

Comparativa de las técnicas de modelado de la previsión de la demanda eléctrica española.

Francisco ESPIGA FERNÁNDEZ

Máster en Ingeniería computacional y Matemática
Inteligencia Artificial

David ISERN ALARCÓN

Carles VENTURA ROYO

06/2016

© Francisco Espiga Fernández
Reservados todos los derechos. Está prohibido la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la impresión, la reprografía, el microfilme, el tratamiento informático o cualquier otro sistema, así como la distribución de ejemplares mediante alquiler y préstamo, sin la autorización escrita del autor o de los límites que autorice la Ley de Propiedad Intelectual.

FICHA DEL TRABAJO FINAL

Título del trabajo:	Comparativa de las técnicas de modelado de la previsión de la demanda eléctrica española.
Nombre del autor:	Francisco ESPIGA FERNÁNDEZ
Nombre del consultor:	David ISERN ALARCÓN
Nombre del PRA:	Carles VENTURA ROYO
Fecha de entrega (mm/aaaa):	06/2016
Titulación:	<i>Master en Ingeniería Computacional y Matemática</i>
Área del Trabajo Final:	Inteligencia Artificial
Idioma del trabajo:	Castellano
Palabras clave	demand forecasting, residual analysis, seasonal models

Resumen del Trabajo (máximo 250 palabras): *Con la finalidad, contexto de aplicación, metodología, resultados i conclusiones del trabajo.*

La previsión de la demanda del consumo eléctrico es una de las piezas clave en la operación del sistema eléctrico de transporte en tiempo real, ya que permite anticiparse para gestionar un escenario de operación lo menos restrictivo posible.

En este trabajo se han examinado distintas técnicas de construcción de modelos de previsión y su aplicación en la demanda eléctrica española. Los modelos lineales, que permiten la construcción de la curva de la demanda diaria como una combinación de valores función de la tipología del día; los de series temporales, que utilizan la relación de la señal de la demanda con sus valores pasados para prever los valores futuros y los modelos de alisado exponencial, que se sirven de la inercia inherente a la demanda y sus características estacionales para obtener su previsión.

Para analizar el desempeño de los previsores propuestos, se confrontarán a los existentes del operador del sistema español, Red Eléctrica y se estudiarán los residuos de la previsión de estos últimos para extraer información adicional que permita mejorar tanto las soluciones propuestas como las propias del operador del sistema.

Abstract (in English, 250 words or less):

Electricity demand forecasting is one of the linchpins of real time transmission system operation, as it allows for manage, by anticipation, the less constrained operation scenario.

In this paper, different techniques for forecasting model construction have been examined and their application to the Spanish electricity demand. Linear models, that allow the construction of the daily demand curve as a linear combination of values depending on the characteristics of the day; time series, that use the relationship between the demand signal with its past values to forecast the future ones and exponential smoothing models, that use the inherent inertia of the electricity demand and its seasonal characteristics.

To evaluate the performance of the proposed forecasters, they will be confronted against the existing forecasting systems of the Spanish TSO, Red Eléctrica and their residuals of the latter be studied to extract additional information allowing for improvement of the proposal and the own systems of the TSO.

Índice

1. Introducción	1
1.1 Contexto y justificación del Trabajo	1
1.2 Objetivos del Trabajo	1
1.4 Planificación del Trabajo	2
1.5 Breve resumen de productos obtenidos	4
1.6 Breve descripción de los otros capítulos de la memoria	4
2. Análisis exploratorio de la demanda eléctrica española	6
2.1 Introducción	6
2.2 Factores de los que depende la demanda	8
3. Estudio del error del previsor de Red Eléctrica	13
3.1 Introducción	13
3.2 Error de previsión en D-1	14
4. Preparación del set de datos	22
4.1 Introducción	22
4.2 Corrección de los días singulares	22
4.3 Modelado de las variables explicativas adicionales	23
4.4 Sets de datos de trabajo	27
5. Modelos lineales de previsión	28
5.1 Introducción	28
5.2 Variantes de modelado	28
5.3 Resultados	29
5.4 Selección del modelado de las distintas componentes	32
5.5 Perfiles	33
6. Modelos de series temporales y alisado exponencial	37
6.1 Introducción	37
6.2 Modelos ARIMA	37
6.3 Modelo de doble estacionalidad de Holt-Winters	40
6.5 Modelo mixto con triple estacionalidad de Holt Winters	41
7. Benchmarking	43
7.1 Introducción	43
7.2 Medidas de centralidad y dispersión	43
7.3 Medidas del error	43
7.4 Test de Diebold-Mariano	46
7.5 Distribución de los residuos	47
7.6 Análisis de los modelos Holt Winters	53
8. Conclusiones	63
9. Glosario	65
10. Bibliografía	66
11. Anexos	68
Anexo 1- Tabla de pesos eléctricos de cada Comunidad Autónoma:	68
Anexo 2 – Tablas de los parámetros de los diferentes modelos lineales	69
Anexo 3 – Tablas de los parámetros de los modelos ARIMA	73

Lista de figuras

Figura 1: variabilidad de la demanda eléctrica española	6
Figura 2: evolución de la demanda en Europa 2004-2015	7
Figura 3: variación mensual de la demanda.....	8
Figura 4: ciclo semanal de la demanda eléctrica (valores medios anuales)	8
Figura 5: evolución anual de la demanda.....	9
Figura 6: variación horaria de la demanda	10
Figura 7: correlogramas total y parcial de la demanda.....	11
Figura 8: estacionalidad anual de la demanda	11
Figura 9: efecto de los días festivos en la demanda	12
Figura 10: error de previsión D REE.....	13
Figura 11: Error de previsión D-1 REE incluyendo atípicos y detalle	14
Figura 12: error de previsión horario	14
Figura 13: error de predictor horario sin atípicos.....	15
Figura 14: relación entre el error cometido y la demanda	15
Figura 15: autocorrelogramas parcial y total del error de previsión.....	16
Figura 16: autocorrelograma diferencia del error de previsión	16
Figura 17: distribución del residuo sin atípicos.....	17
Figura 18: distribución de atípicos según tipología de día.....	18
Figura 19: distribución horaria de atípicos del error	18
Figura 20: distribución mensual de atípicos del error	19
Figura 21: distribución semanal de atípicos del error	19
Figura 22: distribución del error según tipología de día.....	20
Figura 23: distribución del error según día de la semana.....	20
Figura 24: distribución del error según la hora del día	21
Figura 25: distribución del error según mes	21
Figura 26: ejemplo de función de estacionalidad numérica.....	23
Figura 27: ejemplo de función de estacionalidad de variable binaria	24
Figura 28: ejemplo de función de estacionalidad con series de Fourier.....	25
Figura 29 : funciones generatrices de la estacionalidad diaria.....	25
Figura 30: estimación de la tendencia con medias anuales	26
Figura 31: media del error de previsión de los modelos lineales.....	29
Figura 32: desviación típica del error de previsión de los modelos lineales.....	30
Figura 33: cuantiles extremos del error de previsión de los modelos lineales ..	30
Figura 34: R2 ajustado de los modelos lineales	31
Figura 35: perfil semanal de la demanda	34
Figura 36: perfil mensual de la demanda	34
Figura 37: perfil horario de la demanda.....	35
Figura 38: perfil según tipología del día de la demanda.....	35
Figura 39: ejemplo de previsión de demanda con el uso de perfiles.....	36
Figura 40: diagrama de flujo de generación del modelo.....	41
Figura 41: diagrama de flujo de explotación del modelo	42
Figura 42: scores de entrenamiento de los distintos modelos.....	44
Figura 43: scores de validación de los distintos modelos.....	45
Figura 44: distribuciones del error de entrenamiento de los distintos modelos.	47
Figura 45: distribuciones del error de entrenamiento de los modelos lineales..	48
Figura 46: distribuciones del error de entrenamiento de los modelos ARIMA...48	

Figura 47: Error de entrenamiento de los modelos Holt-Winters.....	49
Figura 48: distribuciones del error de validación de los distintos modelos.....	50
Figura 49: distribuciones del error de validación de los modelos lineales.....	50
Figura 50: distribuciones del error de validación de los modelos ARIMA.....	51
Figura 51: distribución del error de validación de los modelos Holt-Winters.....	52
Figura 52: Error de entrenamiento de los modelos Holt-Winters.....	54
Figura 53: distribución del error de validación de los modelos Holt-Winters.....	54
Figura 54: Tendencia estimada mediante Loess y previsión para la validación	55
Figura 55: boxplot según la tipología de día.....	56
Figura 56: boxplot según el mes (enero-abril).....	57
Figura 57: boxplot según el mes (mayo-agosto).....	57
Figura 58: boxplot según el mes (sept-dic).....	58
Figura 59: dispersión de los errores superiores a 2000 MW según modelo.....	60
Figura 60: dispersión de errores superiores a 2000 MW por día de la semana	60
Figura 61: dispersión errores superiores a 2000 MW según tipología de día.....	61

1. Introducción

1.1 Contexto y justificación del Trabajo

La demanda eléctrica en el Sistema Eléctrico español se caracteriza por una gran dependencia de la época del año así como de otros factores como el clima o el comportamiento humano, lo que provoca en ocasiones fuertes desvíos entre la previsión y realización, que se traducen en grandes costes económicos.

Este trabajo ha buscado confrontar los previsores actuales del Operador del Sistema de Transporte (o TSO por sus siglas en inglés Transmission System Operator) contra algoritmos como Holt Winters, ARIMA o la caracterización mediante perfiles-tipo y valorar si es posible una mejora respecto de los sistemas de previsión existentes. En la actualidad, la previsión publicada por Red Eléctrica consiste en una crónica de previsión para el día D y D+1 (día de realización de la previsión y posterior), sin que existan actualizaciones horarias de la misma.

Como objetivo secundario, el trabajo ha intentado dar respuesta a los errores de previsión existentes así como justificar los desvíos de previsión acorde a las diferentes coyunturas que puedan presentarse, como festividades, variaciones de temperatura, etc. y a las épocas de transición entre distintos regímenes de demanda (verano e invierno).

1.2 Objetivos del Trabajo

- Caracterizar la demanda eléctrica en España respecto de sus variables estructurales y coyunturales.
- Analizar el error de previsión de los sistemas existentes así como la variabilidad en los perfiles-tipo.
- Prever la demanda mediante series temporales y perfiles característicos.
- Valorar la mejora de los previsores mediante la utilización de los preexistentes como benchmark.

1.3 Enfoque y método seguido

El campo de la previsión eléctrica es uno de los más ricos en lo que a técnicas utilizadas se refiere. Estas técnicas cubren todo el espectro de transparencia algorítmica, desde modelos lineales fácilmente interpretables hasta las redes neuronales más opacas, pasando por modelos de series temporales o de espacios de estado.

Este trabajo se ha construido en torno a los pilares de la transparencia y la reproducibilidad, buscando mediante las diferentes técnicas utilizadas cómo obtener previsiones de demanda lo suficientemente precisas mediante modelos fácilmente interpretables y con datos de libre disponibilidad.

Del mismo modo, de las cuatro herramientas de partida, R, MATLAB, Python y AMPL, se decidió trabajar con R al ser un lenguaje de código abierto con buenas librerías de desarrollo para el campo de trabajo. MATLAB y AMPL requerían licencias por lo que la reproducibilidad podría verse comprometida y si bien Python también es de código abierto, determinados solvers de optimización para la calibración de los modelos no lo son, por lo que finalmente y por simplicidad, se decidió utilizar R como la única herramienta de desarrollo.

1.4 Planificación del Trabajo

Desglose de los paquetes de trabajo (WP)

WP01

- Contenidos:
 - Introducción.
 - Minería y limpieza de datos.
 - Fecha estimada finalización: semana 1 febrero 2016

WP02

- Contenidos:
 - Análisis exploratorio de la demanda eléctrica española.
 - Estudio del error de previsión de Red Eléctrica.
 - Estudio de los valores atípicos.
- Fecha estimada finalización: semana 1 marzo 2016

WP03

- Contenidos:
 - Generación de los modelos lineales de previsión
 - Benchmark de los distintos modelados de las variables explicativas
 - Elección del mejor modelado de las variables explicativas
- Fecha estimada finalización: semana 1 abril 2016

WP04

- Contenidos:
 - Modelos de previsión mediante series temporales
 - Modelos de previsión mediante alisado exponencial
- Fecha estimada finalización: semana 1 mayo 2016

WP05

- Contenidos:
 - Benchmarking resultados
- Fecha estimada finalización: semana 3 mayo 2016

WP06

- Contenidos:
 - Documentación del trabajo
- Fecha estimada finalización: semana 3 mayo 2016

WP07

- Contenidos:
 - Aplicación R
 - Manual usuario aplicación R
- Fecha estimada finalización: semana 3 mayo 2016

Programación del proyecto

WP	ENERO	FEBRERO
WP_01		
WP_02		

WP	MARZO	ABRIL
WP_02		
WP_03		
WP_04		
WP_05		
WP_06		
WP_07		

WP	MAYO	JUNIO
WP_04		
WP_05		
WP_06		
WP_07		

1.5 Breve resumen de productos obtenidos

- Previsor mediante modelos lineales.
- Aplicación web-R de previsión de la demanda.
- Estudio de caracterización del error de los predictores actuales de Red Eléctrica.
- Benchmark de los predictores propuestos y los existentes.

1.6 Breve descripción de los otros capítulos de la memoria

2. Análisis exploratorio de la demanda eléctrica española

En este capítulo se han estudiado los distintos factores coyunturales que condicionan la demanda española, como puede ser la hora del día o si éste es festivo o no. Además, se analizarán las componentes cíclicas (anual, semanal) de la demanda y la tendencia de la demanda a largo plazo.

El objetivo fue obtener la serie de demanda eléctrica con granularidad horaria desde 2007 (8 años completos de datos) y de la previsión, para los horizontes proporcionados por Red Eléctrica (D, D-1) y comprender qué factores se deben considerar a la hora de modelar los sistemas de previsión.

3. Estudio del error de previsión de Red Eléctrica y datos atípicos

En este capítulo se ha analizado el error de los predictores actuales de Red Eléctrica, focalizándose en el error del predictor D-1 al considerar la víspera como el horizonte de trabajo. Se ha buscado la correlación del error de previsión con los factores coyunturales y comprender la causa detrás de errores de previsión atípicos, dando un valor a la frontera entre ellos y los errores de previsión esperables.

El objetivo del capítulo fue comprender las causas de los errores de previsión y obtener las series temporales de previsión y demanda corregidas para alimentar los modelos de previsión. El objetivo secundario consistió en obtener un testigo contra el que comparar los distintos modelos diseñados para evaluar su bondad.

4. Preparación del set de datos

En este capítulo se justifica la inclusión de las diversas variables explicativas (componentes cíclicas, tendencia, días festivos) y las distintas opciones de modelado disponibles para cada una de ellas.

El objetivo fue el de obtener un set de datos ampliados, de las series temporales de previsión y demanda junto con las variables exógenas finalmente incluidas (retardos, estacionalidades, tipo de día) y sus variantes de modelado (series de Fourier, variables binarias o variables numéricas).

5. Modelos lineales

En este capítulo se han creado y comparado los distintos modelos lineales de previsión. Su determinación en primer lugar obedece por un lado a su simplicidad de realización, ya que es sencillo considerar distintas variables explicativas simultáneamente y de evaluación del impacto de éstas, mediante la contribución individual de cada una de ellas a la varianza explicada. Por otro lado, de este modo las variables explicativas con distintas opciones de modelado pueden ser comparadas para escoger aquel o aquellos que modelen mejor esa característica de la señal para incluirlos en los modelos de series temporales y alisado exponencial.

Además de esta información de gran utilidad, se ha obtenido e implementado en R el algoritmo de generación de los perfiles de demanda basado en modelos lineales.

6. Modelos de series temporales y alisado exponencial

En este capítulo se ha modelado la demanda tanto mediante técnicas de Box-Jenkins, para los modelos ARIMA basados en series temporales como mediante los algoritmos Holt-Winters de doble y triple estacionalidad en el caso de los modelos de alisado exponencial. El modelado de las variables explicativas de estacionalidad y tipología del día se ha restringido a los modelos que han proporcionado los mejores resultados en los modelos lineales.

El uso de estos modelos junto a los lineales permite modelar la demanda desde una perspectiva temporal (ARIMA), una de estabilidad coyuntural (modelos lineales) y por último una inherente a la inercia de los fenómenos físicos relacionados con la demanda (Holt Winters). El objetivo de este capítulo consistió en obtener los modelos arima y holt winters para compararlos a los lineales e integrarlos en la aplicación de previsión.

7. Benchmarking de resultados

En este capítulo se confrontarán todos los modelos diseñados y los existentes mediante la metodología de Diebold-Mariano y la comparación de otras variables escogidas para el scoring.

2. Análisis exploratorio de la demanda eléctrica española

2.1 Introducción

La previsión de la demanda eléctrica es la primera etapa en la operación del sistema de transporte eléctrico y condiciona la programación de la generación y la gestión de la red de transporte, siendo su fin último el de garantizar en todo momento el perfecto equilibrio entre la oferta y la demanda de energía.

Esta demanda no es estática sino que depende tanto de factores estructurales, como el día de la semana, la hora o la estación del año como de variables exógenas como pueden ser la temperatura o la nubosidad. Todas estas posibilidades definen la coyuntura que configurará el perfil de demanda de un día determinado.

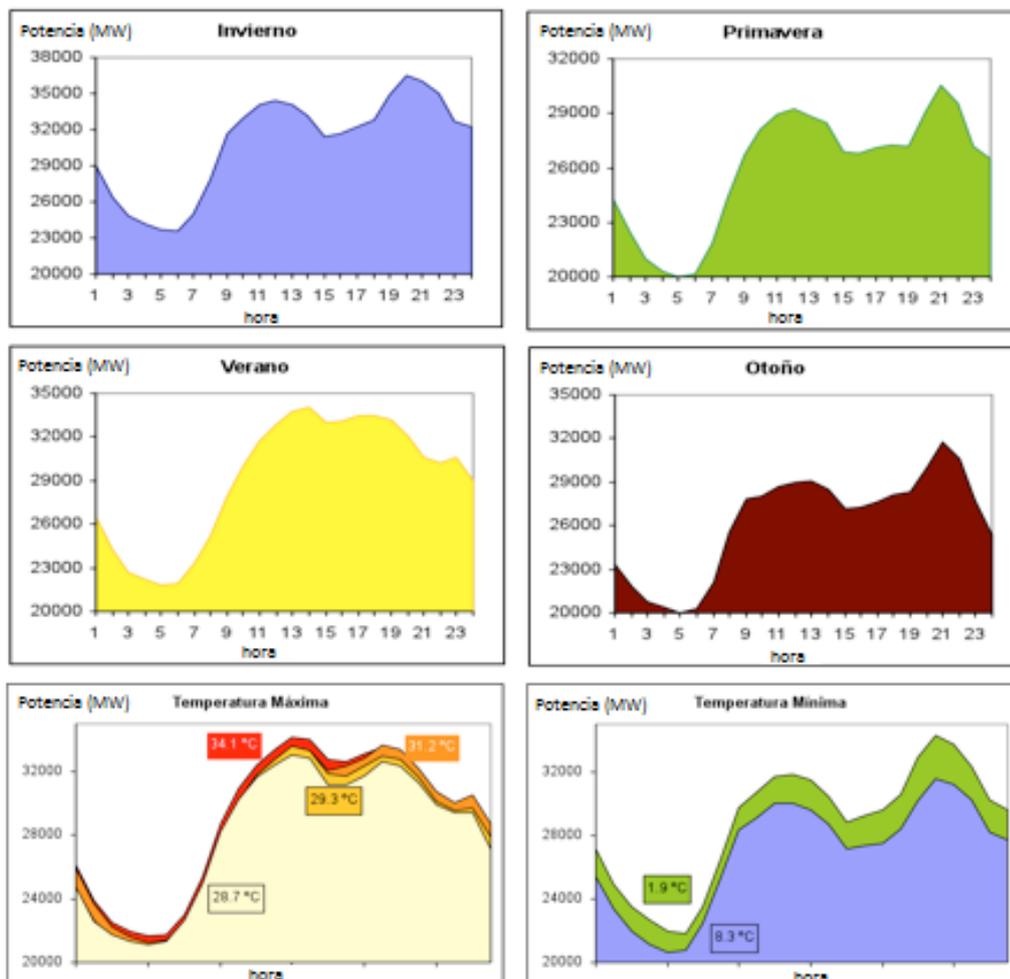


Figura 1: variabilidad de la demanda eléctrica española (*Red Eléctrica de España*)

Además de todos estos factores coyunturales se ha tenido en consideración también la evolución de la tendencia del consumo eléctrico.

Desde la crisis de 2008, en la mayoría de los países europeos se ha visto cómo una tendencia de consumo eléctrico creciente se ha frenado, presentando desde entonces un crecimiento prácticamente nulo, incluso contrayéndose en algunos casos. Debido no solo a una ralentización del tejido industrial, sino también a nuevas tecnologías como la iluminación de bajo consumo y edificios y dispositivos cada vez más eficientes energéticamente, que han hecho que en el último lustro las previsiones alcistas se hayan modificado hacia perspectivas de estabilidad y contracción.

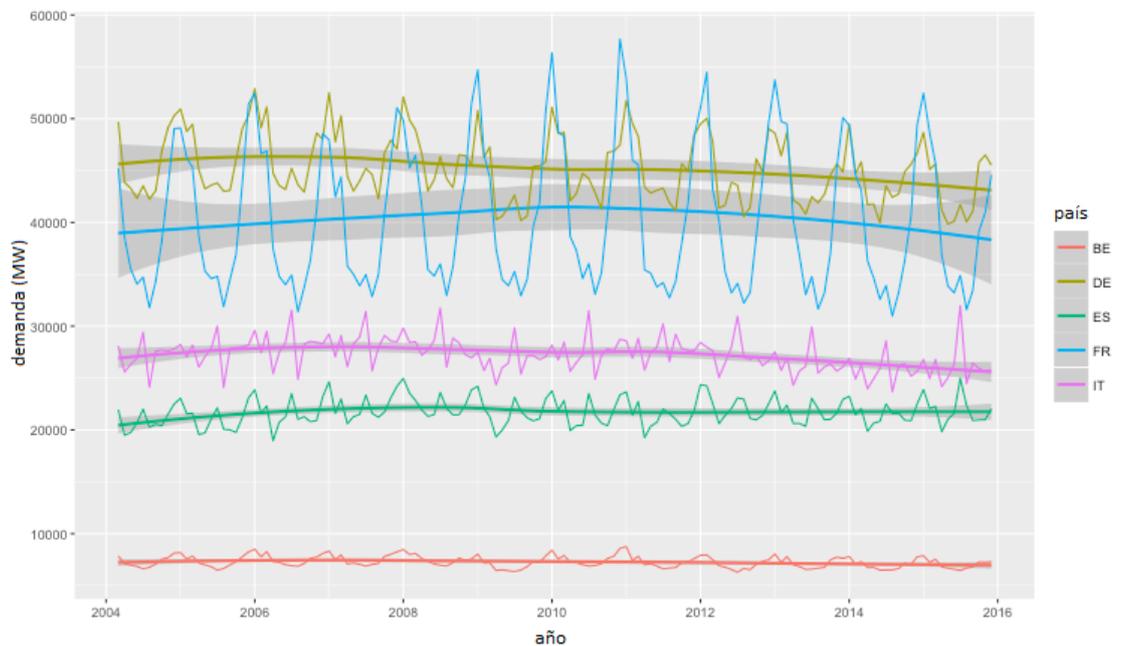


Figura 2: evolución de la demanda en Europa 2004-2015 (ENTSO-E 2016)[1]

2.2 Factores de los que depende la demanda

Estacionalidad anual de la demanda:

En cuanto a la demanda en España, asumiendo que la demanda de cada mes proviene de una distribución normal, se observa en el diagrama de cajas una estacionalidad con picos en torno a los meses más fríos (noviembre-febrero) y el más cálido (julio) con valles entre marzo y junio y de agosto a octubre.

Esto pone en evidencia una sensibilidad de la demanda eléctrica a la temperatura, tanto por los usos de calefacción como de climatización y también a la actividad económica, que provoca una disminución en agosto del consumo eléctrico a pesar de ser éste un mes igualmente cálido.

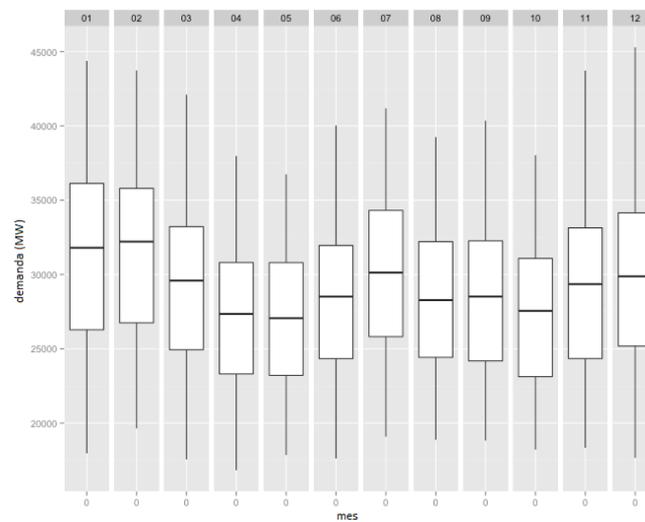


Figura 3: variación mensual de la demanda

El ciclo semanal de la demanda:

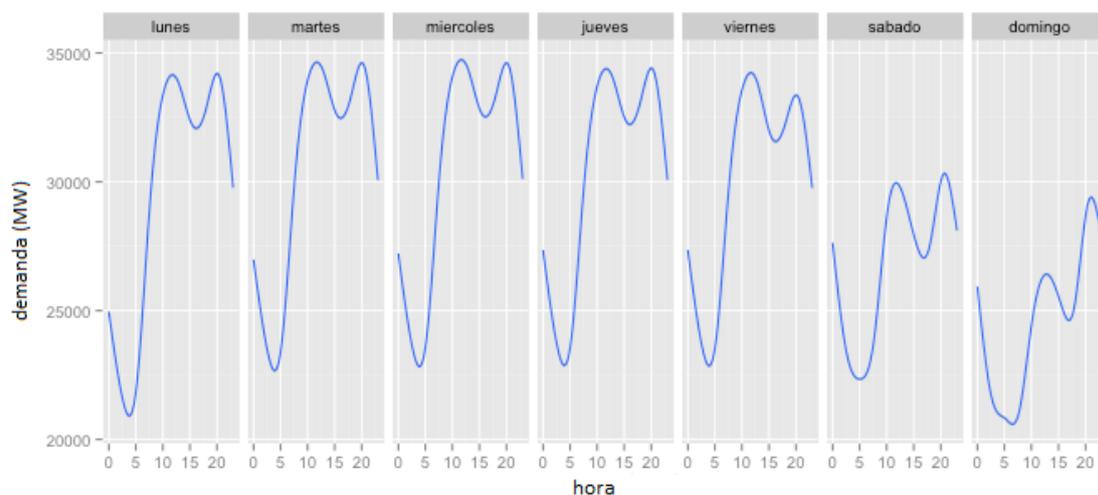


Figura 4: ciclo semanal de la demanda eléctrica (valores medios anuales)

El análisis semanal nos muestra cómo de lunes a viernes el perfil de demanda es bastante similar, salvo por la parte del valle de la madrugada del lunes, que marca la transición entre el fin de semana y la semana laborable y el último tercio de horas del viernes, que presentan un pico más bajo marcando la transición hacia el sábado y el domingo, cada uno con perfiles propios.

Tendencia anual de la demanda:

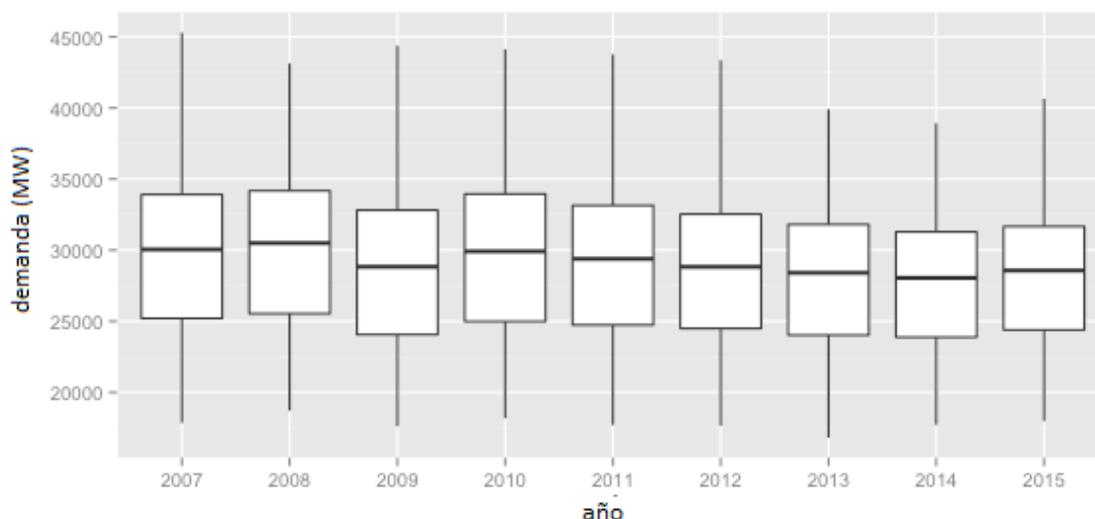


Figura 5: evolución anual de la demanda

Desde la crisis de 2008, se aprecia cómo la demanda presenta una evolución prácticamente nula como refleja la mediana en torno a 30000 MW. Cabe destacar que desde 2012 las puntas de consumo, de en torno a 45000 MW han disminuido a los 40000. Imputar a una causa concreta esta disminución resulta difícil, ya que el consumo registrado en la red de transporte no tiene en cuenta la parte de autoproducción de las industrias o la producción que no abandona la red de distribución (red con un voltaje inferior a 220 kV en España).

Dependencia horaria de la demanda:

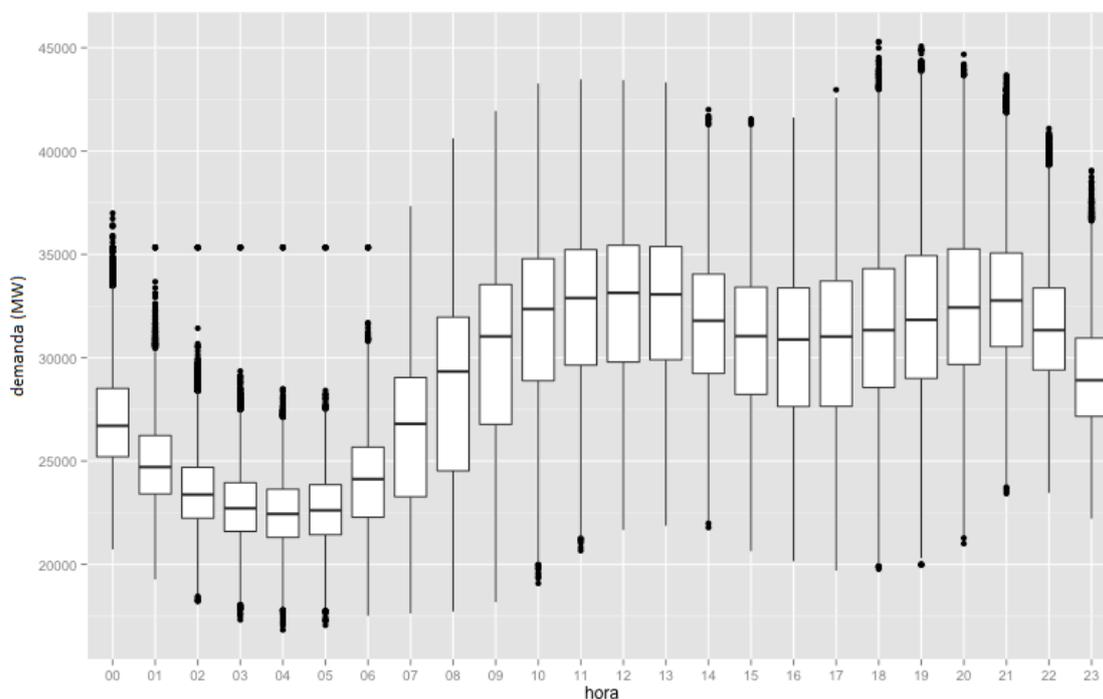


Figura 6: variación horaria de la demanda

La dependencia horaria de la demanda podría separarse en cinco secciones.

(00-06h): un rango cuartil-cuartil pequeño y Q1-Q3 con una amplitud de 5000 MW, contenidos entre los 22 y 28000 MW. Los atípicos superiores corresponden al uso de servicios de sistema, cuando se bombea agua de las cuencas aguas arriba para poder turbinarla durante momentos de mayor consumo.

