD, 6.2TD, 6.3TD y 6.4TD. Estos peajes se aplican habitualmente a los polígonos industriales, pero también en aquellos puntos de suministro alejados de las redes de distribución de baja tensión, que se conectan a redes de alta tensión. Esta circunstancia especial puede darse en zonas rurales.

¿Como se tratan los datos en este caso?

- **Analizar patrones de consumo industrial pesado**
- **Se consideran cambios en la producción, o bien paradas de consumo.**
- **Se estudia la tipología de clientes**
- **Evaluar la demanda de grandes instalaciones comerciales**

La tendencia de consumo de los puntos de suministro de esta tarifa suele ser análoga a la de los 3.0TD, solo que en este caso las tarifas 6.1TD suelen tener consumos superiores, y por tanto la la previsión correcta del consumo de las mismas tiene una mayor importancia.

2.2. Método actual

A continuación, veremos cómo se realizan actualmente las previsiones.

De entrada, avanzamos que se trata de un modelo sencillo (medias móviles), con ajustes manuales en caso de que no sea posible la estimación por medias para aquellos CUPS que tienen curva de carga disponible. Para el estudio, vemos el cálculo de la previsión de la semana del 09/10/2022 al 15/10/2022. Cogemos esta semana en particular para observar como se ajustan las curvas para el festivo del 12/10/2022.

1. Alta tensión (Tarifa 6.1TD)

Para realizar la previsión, disponemos de la curva mensual anterior desde el 11/09 hasta al día anterior a la realización de la previsión, 05/10.

Ajustaremos mediante medias con ceros y unos, dando más prioridad a las últimas semanas sobre las primeras.

Dado que el miércoles es festivo nacional, según información de otros festivos nacionales u otros años, este punto de suministro sabemos que va a cerrar el día previo por la tarde. En este caso, ajustaremos manualmente según el criterio siguiente:

- Para el miércoles ajustamos el consumo según un festivo reciente (*Ilustración 15.*); vemos en la *Ilustración 15.* la diferencia respecto al consumo; para este punto en particular nos hemos desviado un 14,73 %, con un coste de 78,88 €. Como dato adicional, de no haber ajustado manualmente, cogiendo por ejemplo de media las últimas 2 semanas, hubiésemos tenido un desvío semanal de 26,79% y 124,77 € de coste.



Ilustración 13. Visualización gráfica curva mensual

	09/10/2022	10/10/2022	11/10/2022	12/10/2022	13/10/2022	14/10/2022	15/10/2022
	Domingo	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado
HORA	Previsión	Previsión	Previsión	Previsión	Previsión	Previsión	Previsión
1	19,500	16,000	31,000	16,000	36,000	42,667	35,500
2	20,500	16,500	30,000	17,000	35,000	42,667	35,000
3	19,500	16,500	29,000	16,000	34,500	43,667	33,000
4	19,500	16,000	28,000	17,000	33,000	43,333	31,500
5	19,500	16,500	27,500	18,000	34,000	41,333	28,000
6	26,500	36,500	42,500	23,000	50,500	54,667	36,000
7	47,000	112,500	117,500	29,000	132,500	140,000	94,000
8	38,500	131,500	148,000	30,000	163,500	144,000	111,000
9	35,000	148,500	145,000	32,000	146,000	147,333	84,000
10	35,500	142,500	135,500	29,000	164,500	147,000	79,500
11	36,000	169,000	147,000	30,000	163,000	154,000	83,500
12	36,000	155,000	150,500	31,000	161,500	170,667	82,500
13	36,500	160,500	152,500	30,000	153,000	188,000	77,000
14	44,000	198,500	161,000	31,000	154,500	200,333	71,000
15	47,000	173,500	152,000	30,000	137,500	159,000	44,000
16	39,000	122,500	135,000	26,000	149,500	145,000	41,000
17	36,500	109,000	129,500	25,000	156,500	123,000	40,000
18	36,500	102,500	133,000	24,000	137,000	119,000	38,000
19	36,500	104,000	136,500	25,000	140,000	113,667	39,500
20	36,000	102,500	130,500	24,000	136,500	108,333	37,500
21	36,000	104,500	130,000	24,000	137,500	90,000	37,500
22	36,000	99,500	109,000	23,000	118,500	80,000	37,500
23	37,500	58,500	54,500	18,000	57,500	57,333	37,000
24	26,500	38,000	33,000	18,000	40,000	42,667	26,000

Ilustración 14. Valores previsión semanal

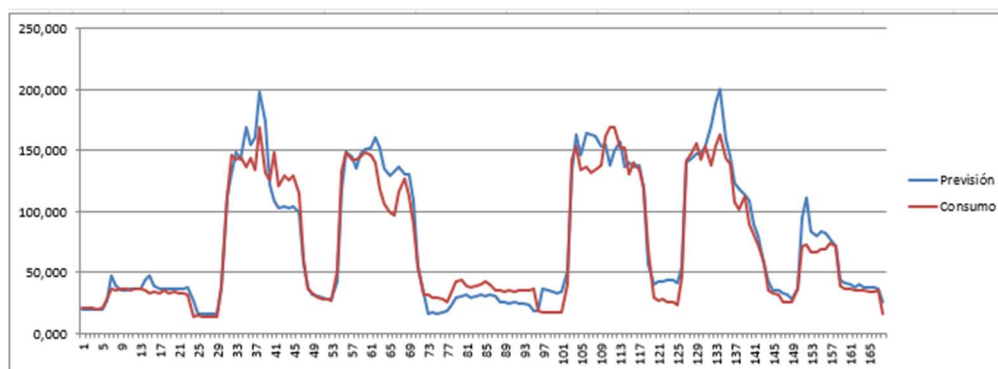


Ilustración 15. Gráfico previsión semanal

2. Media tensión (Tarifa 3.0TD)

Para esta tarifa, dividiremos los puntos de suministro en 2 categorías: puntos con curva de carga y puntos sin curva de carga. Para los que tienen curva de carga, la realización de previsión por CUPS es análoga a la de las tarifas 6.1TD. Para la resta de puntos, perfilaremos el consumo mensual o anual por periodos.

Para los clientes que solo se tiene su consumo anual, éste se pondera mediante un perfilado anual que otorga un peso en tanto por uno a cada hora en función del histórico de demanda energética. Existen también distintos perfiles en función del grupo tarifario de los consumidores finales.

3.0 MES/PER	15/05/2022	16/05/2022	17/05/2022	18/05/2022	19/05/2022	20/05/2022	21/05/2022
HORA	Previsión	Previsión	Previsión	Previsión	Previsión	Previsión	Previsión
1	0,00427851	0,00400424	0,00416739	0,00416739	0,00416739	0,00425979	0,00441357
2	0,00399287	0,00380966	0,00396503	0,00396503	0,00396503	0,00404319	0,00411197
3	0,00385923	0,0037303	0,00386438	0,00386438	0,00386438	0,00393579	0,00398086
4	0,00380974	0,00371181	0,00383047	0,00383047	0,00383047	0,00389606	0,00393288
5	0,00379816	0,00373843	0,0038572	0,0038572	0,0038572	0,00392485	0,00392613
6	0,00382507	0,00386976	0,00400301	0,00400301	0,00400301	0,00405349	0,00397856
7	0,003815	0,00419611	0,00431363	0,00431363	0,00431363	0,00437174	0,0040079
8	0,00342948	0,00462346	0,00476405	0,00476405	0,00476405	0,00479374	0,00376018
9	0,00174635	0,00293997	0,00300544	0,00300544	0,00300544	0,0030077	0,00204116
10	0,00188904	0,00340044	0,00347847	0,00347847	0,00347847	0,00348636	0,00229246
11	0,00200988	0,00355706	0,00364854	0,00364854	0,00364854	0,00367517	0,0025125
12	0,00530007	0,00905096	0,00931739	0,00931739	0,00931739	0,00940363	0,00657344
13	0,00547685	0,00915123	0,00944497	0,00944497	0,00944497	0,00953327	0,00671529
14	0,00549084	0,00878861	0,0090701	0,0090701	0,0090701	0,00912889	0,00656569
15	0,00536857	0,00795171	0,00817889	0,00817889	0,00817889	0,00815988	0,00615948
16	0,00206869	0,00302543	0,00313082	0,00313082	0,00313082	0,003045	0,00236651
17	0,00199976	0,00301419	0,00311835	0,00311835	0,00311835	0,00297327	0,00232569
18	0,00193972	0,00297786	0,00308847	0,00308847	0,00308847	0,00294357	0,00232489
19	0,00191351	0,00284087	0,00295167	0,00295167	0,00295167	0,00287359	0,00231183
20	0,00195224	0,00268704	0,0028136	0,0028136	0,0028136	0,00278879	0,00235944
21	0,00201534	0,00253224	0,00264582	0,00264582	0,00264582	0,00270297	0,00241151
22	0,00225295	0,00251904	0,00262896	0,00262896	0,00262896	0,00272779	0,00254047
23	0,00221295	0,00233758	0,00243104	0,00243104	0,00243104	0,00257156	0,00246788
24	0,00204282	0,00211497	0,00218202	0,00218202	0,00218202	0,00233698	0,00227508

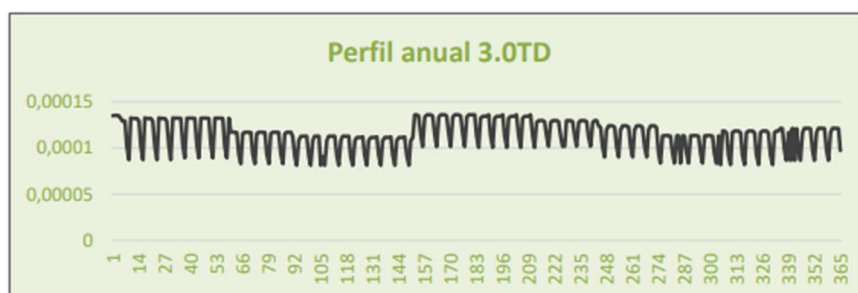


Ilustración 16. Perfil anual 3.0TD

REE es la entidad encargada de generar los perfiles anuales, se publican los perfiles iniciales a principios de año en la web de descarga de ficheros Esios, con la previsión de demanda esperada para el año en cuestión y se va actualizando según pase el año, hasta obtener al finalizar el año los perfiles finales.

