



# **Modelo de Predicción de Demanda Convencional de Gas**

Autor: Alberto Ramirez Varas

Tutor: Victor M. Ruiz Morcillo

Fecha: 06/11/2015

---



# Índice

1	Introducción.....	1
1.1	Datos Demanda de Gas a nivel nacional .....	1
1.2	Necesidad del Estudio .....	4
1.3	Obtención de la Serie Temporal .....	5
1.4	Objetivos del Estudio .....	5
1.5	Limitaciones del Estudio .....	6
2	Estado del Arte .....	6
3	Técnicas de Predicción .....	7
3.1	Modelos ARIMA: Estructura General.....	7
3.2	Modelos ARIMA con variables de Intervención.....	16
3.3	Modelos ARMA con variables externas (ARMAX) .....	16
4	Estudio de la Base de Datos .....	17
4.1	Análisis de la Serie Temporal de Demanda de Gas.....	17
4.1.1	Análisis Gráfico.....	17
4.1.2	Análisis descriptivo y de comportamiento .....	25
4.1.3	Detección de Datos Atípicos .....	27
4.2	Análisis variables exógenas.....	28
4.2.1	Variables Dummy Laboralidad .....	29
4.2.2	Variables Dummy Estacionales.....	30
4.2.3	Variables Temperaturas.....	31
5	Metodología Box-Jenkis.....	34
5.1	Estacionariedad en Varianza .....	35
5.2	Estacionariedad en Media .....	36
5.3	Detección de Estacionalidad .....	38
5.4	Selección del mejor modelo ARIMA .....	39
5.5	Inclusión de variables Dummy de laboralidad o calendario.....	47
5.6	Inclusión de variables de Temperatura.....	49
5.7	Resultados de Predicción.....	51
5.7.1	Resultado solo con modelización ARIMA.....	52
	Gráfica 5.7.1.1. Representación errores.ARIMA .....	52
5.7.2	Resultados incluyendo variables Dummy .....	53
5.7.3	Resultados incluyendo variables Dummy y Temperaturas. ....	54



	5.7.4. Resultados de la predicción en periodos festivos .....	56
6	Conclusiones .....	59
7	Anexo .....	60
8	Bibliografía.....	63

# 1 Introducción

## 1.1 Datos Demanda de Gas a nivel nacional

La predicción de la demanda es un análisis que tiene una gran importancia en el sector gasista, ya que a partir de dichos resultados, los agentes del mercado deben tomar las mejores decisiones sobre el negocio.

El presente trabajo tiene como objetivo la creación de un modelo de predicción para la Demanda de Gas Convencional (GNC). Dicha demanda engloba el consumo en hogares, comercios, industrias y cogeneraciones, que fue en el año 2014 de 249.736Gwh, con un decrecimiento del 9,8% con respecto al mismo dato de 2013.

La Demanda Convencional de Gas compone un alto porcentaje de la Demanda total. La Figura 1.1. demuestra que a final del ejercicio de 2014, la Demanda de Gas Convencional supone un 83% sobre el total de Demanda.