(07-11h): momento de rampa de subida, correspondiente al arranque de grupos térmicos y al aumento de generación para garantizar el paso de la punta del mediodía. El aumento del rango cuartil-cuartil corresponde a la necesidad de reactividad frente a la punta de demanda, que lleva a compensar la estabilidad de la sección de madrugada.

(11-13h): punta de la mañana. Consumo con cuartiles estables, de 20 a 45 GW, que dependerá de la época del año, día de la semana etc.

(14-18h): valle de la tarde. En esta sección es necesario gestionar un descenso desde la punta de la mañana con un aumento paulatino más estable que en rampa de subida para gestionar la punta de la tarde.

(19-23h): paso de la punta de la tarde y disminución hasta la madrugada. La sección de la punta de la tarde también presenta una gran variabilidad, dependiendo del régimen de demanda que corresponda. Esto se debe a que en los meses más fríos la punta diaria es la punta de la tarde (iluminación, calefacción en los hogares) mientras que en los meses cálidos la punta diaria es la del mediodía.

Componentes cíclicas de la demanda:

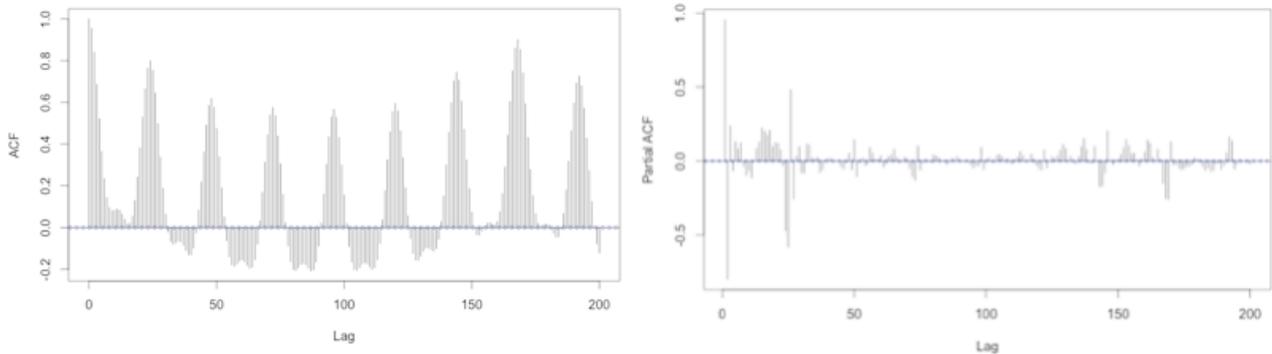


Figura 7: correlogramas total y parcial de la demanda

El análisis del correlograma de la serie temporal de la demanda muestra una fuerte periodicidad cada 24 horas, que se confirma en el correlograma parcial con los picos en el retardo 24, 48 y en menor medida en el resto de múltiplos de 24. Además de esta periodicidad diaria, el hecho de que la correlación aumente a partir del retardo 168 corresponde con una periodicidad semanal.

A estas dos microestacionalidades se une una estacionalidad anual, como observamos en la figura y que corresponde a una senoide con picos en torno a los meses más fríos (enero-febrero) y cálidos (julio-agosto). Las rupturas que se observan corresponden a las fechas que rodean al 15 de agosto y a las vacaciones de navidad en diciembre.

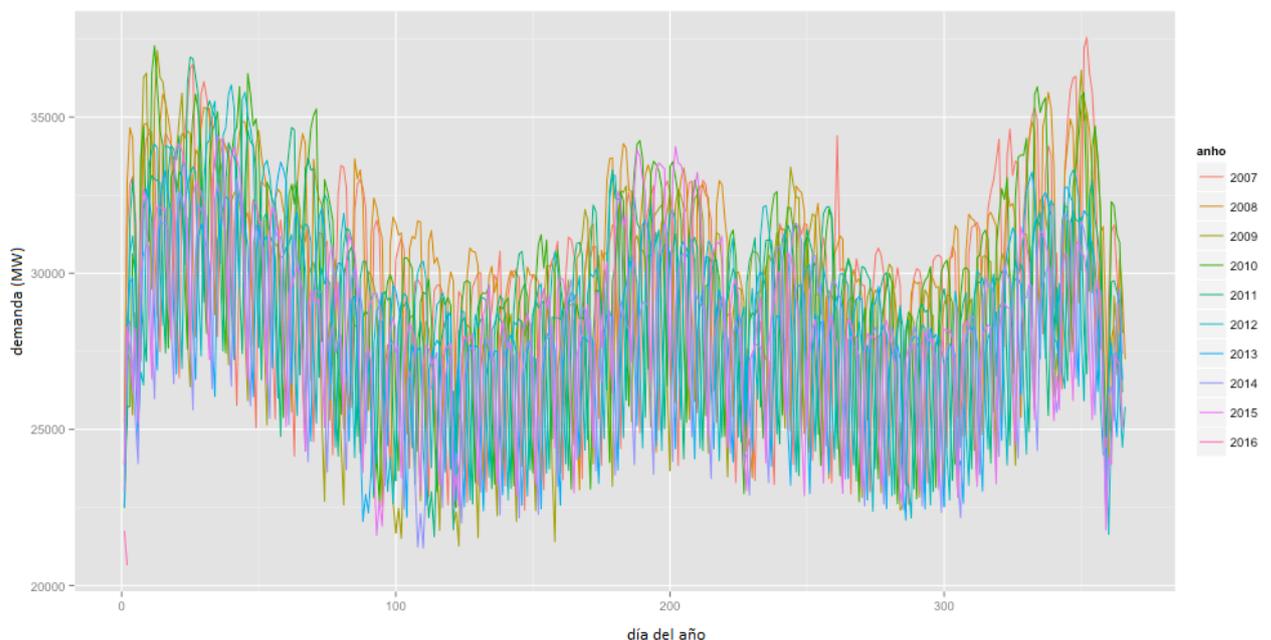


Figura 8: estacionalidad anual de la demanda

Efecto de los días festivos en la demanda:

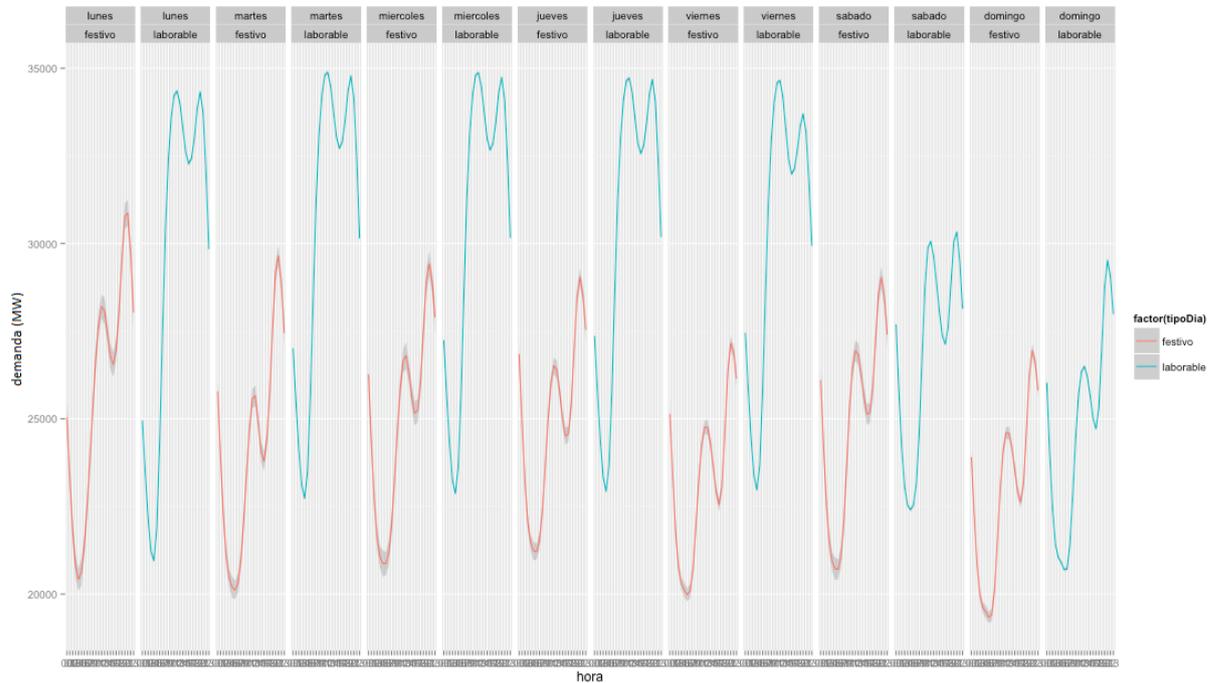


Figura 9: efecto de los días festivos en la demanda

Se constata cómo el perfil de los días festivos difiere del perfil laborable. De todos los perfiles semanales medios, el perfil de un día festivo guarda la misma forma que un domingo, aunque la amplitud en términos de potencia (MW) varía según el día. Los lunes representan la forma con una mayor amplitud y consumo medio, siendo los viernes y domingos aquellos con una menor.

El hecho de que los perfiles de martes a jueves sean menores que el lunes, muestra el efecto de un fin de semana largo, ya que el lunes festivo actúa como transición para la nueva semana, sustituyendo al domingo. Del mismo modo, la amplitud menor de martes a viernes modela el efecto de puentes y vacaciones más largas (Pascua) que indican perfiles de consumo menores, similares al mínimo semanal del domingo por una disminución de la actividad industrial y de servicios, operación salida de vehículos, etc.

Las vísperas y días posteriores a festivo también modifican su perfil aunque en menor medida y únicamente determinadas horas (como sucede lunes y viernes respecto a los días laborales). Por los pocos puntos existentes, se ha preferido excluir esta agrupación en el análisis.

3. Estudio del error del previsor de Red Eléctrica

3.1 Introducción

De los dos previsores de los que Red Eléctrica proporciona información [2] (REE, 2016) y que corresponden a la previsión realizada en D-1 (víspera del día de consumo) y en D, día de la realización misma, se ha estudiado el previsor en D-1 pormenorizadamente al entender que éste podría aportar no solo información sobre las debilidades de la previsión actual sino también un benchmark contra el que comparar los resultados de los previsores contruidos.

A modo de curiosidad, un somero vistazo al error de previsión en D permite comprobar cómo la mayoría de errores de la previsión efectuada el día mismo de la realización (D de ahora en adelante) están en la banda de ± 5000 MW.

Sin embargo, existen valores superiores, correspondientes a días atípicos que se examinarán por separado al analizar el D-1 para juzgar si es necesario tenerlos en cuenta pues es un error del previsor o si por el contrario se corresponden a un cambio súbito del comportamiento de consumo al cual es difícil reaccionar con la suficiente anticipación y que por tanto no conviene dejar contaminen el set de datos entrenamiento.

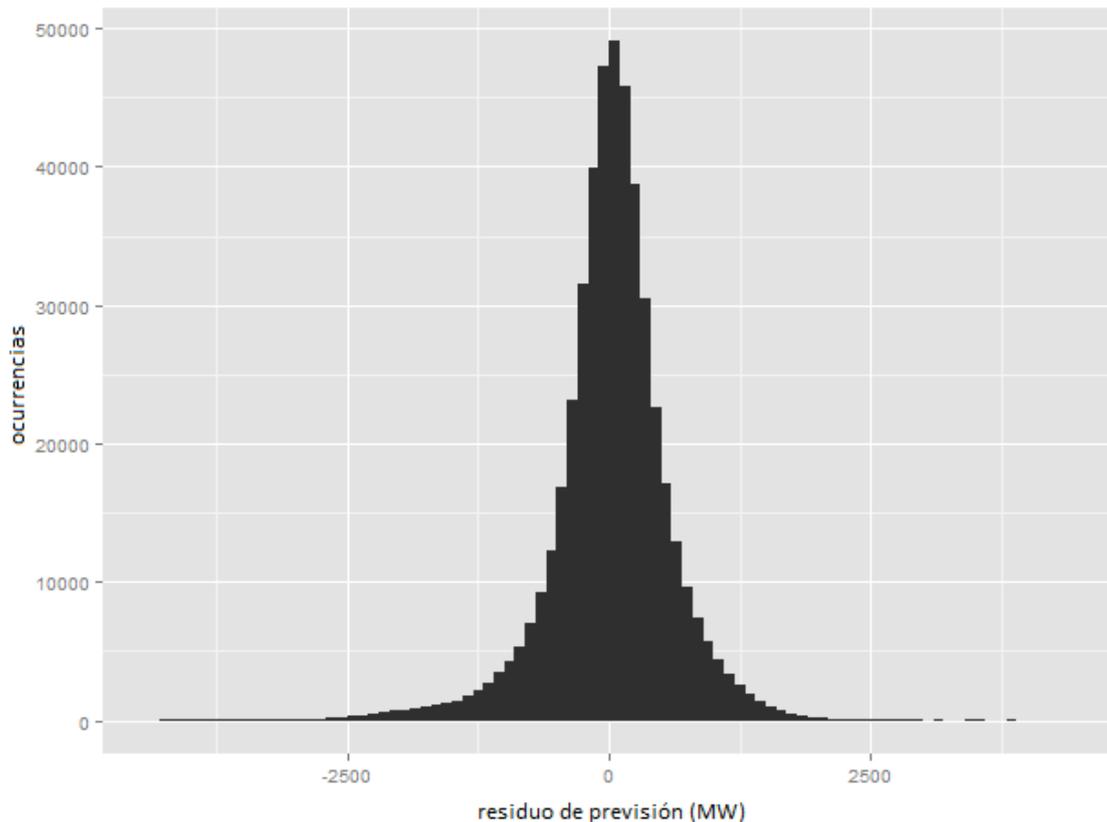


Figura 10: error de previsión D REE

3.2 Error de previsión en D-1

En este caso, si bien se encuentran grandes atípicos, el orden de dos veces superiores a en D, resulta interesante ver cómo la amplitud del error en D-1 una vez restringidos estos atípicos es menor a la del error cometido en D, con unas colas menos densas.

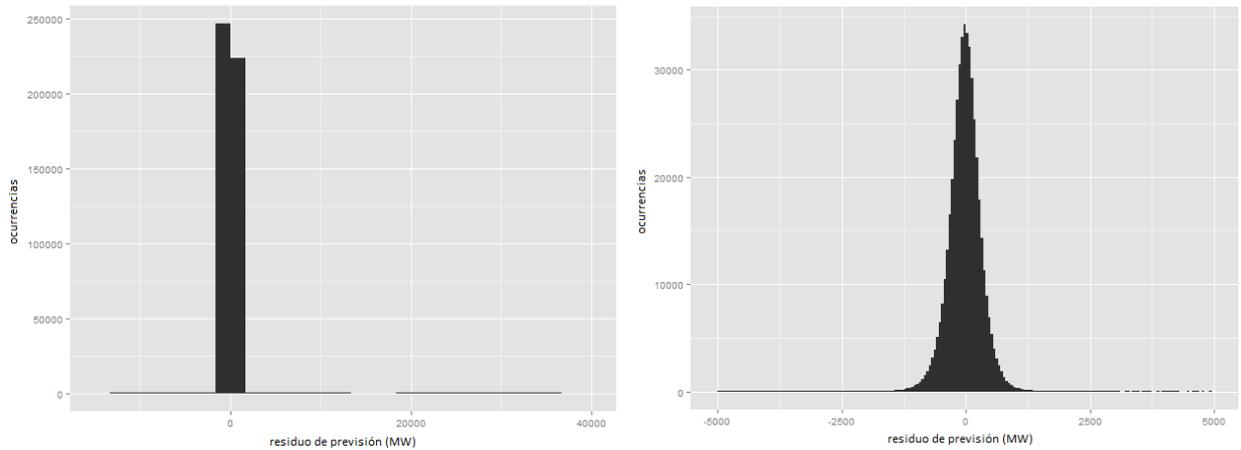


Figura 11: Error de previsión D-1 REE incluyendo atípicos y detalle

En la figura del error en D-1 incluyendo atípicos se observa cómo aparecen errores por encima de los 20000 MW, que corresponden a un error de publicación de la previsión por parte de Red Eléctrica. Asimismo en la figura del detalle, se observa la baja densidad de errores de previsión por encima de los 2500 MW, por lo que se ha optado por trabajar en ese rango del residuo tanto en el análisis exploratorio como posteriormente para calibrar los modelos, de manera que el ajuste por mínimos cuadrados de los coeficientes se vea afectado lo menos posible por estas ocurrencias singulares.

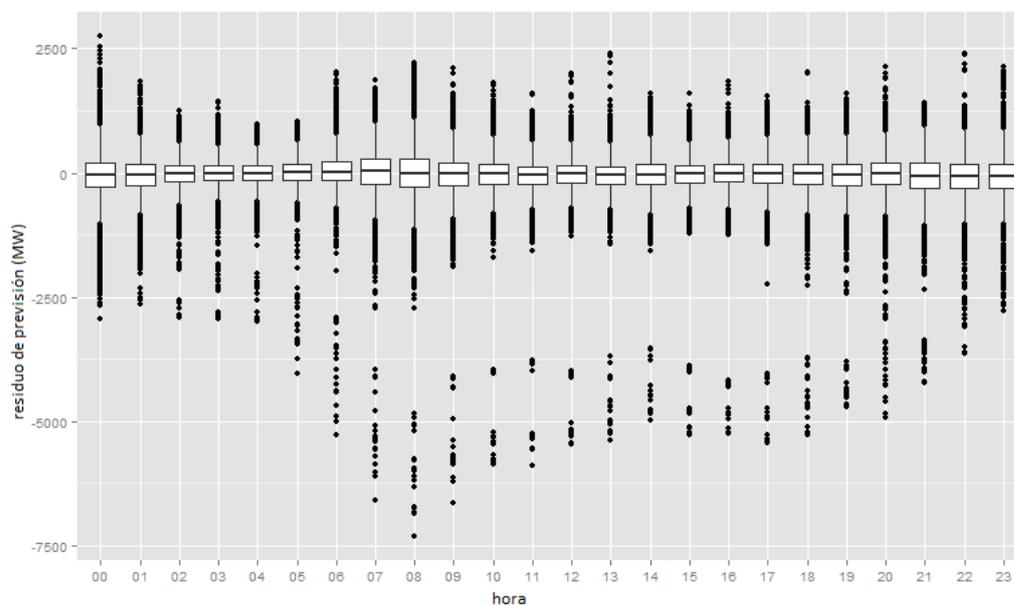


Figura 12: error de previsión horario

Analizando hora a hora, se observa como tanto por defecto como por exceso, los atípicos del diagrama de cajas parecen formar perfiles de demanda, lo cual corrobora la necesidad de eliminar los días atípicos tanto en la etapa de calibración como también en la exploración de los datos

Una vez eliminados esos días, observamos cómo el error presenta una mediana nula y un aumento del rango intercuartil en la rampa de subida (a partir de las 7 de la mañana) y en la punta de la tarde (20-22 horas).

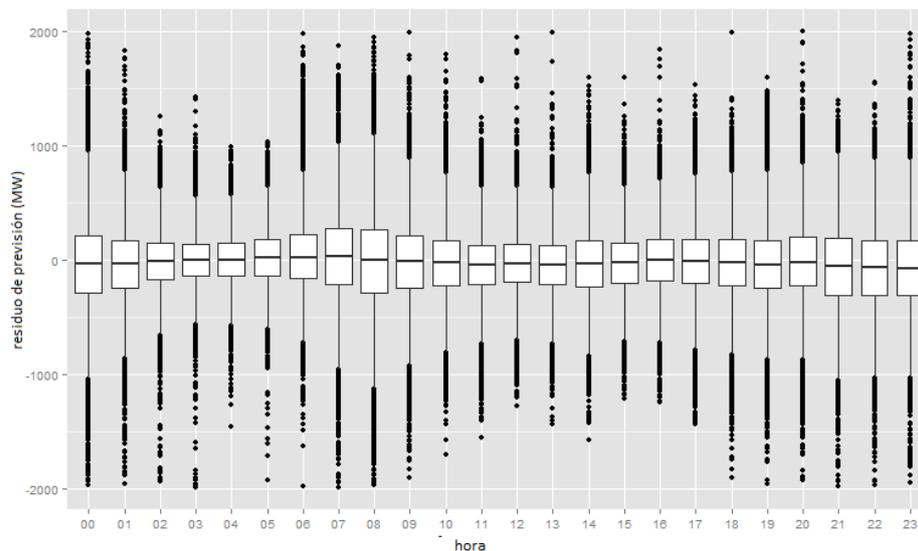


Figura 13: error de previsor horario sin atípicos

Relación del error de previsión con la demanda eléctrica:

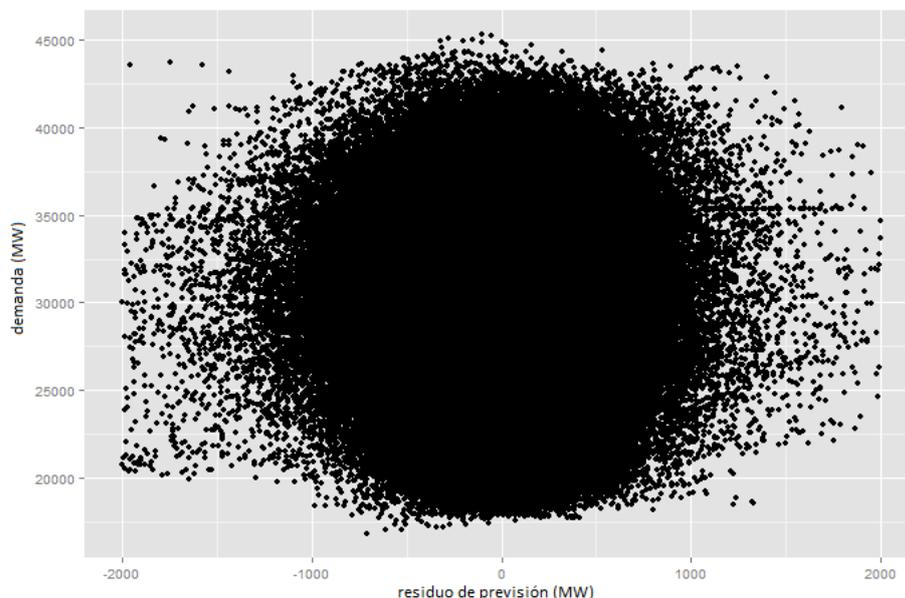


Figura 14: relación entre el error cometido y la demanda

La distribución es uniforme y la correlación de -2.4%, por lo que no parece existir una conexión entre el error de previsión cometido y el valor de la demanda a una hora determinada.

Relación temporal del error:

El error parece tener una relación con las horas previas, como se observan en los correlogramas parcial y total. Diferenciando la serie a un retardo, para eliminar la componente de la tendencia del primer correlograma se observa como existe autoregresión en la serie.

Un ajuste del modelo diferenciado proporciona el menor AIC (criterio de información de Akaike) para el modelo ARIMA (3, 0, 3). Por tanto, es evidente que el error contiene información que se podría extraer para mejorar el modelo de Red Eléctrica, si bien al no conocer cómo está constituido no resulta posible averiguar cómo utilizar este ajuste para crear una etapa de postcondicionamiento de los previsores existentes, a riesgo de amplificar los errores en los atípicos. Sin embargo, resulta útil para comparar la información subyacente en el residuo de los nuevos modelos frente a los existentes para ver cuál de ellos se corresponde mejor a un error de previsión más cercano al ruido blanco.

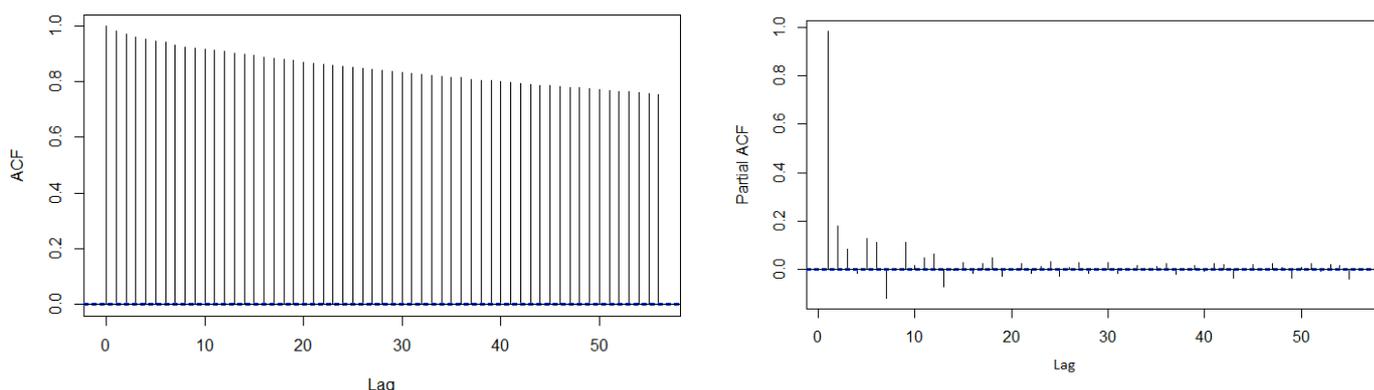


Figura 15: autocorrelogramas parcial y total del error de previsión

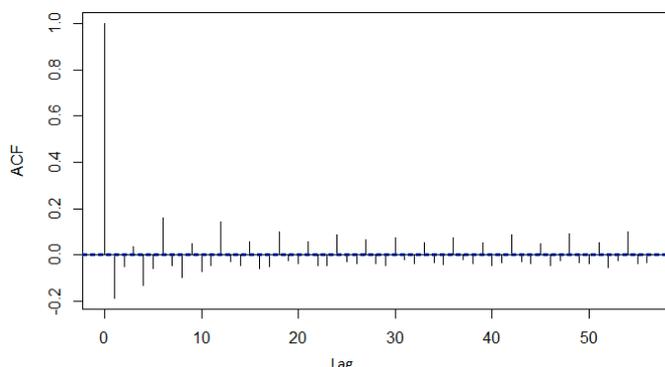


Figura 16: autocorrelograma diferencia del error de previsión

Valores atípicos

Examinando el set de datos disponible en el site de Red Eléctrica se ha comprobado que en determinados momentos de 2007 y 2015 la alimentación de datos no fue correcta (previsiones o demandas nulas), por lo que se ha decidido eliminar esos años por completo del set de datos de entrenamiento, para evitar que corrompan la integridad del modelo. No obstante, 2015 constituirá el set de validación de los modelos ya que resulta interesante al corresponder a un cambio de régimen en la tendencia anual de la demanda.

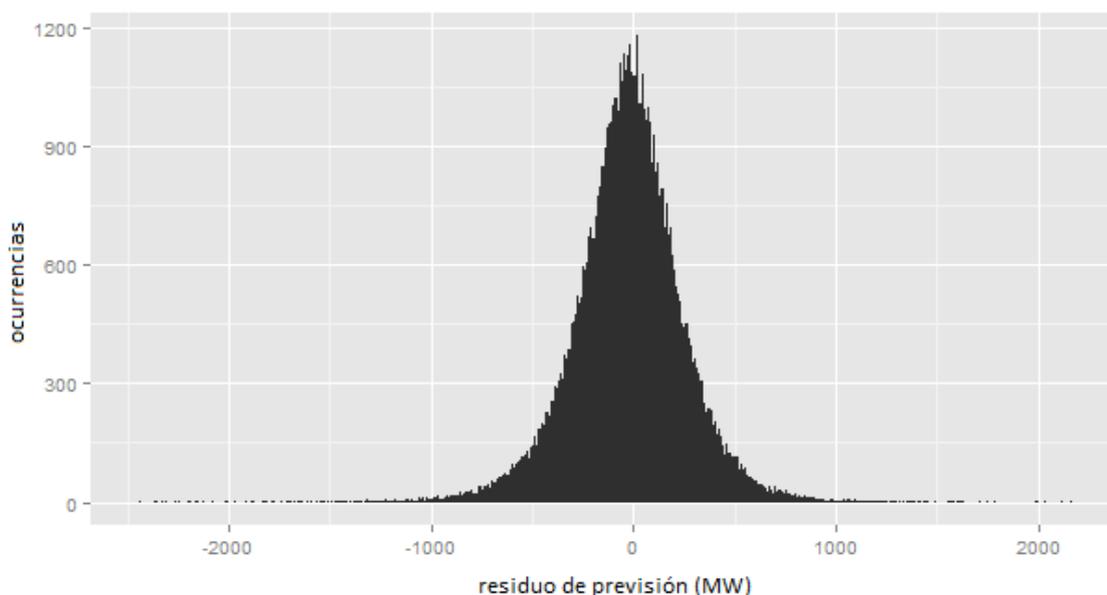


Figura 17: distribución del residuo sin atípicos

Determinando los cuantiles $q_{005} = -980.833$ y $q_{995} = 815.867$, se puede considerar como evento singular cualquier punto horario cuyo error de previsión es superior a 1000 MW en valor absoluto, o lo que es lo mismo, un 2.2 % del máximo registro absoluto de demanda.

Examinando pormenorizadamente todos estos días, se ha definido como día singular aquel que presenta 3 o más horas consecutivas o 4 o más horas no consecutivas con un error de previsión por encima de 1000 MW. Se han identificado las siguientes categorías de día singular:

- Huelgas generales: afectación desde las 22 horas de la víspera hasta el inicio laboral del día posterior a la huelga. Días afectados: 29/09/2010, 29/03/2012, 14/11/2012, 14/11/2014
- Eventos deportivos: el 5-1 del Barcelona-Real Madrid del 29/11/2010, la final de la copa del mundo del 11/07/2010 y las finales de la Eurocopa presentan un error por exceso (demanda real inferior a la prevista) que comienza desde las 2 horas previas al evento y hasta 2-3 horas después de su finalización. La región de afectación es de unas 6 horas, comenzando el evento al final de la hora 2).

Errores según la tipología del día:

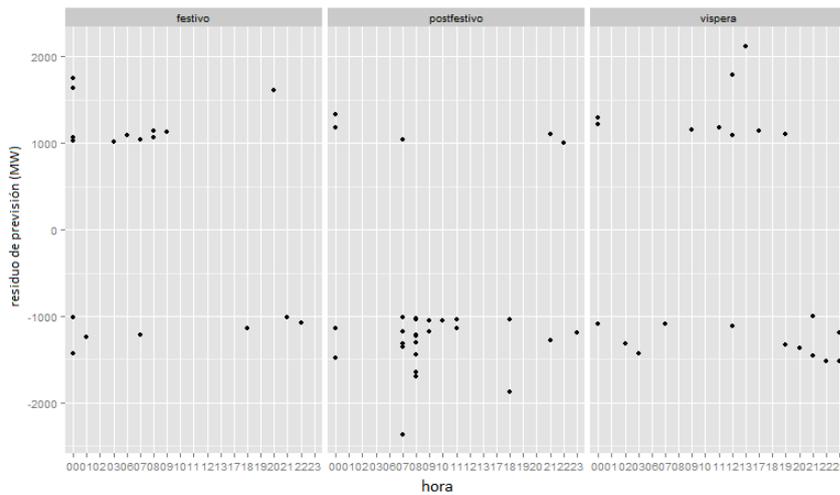


Figura 18: distribución de atípicos según tipología de día

En los días festivos tiende a subestimarse la demanda, notablemente en la primera parte del día. Esto puede corresponder a que el consumo doméstico no varía sus pautas matinales como tampoco lo hace la gran industria.

Los días posteriores a festivos presentan una sobreestimación de la demanda, fruto posiblemente de la no consideración de la inercia de tener un día festivo precediendo y considerando que el día presenta el comportamiento natural de ese día de la semana.

Por último, las vísperas presentan errores uniformes tanto por defecto como por exceso, aunque se tiende a subestimar la demanda de la punta de la mañana y a sobreestimar la última parte del día que es cuando el comportamiento de consumo se ha modificado.

Distribución horaria del error:

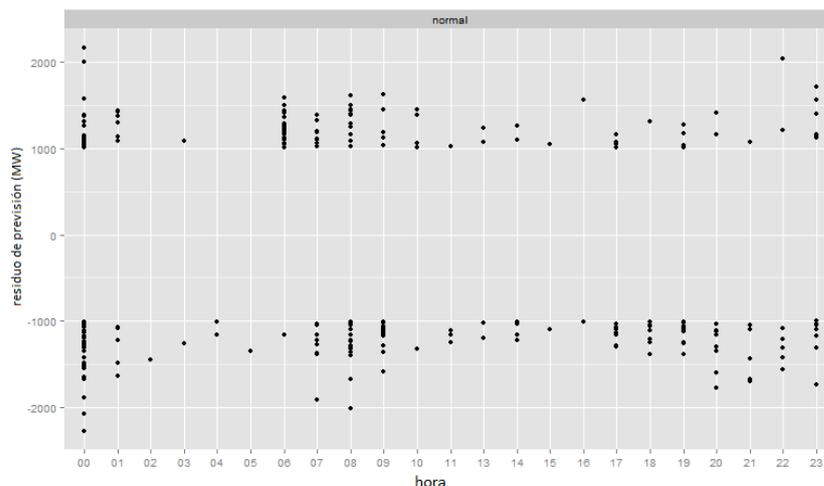


Figura 19: distribución horaria de atípicos del error

En los días normales, el error se concentra en la rampa de subida así como en torno a la medianoche. En el cinturón de la punta de la tarde también se observa un aumento aunque más ligero y que puede corresponder a una sensibilidad a la temperatura. En todo caso, el error no presenta un sesgo hacia la sobre o subestimación.