En el caso de las 3.0TD se obtienen mejores resultados si se normalizan estos perfiles y se distribuyen por meses. Por ejemplo, en el caso de un cliente estacional que tiene la totalidad de su consumo los meses de julio-agosto, si distribuimos su consumo sobre el total anual, los resultados no serán favorables, ya que habremos asignado más consumo del debido sobre meses que no tiene consumo, y menos sobre los que sí tiene.

La metodología en este caso es coger el consumo total por mes/periodo y utilizando el perfil normalizado calculamos la previsión horaria.

3. Baja tensión (Tarifa 2.0TD)

Para los puntos de suministro de tarifas 2.0TD ocurrirá lo mismo que para los anteriores, tendremos puntos con curva de carga o bien puntos que debemos perfilar el consumo anual. Análogamente a las 3.0TD, existen unos perfiles para esta tarifa que REE publica a principios de año.

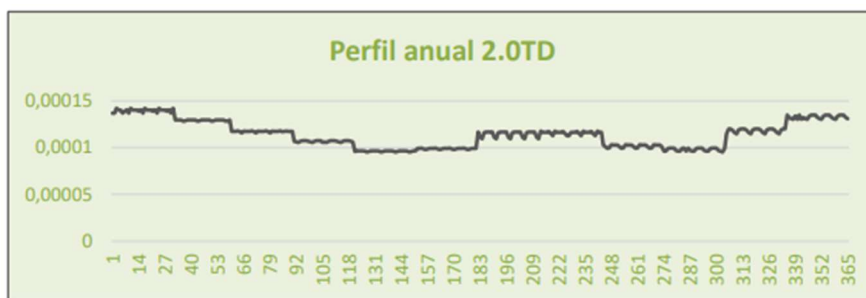


Ilustración 17. Perfil anual 2.0TD

2.3. Del proceso manual hacía la automatización

La precisión en las previsiones se adquiere mediante un enfoque detallado y específico para cada tarifa o tipología de clientes, considerando las características únicas de los diferentes tipos de clientes. ¿Que tenemos en cuenta a la hora de realizar las previsiones?

1. **Análisis de datos históricos** - se estudian los datos históricos de consumo por cada punto de suministro; cogiendo los datos más relevantes en el siguiente orden: *curvas de carga* → *consumo anual por periodo* → *consumo anual*
2. **Identificación de tendencias** – se analizan las curvas de carga, para identificar patrones de crecimiento o decrecimiento
3. **Variables de crecimiento por tarifas** – se consideran y anticipan subidas de temperatura, previsión de aumento de consumo o bajada por variación de previsión de cierre o aumento de producción.
4. **Modelos diferentes por tarifas** – en función de la recogida de datos se utiliza bien un perfilado mensual, anual o bien una media de los datos de las diferentes curvas de carga.

Modelizar las previsiones de electricidad implica ciertos enfoques matemáticos y estadísticos avanzados que no solo consideren el flujo de la demanda, sino que incorporen además los ciertos factores externos que influyen en el consumo energético. En este contexto, la comercializadora se encuentra ante el reto de optimizar sus operaciones, tanto en tiempo como en calidad.

La importancia de este trabajo radica en la necesidad de desarrollar modelos predictivos que no solo sean precisos, sino también adaptables a las diferentes condiciones y nuevas normativas del mercado energético. La identificación de patrones, la consideración de tendencias históricas, y la incorporación de variables relevantes se convierten en aspectos fundamentales para abordar este proyecto.

A continuación, estudiaremos la selección de variables, la evaluación de la precisión del modelo, utilizando cálculos con consumos finales y veremos cómo podemos implementar nuevas estrategias para seguir mejorando el método. A niveles prácticos, este trabajo busca contribuir al avance de la eficiencia operativa de una consultora al proponer soluciones matemáticas sólidas y aplicables a los desafíos específicos de las comercializadoras clientes de la consultora.

3. Modelos y metodología

En este apartado explicaremos los distintos modelos matemáticos que hemos usado para realizar la predicción utilizando las curvas de carga de la comercializadora mencionada anteriormente.

Para comprobar la calidad de cada modelo disponemos de más de 1800 curvas de carga para el año 2022 y nos servirán para evaluar mediante distintas métricas cuál de ellos es el más adecuado para nuestro problema, dando mejores resultados.

3.1. Notación, conceptos básicos y modelos base

Llamaremos **serie temporal** a un conjunto de observaciones de una variable en intervalos de tiempo regulares (horas, días, meses, trimestres etc.). Estos datos representan registros de una variable particular en función del tiempo, por ese motivo anotaremos una serie univariante como:

Y_t , con $1 \leq t \leq T$, donde T es el tamaño de la serie.

Hay que destacar la importancia del orden en el análisis de series temporales. A diferencia de muchos métodos estadísticos convencionales, donde las observaciones se consideran independientes y el orden no importa, en el análisis de series temporales el orden de las observaciones es fundamental y rompe con el supuesto de independencia, por los posibles siguientes motivos:

- Dependencia temporal: Las observaciones en una serie temporal suelen estar correlacionadas en el tiempo, lo que significa que el valor en un momento dado depende de los valores anteriores y posteriores en la serie.
- Efectos de memoria: las observaciones pasadas influyen en las futuras. Esto se debe a la presencia de efectos de memoria en muchos procesos del mundo real.
- Modelado y predicción: hay métodos de predicción que están diseñados específicamente para aprovechar la dependencia lineal tratada en el primer punto. Los modelos de series temporales, como los modelos autorregresivos (AR) y de media móvil (MA), se basan en las observaciones pasadas para predecir el futuro.
- Estacionariedad: La estacionariedad implica que las propiedades estadísticas de la serie no cambian con el tiempo, esto facilita de manera sustancial la modelización de la misma. Por tanto, para aplicar ciertas modelizaciones se requiera que la serie sea estacionaria; si no es así, esta se transforma para que lo sea.

Definiremos el análisis de series temporales como el conjunto de técnicas de estudio de series de observaciones dependientes observadas en el tiempo. En resumen, el análisis de series temporales se diferencia de otros métodos estadísticos debido a la importancia del **orden** y la **dependencia temporal** en las observaciones. Esto permite abordar cuestiones relacionadas con

la tendencia, la estacionalidad, la autocorrelación y la predicción de eventos futuros en base a datos históricos.

A continuación, veremos unos claros ejemplos de series temporales:

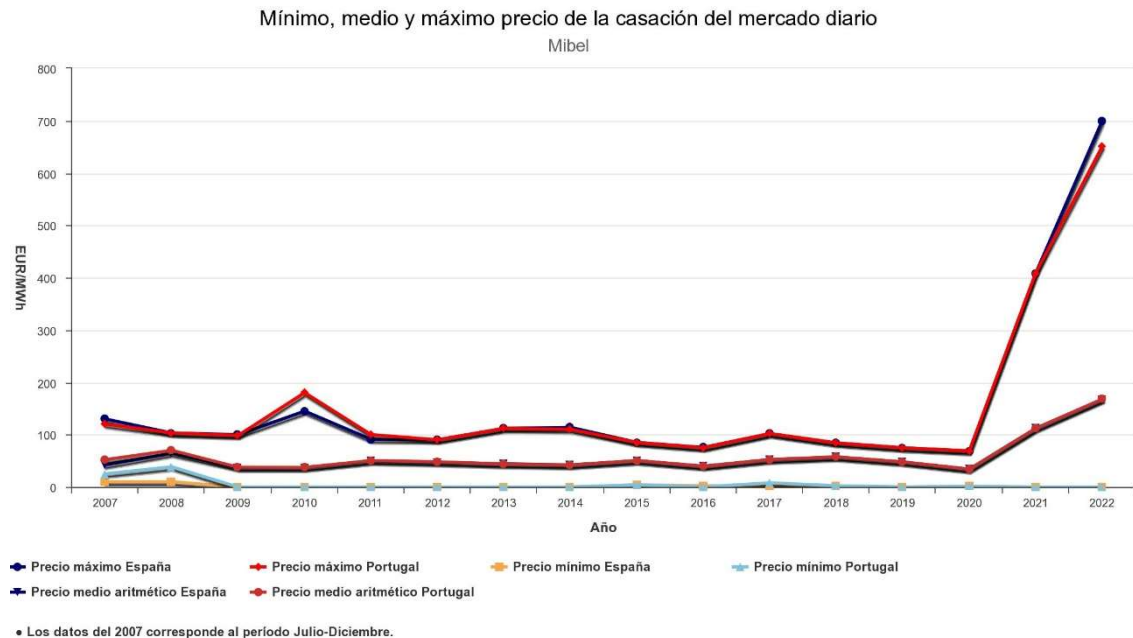


Ilustración 18. Evolución histórica precio OMIE mercado diario

EVOLUCIÓN DE LA DEMANDA (B.C.) (GWh) | SISTEMA ELÉCTRICO: Peninsular

Del 01/2020 al 12/2020

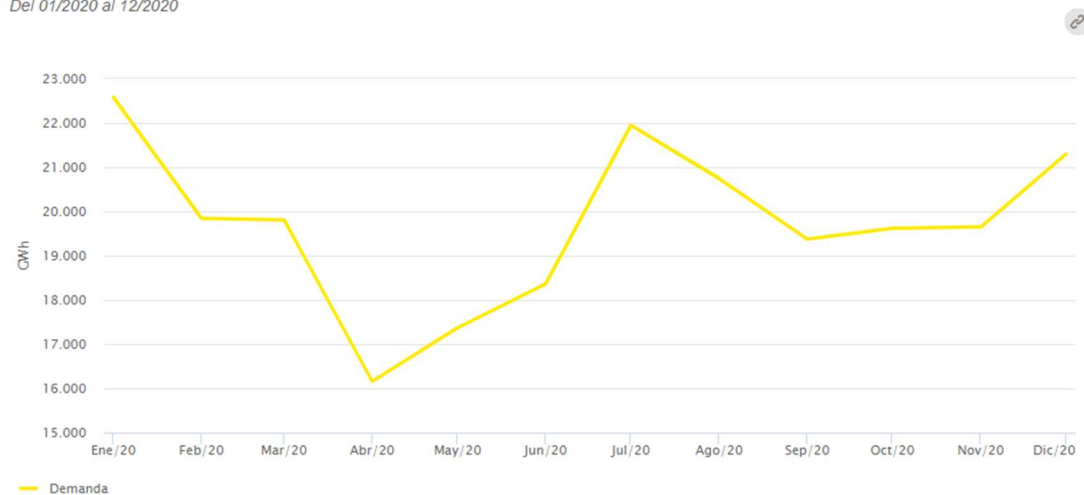


Ilustración 11. Demanda nacional (Territorio peninsular) durante el año 2020 (COVID-19)