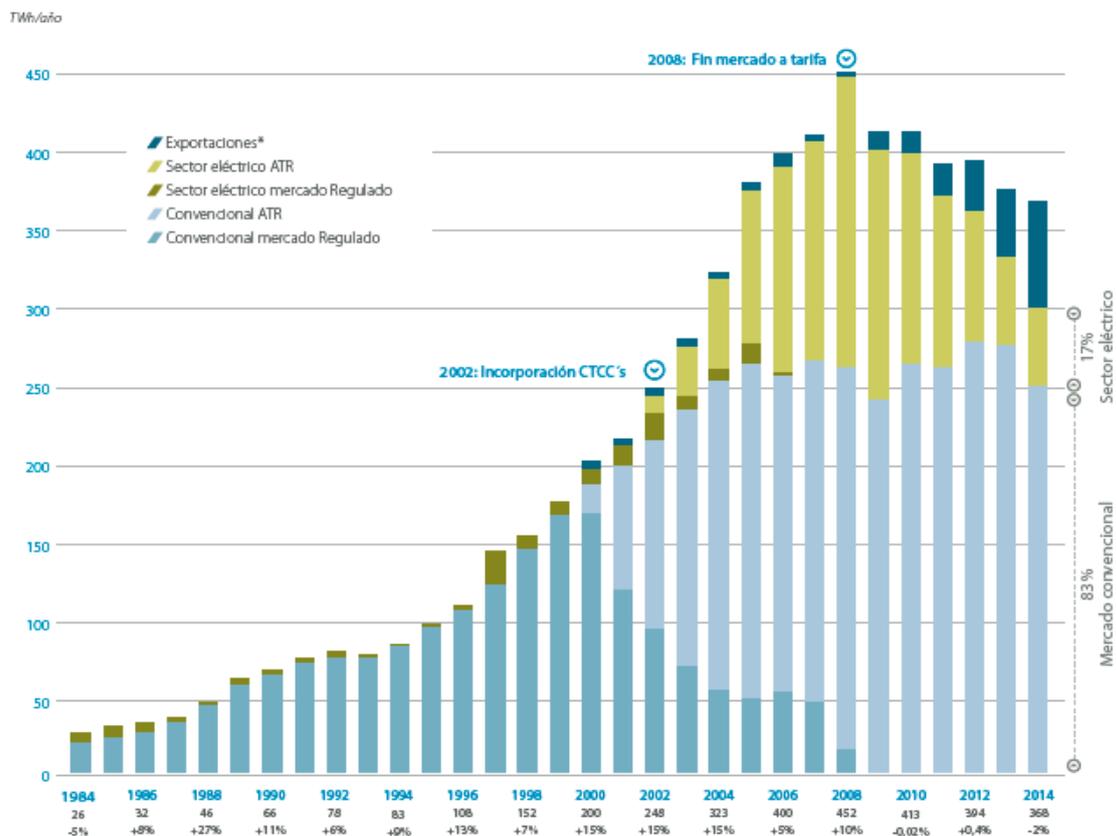


Figura 1.1.1 Porcentaje Demanda Convencional de Gas. Fuente: ENAGAS

Las Comunidades Autónomas con mayor Demanda de Gas fueron Cataluña, Andalucía y Valencia, siendo Extremadura, Baleares y La Rioja las que menos consumo tienen.