Distribución mensual del error:

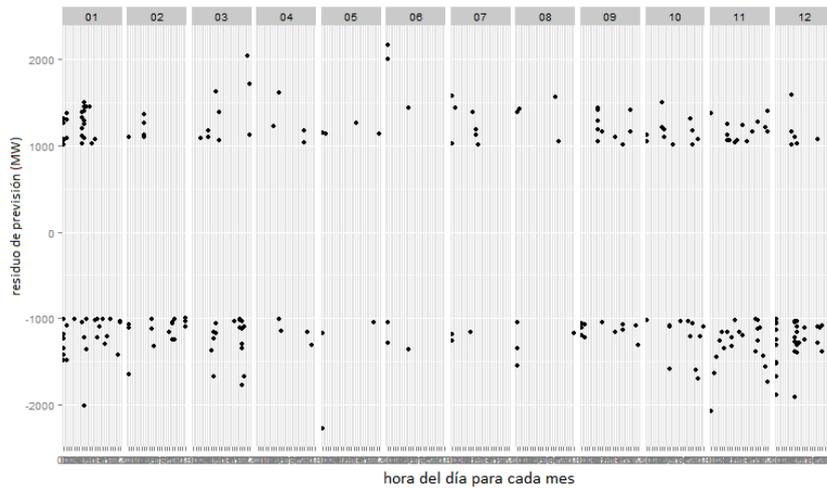


Figura 20: distribución mensual de atípicos del error

Mes a mes, observamos cómo los errores se concentran en los meses fríos. Esto puede corresponder a un mayor consumo eléctrico para calefacción que no está convenientemente modelado en el predictor. Semanalmente, los errores son más frecuentes el lunes, que marca el cambio del régimen semanal notablemente en la segunda parte del día, que debe adaptarse al nuevo consumo semanal teniendo todavía información (inercia) del fin de semana precedente.

Distribución semanal del error:

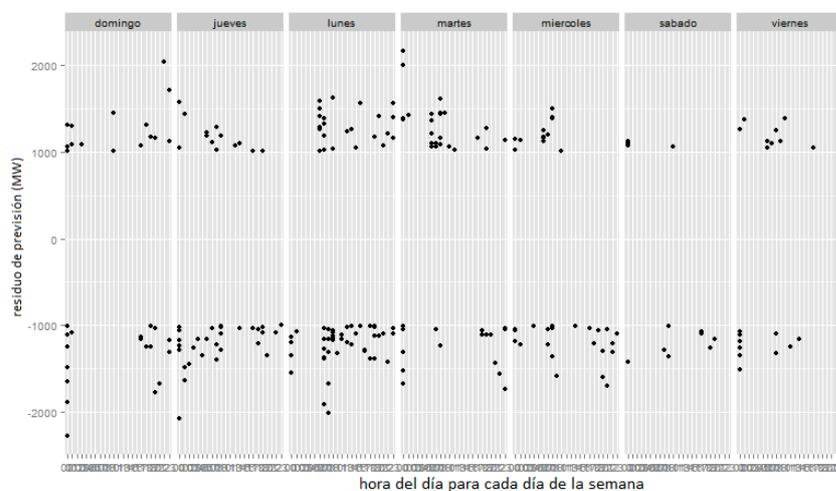


Figura 21: distribución semanal de atípicos del error

En lo que respecta al día de la semana, los errores son más elevados en los días laborables. Esto puede obedecer a varias causas, pero notablemente el impacto de puentes, festividades locales o que se desplazan de un año para otro y que modifican la demanda no solo del día festivo sino también de las componentes estacionales (diaria, semanal, anual) al arrastrar esta información al ciclo siguiente.

Error de puntos no singulares:

Además de los puntos atípicos, con un error por encima de los 1000 MW, es de interés analizar los puntos en los que se concentran los errores. De este modo podrán identificarse sesgos en los previsores actuales así como obtener información interesante con la que alimentar a nuevos modelos

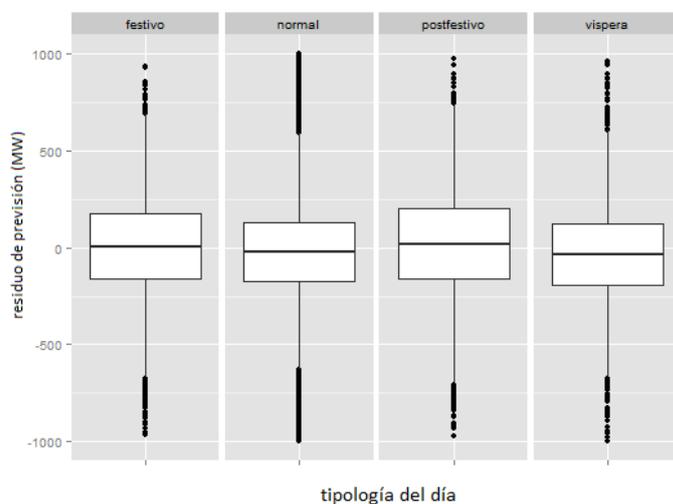


Figura 22: distribución del error según tipología de día

El error entre los distintos tipo de día es bastante uniforme, sin sesgos notables. Lo mismo sucede durante la semana, con un rango intercuartil algo superior para los lunes, pero sin diferencias notables.

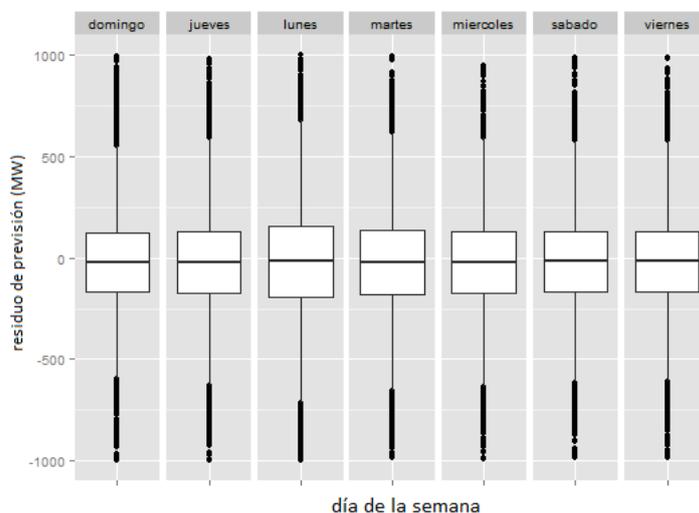


Figura 23: distribución del error según día de la semana

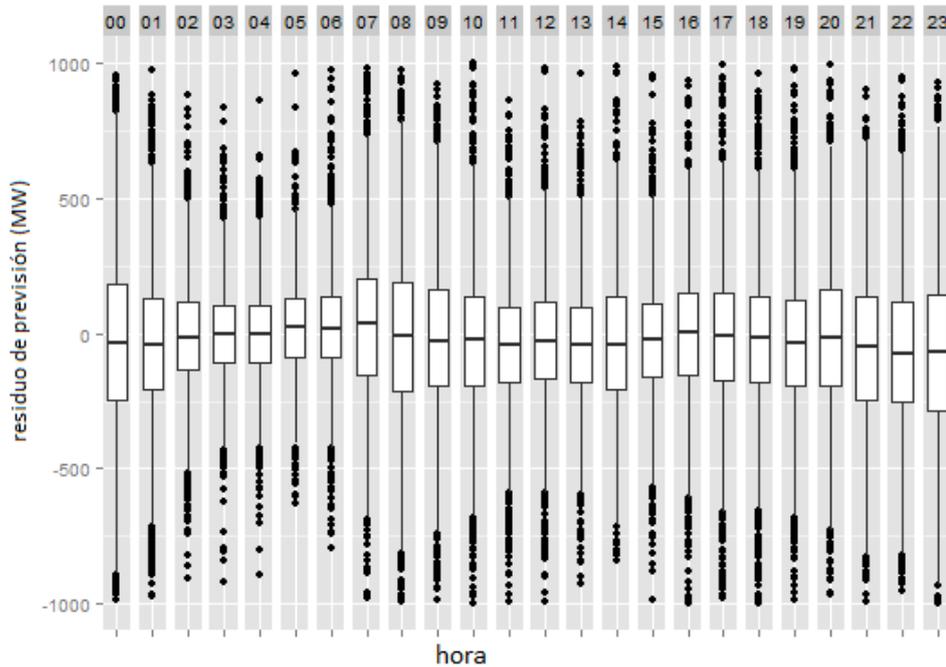


Figura 24: distribución del error según la hora del día

Se puede apreciar cómo el error de previsión cometido se distribuye uniformemente independientemente del día de la semana, mes o tipo de día. Hora a hora se observa una disminución durante el valle de la madrugada, asociada a ser una franja en la que la demanda suele ser bastante constante.

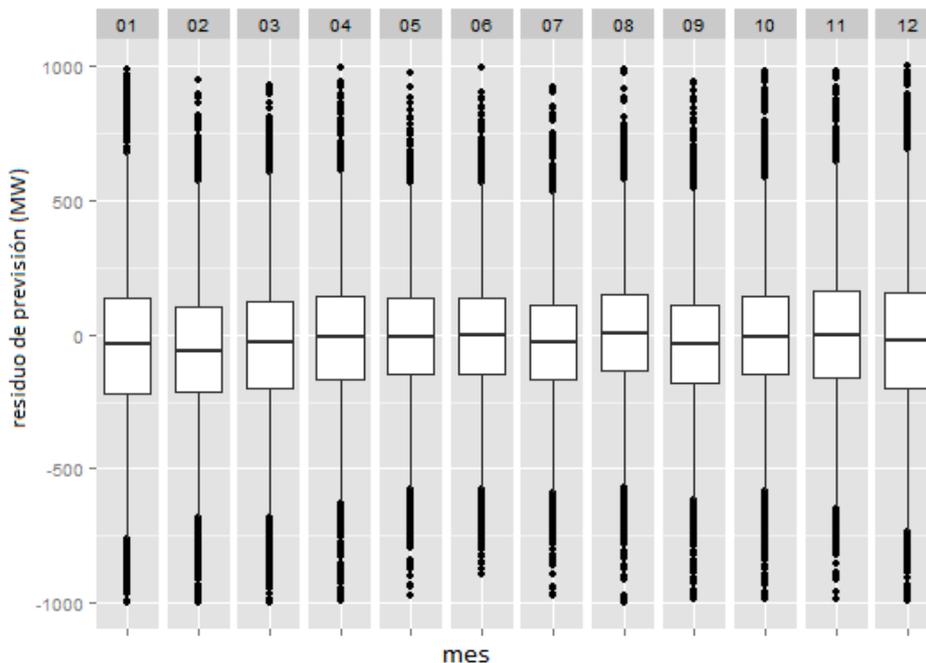


Figura 25: distribución del error según mes

Mes a mes, el rango es mayor en los meses fríos y parece haber un ligero sesgo de la media siguiendo la propia estacionalidad, pero sin llegar a ser significativo. De todo ello, inferimos que el error por debajo de 1000 MW está uniformemente distribuido y sin sesgos.

4. Preparación del set de datos

4.1 Introducción

Una vez analizado el error de previsión e identificados los días singulares cuya demanda es necesario corregir, es necesario acondicionar el set de datos y prepararlo para el proceso de calibración del sistema de previsión modelado.

De no ser así, los comportamientos anómalos en esos días singulares condicionarán el valor de los distintos coeficientes, provocando el sobreaprendizaje del modelo y por tanto que la sensibilidad a ligeras variaciones de las explicativas derive en la obtención de una previsión de demanda con un elevado residuo.

De los datos disponibles, entre 2007 y 2015, se decidió eliminar los años extremos ya que ambos presentaban ausencias notables de valores así como ciertas anomalías de los registros de 2007. Por lo tanto el set de datos de entrenamiento final contiene los años completos de 2008 a 2014, siete años civiles.

4.2 Corrección de los días singulares

Las huelgas generales, eventos deportivos y otros puntos de comportamiento anómalo de la demanda hacen necesario acondicionar la serie de datos para que los algoritmos de calibración sean capaces de encontrar una relación entre los datos, teniendo al mismo tiempo la precaución de no variar la serie de forma que se creen dependencias entre las variables que realmente son inexistentes o muy débiles.

Las rupturas durante las vacaciones de diciembre, que habitualmente provocan errores en el previsor, no se han modificado para intentar forzar en la calibración al sistema de previsión a captar el comportamiento especial de esa región reducida.

El valor del 5 de diciembre de 2013 ha sido reemplazado por su previsión, ya que el efecto del largo puente hacía descender bruscamente la demanda durante todo el día y ese valor podría condicionar el resto de años. Además, la ausencia de puntos que cumplan estas características (domingo, víspera de festivo que además es el comienzo de un largo puente) suponían un riesgo de sobreaprendizaje que era necesario mitigar.

La serie de datos, de manera general, ha sido acondicionada mediante descomposición STL, "stl decomposition using Loess", método propuesto por *Cleveland et al. (1990) [3]*.

4.3 Modelado de las variables explicativas adicionales

Componentes cíclicas:

Las estacionalidades diaria, semanal y anual se han modelado de 3 maneras diferentes, a fin de permitir evaluar distintos modelos y seleccionar la más conveniente. Se ha utilizado como ejemplo la periodicidad semanal de la demanda.

- Como número entero: la variable explicativa tomará valores entre 1 y la periodicidad. Por ejemplo, 1 el lunes, 2 el martes hasta 7 el domingo. Esto permitirá integrar la estacionalidad como una función de números enteros. Este modelado es el más simple y permite plasmar correctamente estacionalidades aditivas de la forma.

$$S(x) = f(x) = \sum a_n x^n$$

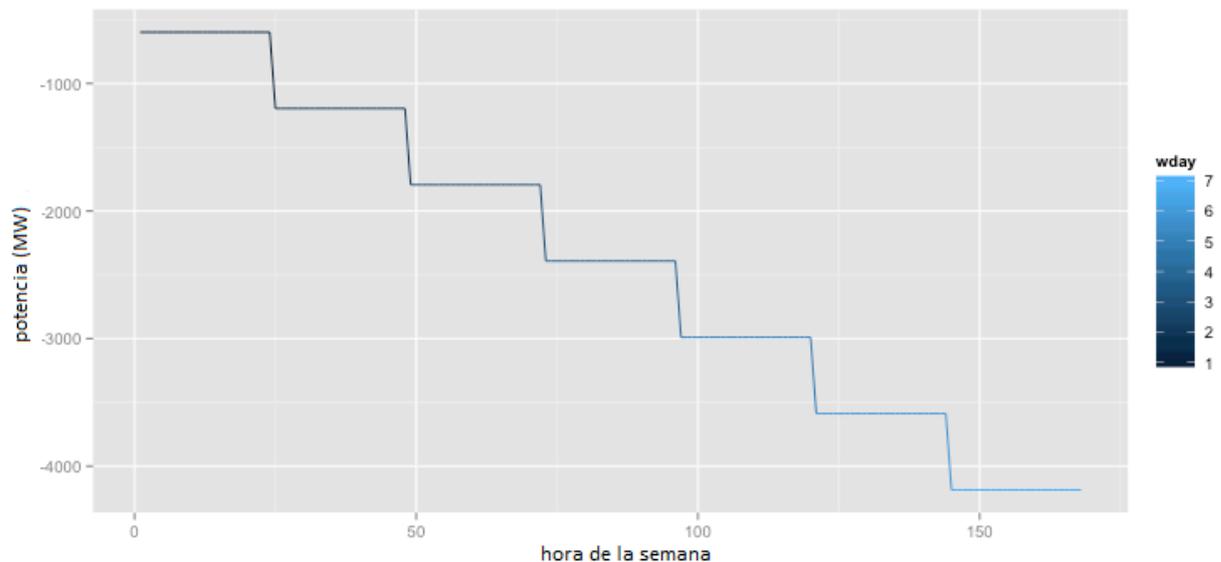


Figura 26: ejemplo de función de estacionalidad numérica

- Como variable categórica: cada estado corresponderá con una configuración binaria y se añadirán n columnas para modelar hasta 2^n estados. En el ejemplo se necesitan 3 columnas y el modelado es:
 - 0 0 1: lunes
 - 0 1 0: martes
 - 0 1 1: miércoles
 - 1 0 0: jueves
 - 1 0 1: viernes
 - 1 1 0: sábado
 - 1 1 1: domingo

Con este modelado, se obtendrán n coeficientes que permitirán la generación de 2^n funciones, una correspondiente a cada categoría. Para el ejemplo de la estacionalidad semanal, un ajuste por mínimos cuadrados proporciona la siguiente función, una aproximación mediante escalones de una función sinusoidal.

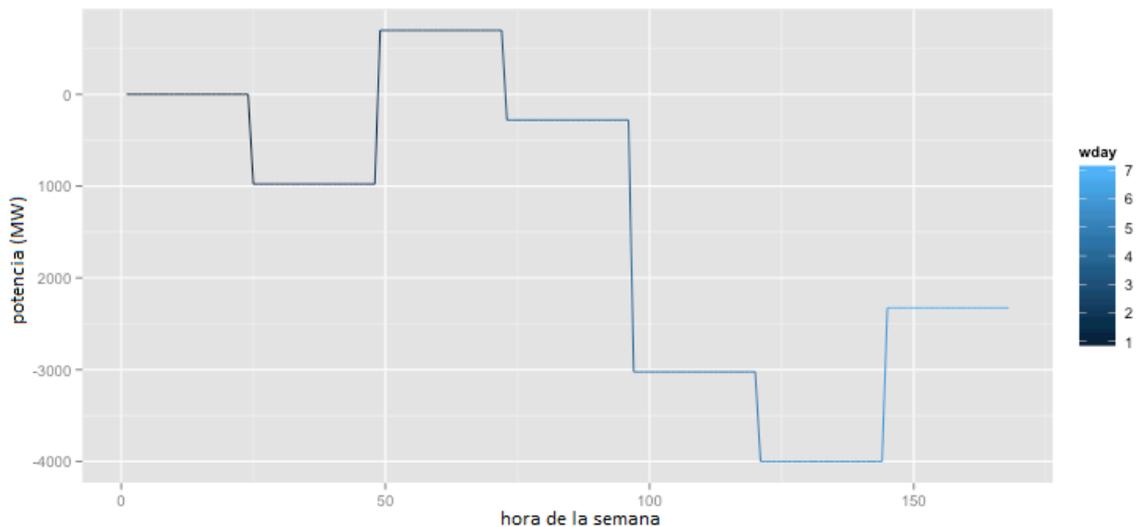


Figura 27: ejemplo de función de estacionalidad de variable binaria

- Mediante series de Fourier:

Las series de Fourier permiten el modelado de estacionalidades de diferentes periodos utilizando distintas series con diferente frecuencia. A diferencia del modelado con números enteros o con variables categóricas binarias, las funciones presentan un perfil liso para órdenes pequeños y son lo suficientemente flexibles para adaptarse a patrones estacionales más ondulados variando el orden K de los coeficientes.

$$\sum_{i=1}^k a_i \cdot \sin \frac{2k\pi t}{T} + b_i \cdot \cos \frac{2k\pi t}{T}$$

En el ejemplo semanal, se observa cómo la combinación lineal de los distintos órdenes generados constituye una manera flexible de modelar estacionalidades complejas.

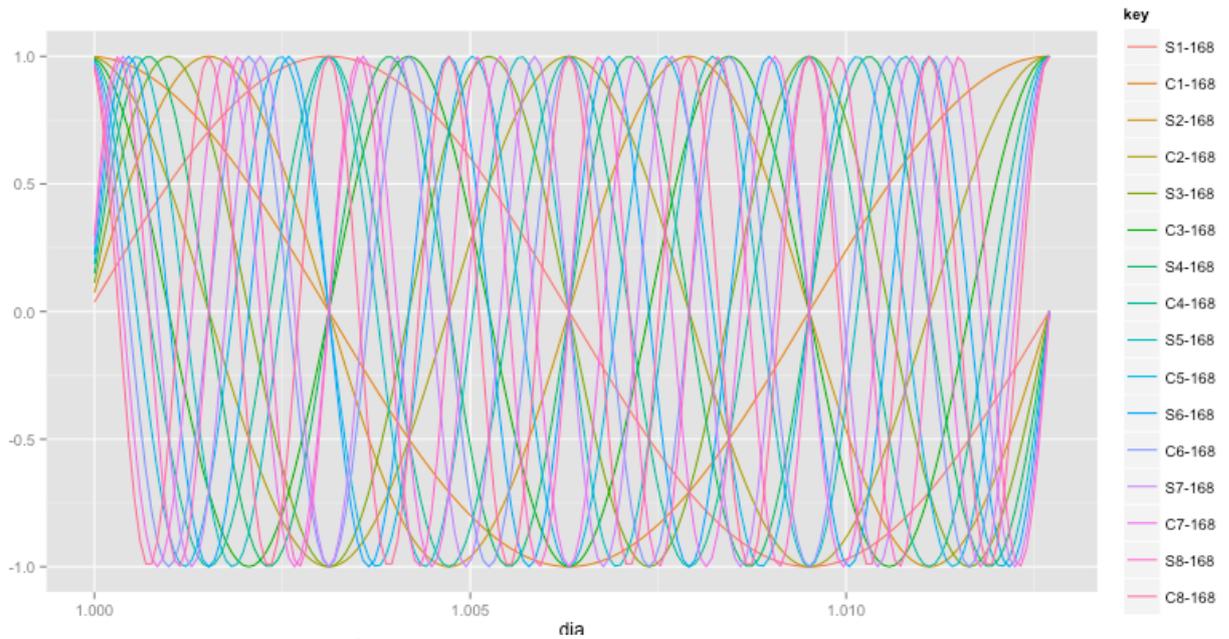


Figura 28: ejemplo de función de estacionalidad con series de Fourier

Cada función de estacionalidad (diaria, semanal, anual) será una combinación lineal de las funciones sinusoidales de Fourier de orden $1 \dots n$. Se han generado las series de Fourier hasta orden 8 aunque se han desechado los órdenes superiores a 2 para la estacionalidad diaria y a 6 para la semanal y anual al ser su p-valor elevado y por tanto ser estos poco significativos.

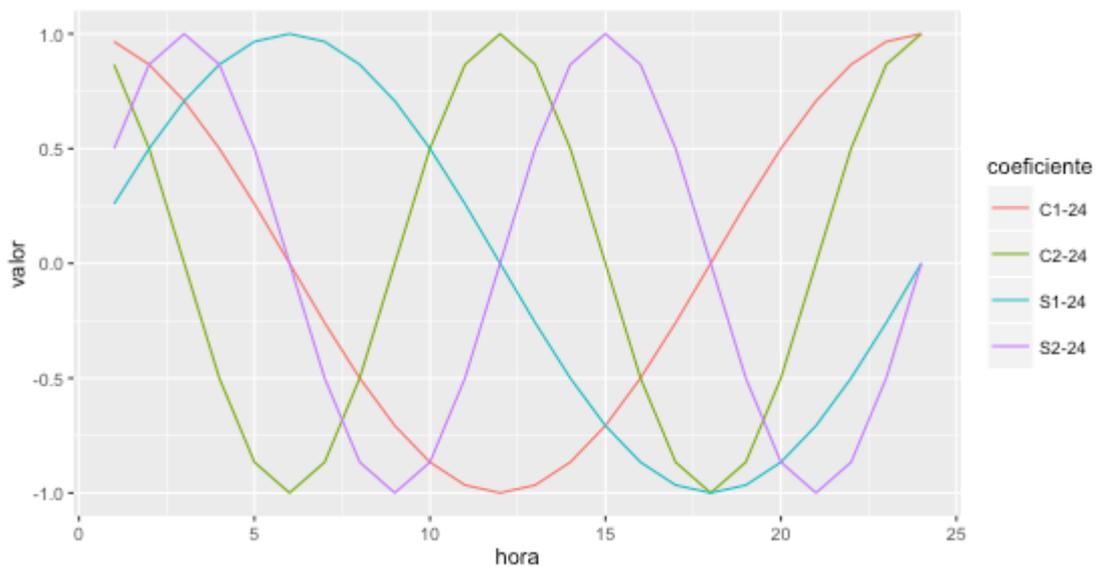


Figura 29 : funciones generatrices de la estacionalidad diaria

Tendencia:

Como se vio en el análisis exploratorio, no existía una tendencia clara. La media de la demanda mensual para cada año exhibe unas pautas similares y aunque parece existir cierta tendencia decreciente, podría del mismo modo tratarse de una estacionalidad sinusoidal macro.

La media anual tampoco arroja más luz al respecto de la tendencia, y como la demanda media anual se mantiene entre los 28000 y 30000 MW, se ha optado por no modelar la tendencia.

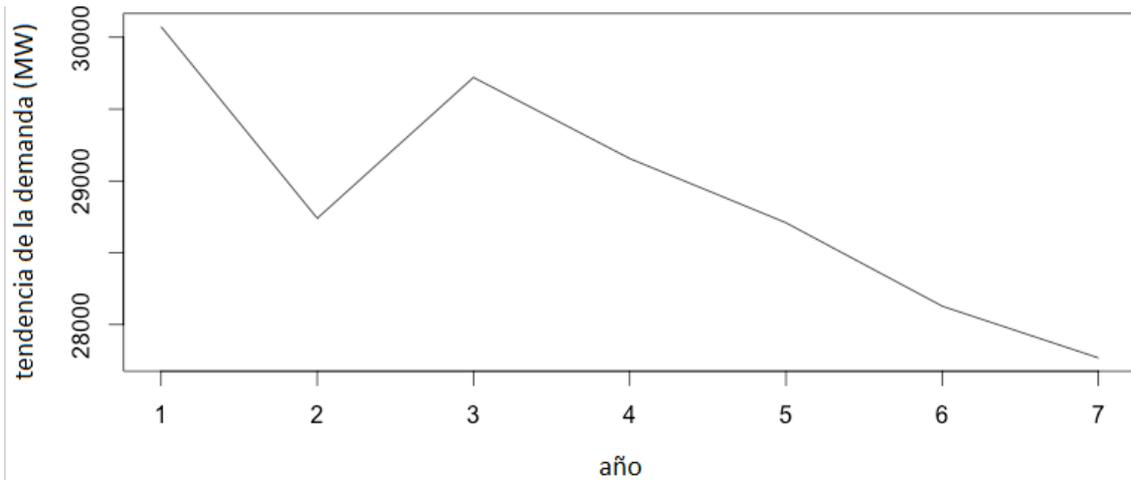


Figura 30: estimación de la tendencia con medias anuales

Dependencia temporal:

Además de las componentes cíclicas, se han incluido los retardos 1,2, 168 (valor la semana previa) y 8760 (año previo).

Tipo de día: se ha decidido distinguir entre días laborables normales, festivos, vísperas y días posteriores a festivo utilizando una variable categórica con dos columnas. Así:

- 0 0: día laborable.
- 0 1: día víspera de festivo.
- 1 0: día festivo.
- 1 1: día posterior a un festivo.

Además, los días que podrían pertenecer a varias clases (por ejemplo el 7 de diciembre) se han etiquetado con el orden de prelación de festivo frente a víspera o día posterior a festivo (es decir, es más importante que un día sea festivo en sí mismo a que sea víspera o día posterior a otro festivo) y en los casos en los que es víspera y también día posterior a festivo, las primeras 16 horas corresponden a una configuración de posterior a festivo (paso de la punta de la mañana) y las 8 últimas a una víspera.

Horarios de invierno y verano:

Para el modelado del cambio de hora, se ha utilizado una variable binaria con dos niveles.

0 1: horario de invierno

1 0: horario de verano

VARIABLES EXÓGENAS:

Dado el influjo de la temperatura en el consumo eléctrico, sería recomendable incluir la temperatura en la previsión de la demanda. La dificultad de acceso a datos públicos de temperatura unido a la problemática en la generación de una temperatura representativa en términos eléctricos de la península ibérica, han hecho que finalmente ésta no se haya incluido.

Sin embargo, se han calculado los pesos eléctricos por Comunidad Autónoma en caso de que en algún momento una serie de datos de temperatura nacional o regional estuviese a disposición de manera pública. Para obtenerlos, se ha trabajado con la demanda eléctrica mes a mes de cada CCAA, al entender que asignar el mismo peso eléctrico en enero y julio por ejemplo a Andalucía no reflejaba adecuadamente la realidad.

La tabla de pesos puede encontrarse en los anexos.

4.4 Sets de datos de trabajo

Set de entrenamiento: constituido por los años completos 2008 a 2014. Los datos base del site de Red Eléctrica han sido ampliados como se indica en las secciones precedentes de este capítulo. El uso de años completos permite considerar no solo las estacionalidades diaria y semanal sino también la macroestacionalidad anual. Además, entre 2008 y 2014 en España se han registrado tanto olas de calor como de frío, como en febrero de 2012 así como veranos e inviernos tanto moderados y agudos, lo cual aumenta su interés de cara a entrenar un modelo flexible capaz de adaptarse a este tipo de variabilidad.

Set de validación: compuesto por el año 2015. Pese a disponer de datos de 2007, al no poder obtener el retardo a un año (datos de 2006), se ha excluido.

5. Modelos lineales de previsión

5.1 Introducción

En los modelos lineales, la respuesta es una combinación lineal de las distintas variables explicativas [4](James) y [5](Hastie)

$$\hat{y} = f(\hat{x}) = \sum \hat{\beta}x$$

Para obtener el valor de los coeficientes Beta, se optimiza la función

$$\min \sum (y_i - f(\hat{x})_i)^2 = \min \sum (y_i - \sum \hat{\beta}x_i)^2$$

Siendo y la serie de demanda del conjunto de datos de entrenamiento y X el vector de las variables explicativas correspondientes a ese instante de tiempo.

Para evaluar los distintos modelos, se ha recurrido al coeficiente R^2 ajustado, que mide el porcentaje de la varianza que el modelo es capaz de reflejar con esas explicativas, penalizando el resultado a medida que se añaden nuevas x para buscar el equilibrio entre la adición de nuevas variables y la ganancia que ello proporciona. Además, se han considerado la desviación típica y la media como indicadores de la amplitud del error y su sesgo respectivamente así como los cuantiles q_{005} y q_{995} para tener una idea del error esperado máximo en el 99% de los casos.

5.2 Variantes de modelado

En todos los modelos lineales se han incluido los días festivos, como variable binaria así como los retardos diario, semanal y anual como valor numérico, para tener ya en cuenta la noción de la dependencia temporal.

Por el contrario, las componentes cíclicas horaria, semanal y diaria se han modelado de tres maneras diferentes, como valor numérico, variable binaria y con coeficientes de Fourier. La tendencia no se ha tenido en cuenta por lo indicado en el capítulo anterior.

Así, el nombre de cada modelo lineal es:

Im.[ciclo horario][ciclo semanal][ciclo anual][perfil base]

Por ejemplo Im.BNFX corresponde al modelo lineal con la estacionalidad horaria modelada como variable binaria, la semanal como valor numérico y la anual con coeficientes de Fourier. La x o su ausencia indican si se ha utilizado un perfil de referencia (una ordenada en el origen) o no. Esto modifica el valor de los coeficientes Beta y, en el caso que nos ocupa, genera bien perfiles independientes unos de los otros (modelo x) o bien modelos incrementales

(modelo sin x) , por ejemplo el de un domingo a medianoche no festivo, sobre el que se construyen el resto de los perfiles.

Considerando las 3 opciones de modelado para cada una de las estacionalidades así como la inclusión o no de un perfil de referencia, se han construido y comparado un total de 54 modelos de previsión lineales.