En el 2020, según podemos observar en la *Ilustración 11.*, la evolución de la demanda se ha visto condicionada por la pandemia. La demanda de energía eléctrica en España durante el año 2020 se ha visto reducida en un 5,5 % respecto al año anterior, 0,4% menos que el registrado durante la crisis de 2009, y con un nivel de demanda similar al del año 2004.

En la siguiente figura *Ilustración 12*, podremos ver que los resultados diarios de los precios medios de OMIE durante el mes de enero 2023 han oscilado entre 0 y 150 €/MWh. El gráfico no muestra ninguna tendencia creciente o decreciente clara.

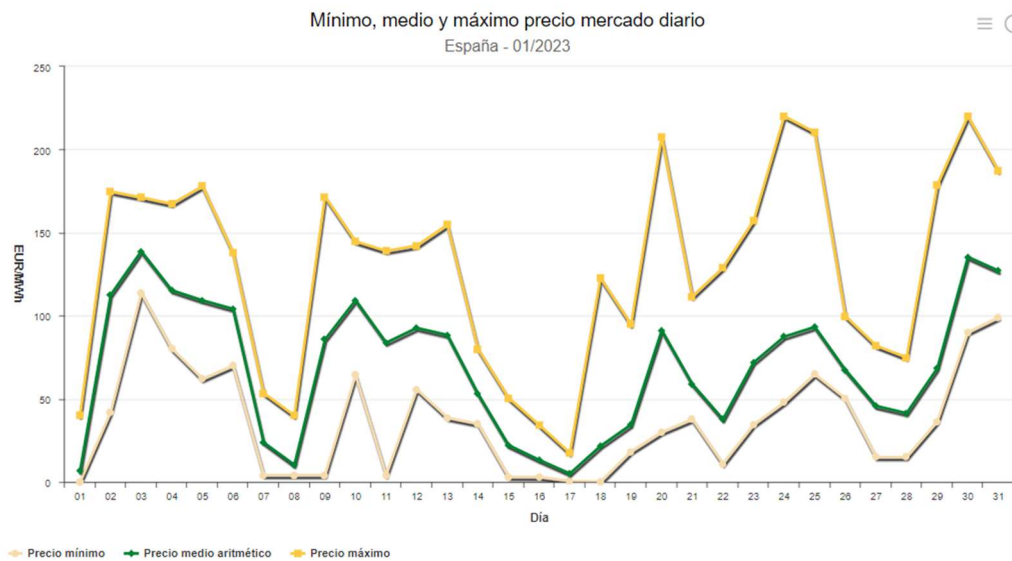


Ilustración 12. Precios resultados del mercado diario durante el mes de enero-2023

El objetivo es construir un modelo que permita representar la evolución de una serie temporal, basado en el histórico de los datos registrados de la serie y permita generar previsiones a futuro.

En nuestro estudio el consumo seguirá el modelo de una serie temporal discreta, ya que los contadores registran horariamente el consumo, y por tanto hay un número finito de observaciones recogidos a intervalos equidistantes en el tiempo.

Igual que la mayoría de las series, en este caso se tratará de un modelo estocástico, es decir solo se pueden determinar parcialmente los valores futuros en función de los pasados. En caso contrario, se trataría de un proceso determinista.

3.1.1. Proceso estocástico

Llamamos **proceso estocástico** a una colección de variables aleatorias $\{Y_t, t \in T\}$ siendo T habitualmente el tiempo. Dependiendo de la naturaleza del proceso estocástico, puede haber diferentes niveles de aleatoriedad y dependencia temporal entre las variables aleatorias. Los procesos estocásticos proporcionan un marco matemático para modelar y entender fenómenos que involucran aleatoriedad y cambio a lo largo del tiempo.

Si tenemos una colección finita de observaciones, $T = \{1, \dots, N\}$, diremos que el proceso estocástico es discreto. Si, por lo contrario, la colección es infinita, eso es $\{Y_t, t > 0\}$, diremos que el proceso estocástico es continuo. Como ya comentamos, en nuestro caso tenemos una colección finita de datos, por tanto, un proceso estocástico discreto.

Dada una observación de la variable Y_t , notaremos por $E(Y_t)$ su esperanza, $Var(Y_t)$ la varianza y $Cov(Y_t, Y_u)$ la covarianza entre dos variables aleatorias. Asumiremos que $E(Y_t^2) < \infty$ (1), ya que esto nos garantiza que la esperanza y la varianza de cualquier variable Y estarán bien definidas.

Veamos que, por la desigualdad de Jensen, si tenemos una función convexa γ , $\gamma(E(Y_t)) \leq E(\gamma(Y_t))$, por tanto $|E(Y_t)| \leq E|Y_t| < \infty$ y por tanto $E(Y_t^2)^{1/2} < \infty$ por Cauchy-Schwarz. Entonces tenemos:

$$Var(Y_t) = E(Y_t^2) - (E(Y_t))^2 < \infty \quad (2)$$

3.1.2. Estacionariedad

Dada una serie temporal, el objetivo es identificar un proceso estocástico que la haya podido generar. El concepto de estacionariedad es fundamental en el análisis de series temporales, ya que este proporciona una estabilidad estadística a lo largo del tiempo siendo esta la clave para identificar y modelar procesos estocásticos.

Si los datos recogidos se observan secuencialmente y tienen cierta dependencia esto implica que es probable que la serie temporal $\{Y_{t-1}, t \in T\}$ nos puede servir para predecir el comportamiento de la próxima observación Y_t . Es decir, el propósito sería descubrir cual es la distribución de Y_t condicionada al conjunto de información $\{Y_{t-1}, t \in T\}$, es decir $f(Y_t|Y_{t-1})$ y, en particular, su media y su varianza.

Se dice que una serie es **estacionaria** si sus propiedades estadísticas permanecen constantes a lo largo del tiempo. Es decir, se requiere que todas las variables aleatorias del proceso tengan la misma media, la misma varianza y que la covarianza entre dos variables cualesquiera del proceso estocástico dependa exclusivamente del número de periodos de distancia entre ellas. En resumen, si Y_t es una serie estacionaria en media y varianza se cumple que:

- $E(Y_t) = \mu, \forall t$
- $Var(Y_t) = E((Y_t - \mu)^2) = \sigma^2, \forall t$
- $Cov(Y_t, Y_{t-k}) = E((Y_t - \mu)(Y_{t-k} - \mu)) = \gamma_k, \forall t, k$, siendo μ, σ^2, γ_k momentos que no dependen del tiempo.

Por (1) y (2), hemos visto que todos estos valores están bien definidos y también finitos.

Denominamos la covarianza γ_k como la **autocovarianza** de orden k , y el conjunto de autocovarianzas del proceso estocástico como la función de autocovarianzas. Estas miden la relación lineal entre observaciones del proceso con una distancia de k periodos de tiempo.

Denominaremos **ruido blanco** al proceso estocástico estacionario con media cero, varianza constante y función de autocovarianzas nula.

Notaremos como $a_t \sim RB(0, \sigma^2)$ y por las definiciones anteriores tenemos:

- $E(a_t) = 0, \forall t$
- $Var(a_t) = \sigma^2, \forall t$
- $Cov(a_t, a_{t-k}) = 0, \forall t, k, k \neq 0$

Igual que en el caso de la autocovarianza, la **función de autocorrelación FAC** del proceso estocástico X se compone por el conjunto de las autocorrelaciones del proceso $\{\rho_k, k \in \mathbb{Z}\}$, siendo la autocorrelación:

$$\rho_k = \text{Corr}(Y_t, Y_{t-k}) = \frac{\gamma_k}{\gamma_0}$$

Para $k \geq 0$, la representación gráfica de ρ_k es el **correlograma**. También tenemos:

- $\rho_0 = \frac{\gamma_0}{\gamma_0} = 1$
- $|\rho_k| \leq 1$
- $\rho_k = \rho_{-k}$

Por último, tenemos la **función de autocorrelación parcial (FACP)**. La función de autocorrelación parcial, constituida por los coeficientes de autocorrelación parcial, también recoge el grado de asociación lineal entre las variables Y_t y Y_{t-k} , igual que la FAC, pero además tiene en cuenta las variables intermedias $Y_{t-1}, \dots, Y_{t-k+1}$. El coeficiente de autocorrelación parcial se denota por ρ_{kk} y es igual al coeficiente p_k de la siguiente ecuación:

$$Y_t = \alpha_0 + p_1 Y_{t-1} + p_2 Y_{t-2} + \dots + p_k Y_{t-k} + w_t \quad (3)$$

Llamaremos estimadores consistentes de los coeficientes FAC y FACP a sus análogos muestrales:

- **Coefficientes de autocorrelación muestrales** r_k , miden la interdependencia entre las observaciones:

$$r_k = \frac{\sum_{t=k+1}^T (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^T (Y_t - \bar{Y})^2}$$

- **Coefficientes de autocorrelación parcial muestrales**, lo denotamos por r_{kk} , y se obtienen análogamente de la estimación de p_k de la regresión (3).