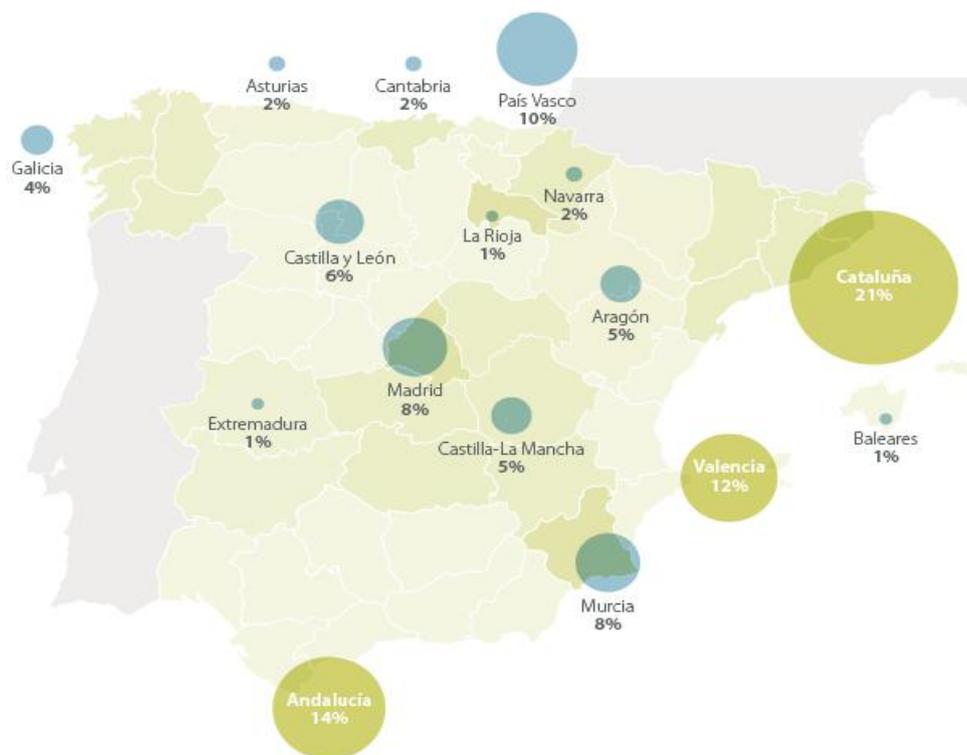


Figura 1.1.2 Distribución del Consumo de Gas a nivel Nacional. Fuente: ENAGAS

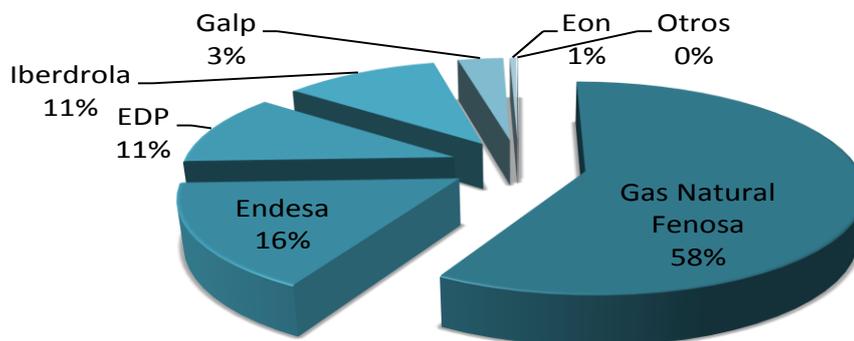
		Crecimiento					Crecimiento		
		2013	2014	vs. 2013			2013	2014	vs. 2013
Andalucía	Convencional	34.588	33.042		Extremadura	Convencional	1.494	1.972	
	CTCC+CT	12.404	9.628			CTCC+CT	0	0	
	<b>Total</b>	<b>46.992</b>	<b>42.670</b>	<b>-9,2%</b>		<b>Total</b>	<b>1.494</b>	<b>1.972</b>	<b>+32,0%</b>
Aragón	Convencional	15.608	13.379		Galicia	Convencional	11.521	11.714	
	CTCC+CT	401	743			CTCC+CT	2.603	1.317	
	<b>Total</b>	<b>16.009</b>	<b>14.122</b>	<b>-11,8%</b>		<b>Total</b>	<b>14.124</b>	<b>13.031</b>	<b>-7,7%</b>
Asturias	Convencional	5.726	5.474		La Rioja	Convencional	2.876	2.329	
	CTCC+CT	677	683			CTCC+CT	827	343	
	<b>Total</b>	<b>6.403</b>	<b>6.157</b>	<b>-3,8%</b>		<b>Total</b>	<b>3.703</b>	<b>2.672</b>	<b>-27,8%</b>
Baleares	Convencional	680	655		Madrid	Convencional	26.573	22.728	
	CTCC+CT	2.764	3.551			CTCC+CT	0	0	
	<b>Total</b>	<b>3.444</b>	<b>4.205</b>	<b>+22,1%</b>		<b>Total</b>	<b>26.573</b>	<b>22.728</b>	<b>-14,5%</b>
Cantabria	Convencional	5.000	4.571		Murcia	Convencional	20.276	18.202	
	CTCC+CT	0	0			CTCC+CT	4.938	5.048	
	<b>Total</b>	<b>5.000</b>	<b>4.571</b>	<b>-8,6%</b>		<b>Total</b>	<b>25.214</b>	<b>23.249</b>	<b>-7,8%</b>
Castilla y León	Convencional	22.314	16.912		Navarra	Convencional	6.508	5.404	
	CTCC+CT	0	0			CTCC+CT	808	844	
	<b>Total</b>	<b>22.314</b>	<b>16.912</b>	<b>-24,2%</b>		<b>Total</b>	<b>7.316</b>	<b>6.248</b>	<b>-14,6%</b>
Castilla-La Mancha	Convencional	13.323	12.296		País Vasco	Convencional	25.532	23.157	
	CTCC+CT	2.793	2.796			CTCC+CT	6.785	6.387	
	<b>Total</b>	<b>16.116</b>	<b>15.092</b>	<b>-6,4%</b>		<b>Total</b>	<b>32.317</b>	<b>29.544</b>	<b>-8,6%</b>
Cataluña	Convencional	57.239	51.326		Valencia	Convencional	27.461	26.574	
	CTCC+CT	13.419	11.874			CTCC+CT	8.364	8.551	
	<b>Total</b>	<b>70.658</b>	<b>63.200</b>	<b>-10,6%</b>		<b>Total</b>	<b>35.825</b>	<b>35.125</b>	<b>-2,0%</b>
<b>Reparto porcentual</b>					<b>Convencional</b>	<b>276.718</b>	<b>249.736</b>		
					<b>CTCC+CT</b>	<b>56.782</b>	<b>51.765</b>		
					<b>Total</b>	<b>333.500</b>	<b>301.500</b>		<b>-9,6%</b>