5.3 Resultados

Media y desviación típica:

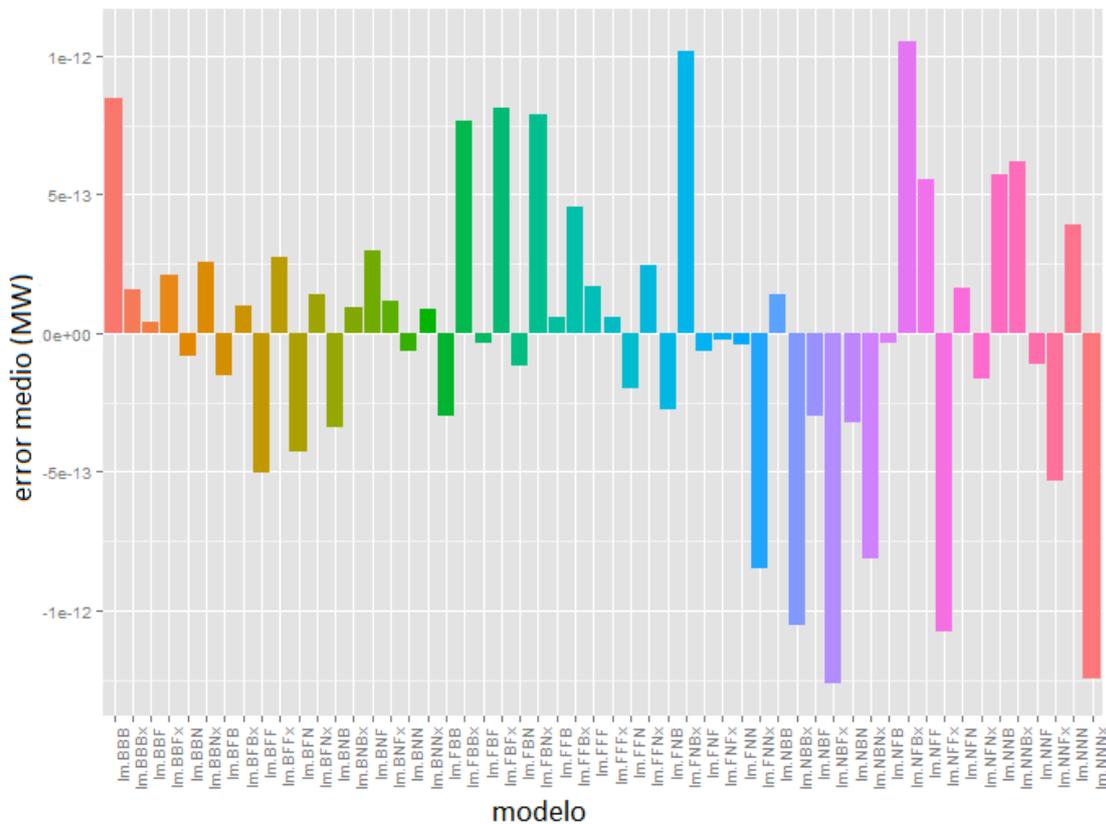


Figura 31: media del error de previsión de los modelos lineales

En lo que respecta a la media, el error de entrenamiento de los modelos lineales no presenta un sesgo acusado, ya que su media es cercana al cero. Se aprecia una mayor amplitud en los modelos en los que el ciclo horario no está modelado como variable binaria, aunque no es significativo.

En cuanto a la desviación típica, todos los modelos en los que la estacionalidad horaria se modela como variable binaria ofrecen mejores resultados, con una desviación típica del orden de 200 MW por debajo del resto.

La utilización o no de un perfil de referencia no afecta a la desviación típica, y a la media de manera muy pequeña.

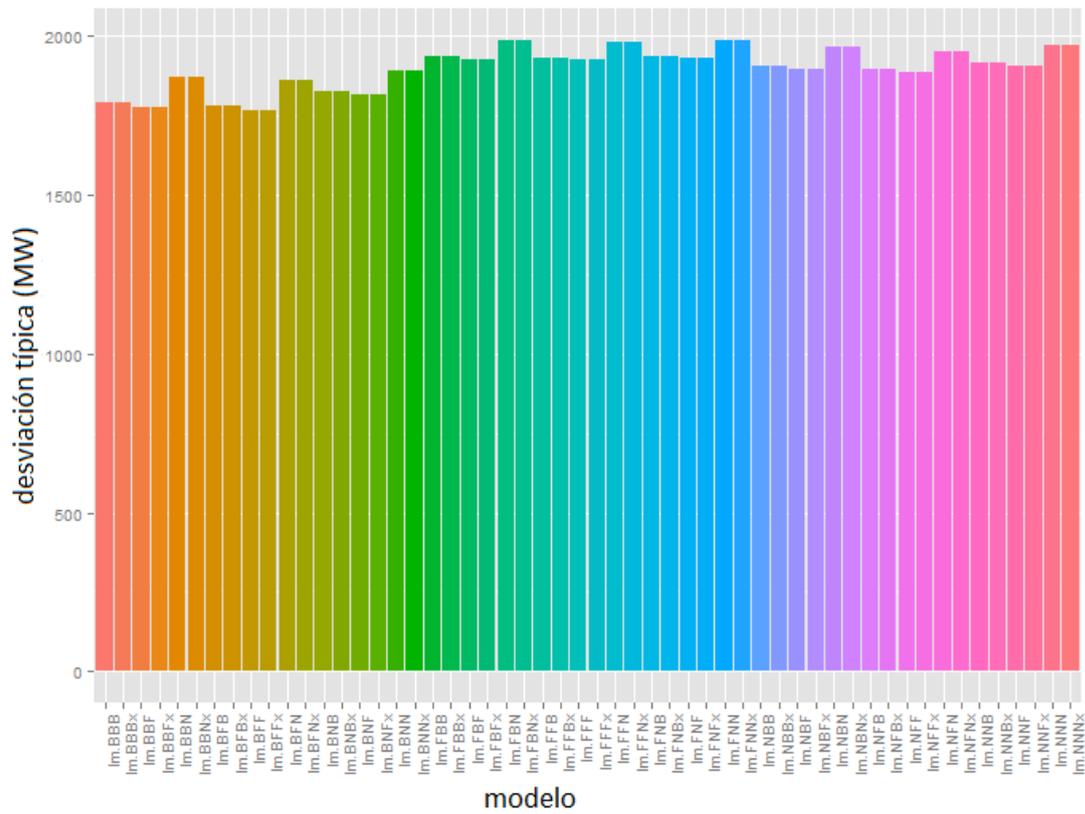


Figura 32: desviación típica del error de previsión de los modelos lineales

Cuantiles q005 y q995

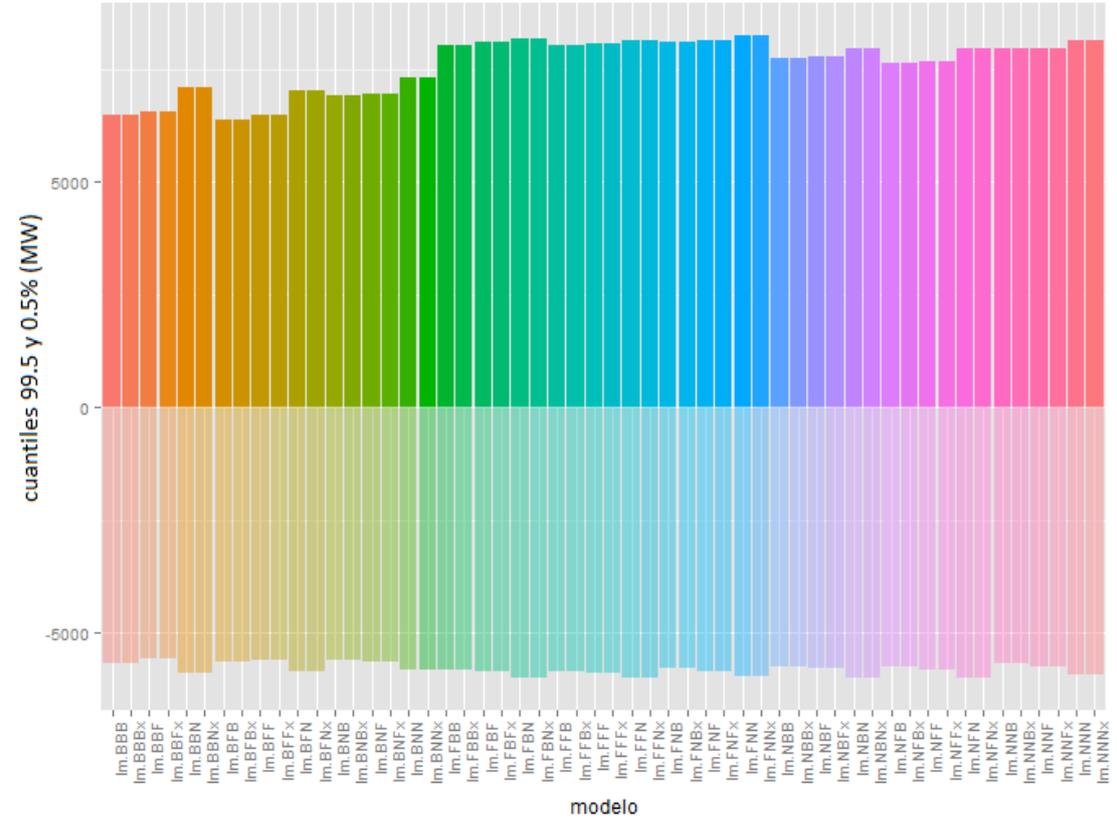


Figura 33: cuantiles extremos del error de previsión de los modelos lineales

En cuanto a los cuantiles, ocurre como con la desviación típica, los modelos con estacionalidad horaria B son los que ofrecen los mejores resultados, seguidos por los numéricos y las series de Fourier en último lugar. En cuanto a las otras estacionalidades, las diferencias no son tan evidentes, por lo que se restringirá el estudio a los modelos Im.B** para discriminar entre los modelados de las otras dos estacionalidades.

R² ajustado

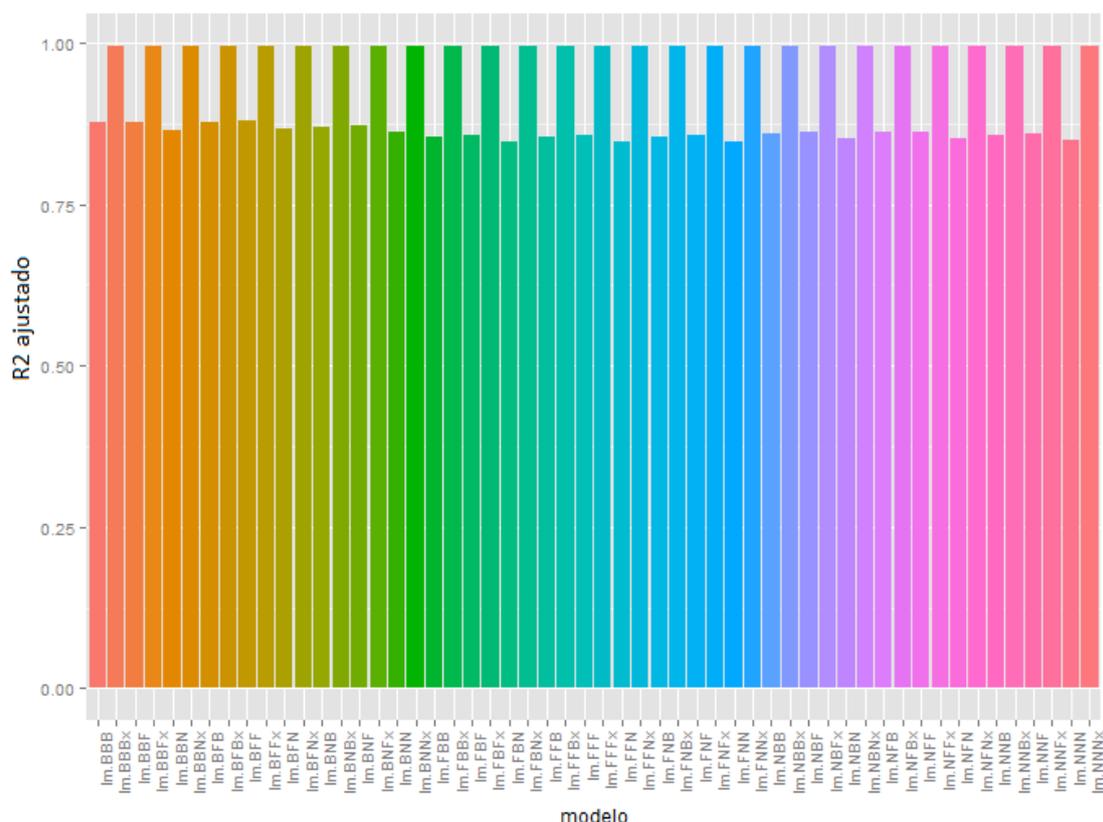


Figura 34: R² ajustado de los modelos lineales

Aunque los modelos sin demanda de referencia (Im.***) presentan un R² superior, se ha preferido guardar un modelado con ordenada en el origen. De este modo, el mayor número de MW corresponderán a esta demanda de referencia y la desviación típica más importante será la misma para todos los puntos previstos. Al ser los MW repercutidos a la tipología del día y estacionalidad menores, esto permitirá que se hagan patentes las diferencias entre ellas durante la estimación de coeficientes y que las desviaciones típicas individuales no tengan tanto impacto en la previsión final, homogeneizando el intervalo de confianza a lo largo de la curva diaria prevista.

Análisis de las estacionalidades semanal y anual.

En cuanto a la desviación típica, todos los modelos con estacionalidad horaria modelados como variable binaria ofrecen mejores resultados que sus análogos con modelado numérico o con series de Fourier. Por ello la estacionalidad horaria será modelada exclusivamente como variable binaria.

Estacionalidad semanal	Set validación		Set entrenamiento	
	Desviación típica	Media	Desviación típica	Media
Modelización				
Binaria	1645.978	129.70778	1813.262	2.396358e-13
Fourier	1711.278	135.02703	1803.515	-9.431153e-14
Numérica	1675.091	48.61074	1846.031	3.285033e-14

Tabla 1 scores de los diferentes modelados de la estacionalidad semanal

Estacionalidad anual	Set validación		Set entrenamiento	
	Desviación típica	Media	Desviación típica	Media
Modelización				
Binaria	1655.679	159.81866	1801.467	1.184805e-13
Fourier	1648.343	165.20073	1785.945	7.315353e-14
Numérica	1723.307	-11.67385	1875.396	-1.345945e-14

Tabla 2 scores de los diferentes modelados de la estacionalidad anual

Sin embargo, para las estacionalidades semanal y anual, existen discrepancias sobre la mejor modelización posible. Si bien el modelado mediante series de Fourier proporciona los mejores resultados en el set de entrenamiento, es la desviación típica del modelado binario y la media numérica la que obtiene un mejor resultado.

5.4 Selección del modelado de las distintas componentes

Al no existir una exclusión justificada de uno u otro modelo, se han conservado los modelados como variable binaria y serie de Fourier para sendas estacionalidades, descartando la estacionalidad numérica.

Se han comparado los siguientes modelos lineales: Im.BBB, Im.BBF, Im.BFB, Im.BFF e integrado ambas opciones de modelado en los modelos de series temporales.

5.5 Perfiles

Introducción

A parte de para el prototipado de modelos sencillos con los que evaluar los diferentes modelados con los que una variable explicativa es abordable, como se ha visto en los apartados anteriores los modelos lineales también ofrecen resultados interesantes, si bien más limitados, en cuanto a su uso como sistemas de previsión.

Estos permiten generar perfiles para cada una de las variables explicativas del modelo y de ese modo, una vez escogida la coyuntura del punto a prever, generar la forma de la demanda diaria que le corresponda.

Matemáticamente, si asumimos que cada uno de los niveles de las variables explicativas (los días de la semana para la estacionalidad semanal, las posibilidades de días festivos, vísperas... para el tipo de día, etc.) proviene de una variable aleatoria que obedece una ley gaussiana, con el valor de calibración óptimo como media y con una desviación típica, tendremos:

Sea $y(h,d,m,t)$ la realización de la demanda una hora (h) de un día (d) de la semana determinado de un mes (m), laborable (t).

$$\begin{aligned} & y_{h,d,m,t} \\ & H \sim N(\mu_h, \sigma_h) \\ & D \sim N(\mu_d, \sigma_d) \\ & M \sim N(\mu_m, \sigma_m) \\ & T \sim N(\mu_t, \sigma_t) \end{aligned}$$

$$\begin{aligned} \text{Así, } Y & \sim N(\mu_y, \sigma_y) = H + D + M + T \\ \mu_y & = \mu_h + \mu_d + \mu_m + \mu_t \\ \text{Var}_y & = \text{Var}_h + \text{Var}_d + \text{Var}_m + \text{Var}_t = \sigma_y^2 \end{aligned}$$

Perfiles

El ejemplo a continuación corresponde a los perfiles del modelo Im.BBB. Las líneas y regiones sombreadas corresponden a una desviación típica.

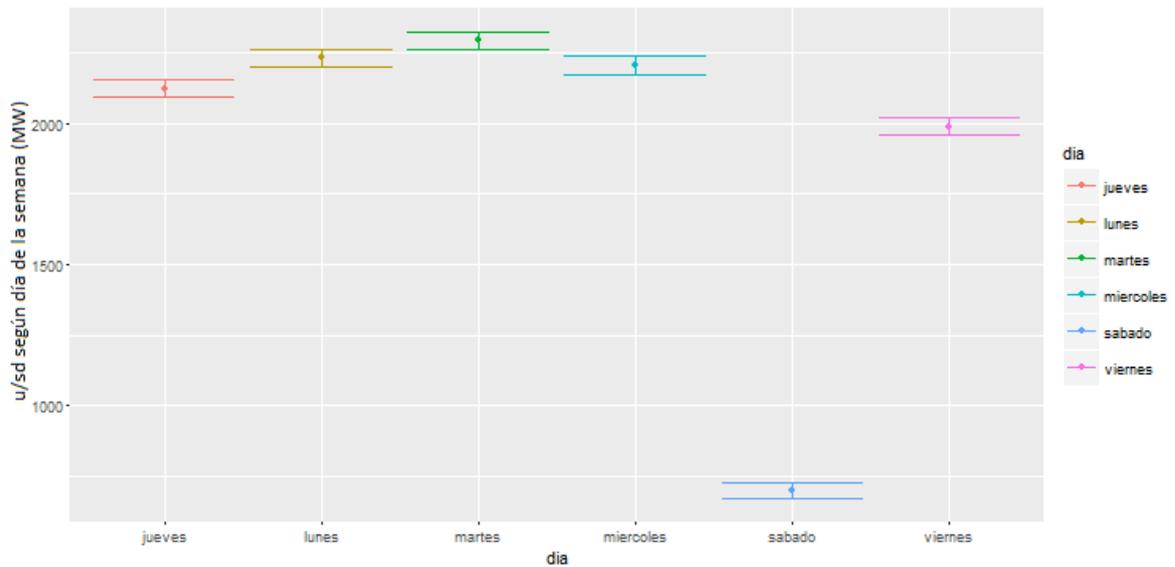


Figura 35: perfil semanal de la demanda

El perfil semanal utiliza como referencia la madrugada (0 horas) del domingo del mes de enero, por ello están ausentes en la gráfica (no hay un estimador para ese valor).

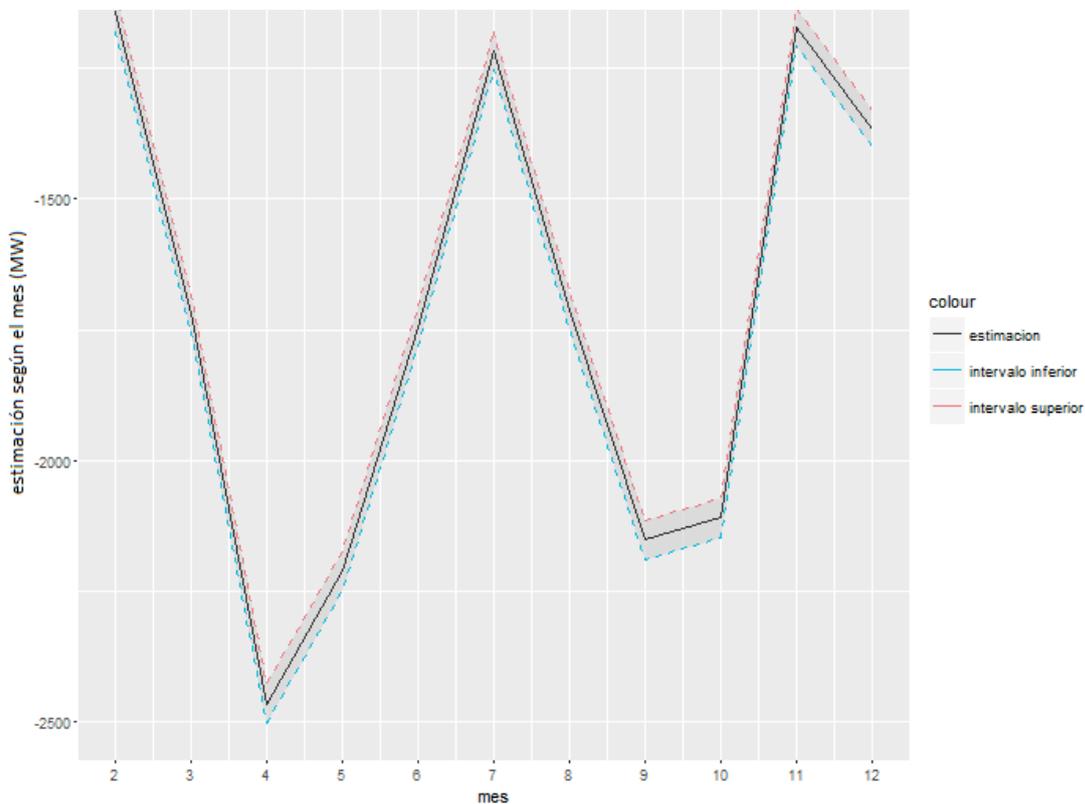


Figura 36: perfil mensual de la demanda

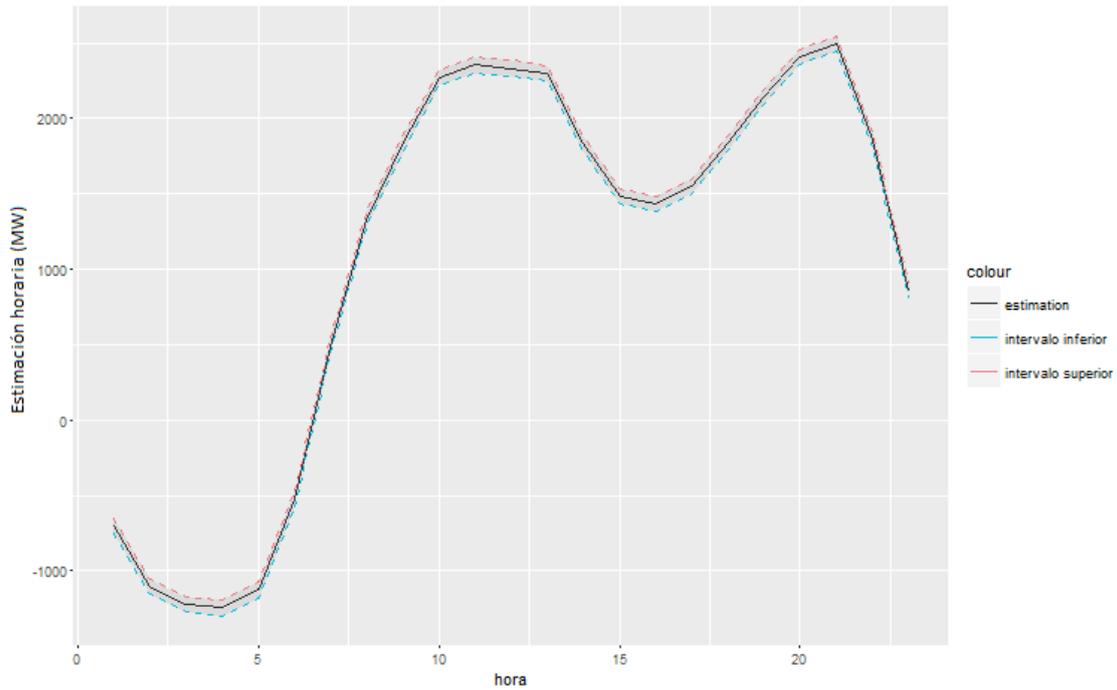


Figura 37: perfil horario de la demanda

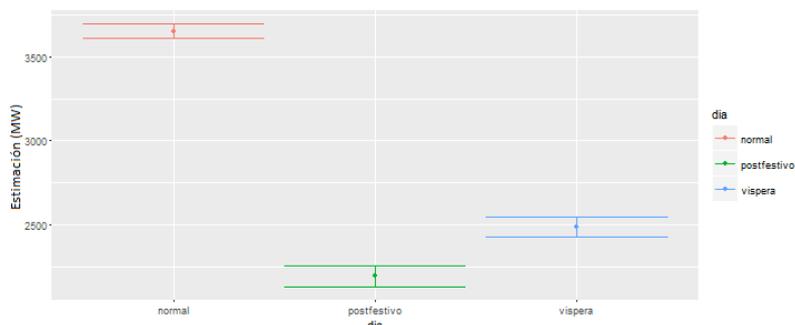


Figura 38: perfil según tipología del día de la demanda

Lo mismo sucede para los días festivos, que sirven de referencia respecto del resto de tipologías. En el perfil horario se observa cómo las rampas de la demanda horaria tienen una amplitud menor que los picos y valles.

El valor de referencia (mañana del domingo de enero) se sitúa en torno a los 5030 MW, con una desviación típica inferior a 200 MW.

Previsión:

Para el ejemplo se ha seleccionado el 10 de noviembre de 2015, que corresponde a un martes laborable. Para calcularlo, se han tomado los retardos a una semana y un año (es decir, el valor que tomó la demanda la semana previa y el año previo) ponderados por el valor de los coeficientes de la calibración y se han sumado a la media de cada uno de los niveles coyunturales de los perfiles. Lo mismo se ha hecho para las varianzas, para determinar la desviación típica.

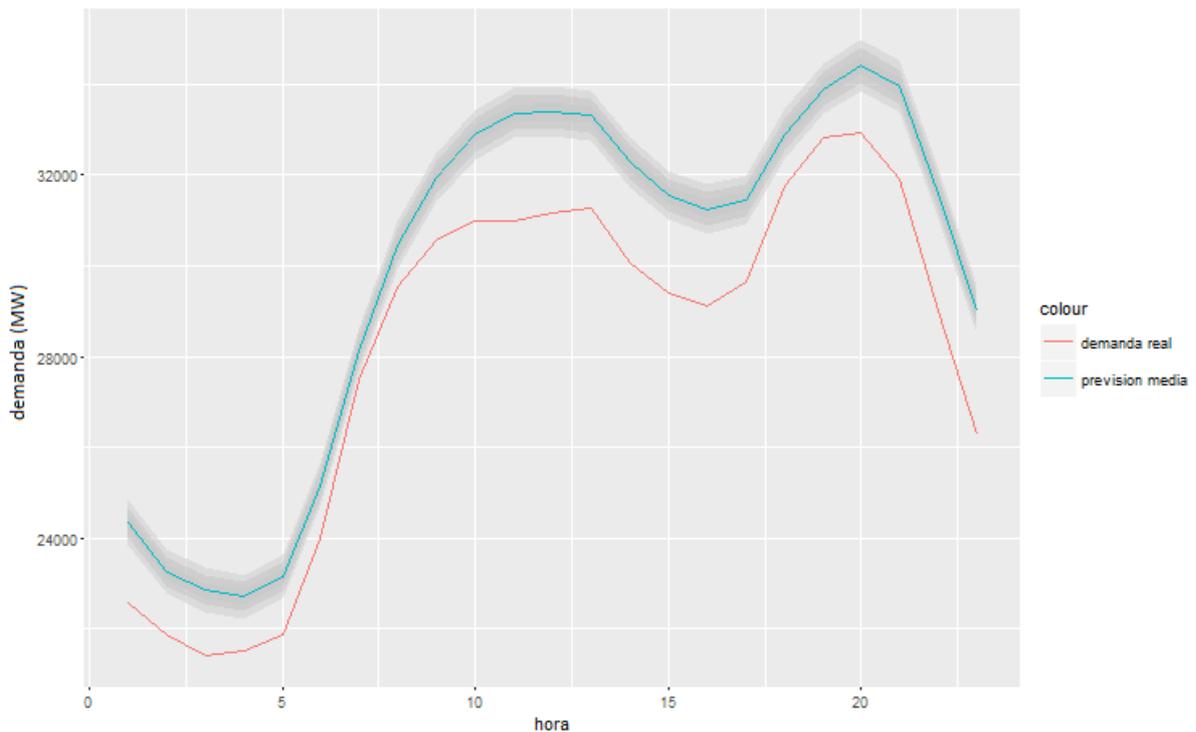


Figura 39: ejemplo de previsión de demanda con el uso de perfiles

Así, en la figura se observa cómo, si bien en la forma las dos curvas se corresponden, no existe un solapamiento por lo que el previsor sobreestima la demanda. Este error en banda puede deberse tanto a la ausencia de la temperatura dentro de las consideraciones de modelado como de la tendencia, que podría sugerir una disminución generalizada de la demanda.

El residuo está en muchos casos por encima de los 2000 MW, lo cual resulta problemático desde el punto de vista de la regulación del sistema eléctrico y pone en evidencia las limitaciones de este tipo de modelos.

6. Modelos de series temporales y alisado exponencial

6.1 Introducción

A diferencia de los modelos lineales, el análisis mediante modelos de series temporales tiene en cuenta el hecho de que los puntos de la serie temporal pueden tener una estructura interna con propiedades como la autocorrelación (dependencia entre realizaciones consecutivas de la señal), tendencia (variación de la señal media con el tiempo) o estacionalidad (repetición del comportamiento de la señal periódicamente) que deben ser tenidas en cuenta.

Si bien en los modelos lineales considerábamos las realizaciones la semana y año previo para modelar de manera aproximada los ciclos estacionales, la arquitectura de modelado en series temporales permite incluirlas en el modelo de manera mucho más rica y precisa.

6.2 Modelos ARIMA

Introducción

Los modelos de series temporales ARIMA son un tipo de modelos de regresión en los que los predictores o variables explicativas son retardos de la respuesta o de los errores de previsión.

Los modelos autoregresivos integrados de media móvil (Auto-Regressive Integrated Moving Average) están compuestos de una parte autoregresiva, AR que toma como entrada las realizaciones recientes de la serie temporal y una parte MA, que considera los valores pasados del error de previsión. La I son el número de diferenciaciones que deben efectuarse sobre la serie temporal para garantizar su estacionaridad.

Los modelos se identifican por 3 letras, ARIMA(p,d,q), que hacen referencia a:

- p: número de términos autoregresivos
- d: orden de diferenciación para garantizar la estacionaridad
- q: número de retardos del error de previsión considerados

Metodología Box-Jenkins

Para determinar el orden del modelo ARIMA, se ha utilizado el marco de trabajo de [6](Box 1970).

Esta metodología se desarrolla en 3 etapas. Una primera de identificación del modelo, garantizando que sea estacionaria y considerando la estacionalidad de la respuesta (en nuestro caso, habrá que considerar las estacionalidades horaria, semanal y anual). La segunda consiste en la estimación de parámetros una vez escogido el modelo, mediante mínimos cuadrados no lineales y por

último una tercera fase de validación del modelo verificando que la estructura del residuo de previsión sea independiente temporalmente y que éste tenga media y varianza constantes en el tiempo.

Modelos considerados

Modelado de las estacionalidades

Dada la importancia de la estacionalidad en la serie temporal de la demanda eléctrica, era necesario su inclusión en el modelo de series temporales.

Para poder tomar las 3 estacionalidades patentes en cuenta, se ha modificado el modelo ARIMA para obtener un modelo SARIMAX (Seasonal Auto-Regressive Integrated Moving Average model Exogenous) en el que la frecuencia dominante definirá la estacionalidad principal mientras que las otras dos estacionalidades serán modeladas como variables independientes.

Para obtener la frecuencia dominante se estima la función de densidad espectral del mejor modelo ARIMA (basando su selección en el AIC, criterio de información de Akaike). En caso de que exista un máximo (local o global) de la función de densidad espectral a la frecuencia f , la función tomará el valor f^1 , en caso contrario tomará el valor 1.

Tras el análisis, se obtiene que la frecuencia dominante es 24, correspondiente a la estacionalidad horaria, por lo que tanto la estacionalidad semanal como anual deberán ser incluidas como variables exógenas. A éstas se añade la variable binaria correspondiente al tipo de día (festivo...) así como si es horario de invierno o de verano.

Como se vio en el capítulo de los modelos lineales, los modelados para estas dos estacionalidades podrán ser como variable binaria o como un conjunto de series de Fourier. Esto origina cuatro familias de modelos, arima.BB, arima.BF, arima.FB y arima.FF. La primera letra referencia la estacionalidad semanal y la segunda la anual. Para cada una de las variables binarias (estacionalidad horaria y/o anual, tipo de día y horario de invierno o verano) ha sido necesario suprimir una de las columnas correspondientes a una categoría. La causa es que si no se tendrían columnas linealmente dependientes y por lo tanto la Hessiana no sería invertible.