Denotaremos **operador de retardo** $L^k Y_t$, para simplificar notación a la observación Y_{t-k} .

3.1.3. Modelo autorregresivo de medias móviles ARMA (p, q)

La idea general es construir un modelo sobre una dependencia temporal, esto implica que tenga una parte estructural y una independiente, con la siguiente estructura de un modelo lineal general:

$$Y_t = \delta + \psi_1 Y_{t-1} + \psi_2 Y_{t-2} + \psi_3 Y_{t-3} + \dots + a_t, \forall t$$

$$Y_t = \delta + \psi_1 L Y_t + \psi_2 L^2 Y_t + \psi_3 L^3 Y_t + \dots + a_t = \delta + Y_t (\psi_1 L + \psi_2 L^2 + \psi_3 L^3 + \dots) + a_t$$

Despejando, conseguimos la forma autorregresiva (4) y la media móvil (5) del modelo:

$$(1 - \psi_1 L - \psi_2 L^2 - \psi_3 L^3 - \dots) Y_t = \delta + a_t = \Psi_\infty(L) Y_t, \quad \text{forma } AR(\infty) \quad (4)$$

$$Y_t = \frac{\delta}{\Psi_\infty(L)} + \frac{a_t}{\Psi_\infty(L)} = \mu + a_t + \omega_1 a_{t-1} + \omega_2 a_{t-2} + \dots, \quad \text{forma } MA(\infty) \quad (5)$$

Por otro lado, denotaremos:

- **Polinomio autorregresivo** $\varphi_p(L) = 1 - \sum_{i=1}^p \varphi_i L^i$
- **Polinomio de medias móviles** $\theta_q(L) = 1 + \sum_{i=1}^q \theta_i L^i$

Aproximaremos $\Psi_\infty(L)$ por $\Psi_\infty(L) \cong \frac{\varphi_p(L)}{\theta_q(L)}$.

De esta manera, utilizando la forma autorregresiva del modelo (4) como $\delta + a_t = \Psi_\infty(L) Y_t$, entonces $\delta + \theta_q(L) a_t = \varphi_p(L) Y_t$, esta siendo la forma ARMA (p, q) del proceso.

3.1.3.1. Modelo autorregresivo AR(p)

El objetivo de este modelo es estudiar la regresión de una variable en función de sus valores anteriores en el tiempo, juntamente con un coeficiente residual, que en este caso le llamaremos ruido blanco. El modelo $AR(p)$, siendo este un caso particular del modelo ARMA (con q=0), se define como:

$$Y_t = \delta + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varphi_3 Y_{t-3} + \dots + \varphi_p Y_{t-p} + a_t, \forall t \text{ y } a_t \sim RB(0, \sigma^2)$$

O bien $\varphi_p(L) Y_t = \delta + a_t$.

El proceso será estacionario si todas las raíces de $\varphi_p(L)$ tienen módulo <1. Si el modelo es estacionario entonces se cumple lo siguiente:

- $\sum_{i=1}^p \varphi_i < 1$
- $E(Y_t) = \frac{\delta}{1 - \sum_{i=1}^p \varphi_i} = \mu$
- El correlograma decrece de manera exponencial, sin anularse
- La función de autocorrelación parcial es $\{\rho_{kk}; \rho_{kk} = 0 \text{ si } k > p\}$; es decir el correlograma parcial es 0 a partir del orden p.

3.1.3.2. Modelo de medias móviles MA(q)

Este modelo se utiliza expresar las predicciones como una combinación lineal de términos de error de las observaciones anteriores en el tiempo.

$$Y_t = \delta + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \theta_3 a_{t-3} + \dots + \theta_q a_{t-q}, \forall t \text{ y } a_t \sim \text{RB}(0, \sigma^2)$$

- El proceso será siempre estacionario
- Una condición necesaria para que Y_{t-k} influya en Y_t menos según k aumente, es decir observaciones cada vez mas lejanas influyen menos en las predicciones, es que las raíces del polinomio $\theta_q(L) = 1 + \sum_{i=1}^q \theta_i L^i$ **deben tener módulos <1**.
- $-\sum_{i=1}^q \theta_i < 1$
- La función de autocorrelación parcial decrece de manera exponencial, sin anularse.
- La función de autocorrelación presenta un corte a partir de un k finito.

Volviendo al modelo general ARMA (p, q), este combina los dos modelos anteriores en un solo modelo para describir la estructura de una serie temporal.

La ecuación de un modelo ARMA (p, q) se expresa como:

$$Y_t = \delta + \varphi_1 Y_{t-1} + \dots + \varphi_p Y_{t-p} + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}, \forall t \text{ y } a_t \sim \text{RB}(0, \sigma^2)$$

Donde resumiendo los modelos anteriores tenemos:

- δ es la constante
- $\varphi_1 \dots \varphi_p$ son los coeficientes asociados a los términos autorregresivos AR de orden p
- $\theta_1 \dots \theta_q$ son los coeficientes asociados a los términos de error de la componente MA de orden q
- $a_t, a_{t-1} \dots a_{t-q}$ son los errores de observaciones anteriores.

Si aplicamos el operador de retardo:

$$Y_t(1 - \varphi_1 L - \dots - \varphi_p L^p) = \delta + (1 + \theta_1 L + \dots + \theta_q L^q) a_t$$

$$\varphi_p(L) Y_t = \delta + \theta_q(L) a_t$$

Las propiedades del modelo son las siguientes:

- El modelo es estacionario si todas las raíces de los polinomios característicos de la parte autorregresiva $AR(p)$, el $\varphi_p(L)$ y el de $MA(q)$, $\theta_q(L)$ están fuera del círculo unitario.
- El modelo es invertible si las raíces del polinomio de medias móviles $MA(q)$ están fuera del círculo unitario.
- Si es estacionario, su media es $E(Y_t) = \frac{\delta}{1 - \sum_{i=1}^p \varphi_i}$, igual que para un modelo $AR(p)$.

Uno de los principales usos del modelo ARMA es la predicción de valores futuros en una serie temporal. Los parámetros del modelo se pueden utilizar para estimar las predicciones futuras, y la precisión de las predicciones se puede evaluar utilizando diversas métricas.

Los parámetros del modelo ARMA se pueden estimar utilizando métodos como la máxima verosimilitud o el método de mínimos cuadrados. La identificación correcta de los parámetros p y q es imprescindible para obtener estimaciones precisas.

Un modelo ARMA se considera causal si los coeficientes de la componente MA son finitos y el proceso es estacionario. La causalidad implica que las observaciones futuras solo dependen de las observaciones pasadas y de los términos de error actuales.

Estas propiedades subrayan la importancia de comprender la estacionariedad, invertibilidad y causalidad al estudiar los modelos ARMA, así como la importancia de técnicas de identificación adecuadas para determinar los órdenes óptimos del modelo.

3.1.4. Modelo autorregresivo integrado de medias móviles ARIMA (p, q, d)

El **modelo autorregresivo integrado de medias móviles** es el modelo resultado de un proceso de diferenciación aplicado a una serie que no es estacionaria en media, es decir, no presenta ninguna tendencia a largo plazo. Esta transformación consiste en aplicar diferencia hasta que esto ocurra y tengamos una serie estacionaria en media y presente una cierta tendencia a largo plazo.

Si la serie no es **estacionaria en media**, esto se debe a su parte autorregresiva, en particular, que alguna de las raíces del polinomio autorregresivo no esté fuera del círculo unidad. Supongamos que el modelo ARMA tiene una única raíz unitaria, entonces tenemos el polinomio autorregresivo:

$$\varphi_p(L) = \varphi_{p-1}(L)(1 - L)$$

Siendo $\varphi_{p-1}(L)$ estacionario, ya que solo había una raíz unitaria. Por tanto, tenemos el modelo ARMA con la siguiente expresión:

$$\varphi_{p-1}(L)(1 - L)Y_t = \delta + \theta_q(L)a_t$$

Y denotaremos por $\Delta Y_t = (1 - L)Y_t$, siendo este un proceso ARMA estacionario de orden $p-1$.

En caso de haber más de una raíz unitaria, la expresión del modelo sería:

$$\varphi_{p-1}(L)\Delta^d Y_t = \delta + \theta_q(L)a_t$$

Donde d es el número de diferencias necesarias para tener un proceso estacionario.

La estacionariedad en varianza supone que la dispersión de las observaciones de la serie se mantiene constante a lo largo de la misma.

Por ejemplo, un motivo de incumplimiento puede ser que la varianza dependa del nivel de la serie. Si esto ocurre, se corrige mediante la transformación Box-Cox, siendo Y_t , para $\lambda = 1$, estacionaria en varianza:

$$Y_t^{(\lambda)} = \begin{cases} Y_t^\lambda - 1, & \text{si } \lambda \neq 0 \\ \ln(Y_t), & \text{si } \lambda = 0 \end{cases}$$

3.2. Modelos propuestos

Los 2 modelos explorados, que pueden proporcionar buenos resultados y se ajustan a nuestro problema son los siguientes:

- Media de consumo de las últimas 3 semanas
- SARIMAX

3.2.1. Media del consumo de las últimas 3 semanas

Actualmente este es el método que se usa para realizar la previsión. Se trata de un modelo sencillo, que requiere de un nivel de revisión relativamente bajo.

Como ya se explicó previamente, la tarea de cada gestor asignado a cada comercializadora revisa mediante Excel el comportamiento de las últimas 3 semanas y decide cuál de ellas usar para realizar la media y obtener la previsión.

Este método será nuestro punto de comparación para determinar si otros modelos matemáticos con más capacidad de automatización pueden mejorar el método manual.