Figura 1.1.3 Variación Consumo de Gas 2013-2014 por Comunidades. Fuente: ENAGAS



La Figura 1.1.3. expone la evolución por Comunidades Autónomas de 2013 a 2014 de Demanda de Gas. Se aprecia un descenso generalizado salvo en Baleares y Extremadura (ambas son las Comunidades Autónomas con menor consumo).

En cuanto a Cuota de mercado por número de clientes de Gas, la comercializadora con mayor porcentaje sigue siendo Gas Natural (con un 58,16% de cuota), seguido de Endesa, con un 40% menos de cuota de mercado.



Gráfica 1.1.1 Porcentaje de Cuota de Mercado por Comercializadora. Fuente:CNMC

En relación a la serie temporal que interesa en este análisis, la Demanda de Gas Convencional se cierra el ejercicio con 26.982Gwh menos que a cierre de 2013. La Figura 1.1.4. muestra como la tendencia de esta Demanda de Gas no es clara:



Figura 1.1.4 Evolución Demanda de Gas Convencional 2004-2014. Fuente: ENAGAS

Basándonos en el Informe Anual de ENAGAS para 2014 el descenso del Consumo en el sector convencional podría deberse principalmente a las temperaturas.

- El año 2014 fue clasificado como un año muy cálido, mientras que 2013 fue clasificado como muy frío, lo cual provocó una fuerte disminución de Demanda a nivel doméstico, comercial o pyme.

## 1.2 Necesidad del Estudio

La necesidad de la realización de un modelo predictivo de Demanda de Gas viene motivada por el cambio regulatorio en el mercado español del gas, donde en la Orden IET/2355/2014, del 16 de diciembre de 2014 (adjunta) se modifica, en la disposición final segunda la norma de gestión técnica del sistema, pasando los desbalances de gas en el AOC a representar un 10% del coste de la materia prima, calculada como el promedio entre el NBP y el HH para ese día.

Por lo tanto, y tal y como expone el documento del BOE (Adjuntado pdf en este apartado), a partir del pasado 1 de Marzo de 2015, las comercializadoras de Gas son sancionadas por sus desbalances (tanto por exceso como por defecto de Gas) en los sistemas de almacenamiento.



BOE-A-2014-13057.  
pdf

A raíz de esta modificación del Mercado, se hace especialmente crítico que las comercializadoras cumplan con los procesos que se exigen diariamente (Día D). Por tanto se convierte en prioritario la necesidad de aplicar un modelo predictivo lo más ajustado posible.

1. La previsión de la demanda agregada del día anterior (D-1) para realizar las objeciones necesarias a la nominación que asigna Enagás
2. La previsión de la demanda en el día posterior (D+1) para que el sobrecoste de desvíos se minimice en la medida de lo posible.



El operador del Sistema gasista español (ENAGAS) obliga a los sujetos de mercado a proveerles con frecuencia diaria de una estimación del consumo diario que tendrán sus consumidores de Gas. Las diferentes comercializadoras están obligadas a mandar la previsión de consumo de gas de sus clientes.