Selección del mejor modelo de cada familia

Para seleccionar el mejor modelo de cada una de las familias se ha recurrido al criterio de Schwarz o también llamado criterio de información Bayesiano (BIC, SBC o SBIC). Dado que una de las condiciones de uso de la metodología Box-Jenkins es que el modelo generado tenga un residuo independiente de media y varianza constantes, en el criterio de información bayesiano se va más allá y se asume que se distribuye idénticamente según una distribución normal.

$$BIC = n \cdot \ln(\hat{\sigma}_e^2) + k \cdot \ln(n)$$

Ecuaciones de los modelos

Comenzamos construyendo el modelo ARMA(p,q) e introduciendo los operadores de retardo y diferenciación:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} - \theta_1 z_{t-1} - \dots - \theta_q z_{t-q} + z_t$$

$$B y_t = y_{t-1}$$

$$\nabla^d B = (1 - B)^d$$

Añadiendo las variables explicativas exógenas nos queda el modelo:

$$y_t = \sum \beta_i x_{i,t} + \sum \phi_i y_{t-1} + - \sum \theta_i z_{t-1} + z_t$$

Escribiéndolo en función del operador retardo y añadiendo el término de diferenciación, nos queda finalmente el modelo:

$$y_t = \sum \frac{\beta(B)_i}{v(B)_i} x_{i,t} + \frac{\theta(B)}{\nabla^d \phi(B)} z_t$$

Por último consideramos la parte estacional, cada 24 horas, un proceso únicamente estacional de frecuencia diaria sería $y_t \sim \text{ARIMA}(p,d,q) \times (P,D,Q)_{24}$.

$$\phi(B)_{24} \phi(B^{24}) \nabla^d \nabla_{24}^d y_t = \theta(B)_{24} \theta(B^{24}) z_t$$

Formalizándolo en una única ecuación, queda finalmente:

$$y_t = \sum \frac{\beta(B)_i}{v(B)_i} x_{i,t} + \frac{\theta(B)_{24} \theta(B^{24})}{\phi(B)_{24} \phi(B^{24}) \nabla^d \nabla_{24}^d} z_t$$

Resultados

MODELO	p	d	q
arima.BB	2	1	3
arima.BF	2	1	3
arima.FB	4	1	3
arima.FF	4	1	3

Tabla 3 configuración de los modelos ARIMA

6.3 Modelo de doble estacionalidad de Holt-Winters

Introducción

El modelo de doble estacionalidad es una modificación del modelo simple de alisado exponencial de Holt-Winters [7](Pr, 1960) propuesta por [8](Taylor J. W., 2003)

La ventaja respecto a los modelos ARIMA o SARIMAX es su mayor simplicidad de formulación. Las desventajas son dificultad de incluir las variables exógenas en el modelado así como el efecto de los puntos atípicos que, dada la inercia de los modelos de alisado pueden repercutir negativamente en las previsiones hasta que su efecto en el modelo se haya disipado.

Ecuaciones del modelo

Se parte de las ecuaciones básicas del modelo de Holt-Winters en las que l es el nivel alisado, w el índice de estacionalidad semanal, λ y ω los parámetros de alisado e $\hat{y}_t(k)$ la previsión en el instante t para $t+1$.

Las ecuaciones básicas:

$$\begin{aligned}\hat{y}_t(k) &= l_t + w_{t-s_2+k} + \phi^k \left(y_t - (l_{t-1} + w_{t-s_2}) \right) \\ l_t &= \lambda(y_t - w_{t-s_2}) + (1 - \lambda)l_{t-1} \\ w_t &= \omega(y_t - l_{t-1}) + (1 - \omega)w_{t-s_2}\end{aligned}$$

Se modifican acorde a [8](Taylor J. W., 2003) el parámetro de alisado δ y el índice d de estacionalidad diaria.

$$\begin{aligned}\hat{y}_t(k) &= l_t + d_{t-s_1+k} + w_{t-s_2-k} + \phi^k \left(y_t - (l_{t-1} + d_{t-s_1} + w_{t-s_2}) \right) \\ l_t &= \lambda(y_t - d_{t-s_1} - w_{t-s_2}) + (1 - \lambda)l_{t-1} \\ d_t &= \delta(y_t - l_{t-1} - w_{t-s_2}) + (1 - \delta)d_{t-s_1} \\ w_t &= \omega(y_t - l_{t-1} - d_{t-s_1}) + (1 - \omega)w_{t-s_2}\end{aligned}$$

Resultados

λ	0.6187614
δ	0.5858187
ω	0.92522
ϕ	0.715507
l_0	31
b_0	0.005855241

Tabla 4 parámetros modelo Holt Winters doble estacionalidad

6.5 Modelo mixto con triple estacionalidad de Holt Winters

Introducción

Este modelo es una combinación de los modelos lineales, arima y Holt-Winters cuyo objetivo es el de lograr un compromiso entre el modelado de las componentes estacionales (Holt-Winters), la tendencia (arima) y variables exógenas como el tipo de día, además de el modelado adicional de la estacionalidad anual que en el modelo de doble Holt-Winters se había excluido privilegiando las estacionalidades semanal y diaria. [9](Taylor J. W., 2010)

En primer lugar, se ha extraído la estacionalidad anual y la tendencia mediante stl, una metodología alternativa al arima X-12 propuesta por [10](Phillip G. Gould 2008) que permite modelar estacionalidades con dinámica controlable y que es más robusta frente a atípicos.

Posteriormente, se ha calibrado el modelo ARIMA de la tendencia y a la serie desprovista de estacionalidad y tendencia se ha ajustado acorde a la tipología del día, extrayendo el efecto de ser festivo, víspera o sucesor de festivo para tener una serie temporal equivalente de días no festivos.

Esta última serie se ha ajustado mediante un modelo de doble estacionalidad de Holt Winters.

Diagrama de flujo de generación del modelo

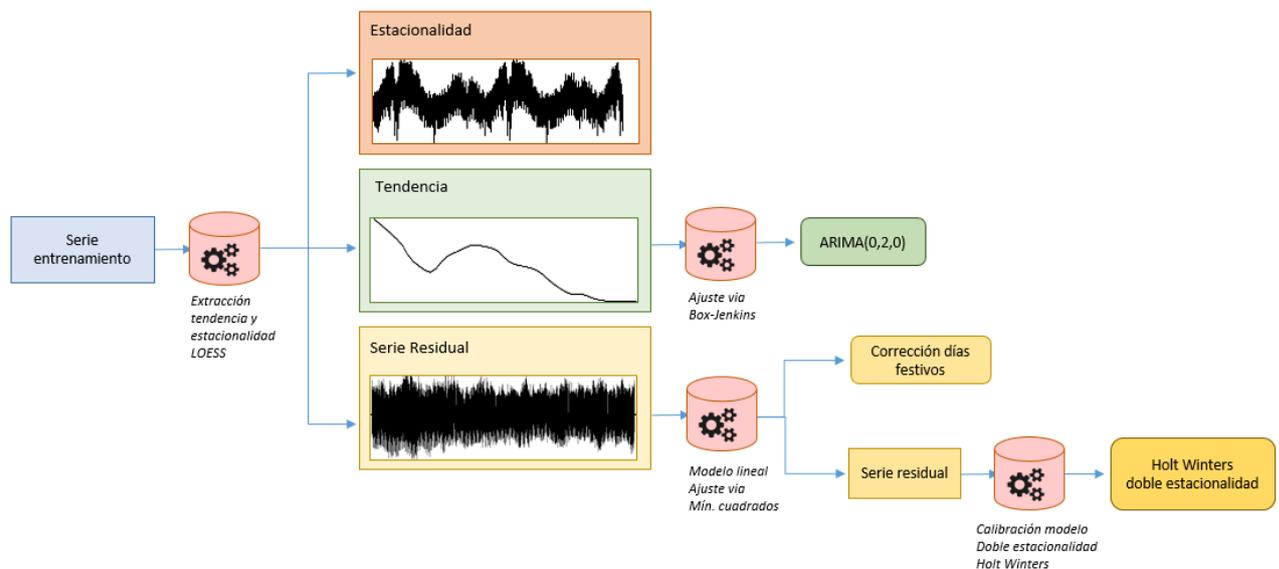


Figura 40: diagrama de flujo de generación del modelo

Diagrama de flujo de explotación del modelo

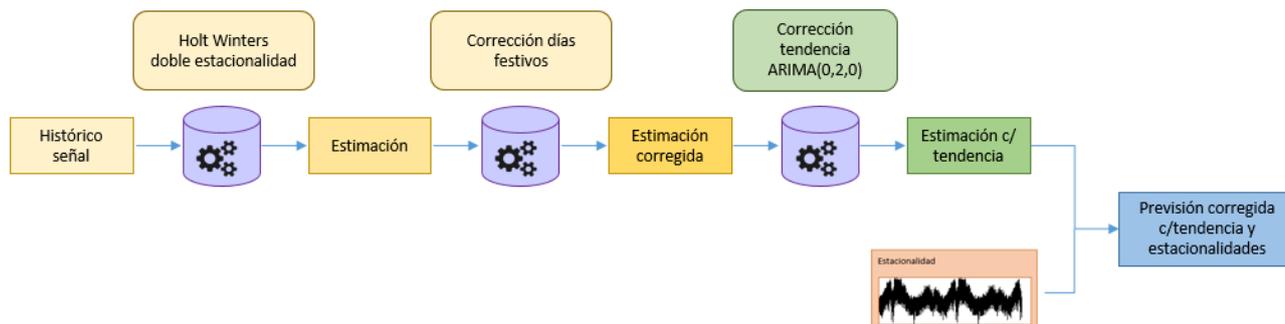


Figura 41: diagrama de flujo de explotación del modelo

Resultados

tipo día	media	desv.típica
festivo	-1001.79	101.10
normal	4.98	18.37
postfestivo	269.32	106.03
víspera	405.75	101.10

Tabla 5 coeficientes de corrección de días festivos

λ	0.4209618
δ	0.3267626
ω	0.8910752
ϕ	0.6913164
l_0	12361.65
b_0	8.275508

Tabla 6 parámetros del modelo de triple estacionalidad sin corrección de días festivos

λ	0.1445291
δ	0.0947925
ω	0.9436823
ϕ	0.8141969
l_0	12998.87
b_0	4.95338

Tabla 7 parámetros del modelo de triple estacionalidad con corrección de días festivos

7. Benchmarking

7.1 Introducción

Puede suceder que los previsores se comporten mejor ante un tipo de escenarios que frente a otros. Por ejemplo los modelos con alisado exponencial que proporcionan buenos resultados con procesos estacionales tienen en contrapartida su inercia que hace que los sucesos atípicos repercutan a posteriori en el error hasta atenuar suficientemente su impacto.

Por ello, aparte de medidas de centralidad, dispersión e indicadores sobre los residuos, se ha evaluado el comportamiento en los días singulares (festivos, vísperas...), vacaciones etc.

7.2 Medidas de centralidad y dispersión

Se han utilizado la media y desviación típica del residuo de previsión $e_t = \hat{y}_t - y_t$. La media proporcionará una medida del sesgo del predictor (tendencia a sobre/subestimar la demanda) mientras que la desviación típica servirá como indicador de la confianza de las previsiones.

7.3 Medidas del error

Para comparar los diferentes modelos de previsión construidos se han utilizado los siguientes indicadores:

Sesgo del modelo:

ME (*Mean Error*)

$$\frac{\sum \hat{y}_t - y_t}{n}$$

Magnitud del error independientemente de su signo:

RMSE (*Root Mean Squared Error*)

$$\frac{\sum (\hat{y}_t - y_t)^2}{n}$$

MAE (*Mean Absolute Error*)

$$\frac{\sum |\hat{y}_t - y_t|}{n}$$

Sesgo porcentual:

MPE (Mean Percentage Error)

$$\frac{\sum \hat{y}_t - y_t / y_t}{n}$$

Error porcentual:

MAPE (Mean Absolute Percentage Error)

$$\frac{\sum |\hat{y}_t - y_t| / y_t}{n}$$

Error escalado:

MASE (Mean Absolute Scaled Error)

$$\frac{n-1}{n} \frac{\sum \hat{y}_t - y_t}{\sum y_t - y_{t-1}}$$

El error escalado [11](Hyndman R. J., 2006) permite evaluar el error cometido por el modelo respecto de la predicción naïve consistente en utilizar la realización precedente de la señal.

Resultados:

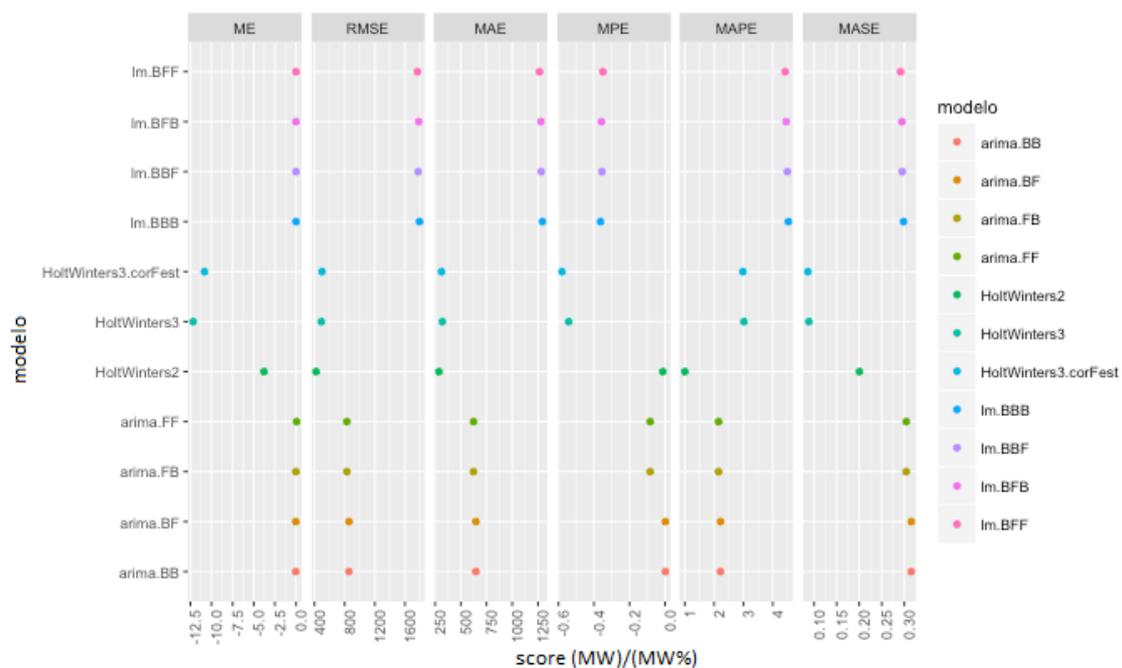


Figura 42: scores de entrenamiento de los distintos modelos

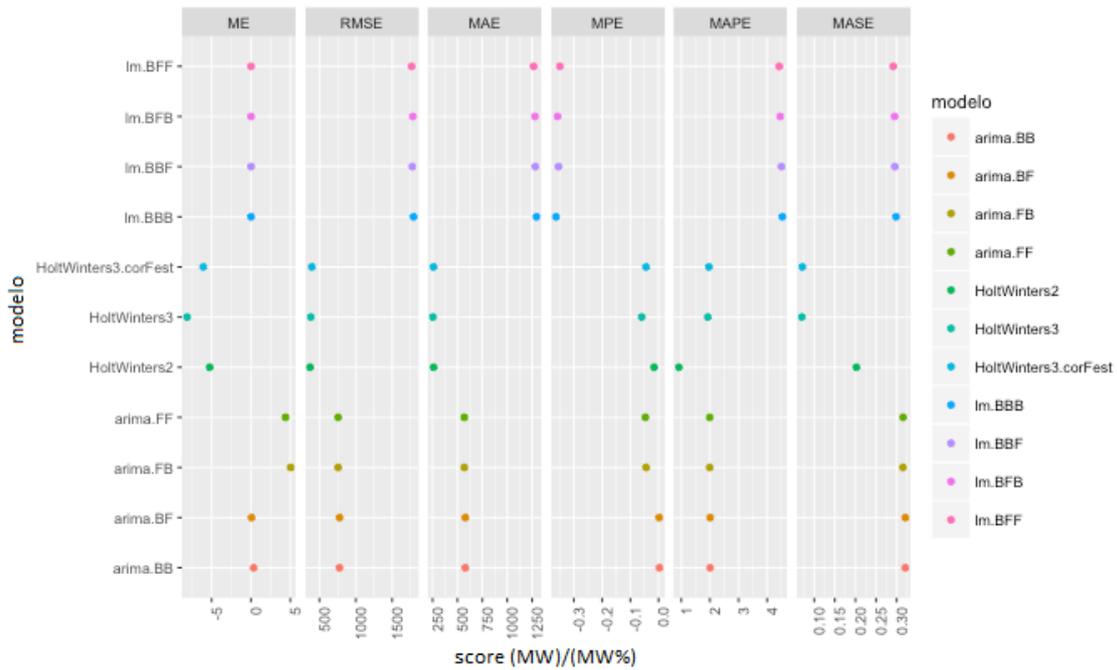


Figura 43: scores de validación de los distintos modelos

El error medio de todos los modelos excepto el Holt-Winters mixto es prácticamente nulo, que tiene un cierto sesgo negativo (-10MW). No obstante, supone menos de un 0.03% de la demanda, por lo que podemos considerar todos los modelos como carentes de sesgo.

En el error absoluto es donde se aprecian más diferencias. Los modelos lineales presentan unos resultados tanto en entrenamiento como en validación pobres, de más de 1500 MW. Los modelos ARIMA están en torno a los 750 MW de error absoluto medio y los modelos Holt Winters, ofrecen en términos absolutos el menor error, de unos 250 MW.

Porcentualmente, el modelo de doble estacionalidad se destaca con resultados del 1% mientras que los arima y Holt Winters triple están en torno al 2%. Como en términos absolutos, los modelos lineales ofrecen los resultados más pobres, por encima del 4% de error. En el error escalado, que ofrece un buen indicador de la inercia del modelo, son los Holt Winters triples los que ofrecen mejores scores seguidos del doble. Los ARIMA y lineales están muy por encima, lo que pone en evidencia su dificultad en la captación de los ciclos estacionales de la demanda.

7.4 Test de Diebold-Mariano

El test de Diebold-Mariano sirve para confrontar dos previsores mediante el análisis de sus series de residuos. La hipótesis nula es que los dos previsores son igualmente precisos, mientras que la alternativa es que uno de ellos es menos preciso.

Formalización:

Sea una serie temporal y_t y sus previsiones asociadas \hat{y}_t para todo t del intervalo temporal considerado. Se define el error de previsión $e_t = \hat{y}_t - y_t$. Se define asimismo la pérdida asociada a la previsión \hat{y}_t como una función del error $g(e_t)$, siendo $g(\cdot)$ una función de pérdidas tal que:

$$\begin{aligned}g(0) &= 0 \\ \forall e \quad g(e) &\geq 0 \\ e_{t1} > e_{t2} &\rightarrow g(e_{t1}) > g(e_{t2})\end{aligned}$$

Considerando dos previsores, definimos sus errores para todo instante e_{1t}, e_{2t} y la función diferencial de pérdidas entre ellos como:

$$d(t) = g(e_{1t}) - g(e_{2t})$$

Los dos previsores tendrán la misma precisión si y solo si la esperanza de la función diferencial de pérdidas es nula para todo instante t .

$$\begin{aligned}H_0: E(d_t) &= 0 \quad \forall t \\ H_1: E(d_t) &= \mu \neq 0\end{aligned}$$

Resultados:

De manera general, los modelos lineales se comportan peor que los arima y a su vez peor que los modelos Holt-Winters.

Dentro de la familia de modelos lineales, es el Im.BFF el que obtiene los mejores resultados (estacionalidad horaria como variable binaria, semanal y anual como series de Fourier). En la familia de arimas, son los modelos con estacionalidad semanal como serie de Fourier, arima.Fx los que proporcionan mejores previsiones respecto a los arima.Bx.

Por último, entre los modelos HoltWinters es el de doble estacionalidad el que mejor se comporta.

Sin embargo, es necesario indicar que las fronteras claras ocurren entre las distintas familias de modelos y no así dentro de ellas, donde el estadístico de Diebold-Mariano (la diferencia media de los residuos de previsión) se sitúa siempre en torno a los 10 MW.

7.5 Distribución de los residuos

Introducción:

Para evaluar el comportamiento de los distintos modelos de previsión utilizados se han confrontado calculando la distribución teórica de los residuos, asumiendo que ésta sigue una ley gaussiana.

El objetivo será discriminar entre los diferentes modelos y escoger el o los que posean una distribución con menor sesgo, es decir, lo más centrada en el cero posible) y cuyas colas sean más pequeñas. Para contextualizarlo, se han confrontado contra el error del previsor de Red Eléctrica.

Error de entrenamiento:

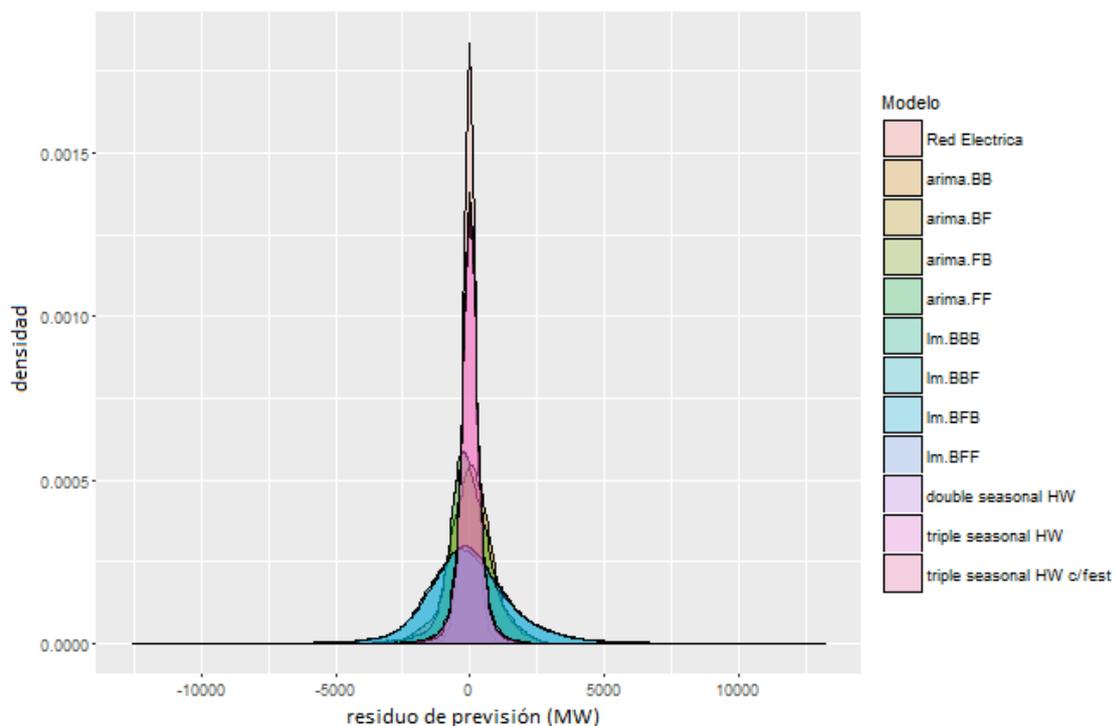


Figura 44: distribuciones del error de entrenamiento de los distintos modelos

Observamos que las tres familias presentan distribuciones claramente diferenciadas. Los modelos lineales presentan alta densidad en las colas, que disminuye en los modelos arima y que es similar en el caso de los modelos Holt-Winters al previsor de Red Eléctrica.

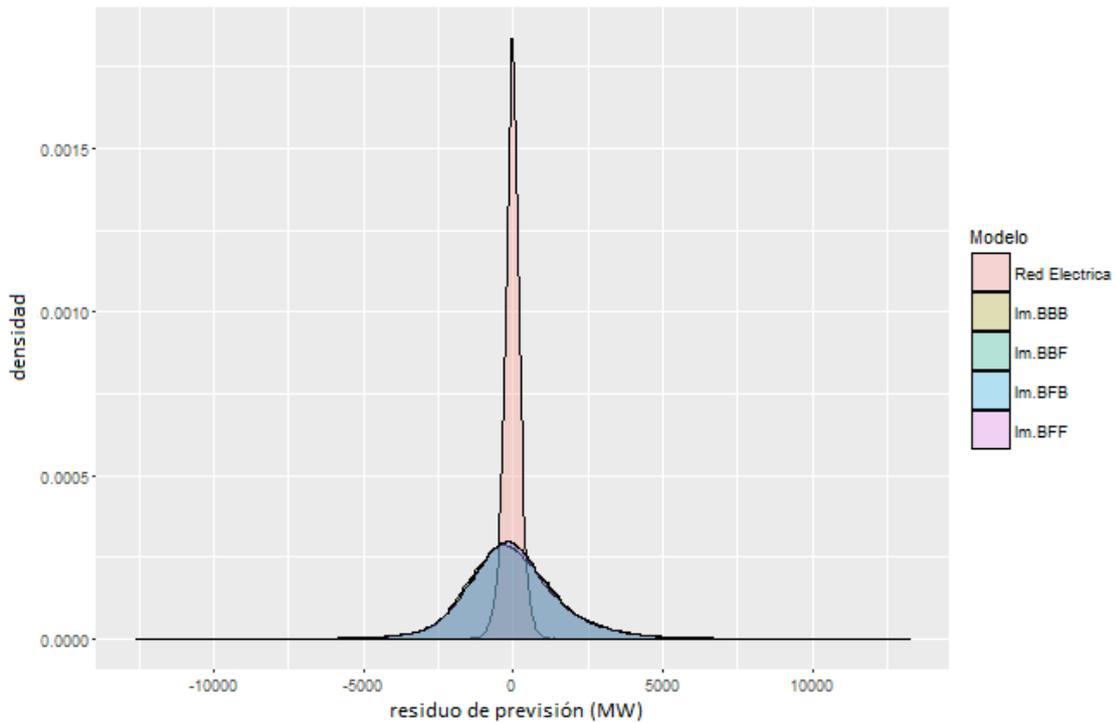


Figura 45: distribuciones del error de entrenamiento de los modelos lineales

Considerando únicamente los modelos lineales, se aprecia una alta densidad en sus colas. Si bien el error parece contenido en el rango de los ± 5000 MW, esto supone más de un 10% de la demanda, y no resulta admisible en términos operacionales.

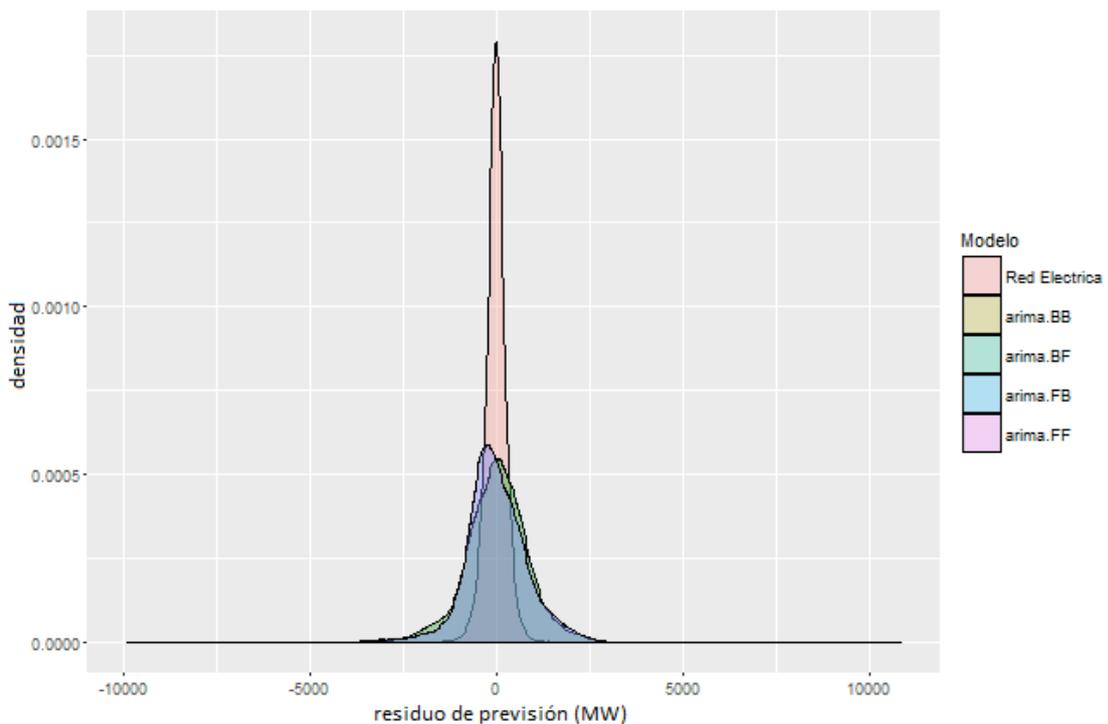


Figura 46: distribuciones del error de entrenamiento de los modelos ARIMA

En los modelos ARIMA el error se distribuye con una mayor densidad central. Los modelos arima.Fx son los que presentan una distribución con menor densidad en las colas, aunque las diferencias dentro de los arima.Fx y arima.Bx no son significativas.

Por último, los modelos Holt-Winters presentan el menor error de entrenamiento. La distribución de los 3 modelos es prácticamente idéntica y, salvo por una menor densidad central que se distribuye homogéneamente en ambas colas, también muy similar al error de Red Eléctrica. Sin embargo es necesario recordar que el error del previsor de Red Eléctrica se trata de un error de validación, ya que se corresponde a las verdaderas previsiones efectuadas.

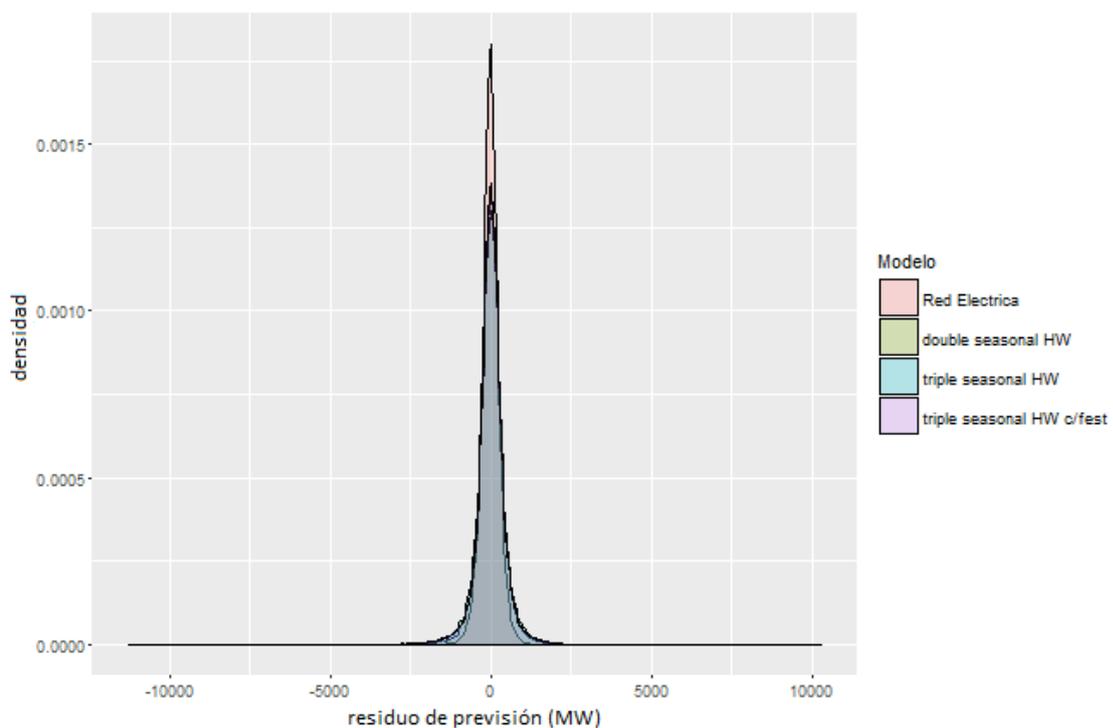


Figura 47: distribuciones del error de entrenamiento de los modelos Holt-Winters

Error de validación:

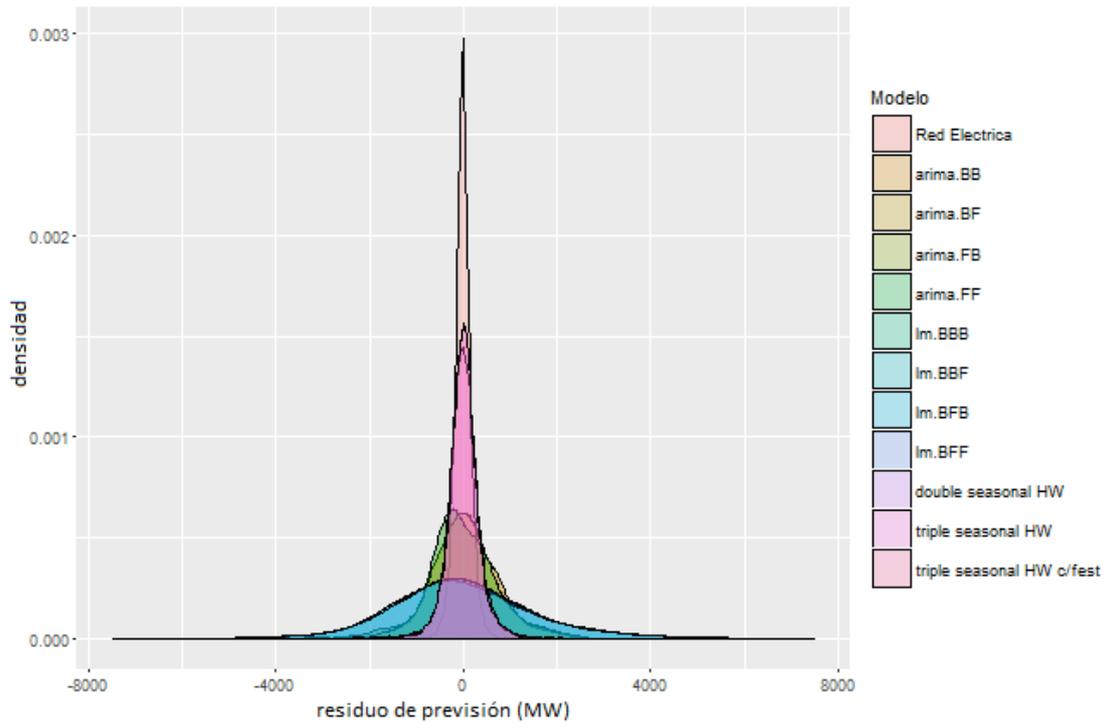


Figura 48: distribuciones del error de validación de los distintos modelos

Tal y como sucedía con el error de entrenamiento, las tres familias de modelos siguen estando diferenciadas, con mejores resultados para los modelos Holt-Winters.