La fórmula para usar será:

$$Prediccion_{w,h} = \frac{C_{w-1,h} + C_{w-2,h} + C_{w-3,h}}{3} \quad (6)$$

Donde:

- $Prediccion_{w,h}$: Es la predicción para cierto día de la semana y cierta hora.
- $C_{w-1,h}, C_{w-2,h}, C_{w-3,h}$: Son los valores de energía de las últimas 3 semanas para el mismo día de la semana y hora.
- Generalizamos este modelo de esta forma, aunque sí puntualizamos que si $\exists x$ tal que $C_{w-x,h}$ es consumo de un día descartable, como por ejemplo un festivo, este puede ser no tomado en cuenta para el cálculo.

Dicho método de manera automatizada al igual que los siguientes modelos deberá tener en cuenta el efecto de festivos o cierres por vacaciones.

¿Cuál es el inconveniente principal de este método?

Tardará más en adaptarse a posibles cambios en tendencias de consumo. Por ejemplo, una primera o segunda semana con temperaturas bajas, no se tendrá en cuenta un incremento de consumo en los puntos de suministro de tipo residencial. De la misma manera, cuando las temperaturas vuelvan a incrementar, se tendrá en cuenta un consumo superior al actual.

Esto puede ser especialmente perjudicial en un tipo de comercializadora con una cartera formada principalmente por clientes domésticos.

3.2.2. SARIMAX (ARIMA)

Un enfoque importante en estadísticas y análisis de series temporales es la metodología de Box y Jenkins para el análisis de series temporales, basada en modelos ARIMA. Los siguientes cuatro pasos son esenciales para comprender y modelar adecuadamente las series temporales:

1. Identificar el modelo ARIMA:

Comienza por decidir qué transformaciones aplicar a la serie temporal para convertirla en una serie estacionaria. La estacionariedad es importante porque facilita la modelización. Esto puede implicar un proceso de selección para eliminar tendencias y estacionalidades.

Luego, se busca un modelo ARMA para la serie estacionaria. Esto implica determinar los órdenes p y q para la estructura autorregresiva y de media móvil, respectivamente.

2. Estimación del modelo ARIMA:

Una vez que se ha identificado provisionalmente un modelo ARMA para la serie estacionaria, se procede a estimar los parámetros AR y MA del modelo utilizando el método de máxima verosimilitud. Esto implica encontrar los valores óptimos de los coeficientes del modelo.

También se obtienen los errores estándar y los residuos del modelo (diferencias entre los valores reales y los valores predichos por el modelo).

3. Diagnóstico:

En esta fase, se realiza un diagnóstico de los residuos para verificar si tienen estructura de dependencia. Se comprueba que los residuos sigan un proceso de ruido blanco, es decir, que sean independientes e idénticamente distribuidos.

Si los residuos muestran estructura de dependencia, se ajusta el modelo para incorporarla. Esto puede implicar cambiar los órdenes p y q del modelo ARMA o aplicar transformaciones adicionales a la serie.

4. Predicción:

Una vez que se ha obtenido un modelo adecuado y validado en la fase de diagnóstico, se puede utilizar para realizar predicciones futuras en la serie temporal.

En resumen, el enfoque de Box y Jenkins es una herramienta ideal para analizar y modelar series temporales, pero es importante aplicar los pasos de manera correcta para obtener unos parámetros adecuados y una serie temporal válida.

- **Estacionariedad:** La suposición de estacionariedad es crucial en esta metodología. La estacionariedad implica que las propiedades estadísticas de la serie no cambian con el tiempo. Si una serie no es estacionaria, es necesario aplicar transformaciones para hacerla estacionaria, como diferenciación, transformaciones logarítmicas o un suavizado exponencial, para series que presentan una tendencia exponencial. Esto asegura que el modelo ARIMA pueda capturar la estructura subyacente de la serie de manera adecuada.
- **Independencia de Residuos:** En la fase de diagnóstico, se verifica que los residuos del modelo no tengan estructura de dependencia y sigan un proceso de ruido blanco. Esto es importante para garantizar que el modelo capture todas las relaciones significativas en los datos y que los errores de predicción son aleatorios.
- **Tamaño de la Serie Temporal:** Box y Jenkins recomiendan que la serie temporal tenga al menos 50 observaciones para aplicar esta metodología de manera efectiva. Esto se debe a que los modelos ARIMA, especialmente los más complejos, requieren un número suficiente de datos para estimaciones precisas. En series más cortas, la precisión de las estimaciones de parámetros puede verse comprometida.

SARIMAX es una extensión del modelo ARIMA el cual permite también usar variables exógenas (tales como podrían ser en nuestro caso por ejemplo la temperatura).

Los parámetros que rigen SARIMAX (p, d, q, P, D, Q, s) son:

- **p, d, q**
 - **p :** Orden de auto regresión (AR) – Representa la cantidad de términos autorregresivos (número de puntos pasados que usar de la time series para aprender), por lo que indica la dependencia de la predicción respecto al valor de sus valores pasados.
 - **d :** Orden de diferenciación (I), es el parámetro que indica cuantas veces se ha diferenciado la serie temporal hasta hacerla estacionaria. Si $d=0$ se presupone estacionariedad en la serie temporal.
 - **q :** Orden de medias móviles (MA), es la cantidad de medias móviles usadas, es decir, indica el número de puntos que usaremos para la media móvil, si el valor

de q es bajo tendremos una gran influencia de los últimos puntos por lo que puede ser más beneficioso en caso de cambios de tendencia.

- **P, D, Q, s**
 - **P, D, Q:** Son los homónimos de p,d,q pero aplicados a calcular la estacionalidad de la time series, por lo que por ejemplo si tenemos un Q alto estaremos suavizando nuestra estacionalidad con las últimas semanas..
 - **s:** Es la longitud del ciclo estacional, por ejemplo, en la venta de bañadores nuestra estacionalidad será anual dado que observaremos los picos de venta en los meses de mayo hasta agosto y no volveremos a observar esos picos hasta el próximo año en esos meses.

La selección de estos parámetros se llevaba a cabo de manera manual o mediante métodos heurísticos lo cual requiere de tiempo y puede estar expuesto a errores. Por lo que realizar este trabajo para más de 100.000 clientes no es posible, más aún cuando la serie temporal puede variar y por lo tanto después de cierto tiempo habría que volver a revisar los parámetros de cada serie temporal.

Por lo que para este trabajo utilizaremos AutoARIMA. Se trata de un algoritmo que realiza la búsqueda automática de estos parámetros. Lo que hace es coger un porcentaje de la serie temporal, en la mayoría de los casos es de un 10%-20% y del final de la serie temporal y lo usa como conjunto de prueba. Con el resto de la serie temporal entrena con diferentes configuraciones de parámetros y realiza la predicción comprobando cual es la configuración de parámetros que ha dado la mejor predicción respecto al conjunto de prueba.

La fórmula que usa SARIMAX para entrenar a partir de los datos conocidos de la serie temporal para luego realizar la predicción es de la siguiente forma:

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \dots + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \dots + a_t$$

Donde:

- Y_t es el valor de la energía en el tiempo t en nuestro caso
- $X_{1,t}, X_{2,t} \dots$ son las variables exógenas (que no utilizaremos en nuestro proyecto)
- $\beta_1, \beta_2 \dots$ son los coeficientes asociados a las variables exógenas
- $\varphi_1, \varphi_2 \dots$ son los coeficientes de autoregresión (p) que capturan la relación entre el valor actual y los valores anteriores
- $\theta_1, \theta_2 \dots$ son los coeficientes de medias móviles (q) que moldean la relación entre el error actual y sus errores anteriores.
- a_t representa el término de error en tiempo y sigue $a_t \sim RB(0, \sigma^2)$

3.3. Prueba retrospectiva

Disponemos de un total de 90 clientes para todo 2022 por lo que para comprobar la eficacia de cada método lanzaremos cada método desde el 1 de febrero de 2022 hasta el 24 de diciembre de 2022. En cada iteración marcaremos una fecha de inicio para realizar la predicción ajustándose a la realidad del negocio y la capacidad de obtener esos datos.

Cada jueves se realizan las previsiones, se comprueban y el viernes se realizan las compras de domingo a sábado de la siguiente semana por lo que para los diferentes tipos de clientes tendremos datos desde:

- 6.1TD - el último dato será del martes de la semana en cuestión.
- 3.0TD - el último dato será del martes de la semana anterior.
- 2.0TD - el último dato será del último día del mes anterior.

Con estas hipótesis de un escenario real realizaremos una predicción a nivel de cliente para cada semana del año siempre teniendo en cuenta el último dato disponible que tendremos para cada tipo de cliente y usando todos los datos disponibles hasta la fecha para realizar una mejor predicción.

La prueba retrospectiva por lo tanto contará con un total de 42 simulaciones por cliente. De esta manera podemos obtener una muestra suficientemente grande para evaluar las métricas durante todo un año pudiendo paliar el efecto que pueden tener los diferentes hábitos de consumo en distintas épocas del año.

Cada predicción será a nivel diario para ahorrar tiempo de computación dado que en un escenario real podemos tener hasta 100.000 clientes y realizar los cálculos a nivel horario sería costoso.

Por ello la predicción a nivel horario será:

$$\hat{Y}(w, h) = \hat{Y}(w) * (\frac{1}{3} \sum_{i=1}^3 C(w, h, i)) \quad (7)$$

Donde:

- $\hat{Y}(w, h)$ es la predicción de la energía el día de la semana w a la hora h
- $\hat{Y}(w)$ es la predicción total de la energía el día de la semana w
- $C(w, h, i)$ es el porcentaje de energía del día de la semana w a la hora h hace i semanas respecto al total de ese día.