### **1.3 Obtención de la Serie Temporal**

Se disponen de datos diarios de Demanda de Gas Convencional obtenidos a partir del Gestor Técnico del Sistema Gasista español (ENAGAS), debiendo realizar la predicción más ajustada posible de uno a diez días vista. Dichos datos son a nivel nacional, por lo que se ofrece un modelo predictivo perfectamente extrapolable a la Demanda de Gas de cada una de las comercializadoras o productoras.

Un total de 881 datos diarios de consumo de Gas a nivel nacional, de los cuales se utilizará el periodo del 1 de Enero de 2013 al 31 de Diciembre de 2014 como periodo de entrenamiento y del 1 de Enero de 2015 al 31 de Mayo de 2015 como periodo de validación.

### **1.4 Objetivos del Estudio**

Los objetivos del trabajo son los siguientes:

1. Obtener pronósticos de la Demanda Convencional de Gas con frecuencia diaria. Proveer la mejor predicción posible en el corto plazo.
2. Visualizar los comportamientos estacionales del consumo de Gas, así como el efecto de festivos sobre este.
3. Comparación de dos modelos de predicción a corto plazo, uno de ellos incluyendo las variables de temperaturas. Ofrece la posibilidad de ver el efecto de dichas variables sobre la demanda de Gas y si son un factor importante a la hora de obtener una buena predicción.

Para ello, se propone utilizar un modelo autorregresivo integrado de media móvil (ARIMA). De estos modelos se tendrán en cuenta los ciclos estacionales (desarrollando modelos SARIMA), así como el efecto de variables exógenas, tales como el efecto de



festivos y de periodos vacacionales. Además se propondrá otro modelo incluyendo variables de temperaturas mínimas y máximas con el fin de observar su efecto en la predicción de la serie de Gas.

Por último se compararán los resultados obtenidos incorporando o no las temperaturas en el periodo de validación. Se propone comparar el comportamiento de los modelos en días específicos (por ejemplo, Semana Santa), así como valorar los resultados en los diez horizontes de predicción mediante los indicadores MAE (Error absoluto medio) , el MAPE (Error absoluto medio en porcentaje) y visualización gráfica.

## 1.5 Limitaciones del Estudio

Se exponen algunas de las limitaciones que presenta el modelo estadístico que se va a desarrollar en este trabajo, y a su vez, se detalla cómo sería una situación “real” en la aplicación del modelo de Demanda de Gas de una Empresa Energética.

1. Se han obtenido datos de Demanda Agregada. Sería más correcto disponerlos por Comunidad Autónoma. De esa forma podría ponderarse las temperaturas en función de la Demanda de cada región.
2. Los datos de temperatura mínima y máxima son reales en las predicciones. En un escenario real se tendrían predicciones de temperaturas para los 10 horizontes de predicción, por lo que los resultados empeorarían en cierto grado.

## 2 Estado del Arte

El desarrollo de este trabajo está basado en el estudio de series temporales aplicado al mercado energético. Existen diversos estudios y artículos basados en este tema, especialmente en la predicción de Demanda eléctrica.

Los primeros intentos de crear modelos de predicción de series temporales se caracterizan por la idea de un mundo determinista. La mayor contribución fue de Yule que aplico el concepto de proceso estocástico a la series de tiempo. A partir de este postulado se fueron



desarrollando los conceptos de modelos lineales de predicción. Los trabajos de Slutsky, Walker, Yaglom o Yule crean los términos conocidos como modelos autorregresivos (AR), modelos de medias móviles (MA). También aportarán a esta metodología el Teorema de descomposición de Wold, o los problemas de predicción lineal de Kolmogorov.

Aunque el punto de inflexión llegó con la publicación de “Time Series Analysis: Forecasting and Control” de Box y Jenkins (1970), que ha tenido un gran impacto tanto teóricamente como prácticamente en la predicción de series temporales. Se popularizó el uso de modelos ARIMA, aplicándose a la predicción de series temporales de diversos campos.