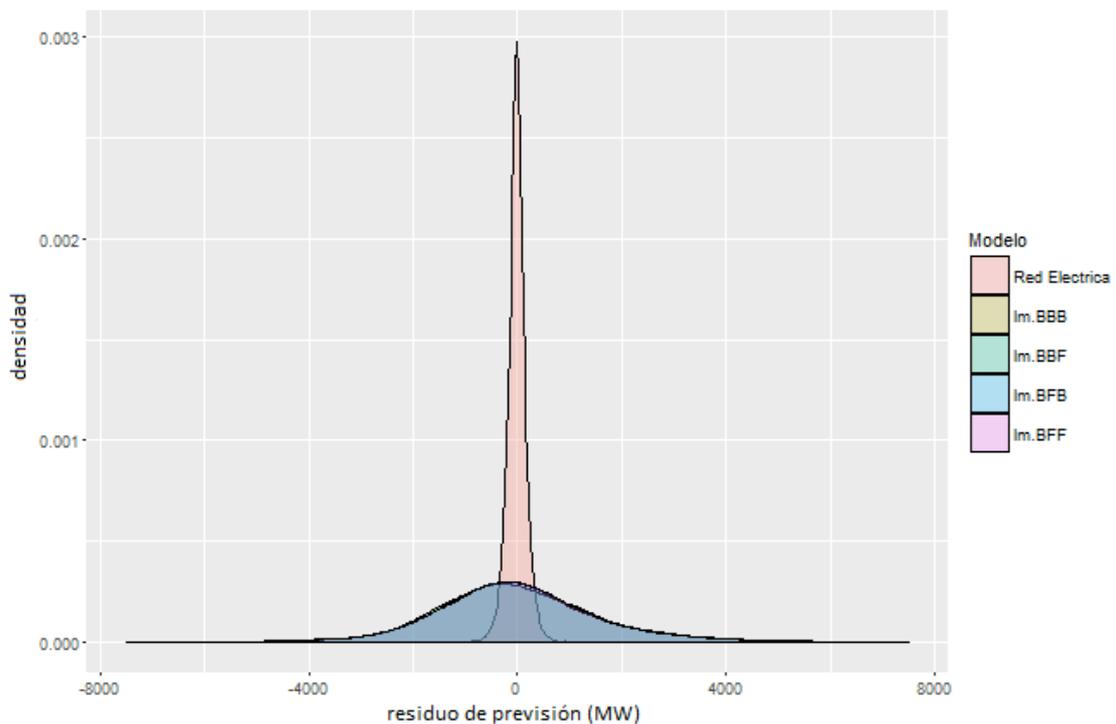


Figura 49: distribuciones del error de validación de los modelos lineales

Los modelos lineales presentan una densidad con menor amplitud, disminuyendo su rango en 1000 MW aunque ésta disminuye en la parte central y aumenta en las colas de la distribución.

En el caso de los modelos ARIMA también sucede algo similar, una menor densidad en el centro que se afecta a las colas y una disminución de la amplitud, aunque menos evidente que en el caso de los modelos lineales.

Ambas familias de modelos presentan un error por encima de lo admisible en términos operacionales.

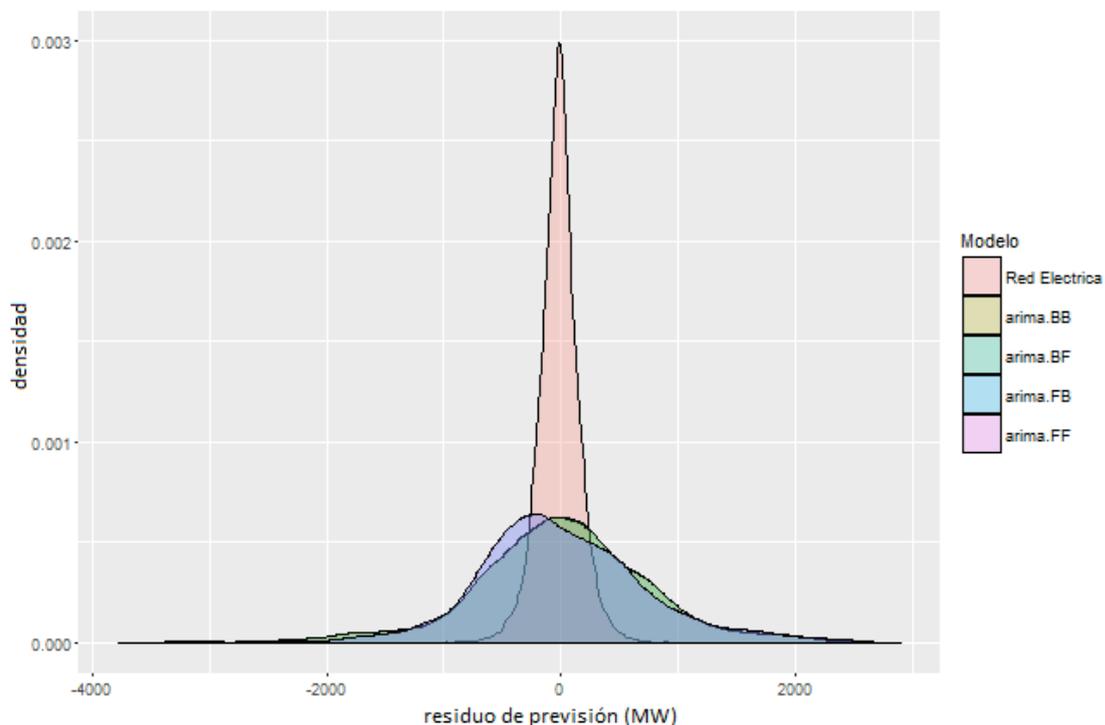


Figura 50: distribuciones del error de validación de los modelos ARIMA

Por último, en los modelos Holt-Winters la amplitud de la distribución del error de validación y de entrenamiento son muy similares. No obstante, el modelo de triple estacionalidad con corrección de días festivos presenta una densidad mayor en la parte central que el doble estacional o el triple sin corrección.

El error se mantiene contenido en los ± 1000 MW, por lo que se podría considerar los modelos Holt-Winters como una solución válida operacionalmente.

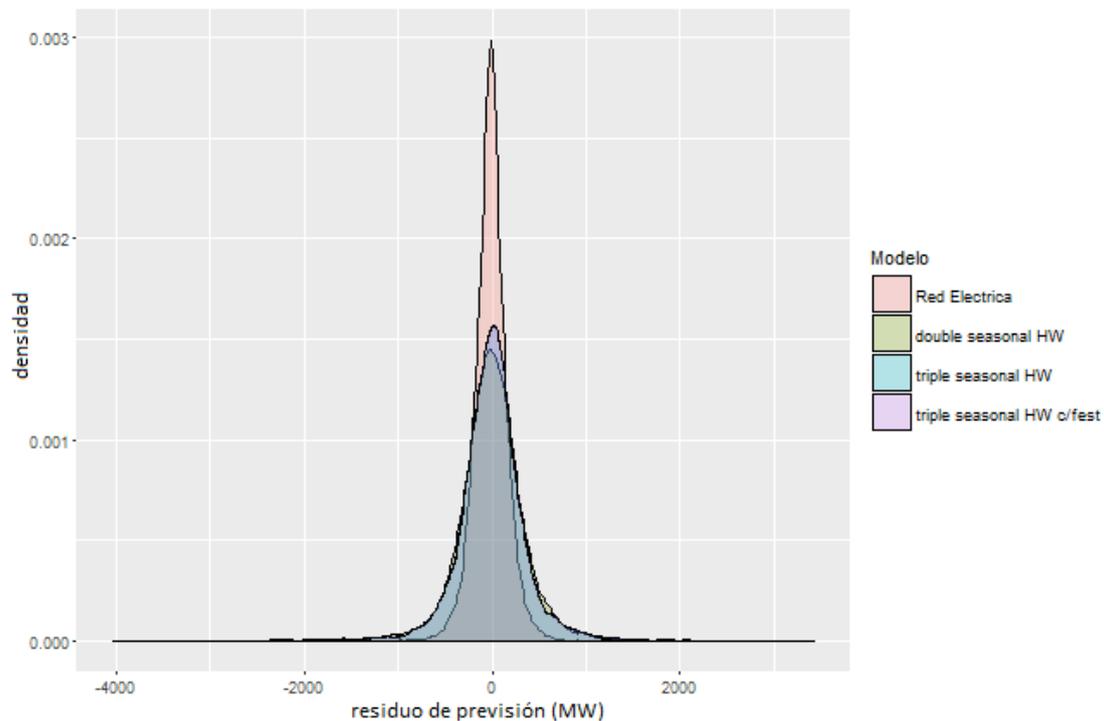


Figura 51: distribución del error de validación de los modelos Holt-Winters

Conclusiones:

De las tres familias de modelos, son los Holt-Winters los que proporcionan los mejores resultados tanto en entrenamiento como en validación. La corrección de los días festivos aporta una mayor precisión y aunque la mejora no es significativa, la complejidad añadida en el modelo se compensa con esta ganancia adicional.

Se ha confrontado el predictor de triple Holt Winters con corrección de días festivos al predictor de Red Eléctrica para analizar pormenorizadamente si globalmente el modelo actual sigue siendo superior o si el modelo propuesto mejora en alguno de los casos los valores obtenidos.

El mejor desempeño de los modelos con alisado exponencial y estacionalidad se debe a que, pese a la ausencia de la temperatura como variable de entrada, la parte termosensible de la demanda es una transformación de ésta mediante una función de transferencia. La mayor inercia de los modelos con alisado exponencial, si bien plantea un riesgo en caso de la ocurrencia de eventos atípicos, permite captar mejor esta dinámica que los modelos lineales o ARIMA.

7.6 Análisis de los modelos Holt Winters

Los tres modelos ofrecen resultados muy parecidos, si bien notablemente con el set de entrenamiento y de manera algo más ligera en el set de validación, los modelos triples, con la estacionalidad anual extraída previamente ofrecen unos mejores resultados.

Scores frente al modelo Red Eléctrica

modelo	ME	RMSE	MAE	MPE	MAPE	MASE
HoltWinters2	-3.794	420.708	287.559	-0.012426	1.00529	0.200378
HoltWinters3	-12.171	489.602	320.043	-0.054245	3.01552	0.088236
HoltWinters3 Corr.	-10.836	499.518	311.547	-0.058015	2.98237	0.085981
REE	-27.100	313.487	209.973	-0.089114	0.73532	0.048424

Tabla 8 scores set de entrenamiento del modelo REE y los modelos HW

modelo	ME	RMSE	MAE	MPE	MAPE	MASE
HoltWinters2	-5.244	371.836	259.887	-0.015557	0.91746	0.202396
HoltWinters3	-8.112	382.347	254.095	-0.059290	1.92790	0.070166
HoltWinters3 Corr.	-6.063	395.225	258.940	-0.044172	1.96776	0.071477
REE	-27.100	313.487	209.973	-0.089114	0.73532	0.048424

Tabla 9 scores set de validación del modelo REE y los modelos HW

Aunque el modelo de Red Eléctrica ofrece los mejores resultados, el sesgo de los modelos Holt Winters es menor, si bien como se expuso antes una diferencia de 10 MW, inferior al 0.03% de la demanda no es significativa.

Este sesgo inferior también se refleja en el error medio porcentual (MPE), aunque en términos absolutos, es el predictor de Red Eléctrica quién ofrece mejores resultados, con 100-150 MW menos de error medio y entre 0.3 y 2% menos en términos porcentuales.

En todos los casos, es el modelo Holt Winters doble el que ofrece los resultados más precisos frente a los otros dos modelos, de triple estacionalidad, excepto por el error escalado. Esto puede evidenciar que, si bien el modelo funciona muy bien de manera general, presenta un ligero sobreentrenamiento que dificultan su adaptación a cambios repentinos, cosa que, los modelos triples pueden hacer al contar con una mayor flexibilidad en detrimento de un error medio general algo más elevado.

Distribuciones del error

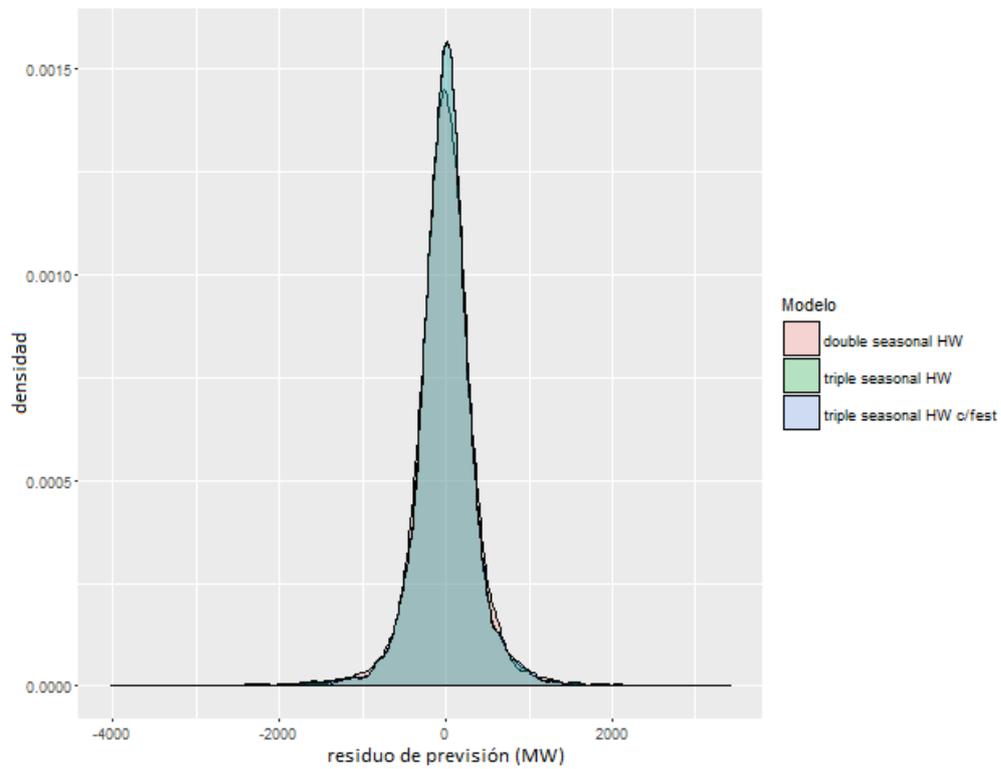


Figura 52: distribución del error de entrenamiento de los modelos Holt-Winters

El error de entrenamiento y de validación es muy similar entre los tres modelos, si bien el modelo triple con corrección de días festivos parece presentar una mayor densidad central y menor en las colas.

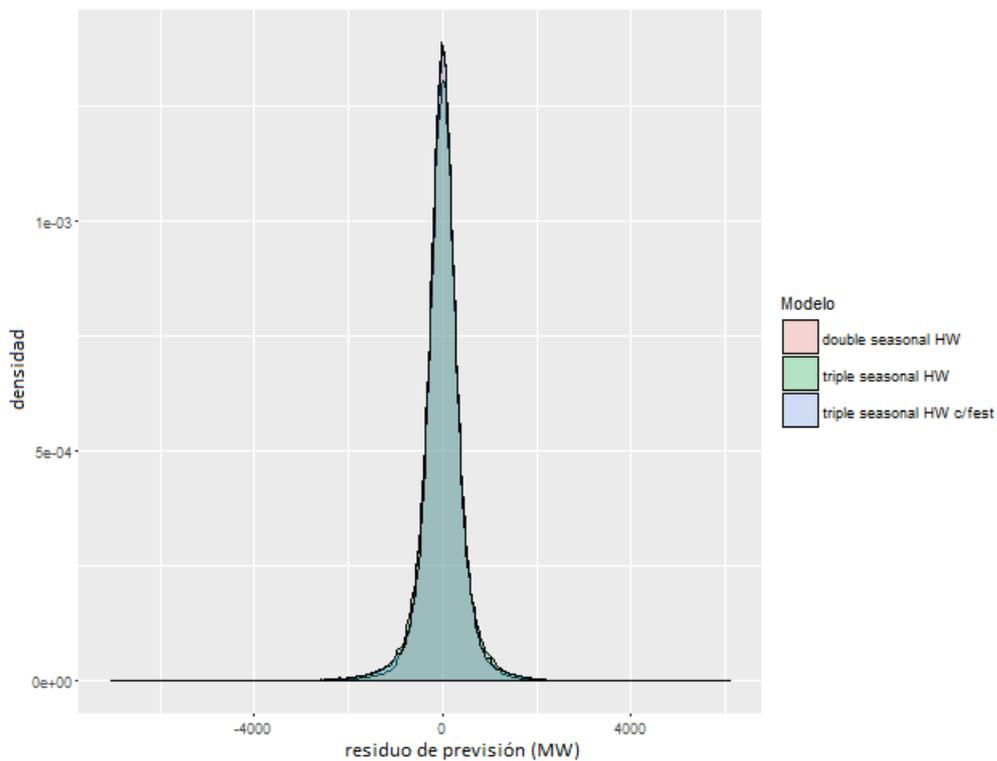


Figura 53: distribución del error de validación de los modelos Holt-Winters

Modelado de la tendencia

La ausencia de grandes mejoras entre el modelo de doble alisado y los de triple radica, sobre todo, en la dificultad de estimación de la tendencia a largo plazo.

En la descomposición y extracción de la estacionalidad anual, se observa que la estimación de la tendencia tiene un perfil que presenta dificultades para estimar a un año vista, y que en caso de sobre o subestimarlos puede imponer variaciones superiores a los 500 MW.

A la hora de hacer previsiones sobre la tendencia, se ha sido bastante conservador obteniendo valores que en ningún caso suponían una variación de más del 1% (280 MW) de la media anual precedente.

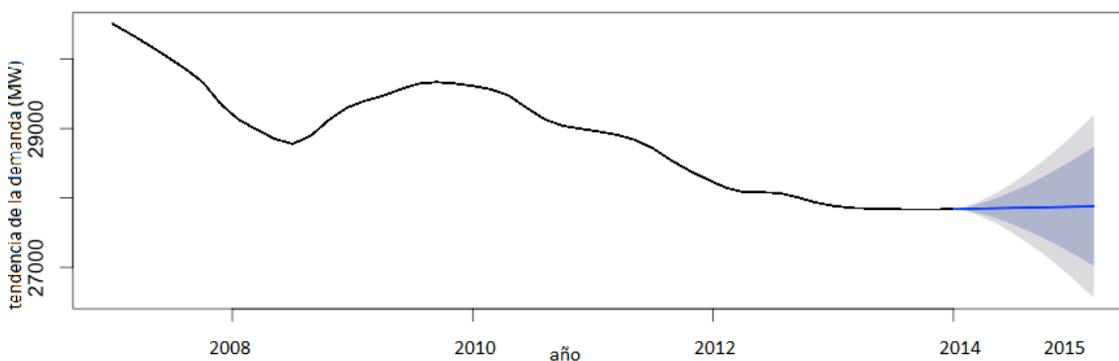


Figura 54: estimación de la tendencia mediante Loess y previsión para la validación

Comparación con el modelo REE según la coyuntura

A pesar de la superioridad del modelo de Red Eléctrica de manera general, puede suceder que los modelos Holt Winters se comporten de forma más adecuada en determinados escenarios. Además, debido a las ligeras diferencias entre los modelos Holt Winters, será necesario analizar por un lado si se justifica la inclusión de la corrección de días festivos o no y por otro si la extracción de la estacionalidad anual supone una diferencia o si por el contrario degrada el modelo al provocar sobreaprendizaje.

De manera general, los modelos se encuentran centrados y contenidos salvo ciertos valores atípicos, notables en el modelo de corrección de días festivos.

Justificación del modelo de corrección de días festivos:

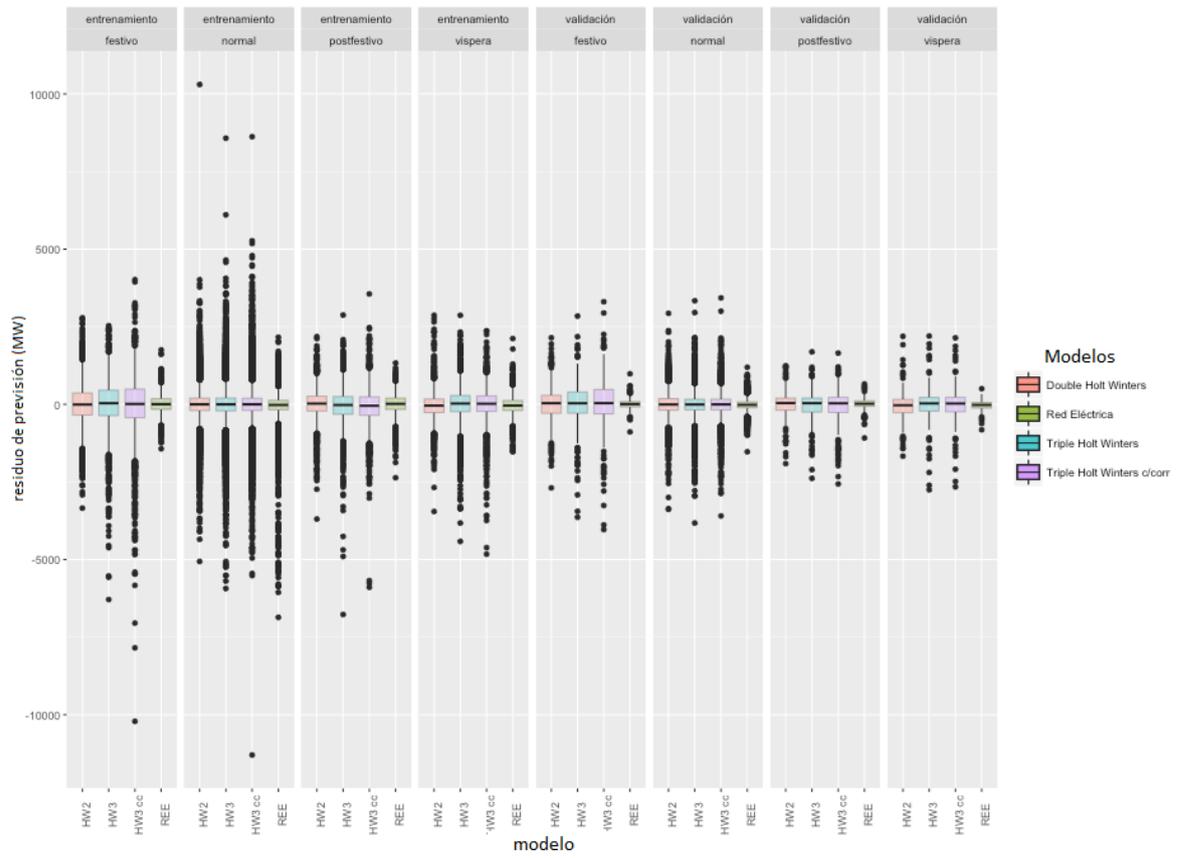


Figura 55: boxplot según la tipología de día

Analizando el boxplot, el modelo de triple estacionalidad de Holt Winters con corrección de días festivos aporta peores resultados que el modelo sin corrección, por lo que conservaremos el modelo sin corrección dentro de los modelos Holt Winters de triple estacionalidad.

Esta degradación puede estar causada porque una variable binaria no consiga modelar adecuadamente las variaciones de la dinámica de los días festivos y sus vísperas y sucesores. La complejidad adicional, el sobreaprendizaje y la incertidumbre no compensan su inclusión en el modelo, mientras que por el contrario sí puede ser interesante la consideración en una etapa posterior de corrección externa al modelo.

Justificación de los modelos con extracción de la estacionalidad anual:

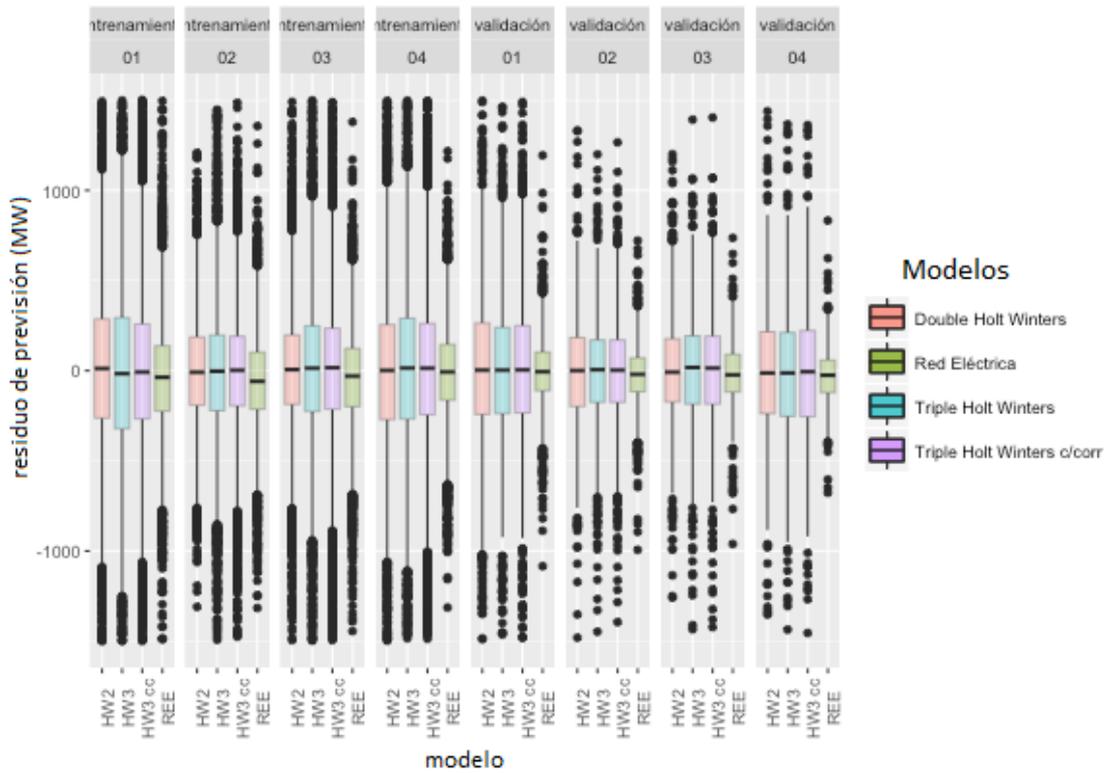


Figura 56: boxplot según el mes (enero-abril)

En el primer tercio del año los modelos de triple estacionalidad no aportan una mejora notable respecto al de doble e incluso degradan los resultados.

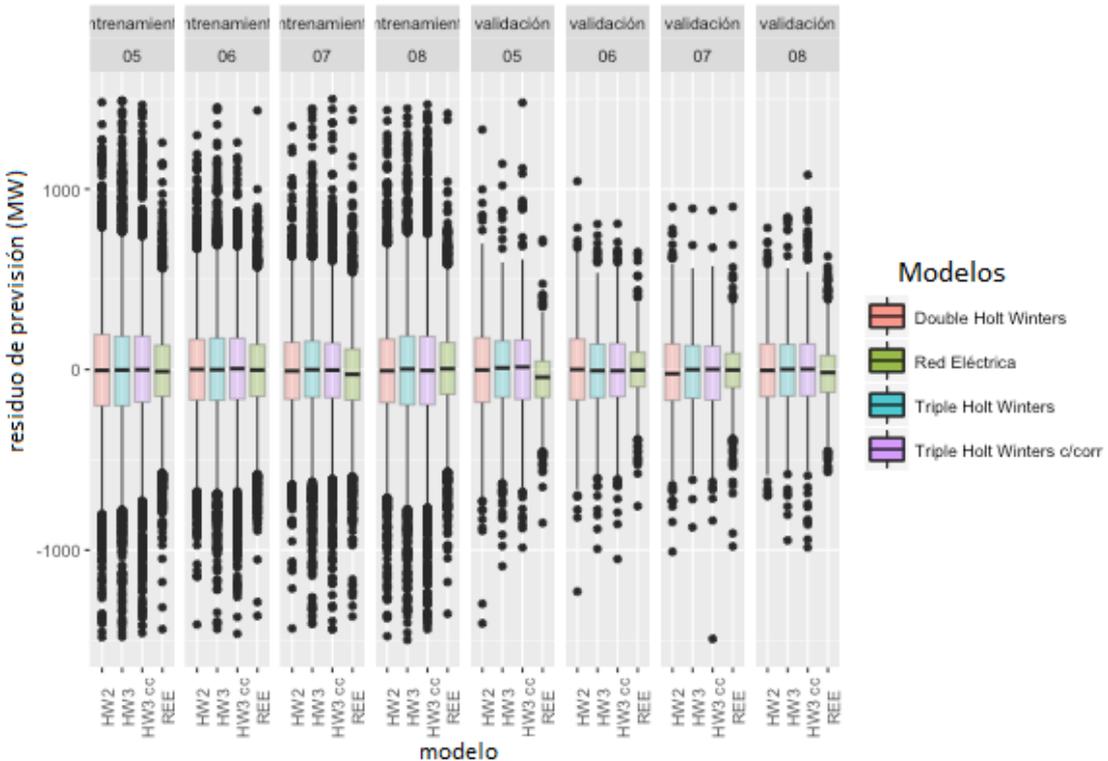


Figura 57: boxplot según el mes (mayo-agosto)

Sin embargo, en los meses cálidos los modelos de triple estacionalidad ofrecen resultados sensiblemente mejores.

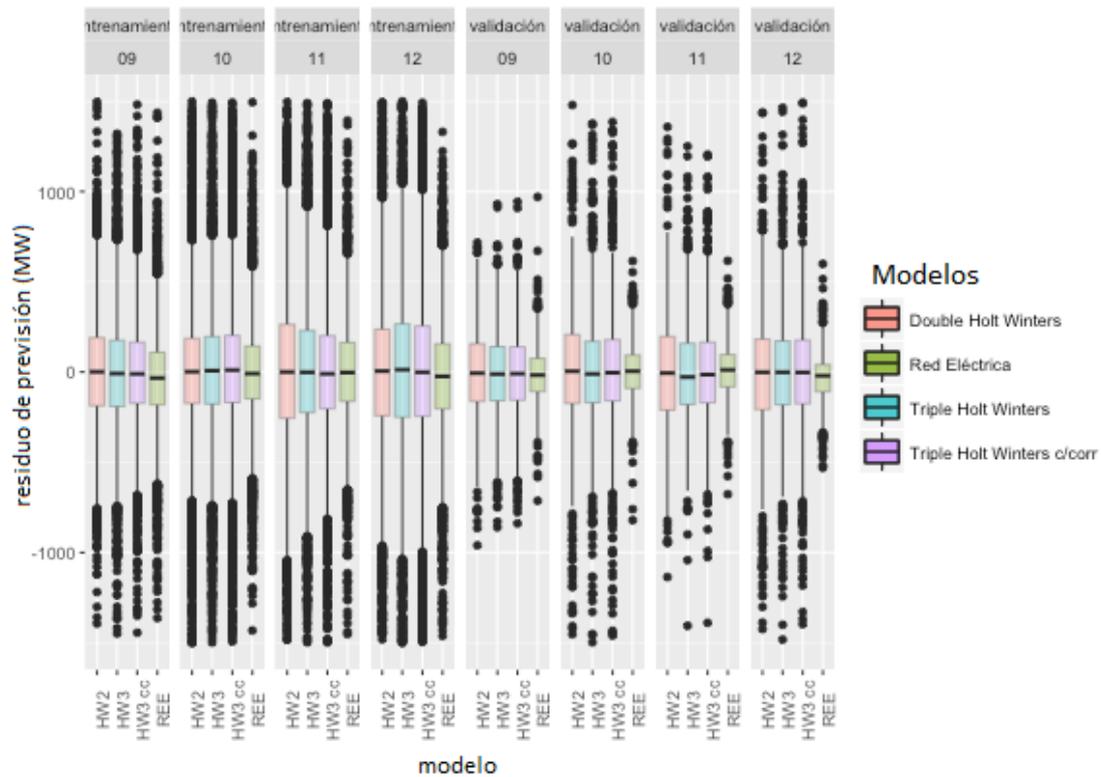


Figura 58: boxplot según el mes (sept-dic)

Por último, en los últimos cuatro meses del año es donde obtenemos la mayor mejora relativa con los modelos de triple estacionalidad.