3.4. Métricas

3.4.1. Cups APE (Error porcentual absoluto)

La primera métrica será a nivel de cup para evaluar la performance individual de cada modelo en cada cliente, para ello utilizaremos la siguiente formula:

$$APE = 100 * \frac{1}{Y_{total}} \sum_{w=1}^7 \sum_{h=1}^{24} |\hat{Y}(w, h) - Y(w, h)| \quad (8)$$

Donde:

- $\hat{Y}(w, h)$ es la predicción de la energía el día de la semana w a la hora h
- $Y(w, h)$ es la energía el día de la semana w a la hora h
- Y_{total} es la energía total de toda la semana.

La métrica nos dará la diferencia porcentual de desvío respecto a la suma de los desvíos horarios respecto a la energía total de la semana. Esta métrica será calculada a nivel de **CUPS – Simulación** por lo que para cada CUPS tendremos un total de 42 métricas las cuales las visualizaremos en media.

3.4.2. Cartera APE (Error porcentual absoluto)

La segunda métrica utilizada será la misma pero esta vez agregaremos todas las predicciones de cada prueba retrospectiva para simular un caso real. La compra en una comercializadora no se realiza cliente a cliente, si no se compra la energía total para todos los clientes que suministra la comercializadora.

Por ello realizaremos:

$$\hat{Y}_{comercializadora}(w, h) = \sum_{c=1}^C \hat{Y}(w, h, c) \quad (9)$$

Donde:

- $\hat{Y}_{comercializadora}(w, h)$ es la predicción de la energía total de la comercializadora el día de la semana w a la hora h
- $\hat{Y}(w, h, c)$ es la predicción de la energía el día de la semana w a la hora h del cliente c

Tras esto usaremos la formula (8) para calcular la métrica.

Esta distinción de agregar las predicciones de todos los clientes ya no solo es a nivel de escenario real, si no que puede ocurrir que alguno de los métodos a pesar de ser peor a nivel individual puede ser mejor a nivel agregado. Si tenemos una sobreestimación en algunos casos y subestimación en otros compensando el resultado final. En el sector, esto se llama apantallamiento de desvíos o bien efecto cartera.

Esto sería muy válido si tenemos una cartera equilibrada, si por lo contrario un gran porcentaje de la energía de nuestra cartera la componen pocos clientes tendríamos que usar el método que mejor se ajuste a estos, por ejemplo.

4. Resultados

4.1. Tarifas 6.1TD

Se realiza el test sobre 10 clientes con tarifa 6.1TD para un total de 42 fechas analizando una semana de previsión, según los datos disponibles indicados en el punto 3.3. Los resultados los vamos a estudiar a nivel cliente y a nivel fecha realización previsión (fcst_init).

Para cada cliente calculamos la métrica APE (8) en valor absoluto para el modelo SARIMAX y para el modelo actual por lo que tenemos un total de 420 métricas con los siguientes resultados a nivel cliente - fecha:

Tabla 1. Estadísticas de la métrica 6.1TD a nivel cliente - fecha

	Media	Mediana	1r cuantil	2o cuantil	3r cuantil
APE SARIMAX	68,20%	33,41%	17,12%	33,41%	57,39%
APE Actual	88,00%	29,17%	14,78%	29,17%	54,90%

El método actual es ligeramente mejor en todos los cuantiles, aunque observamos que la media aritmética para ambos es bastante alta. Esto es debido a que en alguna fecha – cups tenemos época de vacaciones y por lo tanto ambos métodos fallan, y este porcentaje afecta bastante en la media.

Vamos a observar los resultados a nivel de cliente:

Tabla 2. Mediana de la métrica 6.1TD a nivel cliente

Nº	CUPS	APE SARIMAX	APE Actual
1	ES0031405102099001DQ	15,65%	12,91%
2	ES0021000002506068XP	31,89%	28,16%
3	ES0021000009863359PJ	21,44%	21,04%
4	ES0021000010886136HA	61,71%	55,14%
5	ES0021000011335603XG	67,79%	62,68%
6	ES0031104233916001MC	21,93%	20,17%
7	ES0031405107231001VA	22,54%	18,75%
8	ES0031405312309001HK	59,17%	57,59%
9	ES0031406301651001NB	49,62%	43,22%
10	ES0031408107916001AP	8,57%	9,63%

Vemos que la mediana de los test a nivel de cups muestra un comportamiento similar, pero es ligeramente mejor para el método actual.

Vamos a estudiar algunos casos particulares para ver cómo funciona el modelo.

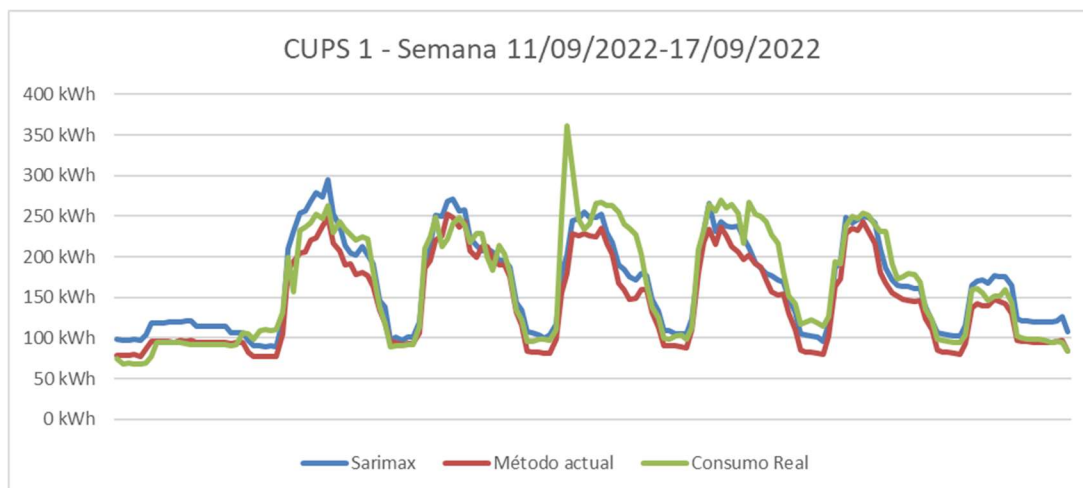


Ilustración 19. Resultados CUPS 1

En la *Ilustración 19*, podemos ver los resultados del CUPS n°1 de la Tabla 1., siendo este uno de los que mejores resultados tiene, tanto con el método actual como con el propuesto.

Podemos ver que, entre semana, el modelo propuesto se ajusta mejor a la curva real, pero vemos que los fines de semana se hace una sobreestimación. En particular para este CUPS, tenemos un desvío de la semana en cuestión del 12,34% con el método propuesto, mientras que con el método actual un 13,95%.

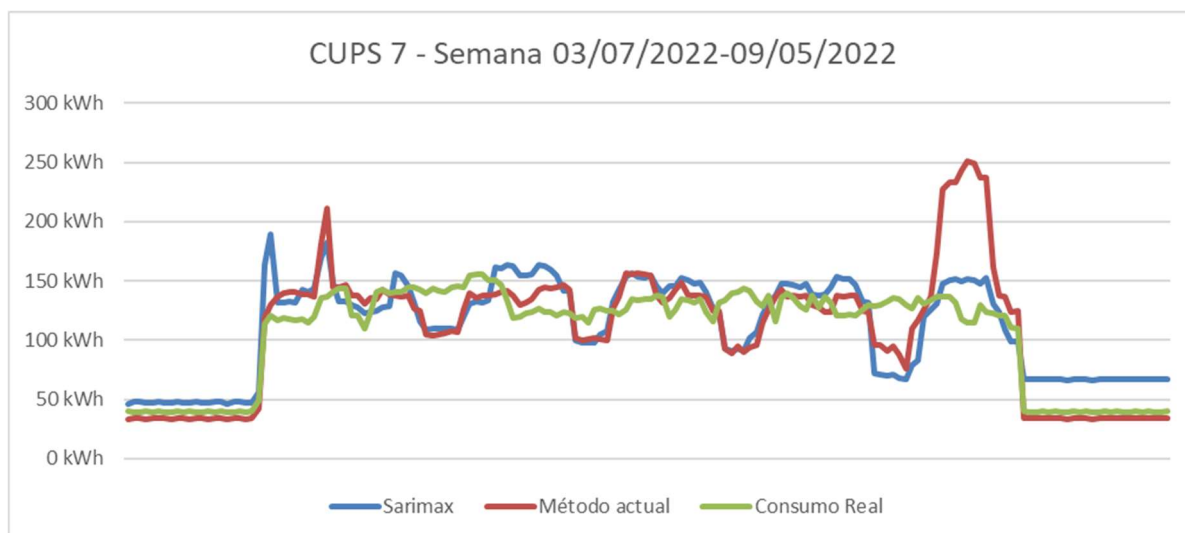


Ilustración 20. Resultados CUPS 7

En la *Ilustración 20*, podemos ver los resultados del CUPS n°7 de la Tabla 1. Para este CUPS, tenemos un desvío de la semana en cuestión del 20,73% con el método propuesto, mientras que con el método actual un 18,72%.

En este caso, igual que en el anterior podemos ver que de nuevo los fines de semana se hace una sobreestimación, pero que entre semana se ajusta mejor. Por otro lado, podemos ver claramente la desvirtuación del método actual del día 08/07 (viernes).

Finalmente veamos los resultados de la semana del 17 al 22 de abril en la *Ilustración 21*. Por un lado, vemos que para el día 18 de abril, ni el método actual ni el propuesto han podido predecir la bajada de consumo, que se debe a la presencia de un festivo. En este caso parece que el método actual se ajusta más, y además podemos decir que claramente tenemos un problema de sobreestimación de Sarimax los fines de semana. Obtenemos un desvío del 10,63% con el método actual, mientras que con el propuesto un 14,63%.