En la aplicación de modelos de series temporales al mercado eléctrico, destacar el trabajo “Forecasting the electricity load from one day to one week ahead for the Spanish system operator” de Cancelo JR, A Espasa y R Grafe (2008). Modelos de predicción de series temporales a corto plazo.

En el estudio de la relación de demanda energética y temperaturas, el trabajo “Temperature and seasonality influences on Spanish electricity load” de Angel Pardo, Vicente Meneu y Enric Valor (2002), relaciona la energía eléctrica con el efecto de las temperaturas, así como una fuerte estacionalidad en la serie temporal.

En el presente trabajo se propone aplicar metodologías de los diferentes estudios expuestos, en de serie temporal diaria de Demanda de Gas Convencional.

## **3 Técnicas de Predicción**

### **3.1 Modelos ARIMA: Estructura General**

Los modelos ARIMA son modelos estadísticos desarrollados por Box and Jenkins para series temporales que tienen en cuenta la dependencia existente entre los datos, es decir, la observación en el momento dado es modelada en función de datos anteriores.



Esta metodología permite describir el valor como una función lineal de los datos anteriores y errores debidos al azar (y en el que se puede incluir un componente cíclico o estacional).

Recibe el nombre en base a los componentes de los que está formado:

1. AR (Autorregresivo)
2. I (Integrado, diferenciación de la variable)
3. MA (Medias Móviles)

Fases para la predicción mediante la metodología ARIMA:

- 1. Primera fase:** Formulación del modelo. Explicación de los componentes de un modelo ARIMA.
- 2. Segunda fase:** Identificar el posible modelo ARIMA
  - a. Análisis de la serie temporal. Decidir las transformaciones a aplicar para convertir la serie en estacionaria
  - b. Decidir un modelo ARIMA para la serie estacionaria, indicando los valores necesarios para la parte autorregresiva y la media móvil.
- 3. Tercera fase:** Una vez que ha sido seleccionado un modelo ARIMA, se estiman los parámetros por máxima verosimilitud y se obtienen los errores estándar y los residuos del modelo.
- 4. Cuarta fase:** Diagnóstico del modelo, en esta fase se comprueba que los residuos no tienen dependencia y siguen un proceso de ruido blanco. Si los residuos mostrasen dependencia, habría que realizar de nuevo las dos primeras etapas y reformular el modelo, hasta obtener uno adecuado.
- 5. Quinta fase:** Una vez que tenemos un modelo adecuado realizamos la predicción.

A continuación se detalla cada una de las fases expuestas anteriormente:

### 1. Formulación de un modelo ARIMA



### Modelos Autorregresivos

La parte AR del modelo describe que las observaciones en un determinado momento son predecibles a partir de las observaciones previas del modelo más un término de error. Se muestra la representación de un modelo Autorregresivo de orden p (AR (p)):

$$X_t = \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_p X_{t-p} + e_t$$

Esta ecuación puede ponerse mediante un operador B, siguiendo la forma:

$$(1 - \beta_1 B - \beta_2 B^2 - \dots - \beta_p B^p) X_t = e_t \quad B^k(X_t) = X_{t-k}$$

### Modelo de Medias Móviles

El valor actual puede predecirse a partir de la componente aleatoria de ese momento, y en menor medida, de los impulsos aleatorios anteriores. Un modelo de Media Móvil de orden q (MA (q)), tiene la siguiente forma:

$$X_t = e_t - \alpha_1 e_{t-1} - \alpha_2 e_{t-2} - \dots - \alpha_q e_{t-q}$$

Expresado mediante un operador B:

$$X_t = (1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_q B^q) e_t$$

### Modelo ARMA

Este modelo es la unión entre un modelo Autorregresivo de orden p y modelos de Medias Móviles de orden q. La ecuación se representa de la siguiente manera:

$$X_t = \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_p X_{t-p} + e_t - \alpha_1 e_{t-1} - \alpha_2 e_{t-2} - \dots - \alpha_q e_{t-q}$$

Reformulando la ecuación mediante el operador B:

$$(1 - \beta_1 B - \beta_2 B^2 - \dots - \beta_p B^p) X_t = (1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_q B^q) e_t$$