Conclusiones del análisis coyuntural:

Si bien el modelo de triple estacionalidad con corrección de días festivos no resulta interesante por la complejidad que añade y porque el sobreaprendizaje degrada el modelo, la extracción de la estacionalidad anual en los modelos de triple estacionalidad ofrece buenos resultados, concretamente a partir de la segunda mitad del año.

Esto pone de manifiesto que si bien la estacionalidad anual tiene una gran importancia en la demanda eléctrica, también lo tiene la termosensibilidad y que al no contar con la serie de temperatura como entrada, se repercute en la estacionalidad una parte de ésta, provocando un sobreaprendizaje y que ante inviernos no convencionales (olas de frío o inviernos particularmente suaves), la extracción previa de la estacionalidad provoque sobre o subestimaciones de la demanda.

Comportamientos extremos y atípicos

Del mismo modo que es necesario analizar el comportamiento habitual de los modelos, también lo es estudiar qué sucede en los extremos, puntos con un error de previsión muy elevado.

Esto tiene un doble objetivo, el primero es el de analizar las fortalezas y debilidades de cada modelo y el segundo, si existen un modelo con un comportamiento degradado en un determinado tipo de ocasiones y otro que proporciona buenas previsiones en ese tipo de escenarios, construir modelos que sean la combinación ponderada de dos o más de ellos.

Cuantiles

0.1%	0.5%	1%	99%	99.5%	99.9%	modelo
-2438.369	-1618.630	-1238.169	1145.160	1418.655	2090.663	Holt Winters 2
-3259.186	-1986.006	-1545.848	1294.290	1603.696	2381.292	Holt Winters 3
-3569.270	-2090.250	-1583.380	1361.762	1695.918	2800.425	Holt Winters 3 c/corrección
-2361.897	-980.833	-796.673	688.500	816.123	1174.725	Red Eléctrica

Tabla 10 cuantiles del error de entrenamiento

0.1%	0.5%	1%	99%	99.5%	99.9%	modelo
-2202.516	-1342.731	-1083.913	1008.365	1223.812	1844.138	Holt Winters 2
-2486.310	-1585.226	-1174.665	1005.278	1290.267	1942.261	Holt Winters 3
-2565.600	-1685.159	-1222.356	1063.989	1360.941	2035.019	Holt Winters 3 c/corrección
-2361.897	-980.833	-796.673	688.500	816.123	1174.725	Red Eléctrica

Tabla 11 cuantiles del error de validación

Tanto en validación como en entrenamiento es el modelo de doble estacionalidad el que proporciona los mejores resultados, siendo incluso inferiores al modelo de Red Eléctrica en algún caso. Esto refleja que la estacionalidad anual no es meramente aditiva, como se ha supuesto al extraerla en los modelos triples, sino que influye en las otras dos estacionalidades a lo largo del año.

Errores superiores a 2000 MW

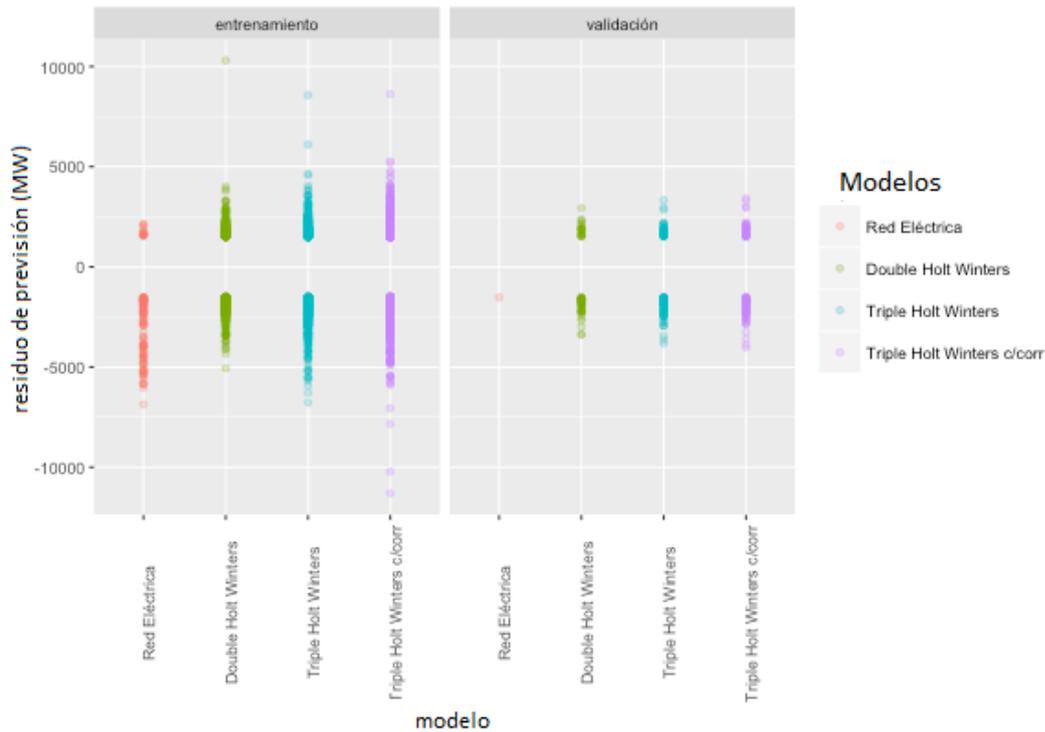


Figura 59: dispersión de los errores superiores a 2000 MW según modelo

Se observa que a diferencia del modelo de Red Eléctrica, que tiene una clara tendencia a la sobreestimación de la demanda en los errores extremos, los modelos propuestos no presentan ese sesgo, salvo el modelo de triple estacionalidad con corrección de días festivos. De manera general, el modelo de doble estacionalidad ofrece errores extremos de menor amplitud que los de triple, aunque superiores a los del modelo Red Eléctrica.

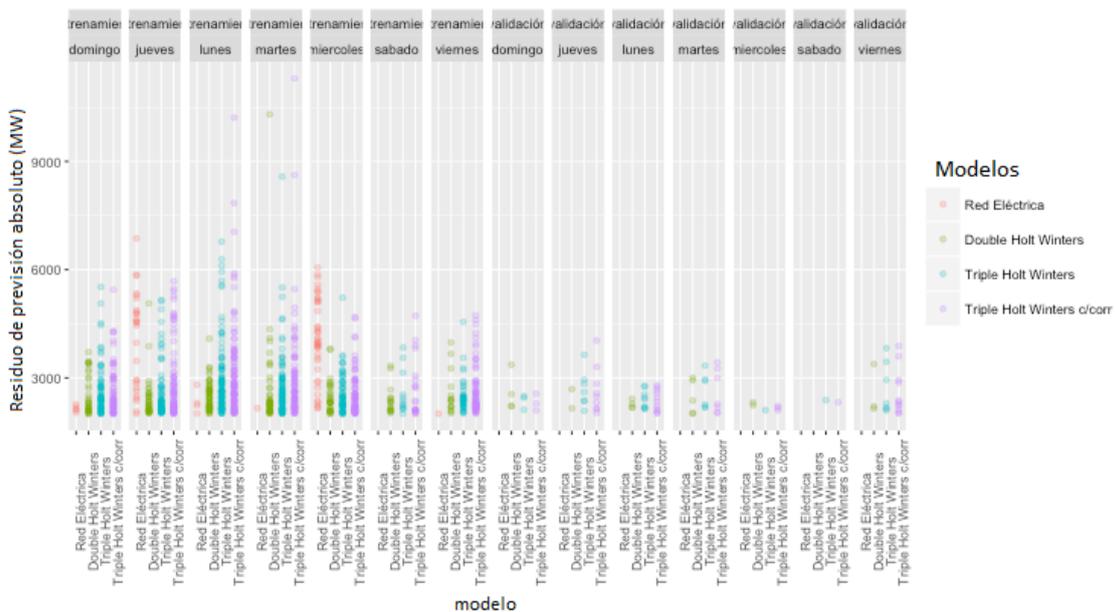


Figura 60: dispersión de errores superiores a 2000 MW por día de la semana

Semanalmente, el número de puntos extremos en los que los modelos propuestos proporcionan mejores resultados que el modelo de Red Eléctrica no es significativo como para poder validar una superioridad de los modelos Holt Winters los miércoles y jueves frente al modelo REE.

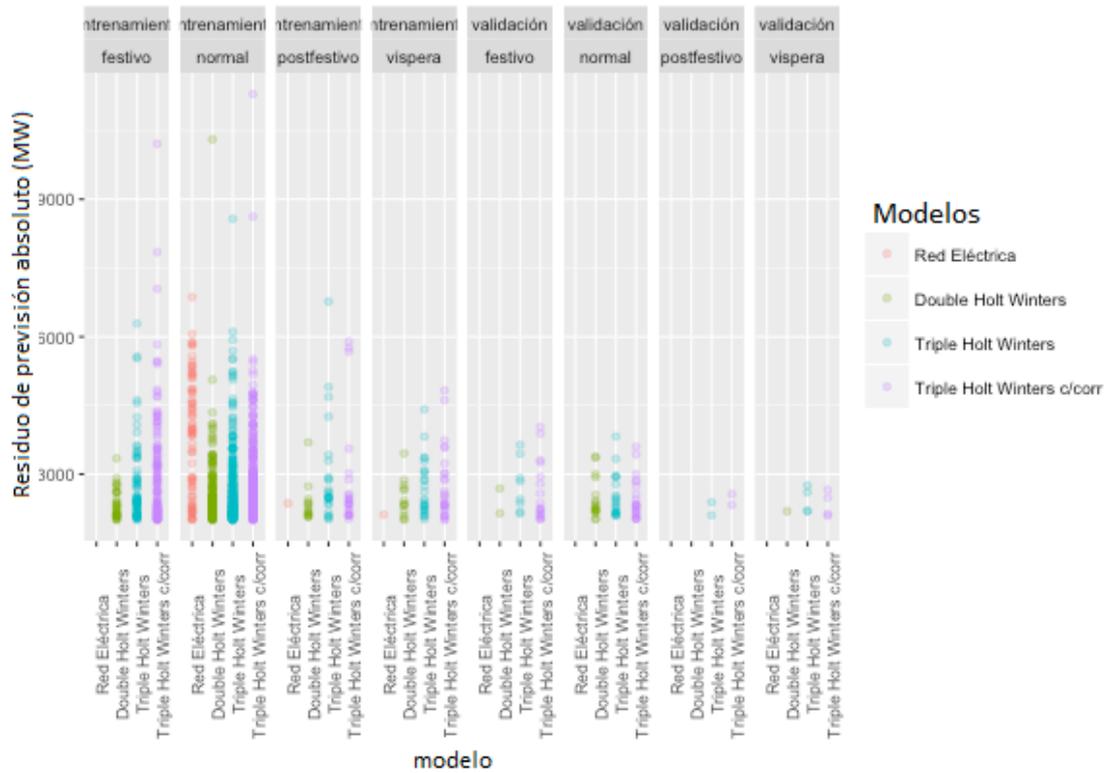


Figura 61: dispersión errores superiores a 2000 MW según tipología de día

En cuanto a la tipología de día, el modelo Holt Winters de triple estacionalidad con corrección se comporta sensiblemente mejor en los extremos que el modelo sin ella. Sin embargo, el comportamiento presenta una mayor degradación que en el modelo de doble estacionalidad. El modelo REE no presenta en festivos, vísperas o posteriores a festivos errores superiores a los 2000 MW, por lo que su degradación respecto a este factor es menor que los modelos propuestos.

Valoración en términos de energía

A parte de valorar la previsión en términos de centralidad, dispersión y extremos, es necesario relativizar la previsión respecto del volumen de ajuste que el operador del sistema ha necesitado movilizar para garantizar en todo momento el balance entre la generación y la demanda eléctrica.

Red Eléctrica	Holt Winters 2	Holt Winters 3	Holt Winters 3 c/corr
12878278	17619262	19629237	19108150
-	+36.8138 %	+52.4213 %	+48.3750 %

Tabla 12 volúmenes de ajuste del set de entrenamiento

Red Eléctrica	Holt Winters 2	Holt Winters 3	Holt Winters 3 c/corr
4898180	2654951	2598380	2647925
-	-45.79719 %	-46.95213 %	-4594063 %

Tabla 13 volúmenes de ajuste del set de validación

Si bien el modelo de Red Eléctrica pone en evidencia su buen funcionamiento respecto a la familia Holt Winters en el set de entrenamiento, requiriendo estos un 36-52% más de volumen de ajuste que éste, en el set de validación, correspondiente a 2015, estos ofrecen mejores resultados requiriendo prácticamente la mitad de volumen de ajuste.

Esto es signo de que si bien en escenarios de tendencia cambiante, como en el periodo 2008-2014, debidos a la crisis económica y los sucesivos repuntes, los modelos Holt Winters no son capaces de captar completamente la dinámica inherente a los ciclos económicos.

Por el contrario, en años como 2015 donde la tendencia es plana o presenta incluso una leve contracción, estos pueden ser una buena solución.

8. Conclusiones

En este trabajo se han construido modelos de previsión basados en tres técnicas distintas (modelos lineales, de series temporales y de alisado exponencial).

Se ha observado la dependencia de la demanda eléctrica tanto de las pautas de consumo doméstico e industrial, reflejado en las estacionalidades diaria, semanal y anual como de la temperatura, cuyo efecto más notable es el desplazamiento de la punta de la demanda en los días fríos del mediodía a la tarde. Esta dependencia de la temperatura se ha hecho patente como la limitación más importante en lo que respecta a la precisión de los modelos.

Si bien la familia de modelos lineales era capaz de modelar adecuadamente la forma de la curva, ésta se encontraba desplazada respecto del verdadero valor, debido a la exclusión del consumo termosensible así como a la dificultad de extraer una pauta del comportamiento de la tendencia anual.

Los modelos de series temporales aumentan la precisión al incluir el pasado reciente de la señal de la demanda en las previsiones, mejorando el modelado de los retardos respecto a los modelos lineales, sin embargo su aptitud para prever el aspecto de la curva diaria es inferior al utilizar las previsiones de los horizontes cortos para construir las de los largos.

Por último, los modelos Holt Winters de doble y triple estacionalidad son los que proporcionan los mejores resultados al encontrar el compromiso entre la precisión y la capacidad de proporcionar una forma de la curva adecuada para los horizontes largos de previsión. Además, estos modelos son capaces de recoger en el alisado exponencial la inercia inherente a los procesos físicos que gobiernan la demanda eléctrica, entre los que se encuentra el consumo eléctrico termosensible.

De entre todos los modelos candidatos, la familia Holt Winters es la que proporciona los mejores resultados. Esta elección se alinea con los resultados de Taylor [8] y [9].

Entre los tres modelos de alisado exponencial, el de doble estacionalidad resulta el más adecuado de implementar debido a que la complejidad añadida de los de triple estacionalidad o triple estacionalidad con corrección de días festivos no se ve justificada con un aumento significativo de la precisión de los modelos. Sin embargo, conviene recordar que la implementación de los de triple estacionalidad en R es aproximada, y que de existir implementaciones del algoritmo en el futuro formales, convendría explorarlas ya que podrían arrojar mejores resultados.

El interés de este trabajo ha sido el de demostrar que, únicamente con datos públicos y herramientas de código abierto es posible obtener previsiones lo suficientemente precisas como para ser utilizadas en entornos de producción y

valorar la importancia de la integración en los modelos de datos externos, concretamente meteorológicos, para aumentar la precisión de los modelos.

Como comentario final, de disponer de estos datos así como de herramientas más sofisticadas de modelado y calibración de los parámetros (solvers y modeladores como AIMMS, AMPL, GUROBI...) sería interesante la creación de un modelo híbrido con un primer bloque de alisado exponencial con triple estacionalidad, asumiendo que la forma de la curva de la demanda es homogénea de lunes a domingo y que integra temperatura, nebulosidad, etc. y un segundo bloque correctivo para adaptarse a tipologías singulares como las de sábados, domingos, festivos o vísperas, entre otros. Mediante una combinación lineal de ambos bloques, para cada paso de tiempo (hora) previsto, podría regularse el peso de cada bloque de modo que se pudiese captar también los cambios de régimen que se hicieron patentes en la estacionalidad semanal los lunes y viernes así como las modificaciones en puentes, vísperas o posteriores a festivo.

9. Glosario

- Correlación: medida de la relación entre dos variables estadísticas, tanto de su dirección (directa o inversa, dependiendo del signo positivo o negativo), como de la fuerza de la relación (0 indicando una relación nula y ± 1 como la máxima relación).
- Autocorrelación: correlación de una señal con ella misma desplazada un número t de instantes de tiempo.
- Autoregresión: relación de una señal en un determinado instante de tiempo con sus valores precedentes.
- Serie estacionaria: una serie temporal se dice estacionaria cuando su media, varianza y estructura de autocorrelación no cambian a lo largo del tiempo.
- Estacionalidad: variación periódica y predecible de una señal.
- Correlograma: representación gráfica de la correlación de una señal y sus retardos a instantes fijos en el tiempo.
- Ruptura: cambio repentino de régimen de una señal.
- Criterio de Información de Akaike (AIC): medida del ajuste de un modelo estadístico a un conjunto de datos, consistente en medir la verosimilitud de que el modelo obtenga la misma variable dependiente dadas sus independientes, penalizando la inclusión de cada variable explicativa considerada. De este modo se favorecerá la selección de modelos parsimoniosos en detrimento del sobreentrenamiento.
- Parsimonia: de un modelo estadístico, cualidad de considerar el menor número posible de variables explicativas de modo que la varianza que éstas consigan explicar sea significativa.
- Sobreentrenamiento: de un modelo estadístico, defecto de calibración originado por la elevada complejidad de un modelo al incluir un gran número de variables explicativas, relativas al número de muestras de datos disponibles y que provoca un mayor error de previsión en producción al ser el modelo muy sensible a fluctuaciones menores dentro de las muestras de datos.

- REE: Red Eléctrica de España
- TSO: Transmission System Operator.
- HW: Holt Winters
- LM: Linear Model
- ARIMA: Autoregressive Integrated Moving Average

10. Bibliografía

- [1] ENTSO-E. *Entso-e Statistical Database*.
<https://www.entsoe.eu/data/data-portal/Pages/default.aspx> (acc. 13/02/2016)
- [2] Red Eléctrica de España. *Demanda eléctrica en tiempo real*.
<https://demanda.ree.es/> (acc. 01/03/2016).
- [3] Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. «STL: A seasonal-trend decomposition procedure based on loess.» *Journal of Official Statistics* 6, no. 1 (1990): 3-73.
- [4] Hastie, Trevor, Tibshirani, Robert, Friedman, Jerome. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Berlin: Springer, 2009.
- [5] James, G., Witten, D., Hastie, T., Tibshirani, R. *An Introduction to Statistical Learning with Applications in R*. New York: Springer, 2013.
- [6] Box, G, Jenkins GM. *Time Series Analysis: forecasting and control*. San Francisco: Holden-Day, 1970.
- [7] Pr, Winters. «Forecasting sales by exponentially weighted moving averages.» *Management Science*, 1960: 324-342 .
- [8] Taylor, J. W. « Short-Term Electricity Demand Forecasting Using Double Seasonal Exponential Smoothing .» *The Journal of the Operational Research Society*, 2003: 799-805 .
- [9] Taylor, James W. «Triple Seasonal Methods for Short- Term Electricity Demand Forecasting .» *European Journal of Operational Research*, 2010: 139-152.
- [10] Phillip G. Gould, Anne B. Koehler, Keith Ord, Ralph D. Snyder, Rob J. Hyndman and Farshid Vahid-Araghi. «Forecasting time series with multiple seasonal patterns.» *European Journal of Operational Research*, 2008: 191(1), 207–220.
- [11] Hyndman, R. J., & Koehler, A. B. «Another look at measures of forecast accuracy. .» *International journal of forecasting* 22, no. 4 (2006): 679-688.
- [12] Blaconá, M.T, Andreozzi, L. and Magnano, L. «Time series models for different seasonal patterns.» *The 34th International Symposium on Forecasting*. Rotterdam: International Institute of Forecasters, 2014. -.

- [13] Diebold, Francis X. «Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests.» *Journal of Business and Economic Statistics*, 2014: -.
- [14] Hyndman, R.J., Koehler, A.B., Ord, J.K., Snyder, R.D. *Forecasting with Exponential Smoothing The State Space Approach*. New York: Springer, 2008.
- [15] Hyndman, Stephan Kolassa and Rob J. «Free open-source forecasting using R.» *Foresight: the International Journal of Applied Forecasting* , 2010: 17, 19-24.
- [16] REE. *Informe del Sistema Eléctrico Español 2014*. Red Eléctrica de España SA, 2015.

11. Anexos

Anexo 1- Tabla de pesos eléctricos de cada Comunidad Autónoma:

CCAA/MES	Enero	Febrero	Marzo	Abril	Mayo	Junio
Andalucía	0.154	0.151	0.151	0.151	0.151	0.157
Aragón	0.039	0.040	0.041	0.042	0.042	0.041
Asturias	0.042	0.043	0.044	0.045	0.044	0.041
C. Valenciana	0.106	0.105	0.104	0.103	0.104	0.109
Cantabria	0.016	0.017	0.018	0.019	0.018	0.017
Castilla la Mancha	0.047	0.046	0.046	0.045	0.046	0.047
Castilla y León	0.054	0.054	0.054	0.053	0.053	0.052
Cataluña	0.186	0.186	0.188	0.190	0.191	0.191
Extremadura	0.019	0.018	0.018	0.017	0.017	0.018
Galicia	0.080	0.079	0.081	0.082	0.080	0.075
La Rioja	0.007	0.007	0.007	0.007	0.007	0.007
Madrid	0.130	0.128	0.124	0.120	0.118	0.120
Murcia	0.032	0.032	0.032	0.033	0.034	0.035
Navarra	0.019	0.019	0.020	0.020	0.020	0.020
País Vasco	0.069	0.072	0.074	0.073	0.074	0.071

CCAA/MES	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
Andalucía	0.164	0.174	0.159	0.151	0.150	0.157
Aragón	0.039	0.040	0.041	0.041	0.041	0.039
Asturias	0.038	0.038	0.040	0.043	0.043	0.042
C. Valenciana	0.114	0.111	0.110	0.105	0.104	0.106
Cantabria	0.016	0.017	0.017	0.018	0.018	0.017
Castilla la Mancha	0.048	0.049	0.045	0.042	0.044	0.048
Castilla y León	0.052	0.051	0.051	0.053	0.054	0.054
Cataluña	0.194	0.188	0.193	0.193	0.189	0.184
Extremadura	0.018	0.021	0.019	0.018	0.018	0.019
Galicia	0.071	0.075	0.077	0.081	0.081	0.081
La Rioja	0.007	0.006	0.007	0.007	0.007	0.007
Madrid	0.121	0.119	0.117	0.120	0.125	0.132
Murcia	0.036	0.037	0.035	0.034	0.033	0.033
Navarra	0.018	0.018	0.019	0.020	0.020	0.018
País Vasco	0.062	0.055	0.069	0.073	0.072	0.063

Tabla 14 - pesos eléctricos mensuales por CCAA

Anexo 2 – Tablas de los parámetros de los diferentes modelos lineales

Modelo lineal Im.BBB

Coefficiente	Estimate	Std. Error	t value	Pr(> t)
referencia	5.021436e+03	1.090170e+02	46.061054	0.000000e+00
hora01	-7.033980e+02	5.069543e+01	-13.874977	1.046117e-43
hora02	-1.105058e+03	5.197457e+01	-21.261519	5.934037e-100
hora03	-1.222066e+03	5.319698e+01	-22.972470	2.730722e-116
hora04	-1.245707e+03	5.410534e+01	-23.023743	8.464417e-117
hora05	-1.124546e+03	5.428947e+01	-20.713883	5.502213e-95
hora06	-5.279064e+02	5.364367e+01	-9.840983	7.795668e-23
hora07	4.904255e+02	5.295932e+01	9.260419	2.099733e-20
hora08	1.324210e+03	5.249069e+01	25.227510	1.037103e-139
hora09	1.843492e+03	5.219113e+01	35.321948	1.435786e-270
hora10	2.271027e+03	5.247630e+01	43.277195	0.000000e+00
hora11	2.353919e+03	5.265546e+01	44.704170	0.000000e+00
hora12	2.333218e+03	5.276848e+01	44.216137	0.000000e+00
hora13	2.299033e+03	5.282595e+01	43.520904	0.000000e+00
hora14	1.836762e+03	5.203677e+01	35.297384	3.361731e-270
hora15	1.481100e+03	5.146270e+01	28.780071	6.142969e-181
hora16	1.433348e+03	5.124077e+01	27.972797	4.169942e-171
hora17	1.551613e+03	5.130398e+01	30.243514	1.877110e-199
hora18	1.834902e+03	5.172207e+01	35.476187	6.784888e-273
hora19	2.141367e+03	5.217364e+01	41.043086	0.000000e+00
hora20	2.407795e+03	5.269194e+01	45.695688	0.000000e+00
hora21	2.496374e+03	5.314630e+01	46.971744	0.000000e+00
hora22	1.857823e+03	5.201478e+01	35.717208	1.506949e-276
hora23	8.564216e+02	5.064821e+01	16.909216	5.377950e-64
diaSemanajueves	2.123667e+03	3.237431e+01	65.597291	0.000000e+00
diaSemanalunes	2.232239e+03	3.162745e+01	70.579161	0.000000e+00
diaSemanamartes	2.293127e+03	3.189524e+01	71.895597	0.000000e+00
diaSemanamiercoles	2.205704e+03	3.244420e+01	67.984526	0.000000e+00
diaSemanasabado	7.022151e+02	2.848210e+01	24.654608	1.475195e-133
diaSemnaviernes	1.989232e+03	3.186414e+01	62.428536	0.000000e+00
mes02	-1.141774e+03	3.663883e+01	-31.162965	1.535622e-211
mes03	-1.718410e+03	3.568929e+01	-48.149186	0.000000e+00
mes04	-2.466605e+03	3.801934e+01	-64.877627	0.000000e+00
mes05	-2.212680e+03	3.796882e+01	-58.276245	0.000000e+00
mes06	-1.744088e+03	3.724050e+01	-46.833083	0.000000e+00
mes07	-1.215140e+03	3.546865e+01	-34.259557	8.151845e-255
mes08	-1.713560e+03	3.625113e+01	-47.269138	0.000000e+00
mes09	-2.153118e+03	3.686911e+01	-58.398958	0.000000e+00
mes10	-2.111062e+03	3.763169e+01	-56.097976	0.000000e+00
mes11	-1.171415e+03	3.642428e+01	-32.160291	4.785290e-225
mes12	-1.365100e+03	3.557161e+01	-38.376120	2.023643e-318
tipoDianormal	3.654014e+03	4.432468e+01	82.437478	0.000000e+00
tipoDiapostfestivo	2.190722e+03	6.074549e+01	36.063938	7.598693e-282
tipoDiavispera	2.484394e+03	5.954890e+01	41.720244	0.000000e+00
retardoSemanal	5.852801e-01	2.960381e-03	197.704286	0.000000e+00
retardoAnual	8.204294e-02	2.398826e-03	34.201284	5.781193e-254

Modelo lineal Im.BBF

Coefficiente	Estimate	Std. Error	t value	Pr(> t)
Referencia	4.331561e+03	1.054180e+02	41.089386	0.000000e+00
hora01	-7.648754e+02	5.026959e+01	-15.215469	3.479125e-52
hora02	-1.219617e+03	5.166966e+01	-23.604124	1.236516e-122
hora03	-1.371582e+03	5.300190e+01	-25.877968	7.289290e-147
hora04	-1.415950e+03	5.398060e+01	-26.230720	8.112219e-151
hora05	-1.297572e+03	5.416818e+01	-23.954517	3.165048e-126
hora06	-6.789608e+02	5.342042e+01	-12.709762	5.812068e-37
hora07	3.827980e+02	5.258129e+01	7.280117	3.375177e-13

hora08	1.270289e+03	5.201354e+01	24.422281	4.214608e-131
hora09	1.845719e+03	5.170510e+01	35.697030	3.055611e-276
hora10	2.324050e+03	5.205485e+01	44.646181	0.000000e+00
hora11	2.437758e+03	5.230238e+01	46.608935	0.000000e+00
hora12	2.435435e+03	5.246411e+01	46.420970	0.000000e+00
hora13	2.409555e+03	5.254498e+01	45.857001	0.000000e+00
hora14	1.934489e+03	5.171060e+01	37.409902	7.143904e-303
hora15	1.566065e+03	5.109739e+01	30.648635	9.795222e-205
hora16	1.505314e+03	5.084672e+01	29.604948	2.897080e-191
hora17	1.616362e+03	5.089779e+01	31.757022	1.551803e-219
hora18	1.906436e+03	5.133350e+01	37.138243	1.435954e-298
hora19	2.224434e+03	5.181479e+01	42.930489	0.000000e+00
hora20	2.504544e+03	5.237220e+01	47.822008	0.000000e+00
hora21	2.605868e+03	5.286604e+01	49.291905	0.000000e+00
hora22	1.949080e+03	5.167303e+01	37.719490	8.193333e-308
hora23	9.097589e+02	5.021323e+01	18.117910	3.576355e-73
diaSemanajueves	2.260531e+03	3.256026e+01	69.426070	0.000000e+00
diaSemanalunes	2.272550e+03	3.145546e+01	72.246593	0.000000e+00
diaSemanamartes	2.401327e+03	3.198291e+01	75.081590	0.000000e+00
diaSemanamiercoles	2.340425e+03	3.265502e+01	71.671233	0.000000e+00
diaSemanasabado	7.744564e+02	2.834347e+01	27.323977	2.113437e-163
diaSemnaviernes	2.115995e+03	3.199916e+01	66.126583	0.000000e+00
`S1-8760`	-5.316602e+01	1.023466e+01	-5.194702	2.057030e-07
`C1-8760`	5.127908e+02	1.113200e+01	46.064560	0.000000e+00
`S2-8760`	2.871674e+02	1.135155e+01	25.297648	1.791098e-140
`C2-8760`	6.505013e+02	1.088622e+01	59.754558	0.000000e+00
`S3-8760`	-1.843142e+01	1.026072e+01	-1.796309	7.245037e-02
`C3-8760`	-6.628632e+01	1.027663e+01	-6.450201	1.125308e-10
`S4-8760`	2.615341e+01	1.020337e+01	2.563214	1.037317e-02
`C4-8760`	-1.277540e+02	1.020302e+01	-12.521197	6.326567e-36
`S5-8760`	1.078514e+02	1.028678e+01	10.484469	1.070503e-25
`C5-8760`	-2.056337e+02	1.052834e+01	-19.531446	1.075026e-84
`S6-8760`	1.928224e+02	1.018773e+01	18.926914	1.156534e-79
`C6-8760`	4.629224e+01	1.024042e+01	4.520540	6.179748e-06
`S7-8760`	9.327693e+01	1.018213e+01	9.160850	5.302568e-20
`C7-8760`	-2.362958e+02	1.018604e+01	-23.197996	1.550767e-118
`S8-8760`	2.444533e+02	1.018993e+01	23.989688	1.370858e-126
`C8-8760`	-1.551591e+02	1.025056e+01	-15.136644	1.151545e-51
tipoDianormal	3.522278e+03	4.412306e+01	79.828508	0.000000e+00
tipoDiapostfestivo	2.306994e+03	6.025446e+01	38.287524	5.594660e-317
tipoDiavispera	2.288644e+03	5.866990e+01	39.008828	0.000000e+00
retardoSemanal	5.689357e-01	3.029125e-03	187.821796	0.000000e+00
retardoAnual	6.847657e-02	2.418368e-03	28.315204	3.042638e-175