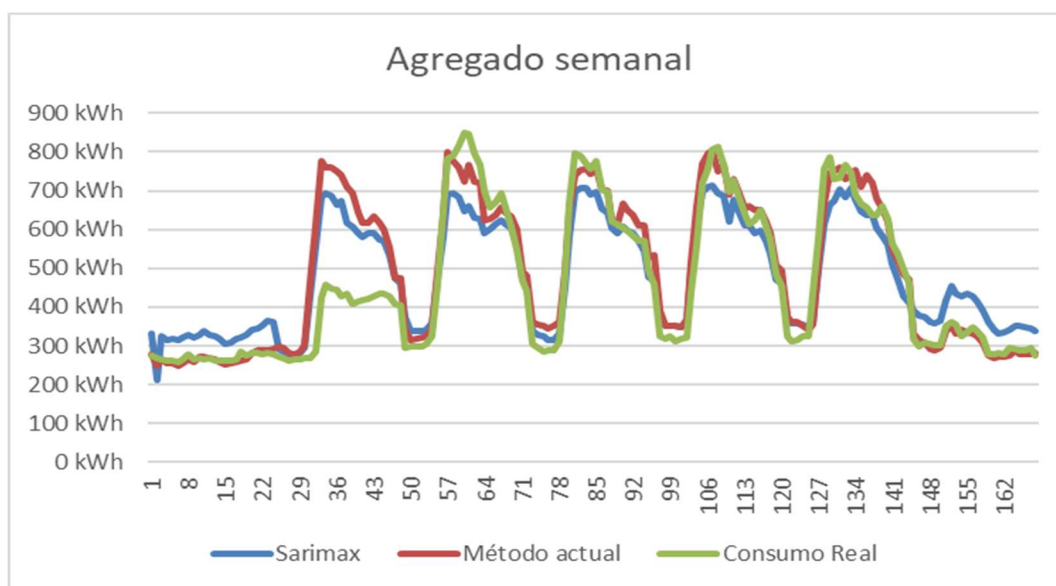


Ilustración 21. Resultados CUPS agregados

A nivel de fecha previsión aplicaremos la fórmula (9), donde sumamos la energía predicha de todos los CUPS para evaluar los resultados agregados, tal y como se haría en la realidad. Vemos que igual que a nivel CUPS, los resultados estadísticos son similares con los dos métodos, siendo el actual ligeramente mejor. Observamos en la *Ilustración 22*, que ambos métodos se equivocan más en los meses de verano, y algunos festivos, como Semana Santa, festivos del 12 de octubre, 1 de noviembre.

Para esta problemática, una solución sería incluir los festivos como variable exógena o bien usar el consumo de los fines de semana cuando están cerrados para sustituir los valores en esos días.

fcst_init	APE SARIMAX	APE Actual
2022-02-02	17,22%	16,71%
2022-02-09	21,92%	18,25%
2022-02-16	15,45%	13,37%
2022-02-23	18,39%	15,61%
2022-03-02	21,06%	22,12%
2022-03-09	20,77%	21,11%
2022-03-16	18,95%	17,79%
...		

Tabla 3. Resultados 6.1TD agregados a nivel fecha

	Media	Mediana	1r cuantil	2o cuantil	3r cuantil
APE SARIMAX	28,55%	25,03%	20,84%	25,03%	33,98%
APE Actual	27,89%	22,11%	18,71%	22,11%	32,24%

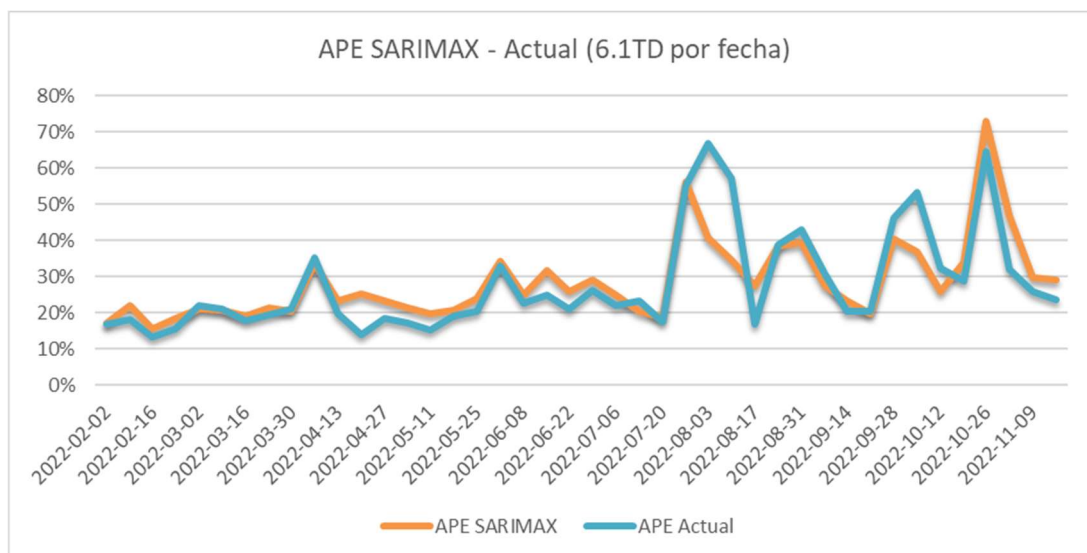


Ilustración 22. Resultados 6.1TD a nivel fecha

4.1.1. Tarifa 6.1TD (días laborables)

Según observamos en el apartado anterior, el modelo Sarimax sobreestima en la mayoría de los casos los fines de semana. Por ese motivo, una de las propuestas para mejorar el método sería introducir los comportamientos de los fines de semana como una variable exógena en el modelo o coger la mediana de los últimos 3 fines de semanas.

Veamos cuales serían los resultados en el supuesto caso de no tener en cuenta los sábados y los domingos. Podemos observar en la *Tabla 4*. una clara mejora en las métricas de Sarimax, siendo estas ya muy cercanas al método actual. Si lo observamos a nivel agregado por fecha en la *Tabla 5*. también se observa una clara mejora.

Tabla 4. Estadísticas de la métrica 6.1TD a nivel cliente - fecha (días laborables)

	Media	Mediana	1r cuantil	2o cuantil	3r cuantil
APE SARIMAX	75,53%	30,46%	15,40%	30,46%	54,84%
APE Actual	96,09%	28,61%	14,75%	28,61%	53,67%

Tabla 5. Resultados 6.1TD agregados a nivel fecha (días laborables)

	Media	Mediana	1r cuantil	2o cuantil	3r cuantil
APE SARIMAX	27,30%	24,17%	19,52%	24,17%	31,66%
APE Actual	28,78%	21,76%	19,21%	21,76%	32,72%

4.2. Tarifas 3.0TD

Se realiza la prueba sobre 30 clientes con tarifa 3.0TD para un total de 44 fechas analizando una semana de previsión, según los datos disponibles indicados en el punto 3.3. Igual que en el caso de las 6.1TD, los resultados los vamos a estudiar a nivel cliente y a nivel fecha realización previsión (fcst_init). Observamos que en el caso de las 3.0TD varias curvas presentan huecos, con lo cual no nos es posible realizar la comparativa de los modelos, por lo que no tendremos en cuenta estos días para los resultados.

Para cada cliente calculamos la métrica APE (8) en valor absoluto para el modelo SARIMAX y para el modelo actual.

Tabla 6. Estadísticas de la métrica 3.0TD a nivel cliente - fecha

	Media	Mediana	1r cuantil	2o cuantil	3r cuantil
APE SARIMAX	141,82%	32,41%	16,52%	32,41%	55,23%
APE Actual	138,40%	23,99%	12,38%	23,99%	43,66%

Volvemos a observar una media alta, mucho más que en el caso de las 6.1TD. En cuanto a los cuantiles aquí la diferencia es mayor entre Sarimax y el método actual, esto es debido a que para estos clientes 3.0TD solo tenemos datos de hasta 2 semanas previas por lo que al estar usando AutoArima, el cual identifica los parámetros automáticamente, puede estar interfiriendo una pendiente que no existe y extrapolándola a las siguientes semanas.

Tabla 7. Mediana de la métrica 3.0TD a nivel cliente

Nº	CUPS	APE SARIMAX	APE Actual
1	ES0031405525271001BF	10,74%	9,94%
2	ES0031408166028001QR	15,30%	17,63%
3	ES0031408574490021ZF	15,70%	15,04%
4	ES0031408166031001PB	17,12%	9,48%
5	ES0031405188803001WL	0,00%	0,00%
6	ES0031408166033001AX	22,53%	16,14%
7	ES0031405789629001RD	26,95%	26,29%
8	ES0031405116186001CE	23,06%	18,11%
9	ES0031405707146001JT	23,52%	17,82%
10	ES0031405268224001FH	0,00%	0,00%

Observamos que en general ambas métricas son similares y es más la tipología de cliente lo que hace que ambos métodos tengan un error mayor en la predicción. Aun así, vemos que el método actual, por lo general da mejores resultados.

Veamos algunos ejemplos:

En la *Ilustración 23*, vemos los resultados que obtenemos para una semana aleatoria para el CUPS 6 de la *Tabla 5*. Observamos cierta sobreestimación por parte de ambos métodos, sin embargo, vemos que igual que en el caso de las 6.1TD hay una sobreestimación del Sarimax los fines de semana. Para esta semana, tenemos un desvío del 30,08% con el método actual, y un 30,91% con Sarimax.