### Modelo ARIMA (p,d,q)

Este proceso se sigue componiendo de una parte autorregresiva y otra de medias móviles, la única diferencia es que se incluye un nuevo operador B en la parte autorregresiva (correspondiente a la diferenciación d).



$$(1 - \beta_1 B - \beta_2 B^2 - \dots - \beta_p B^p)(1 - B)^d X_t = (1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_q B^q) e_t$$

Modelo ARIMA (p,d,q)(P,D,Q)

En el análisis de las series temporales utilizando los modelos ARIMA también es posible identificar estacionalidades cíclicas. Por lo tanto, el primer paréntesis (p,d,q) pertenecería a la parte regular de la serie temporal, y el segundo paréntesis (P,D,Q) pertenecería a la parte estacional.

A continuación se muestran algunas expresiones para este modelo teniendo en cuenta las ecuaciones utiliza en el modelo ARIMA(p,d,q) y teniendo en cuenta la ampliación del modelo con una parte cíclica (P,D,Q). Los ejemplos están aplicados sobre datos mensuales por lo que la diferenciación estacional será de 12 periodos.

- ARIMA(0,1,1)(0,1,1):

$$(1 - B)(1 - B^{12})X_t = (1 - B)(1 - B^{12})e_t$$

- ARIMA(2,1,0)(1,0,0):

$$(1 - B)(1 - B^{12})(1 - B - B^2)X_t = e_t$$

Caso especial del *Ruido Blanco*:

Se trata de un proceso puramente aleatorio, y que sigue las siguientes condiciones:

1. Media cero:  $\mu = E(X_t) = 0$
2. Varianza constante:  $\sigma^2 = var(X_t)$
3. Covarianza cero:  $\sigma = cov(X_t, X_{t+k})$  donde  $k = \dots, -2, -1, 0, 1, 2, \dots$

El interés en la predicción con modelos ARIMA es conseguir que los residuos de la estructura creada se comporten como un Ruido Blanco.

## 2. Segunda Fase: Identificación del modelo

### Análisis de la serie

En la primera fase del estudio de la serie temporal se deben considerar las diferentes características de la serie y la forma de identificarlas y corregirlas.



- Alta frecuencia
- Estacionariedad
- Comportamientos estacionales

La Frecuencia se mide dependiendo del espacio temporal transcurrido entre los datos de la serie temporal y es una característica de la serie temporal que no se puede corregir. La frecuencia de la serie de estudio son los días laborables (lunes a viernes), por lo que tiene una frecuencia alta.

Para poder predecir una serie temporal, es necesario que nuestra serie sea estacionaria, entendiendo un proceso estacionario como un proceso estocástico cuya distribución de probabilidad en un tiempo concreto es la misma para el resto de momentos. El caso del ruido blanco (se mencionará más adelante) es un proceso estacionario.

Metodologías para eliminar la no estacionariedad:

- Para estabilizar la varianza utilizamos metodologías de Box-Cox, como el logaritmo, la raíz cuadrada, etc.
- Para estabilizar la media se toman diferencias:

$$\nabla_s X_t = (1 - B)_s X_t = (1 - B^s) X_t = X_t - X_{t-s}$$

Dependiendo del valor de  $s$  estaremos tomando una diferenciación. Si los datos fuesen diarios:

- $s=1$  Diferenciación diaria
- $s=7$  Diferenciación semanal
- $s=30$  Diferenciación mensual

En principio será suficiente con realizar la transformación de la serie y aplicar una primera diferencia regular.

Para comprobar la estacionariedad de la serie temporal se usa el test de Dickey-Fuller. Este test confirma si una raíz unitaria está presente en el modelo autorregresivo.

Por último hay que tratar la estacionalidad, que representan que los datos que componen la serie temporal se comportan de forma cíclica o periódica. A modo de ejemplo, si



decimos que los datos presentan estacionalidad anual es que los datos de Enero de los diferentes años tienen el mismo comportamiento.