Modelo lineal Im.BFB

Coeficiente	Estimate	Std. Error	t value	Pr(> t)
Referencia	6.893058e+03	1.180052e+02	58.413164	0.000000e+00
hora01	-7.307601e+02	5.046638e+01	-14.480136	1.938101e-47
hora02	-1.148106e+03	5.178240e+01	-22.171736	1.723372e-108
hora03	-1.275333e+03	5.307215e+01	-24.030172	5.225233e-127
hora04	-1.301090e+03	5.406157e+01	-24.066822	2.179016e-127
hora05	-1.175249e+03	5.435469e+01	-21.621852	2.727496e-103
hora06	-5.590320e+02	5.384838e+01	-10.381594	3.155650e-25
hora07	4.921001e+02	5.334544e+01	9.224783	2.928485e-20
hora08	1.357330e+03	5.310123e+01	25.561177	2.337500e-143
hora09	1.902235e+03	5.303786e+01	35.865615	8.263976e-279
hora10	2.352174e+03	5.351865e+01	43.950538	0.000000e+00
hora11	2.447969e+03	5.381534e+01	45.488318	0.000000e+00
hora12	2.435720e+03	5.396753e+01	45.133062	0.000000e+00
hora13	2.407653e+03	5.399605e+01	44.589423	0.000000e+00
hora14	1.938645e+03	5.311458e+01	36.499295	1.447140e-288
hora15	1.575695e+03	5.240695e+01	30.066526	3.632514e-197
hora16	1.525905e+03	5.204730e+01	29.317670	1.231929e-187
hora17	1.644436e+03	5.196640e+01	31.644219	5.253421e-218
hora18	1.930679e+03	5.222190e+01	36.970678	6.262861e-296

hora19	2.237571e+03	5.249548e+01	42.624071	0.000000e+00
hora20	2.503012e+03	5.284386e+01	47.366185	0.000000e+00
hora21	2.588274e+03	5.315177e+01	48.695923	0.000000e+00
hora22	1.925750e+03	5.188493e+01	37.115787	3.249194e-298
hora23	8.886942e+02	5.042567e+01	17.623846	2.398753e-69
`S1-168`	9.244342e+02	1.285561e+01	71.909006	0.000000e+00
`C1-168`	3.795611e+02	1.137403e+01	33.370865	5.430267e-242
`S2-168`	-2.886674e+02	1.063975e+01	-27.131021	3.823776e-161
`C2-168`	5.388491e+02	1.087460e+01	49.551167	0.000000e+00
`S3-168`	-2.338242e+02	1.032406e+01	-22.648478	4.217412e-113
`C3-168`	-7.127126e+01	1.020799e+01	-6.981910	2.941455e-12
`S4-168`	-1.428174e+02	1.025338e+01	-13.928815	4.942077e-44
`C4-168`	-1.311702e+02	1.023070e+01	-12.821233	1.393953e-37
`S5-168`	1.080858e+02	1.026852e+01	10.525936	6.903312e-26
`C5-168`	-2.183153e+02	1.034743e+01	-21.098504	1.841970e-98
`S6-168`	1.647556e+02	1.025996e+01	16.058114	6.587622e-58
`C6-168`	9.942055e+01	1.036335e+01	9.593476	8.822336e-22
`S7-168`	-3.970369e+01	1.176903e+01	-3.373573	7.424542e-04
`C7-168`	5.571424e+01	1.190574e+01	4.679612	2.880317e-06
`S8-168`	1.337478e+01	1.020880e+01	1.310123	1.901592e-01
`C8-168`	-9.637226e+01	1.021212e+01	-9.437044	3.963300e-21
mes02	-1.119289e+03	3.646899e+01	-30.691519	2.678994e-205
mes03	-1.732956e+03	3.553577e+01	-48.766521	0.000000e+00
mes04	-2.494234e+03	3.793001e+01	-65.758850	0.000000e+00
mes05	-2.248621e+03	3.788999e+01	-59.346050	0.000000e+00
mes06	-1.773258e+03	3.713416e+01	-47.752752	0.000000e+00
mes07	-1.219364e+03	3.530382e+01	-34.539149	6.461980e-259
mes08	-1.735404e+03	3.612377e+01	-48.040500	0.000000e+00
mes09	-2.164534e+03	3.674585e+01	-58.905541	0.000000e+00
mes10	-2.140154e+03	3.753478e+01	-57.017887	0.000000e+00
mes11	-1.200772e+03	3.628901e+01	-33.089131	5.399277e-238
mes12	-1.355328e+03	3.540407e+01	-38.281717	6.957091e-317
tipoDianormal	3.689395e+03	4.415855e+01	83.548834	0.000000e+00
tipoDiapostfestivo	2.227988e+03	6.043622e+01	36.865107	2.843160e-294
tipoDiavispera	2.519208e+03	5.928050e+01	42.496404	0.000000e+00
retardoSemanal	5.739174e-01	2.980733e-03	192.542368	0.000000e+00
retardoAnual	8.347497e-02	2.443368e-03	34.163903	2.029590e-253

Modelo lineal lm.BFF

Coficiente	Estimate	Std. Error	t value	Pr(> t)
Referencia	6.321799e+03	1.151761e+02	54.8881291	0.000000e+00
hora01	-7.940208e+02	5.002693e+01	-15.8718687	1.287448e-56
hora02	-1.265680e+03	5.146415e+01	-24.5934337	6.574540e-133
hora03	-1.429045e+03	5.286466e+01	-27.0321390	5.412058e-160
hora04	-1.475688e+03	5.392209e+01	-27.3670333	6.596416e-164
hora05	-1.352024e+03	5.421403e+01	-24.9386463	1.366693e-136
hora06	-7.124430e+02	5.360107e+01	-13.2915814	2.945766e-40
hora07	3.840871e+02	5.294081e+01	7.2550290	4.063112e-13
hora08	1.305445e+03	5.259731e+01	24.8196164	2.576095e-135
hora09	1.909044e+03	5.252769e+01	36.3435824	3.744429e-286
hora10	2.411758e+03	5.307654e+01	45.4392508	0.000000e+00
hora11	2.539090e+03	5.344272e+01	47.5104824	0.000000e+00
hora12	2.545028e+03	5.364275e+01	47.4440319	0.000000e+00
hora13	2.524309e+03	5.369217e+01	47.0144767	0.000000e+00
hora14	2.040589e+03	5.276286e+01	38.6747286	1.976263e-323
hora15	1.663003e+03	5.201462e+01	31.9718352	1.836580e-222
hora16	1.598169e+03	5.162473e+01	30.9574362	8.297971e-209
hora17	1.707923e+03	5.153099e+01	33.1436101	9.163190e-239
hora18	2.000513e+03	5.180366e+01	38.6172228	2.371515e-322
hora19	2.319271e+03	5.210759e+01	44.5092844	0.000000e+00
hora20	2.599291e+03	5.249696e+01	49.5131683	0.000000e+00
hora21	2.698579e+03	5.284788e+01	51.0631478	0.000000e+00
hora22	2.018241e+03	5.152489e+01	39.1702085	0.000000e+00
hora23	9.431453e+02	4.997611e+01	18.8719221	3.259852e-79
`S1-168`	9.890808e+02	1.299061e+01	76.1381518	0.000000e+00
`C1-168`	3.598400e+02	1.127431e+01	31.9167989	1.036976e-221
`S2-168`	-2.802462e+02	1.053692e+01	-26.5966009	5.668520e-155
`C2-168`	5.571625e+02	1.080240e+01	51.5776697	0.000000e+00
`S3-168`	-2.371711e+02	1.022516e+01	-23.1948593	1.667298e-118
`C3-168`	-6.994383e+01	1.010696e+01	-6.9203614	4.549024e-12
`S4-168`	-1.436382e+02	1.015219e+01	-14.1484936	2.249777e-45
`C4-168`	-1.334534e+02	1.013078e+01	-13.1730657	1.419613e-39
`S5-168`	1.187688e+02	1.017549e+01	11.6720483	1.912303e-31
`C5-168`	-2.185465e+02	1.024596e+01	-21.3300084	1.390840e-100
`S6-168`	1.725393e+02	1.016475e+01	16.9742841	1.788469e-64
`C6-168`	1.160019e+02	1.027852e+01	11.2858592	1.648310e-29
`S7-168`	-4.281694e+01	1.165306e+01	-3.6743103	2.386959e-04
`C7-168`	4.403190e+01	1.179412e+01	3.7333777	1.890997e-04
`S8-168`	9.475136e+00	1.010809e+01	0.9373814	3.485661e-01
`C8-168`	-1.011787e+02	1.011220e+01	-10.0056003	1.501421e-23
`S1-8760`	-4.972259e+01	1.018218e+01	-4.8832950	1.045899e-06
`C1-8760`	5.265244e+02	1.111212e+01	47.3828998	0.000000e+00
`S2-8760`	3.040376e+02	1.133076e+01	26.8329456	1.094370e-157
`C2-8760`	6.607187e+02	1.086086e+01	60.8348511	0.000000e+00
`S3-8760`	-2.002360e+01	1.020711e+01	-1.9617297	4.979848e-02
`C3-8760`	-6.951668e+01	1.022474e+01	-6.7988707	1.064050e-11
`S4-8760`	2.308844e+01	1.015046e+01	2.2746193	2.293222e-02
`C4-8760`	-1.288479e+02	1.014979e+01	-12.6946330	7.048403e-37
`S5-8760`	1.021378e+02	1.023529e+01	9.9789832	1.963158e-23
`C5-8760`	-2.095632e+02	1.047823e+01	-19.9998600	1.061324e-88
`S6-8760`	1.932320e+02	1.013558e+01	19.0647205	8.508702e-81
`C6-8760`	4.343996e+01	1.018757e+01	4.2640144	2.010868e-05
`S7-8760`	8.966123e+01	1.012956e+01	8.8514440	8.863966e-19
`C7-8760`	-2.369558e+02	1.013413e+01	-23.3819641	2.201851e-120
`S8-8760`	2.418544e+02	1.013686e+01	23.8589128	3.059392e-125
`C8-8760`	-1.591933e+02	1.020029e+01	-15.6067461	8.354185e-55
tipoDianormal	3.560221e+03	4.393247e+01	81.0384796	0.000000e+00
tipoDiapostfestivo	2.346918e+03	5.993113e+01	39.1602574	0.000000e+00
tipoDiavispera	2.329806e+03	5.838474e+01	39.9043618	0.000000e+00
retardoSemanal	5.569601e-01	3.048709e-03	182.6871881	0.000000e+00
retardoAnual	6.887245e-02	2.465787e-03	27.9312190	1.316457e-170

Anexo 3 – Tablas de los parámetros de los modelos ARIMA

Modelo arima.BB:

Términos temporales y variables exógenas

ar1	1.713716821	dia normal	-0.048673961
ar2	-0.980260041	dia festivo	-0.029697268
ma1	-0.929359317	víspera	-0.047235377
ma2	-0.132330067	cambio hora	0.989435880
ma3	0.605341223		

Estacionalidad semanal

diaSemanajueves	0.007727446
diaSemanalunes	-0.208765003
diaSemanamartes	-0.181736411
diaSemanamiercoles	-0.089654140
diaSemanasabado	0.156699213
diaSemanaviernes	0.049983457

Estacionalidad anual

Semana 1	-0.079774142
Semana 2	-0.336923179
Semana 3	-0.673498306
Semana 4	-1.202835660
Semana 5	-1.532203635
Semana 6	-2.110697969
Semana 7	-2.624392319
Semana 8	-3.137651331
Semana 9	-3.392222035
Semana 10	-3.548352124

Semana 11	-3.744720831
Semana 12	-3.776597173
Semana 13	-4.519973173
Semana 14	-5.294499730
Semana 15	-5.965643698
Semana 16	-6.383904447
Semana 17	-6.622610920
Semana 18	-6.796814325
Semana 19	-6.657688504
Semana 20	-6.386083557

Semana 21	-5.806582772
Semana 22	-5.247626345
Semana 23	-4.634193914
Semana 24	-3.856847750
Semana 25	-3.197476446
Semana 26	-2.434179761
Semana 27	-1.865941271
Semana 28	-0.995103869
Semana 29	-0.220805714
Semana 30	0.411915933

Semana 31	1.068676077
Semana 32	1.505986153
Semana 33	1.581101657
Semana 34	1.532226716
Semana 35	1.242785907
Semana 36	0.712139924
Semana 37	0.270769566
Semana 38	0.107764524
Semana 39	0.153026631
Semana 40	0.234168242

Semana 41	0.321280572
Semana 42	0.376118537
Semana 43	0.489849372
Semana 44	0.356431543
Semana 45	0.140715023
Semana 46	0.028025143
Semana 47	0.157231374
Semana 48	0.254077056
Semana 49	0.264909810
Semana 50	0.253713427
Semana 51	0.260634820
Semana 52	0.241352130

Modelo arima.BF

Términos temporales y variables exógenas

ar1	1.713602747
ar2	-0.980076399
ma1	-0.933706424
ma2	-0.124492117
ma3	0.600345760

día Normal	-0.030980712
día festivo	-0.023079601
víspera	-0.035906195
cambio hora	0.983132649

Estacionalidad semanal

diaSemanajueves	0.002174617
diaSemanalunes	-0.207441072
diaSemanamartes	-0.179744711
diaSemanamiercoles	-0.093198999
diaSemanasabado	0.153399608
diaSemanaviernes	0.043278137

Estacionalidad anual

S1-8760	-1.667665137
C1-8760	-4.636806234
S2-8760	0.837257176
C2-8760	0.705224281
S3-8760	-0.015191798
C3-8760	-0.319379774
S4-8760	-0.292274383
C4-8760	-0.474706858

Modelo arima.FB

Términos temporales y variables exógenas

ar1	0.315897449
ar2	0.622808726
ar3	0.383398907
ar4	-0.728462935
ma1	0.520423489
ma2	-0.526872295
ma3	-0.944502245

Día Normal	0.467620482
día festivo	0.200740226
víspera	0.250425911
cambio hora	0.761088694

Estacionalidad semanal

S1-168	1.344065527
C1-168	2.115929262
S2-168	-1.249880027
C2-168	-0.465105660

S3-168	0.421669485
C3-168	-0.254993537
S4-168	0.274143374
C4-168	0.270634262

Estacionalidad anual

Semana 1	-0.158093692
Semana 2	-0.072523634
Semana 3	-0.193925014
Semana 4	-0.345449757
Semana 5	-0.210478585
Semana 6	-0.218858477
Semana 7	-0.328750784
Semana 8	-0.431664635
Semana 9	-0.341665144
Semana 10	-0.375678618

Semana 11	-0.628371235
Semana 12	-0.572744630
Semana 13	-0.942650235
Semana 14	-1.324031242
Semana 15	-1.455899527
Semana 16	-1.599233929
Semana 17	-1.674806242
Semana 18	-1.605172136
Semana 19	-1.316362822
Semana 20	-1.069187544

Semana 21	-0.931952755
Semana 22	-0.529467543
Semana 23	-0.467255882
Semana 24	-0.195273427
Semana 25	0.007904307
Semana 26	0.316422214
Semana 27	0.275319493
Semana 28	0.511972797
Semana 29	0.550123192
Semana 30	0.502454550

Semana 31	0.536769235
Semana 32	0.474255786
Semana 33	0.358876218
Semana 34	0.543751293
Semana 35	0.482992021
Semana 36	0.408934055
Semana 37	0.458154739
Semana 38	0.663187004
Semana 39	0.705310701
Semana 40	0.814771042

Semana 41	0.590564228
Semana 42	0.896457998
Semana 43	1.205513944
Semana 44	1.263863752
Semana 45	1.340343288
Semana 46	1.403601146
Semana 47	1.652368837
Semana 48	1.810193711
Semana 49	1.282514355
Semana 50	1.383664587
Semana 51	0.799186354
Semana 52	0.319712515

Modelo arima.FF

Términos temporales y variables exógenas

ar1	0.3155453	dia Normal	0.4874438
ar2	0.6232680	dia festivo	0.2015668
ar3	0.3836715	víspera	0.2647297
ar4	-0.7284998	cambio hora	0.807929
ma1	0.5198130		
ma2	-0.5271836		
ma3	-0.9443871		

Estacionalidad semanal

S1-168	1.3432168
C1-168	2.1170825
S2-168	-1.2507747
C2-168	-0.4657776
S3-168	0.4217748
C3-168	-0.2551866
S4-168	0.2740135
C4-168	0.2711933

Estacionalidad anual

S1-8760	0.2065635
C1-8760	1.5834451
S2-8760	1.2528207
C2-8760	1.2626873
S3-8760	-0.0790569
C3-8760	-0.4645351
S4-8760	-0.2893316
C4-8760	-0.3833215

Anexo 4 –Manual de usuario de la aplicación de previsión

Introducción:

La aplicación de previsión de la demanda eléctrica española permite el cálculo de la previsión de demanda utilizando diferentes modelos de previsión.

- Modelos lineales
- Modelos ARIMA de series temporales
- Modelos de alisado exponencial de Holt-Winters

Configuración y requisitos mínimos:

Para poder ejecutar la aplicación correctamente es necesario disponer en la máquina de:

- R versión 3.2.1 o posterior
- R-Studio versión 32/64 bits
- Los siguientes paquetes R:
 - magrittr
 - ggplot2
 - plyr
 - dplyr
 - tidyr
 - forecast
 - shiny
 - DT

Todos ellos están disponibles en CRAN

En caso de estar en una red con proxy, es necesario configurarlo en el fichero de configuración correspondiente, que se encuentra en la ruta /dist/APPREV/www con nombre proxyOpts.csv

```
proxy,163.104.40.34
username,
password,
port,3128
```

Figura anexo 1: configuración del proxy

Como precaución, no editar el fichero con Excel ya que cambiaría el separador de “,” a “;”. Es necesario asimismo un salto de línea una vez modificado el puerto del proxy.

Con esto, la configuración estaría terminada

Experiencia de usuario:



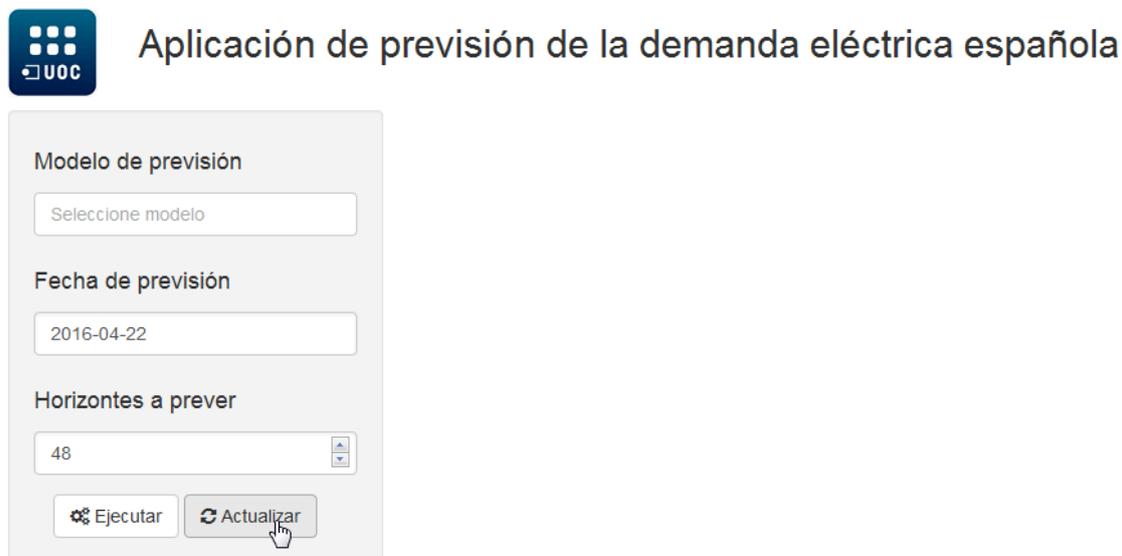
Aplicación de previsión de la demanda eléctrica española

The screenshot shows a web interface for an electricity demand forecasting application. It features three main input sections: 'Modelo de previsión' with a dropdown menu containing 'Seleccione modelo'; 'Fecha de previsión' with an empty text input field; and 'Horizontes a prever' with a dropdown menu showing '24'. At the bottom, there are two buttons: 'Ejecutar' (Execute) with a gear icon and 'Actualizar' (Update) with a refresh icon.

Figura anexo 2: pantalla principal aplicación

La aplicación tiene una única pantalla desde la que se pueden lanzar las previsiones así como actualizar los datos disponibles directamente desde la web de *Red Eléctrica de España*.

Actualización de datos:



Aplicación de previsión de la demanda eléctrica española

Modelo de previsión
Seleccione modelo

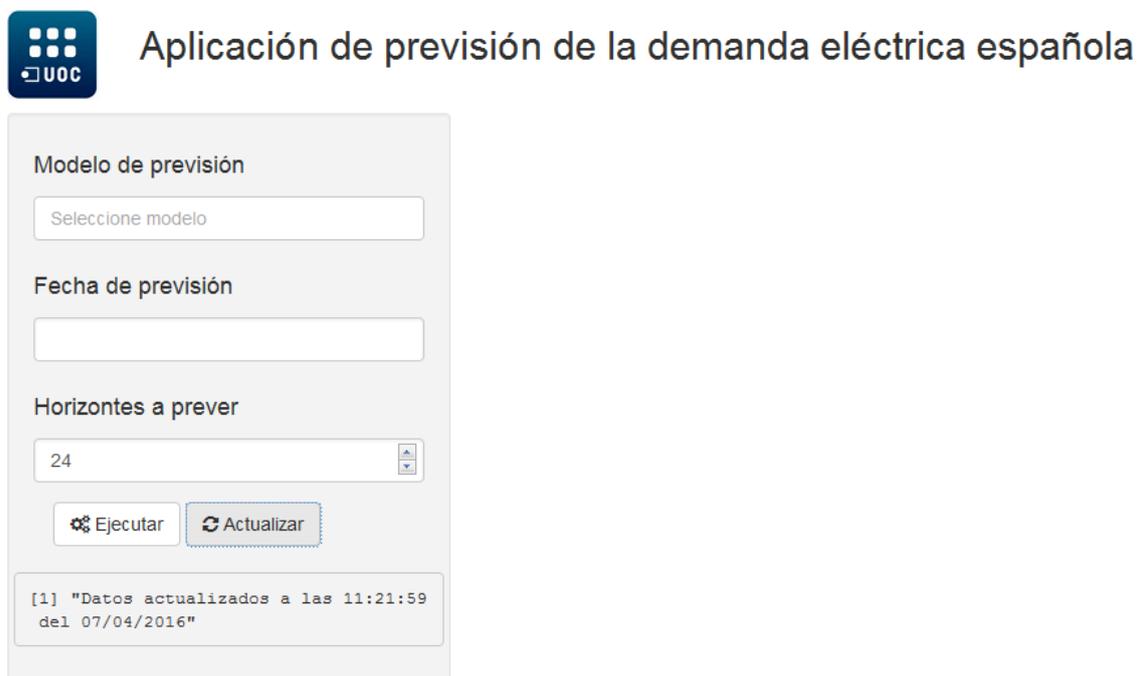
Fecha de previsión
2016-04-22

Horizontes a prever
48

Ejecutar Actualizar

Figura anexo 3: actualización de datos

Bastará con hacer click en el botón de actualizar para lanzar el proceso en segundo plano. Si bien podrían configurarse las previsiones mientras se obtienen los datos no es recomendable pues el fichero estructural se encuentra abierto y en proceso de modificación.



Aplicación de previsión de la demanda eléctrica española

Modelo de previsión
Seleccione modelo

Fecha de previsión

Horizontes a prever
24

Ejecutar Actualizar

```
[1] "Datos actualizados a las 11:21:59 del 07/04/2016"
```

Figura anexo 4: actualización de datos II

Una vez la actualización ha terminado, el siguiente mensaje se muestra, junto con la hora de finalización de la actualización.

Lanzamiento de previsiones:



Aplicación de previsión de la demanda eléctrica española

Modelo de previsión

ARIMA FF LINEAR BFB
LINEAR BBB
HOLT WINTERS DOBLE
ARIMA BB ARIMA FB

Fecha de previsión

2016-04-22

Horizontes a prever

48

Ejecutar Actualizar

Visualizar previsión

ARIMA BB
 ARIMA FF
 ARIMA FB
 LINEAR BBB
 LINEAR BFB
 HOLT WINTERS DOBLE

Figura anexo 5: lanzamiento de una previsión

Para lanzar el cálculo de previsiones el proceso es el siguiente:

- 1) Seleccionar de la lista desplegable “Modelo de previsión” aquellos modelos que se quiere utilizar
- 2) Seleccionar la fecha a partir de la cual prever (por defecto la medianoche). En caso de seleccionar una fecha superior a la de ejecución de la aplicación, al no existir datos de demanda disponibles se calculará la demanda a partir del último instante de tiempo disponible en los datos estructurales (por ejemplo si los datos actualizados son hasta el 10 de abril a las 11:00 y seleccionamos el 15 de abril, la primera hora prevista serán las 12 del 10 de abril).
- 3) Seleccionar el número de horizontes (horas) a partir de la fecha de inicio para prever.
- 4) Seleccionar al menos uno de los modelos de la checklist “Visualizar previsión” para que se grafique inmediatamente después del cálculo y así saber que ha concluido.
- 5) Pulsar sobre el botón ejecutar.

Visualización de resultados:



Aplicación de previsión de la demanda eléctrica española

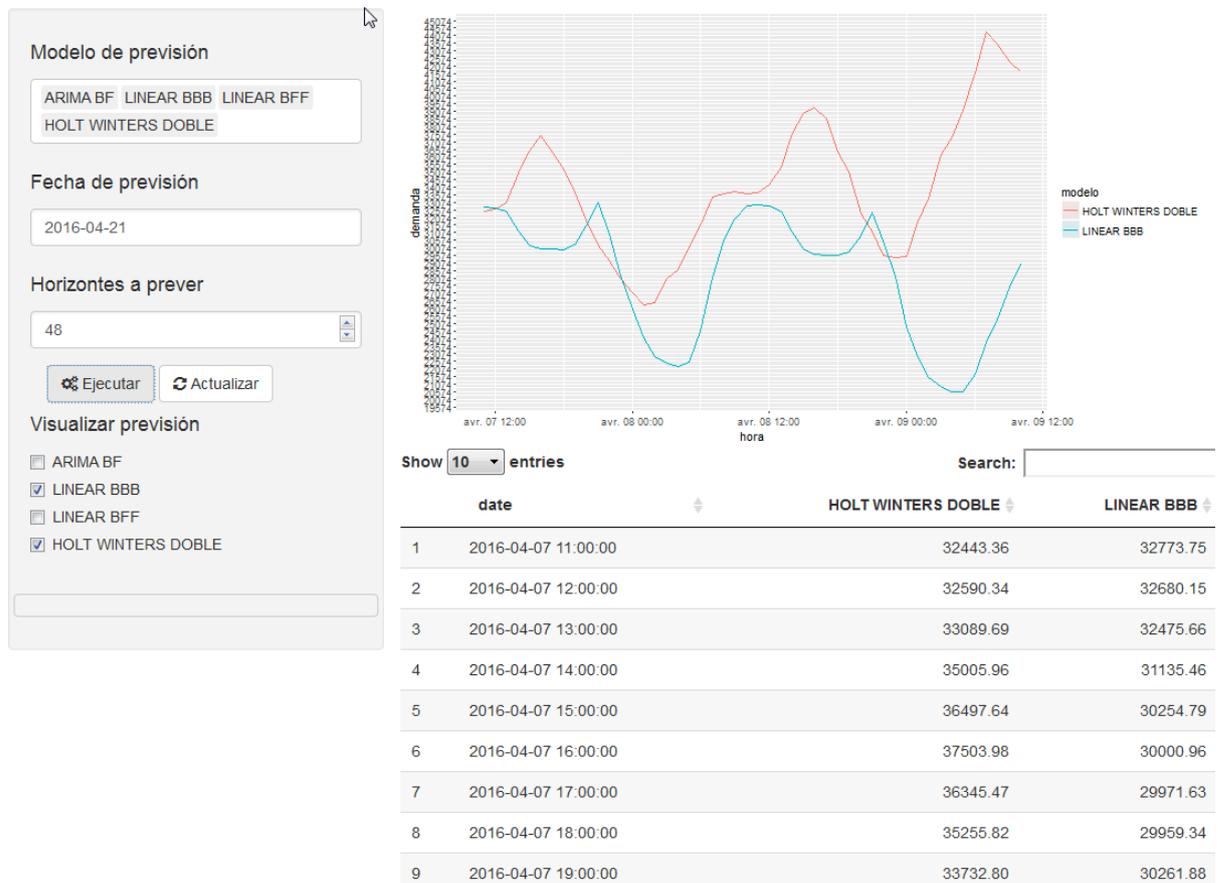


Figura anexo 6: resultado de la previsión

El cálculo de la previsión se obtiene como gráfico, para el análisis de la forma de la curva así como una tabla de valores. Desde la checklist “Visualizar previsión” se puede modificar la visualización escogiendo ocultar/mostrar los diferentes modelos.



Aplicación de previsión de la demanda eléctrica española

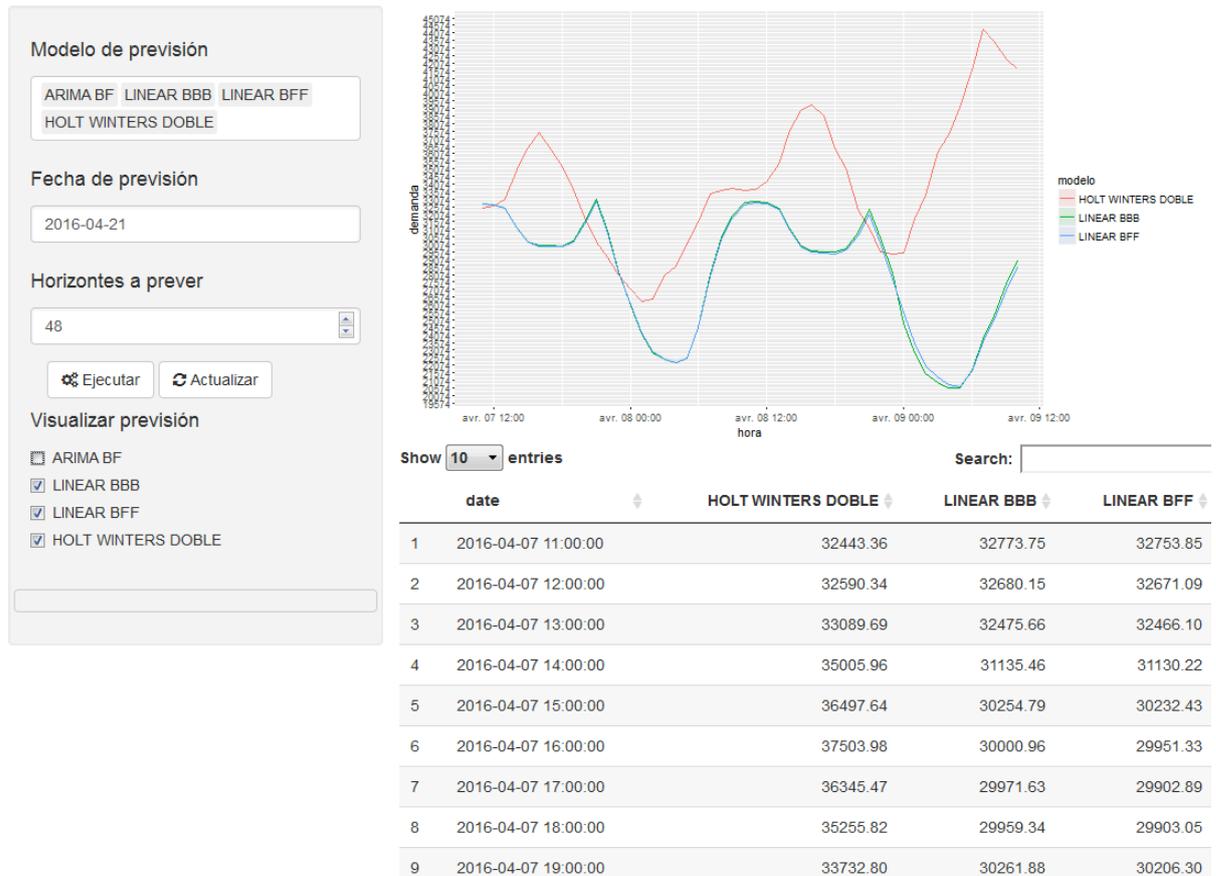


Figura anexo 7: opciones de visualización

En la imagen se observa la gráfica con los tres modelos, en la que la región sombreada en cada línea corresponde al intervalo de confianza de la previsión (excluido para Holt Winters). Las columnas de la tabla corresponden a la serie temporal de previsiones de cada modelo, por orden creciente de fecha.