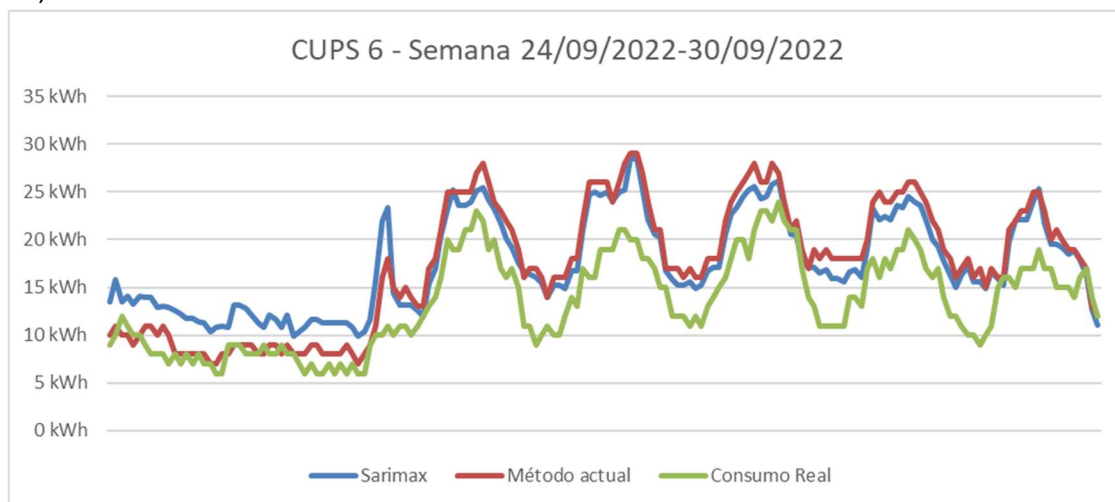


Ilustración 23. Resultados CUPS 6

En el siguiente gráfico, en la *Ilustración 24*. Tenemos los resultados agrupados por días del CUPS nº 8. De nuevo, podemos ver una sobreestimación del Sarimax el fin de semana.

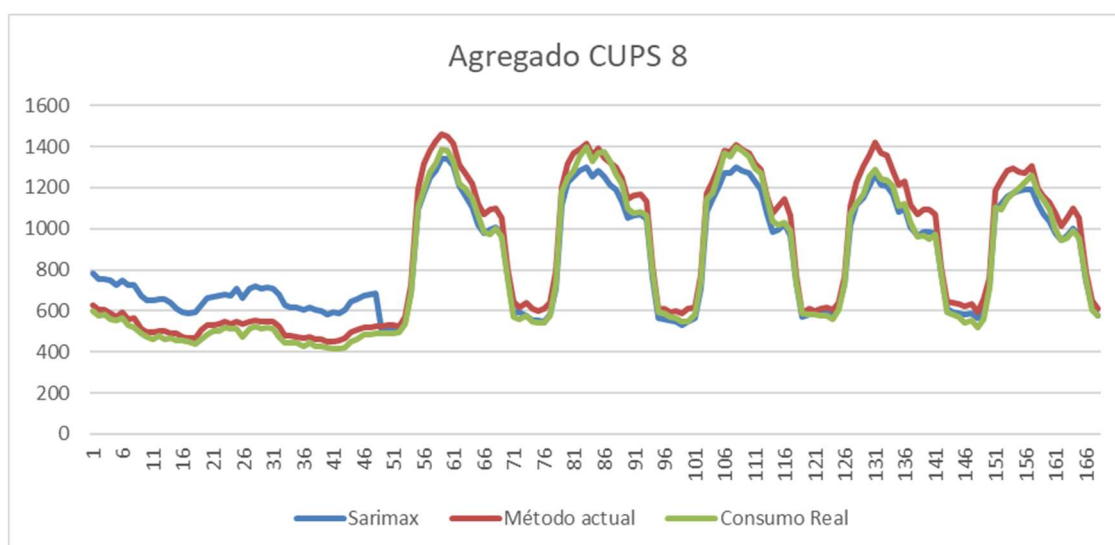


Ilustración 24. Resultados agregados CUPS 8

A nivel agregado por de fecha previsión, vemos que igual que a nivel CUPS, el actual es mejor, aunque ambos son mejorables. Observamos en la *Ilustración 25*. que ambos métodos se equivocan más en los meses de verano, y algunos ciertos festivos, como Semana Santa.

fcst_init	APE SARIMAX	APE Actual
2022-01-22	47,55%	23,23%
2022-01-29	38,80%	19,29%
2022-02-05	38,22%	20,73%
2022-02-12	34,30%	20,21%
2022-02-19	32,25%	21,14%
2022-02-26	30,28%	19,12%
2022-03-05	30,11%	20,88%
2022-03-12	29,66%	22,24%
2022-03-19	30,24%	21,75%
...		

Tabla 8.Resultados 3.0TD a nivel fecha

	Media	Mediana	1r cuantil	2o cuantil	3r cuantil
APE SARIMAX	41,41%	36,94%	31,60%	36,94%	44,79%
APE Actual	34,40%	27,02%	21,88%	27,02%	37,45%

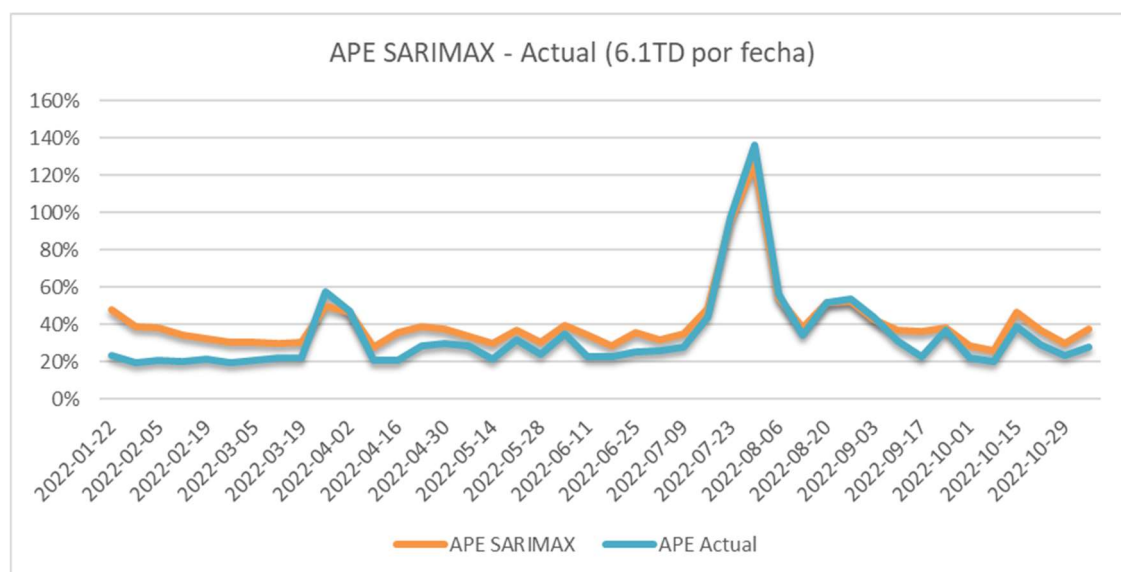


Ilustración 25. Resultados 3.0TD a nivel fecha

4.2.1. Tarifa 3.0TD (días laborables)

Igual que en el anterior caso, veamos cuales serían los resultados en el supuesto caso de no tener en cuenta los sábados y los domingos. A nivel cups-fecha, volvemos a ver una clara mejora usando SARIMAX pero sigue siendo mejor el método actual. A nivel agregado por fecha mejora respecto la previsión anterior, aunque no respecto el método actual.

Tabla 9. Estadísticas de la métrica 3.0TD a nivel cliente - fecha (días laborables)

	Media	Mediana	1r cuantil	2o cuantil	3r cuantil
APE SARIMAX	180,01%	28,23%	14,57%	28,23%	47,90%
APE Actual	197,00%	22,84%	11,36%	22,84%	42,82%

Tabla 10 Resultados 3.0TD agregados a nivel fecha (días laborables)

	Media	Mediana	1r cuantil	2o cuantil	3r cuantil
APE SARIMAX	38,55%	33,25%	28,63%	33,25%	38,86%
APE Actual	35,58%	26,44%	21,08%	26,44%	41,20%

5. Conclusiones

En resumen, vemos que el método actual funciona muy bien, debido a la naturaleza de los datos, ya que el consumo de un punto de suministro suele ser constante y periódico, sin tendencias crecientes o decrecientes sostenidas.

Los resultados del método propuesto Sarimax se ven afectados por esta falta de tendencias, sobre todo en las tarifas 3.0TD. En este caso, al hacerse la previsión a 3 semanas vista, si el modelo interpreta que hay una tendencia, esta se verá arrastrada semana tras semana. Este error, no obstante, no lo vemos en las tarifas 6.1TD, al tener datos más recientes a la fecha de predicción. Los resultados en las 6.1TD son prometedores, ya que, si se corrige el error de los fines de semana detectado, este se convierte en un modelo competitivo.

En un escenario real se podría usar un método u otro en función de la antigüedad de los datos de los que dispongamos para cada cliente.

Como propuesta de mejora, se podrían fijar los parámetros de Sarimax para evitar los problemas de tendencia y además integrar otras variables exógenas, como día de la semana, festivos, cambios de temperatura, etc. que puedan ayudar a una mejor previsión.

6. Bibliografía

Daniel Peña: Análisis de Series Temporales (2002). Alianza Editorial

Rob J Hyndman and George Athanasopoulos: Forecasting: Principles and Practice (3rd ed)
<https://otexts.com/fpp3/index.html>

Box, George E.P. y Jenkins, Gwilym M. (1973). Time Series Analysis: Forecasting and Control. Holden Day

<https://www.esios.ree.es/es/descargas>

[¿Cómo saber mi distribuidora de luz? - Audinfor System - Software sector energético](#)

<https://www.endesa.com/es/la-cara-e/sector-energetico/como-funciona-el-mercado-electrico-en-espana>

<https://www.factorenergia.com/es/blog/factura-luz/agentes-del-mercado-electrico/>

<https://es.support.somenergia.coop/article/1004-la-tarifa-2-0td>

<https://es.support.somenergia.coop/article/1116-la-tarifa-3-0td-periodos>

<https://solarfam.com/como-seran-los-nuevos-peajes-tarifarios-en-espana-a-partir-de-junio-para-el-autoconsumo/>

[Informe del Sistema Eléctrico Español 2020. \(ree.es\)](#)