La Estacionalidad de la serie temporal se comprobará de forma gráfica y mediante tres metodologías diferentes:

1. Visualizar la serie temporal y detectar picos estacionales: En el estudio de la variable objetivo se ha podido observar que no presente comportamientos estacionales.
2. Comprobar los resultados de la Función de Autocorrelación y de la Función de Autocorrelación Parcial. Como se verá en el siguiente apartado (Selección del modelo ARIMA), los correlogramas no presentan presencia comportamiento estacional.
3. Realización de un periodograma.

Estudio de la función de autocorrelación (ACF) y función de autocorrelación parcial (PACF)

La función de autocorrelación (ACF) y la función de autocorrelación parcial (PACF) calculan la relación estadística lineal entre las observaciones de una serie temporal. Mediante la visualización de las gráficas y sus correspondientes resultados podremos valorar que ordenes (p,d,q) tomara el modelo ARIMA y que ordenes (P,D,Q) tomara la parte estacional.

Primero destacar el índice de correlación entre dos variables:

$$\rho_1 = \frac{cov(x_t, x_{t-1})}{\sqrt{var(x_t) \times var(x_{t-1})}}$$

Por el supuesto de estacionariedad de la serie temporal, se entiende que  $var(x_t) = var(x_{t-1})$  por lo que la función de correlación sería:

$$\rho_1 = \frac{cov(x_t, x_{t-1})}{var(x_t)}$$

En general, para cualquier periodo s, podremos sacar las funciones de autocorrelación:



$$\rho_s = \frac{cov(x_t, x_{t-s})}{var(x_t)}$$

Evidentemente, si el valor de  $s=0$ :

$$\rho_s = \frac{cov(x_t, x_t)}{var(x_t)} = \frac{var(x_t)}{var(x_t)} = 1$$

Una vez que se han obtenido las funciones de correlación, se debe comparar el valor de dichas funciones entre los diferentes modelos. El estimador muestral de la función ACF viene dado por la siguiente expresión:

$$r_s = \frac{\sum_{t=1}^{n-s} (x_t - \bar{x}_t)(x_{t-s} - \bar{x}_t)}{\sum_{t=1}^n (x_t - \bar{x}_t)^2}$$

Donde  $\bar{x}_t$  es el promedio de  $x_t$ .

Para realizar la expresión matemática de la PACF son necesarias dos ecuaciones:

1º Ecuación:

$$v_t = x_t - \alpha_1 x_{t-1} - \alpha_2 x_{t-2} - \dots - \alpha_{k-2} x_{t-k-2} - \alpha_{k-1} x_{t-k-1}$$

Donde:

$$t = k + 1, \dots, T$$

De la ecuación deducimos que cada uno de los  $\alpha$  son los parámetros estimados para cada una de las series  $x_t$  y las sucesivas, tal como  $x_{t-1}, x_{t-2}, \dots, x_{t-k-2}, x_{t-k-1}$ . Siendo  $v_t$  la serie que recoge la parte  $x_t$  de no explicada por cada una de las series  $x_{t-1}, x_{t-2}, \dots, x_{t-k-2}, x_{t-k-1}$ .

2º Ecuación:

$$w_t = x_{t-k} - \delta_1 x_{t-1} - \delta_2 x_{t-2} - \dots - \delta_{k-2} x_{t-k-2} - \delta_{k-1} x_{t-k-1}$$

Donde:

$$t = k + 1, \dots, T$$

De la ecuación deducimos que cada uno de los  $\delta$  son los parámetros estimados para cada una de las series  $x_{t-k}$  y las sucesivas, tal como  $x_{t-1}, x_{t-2}, \dots, x_{t-k-2}, x_{t-k-1}$ . Siendo  $w_t$



la serie que recoge la parte  $x_{t-k}$  de no explicada por cada una de las series  $x_{t-1}, x_{t-2}, \dots, x_{t-k-2}, x_{t-k-1}$ .

Los coeficientes de autocorrelación parcial de orden  $k$  es el coeficiente de correlación entre  $v_t$  y  $w_t$ , ya que ambos se han calculado con una separación  $k$ .

Por tanto, definimos el coeficiente de autocorrelación parcial de orden  $k$ :

$$\tau_{v,w} = \frac{\sum_{t=k+1}^n (v_t - \bar{v}_t)(w_{t-k} - \bar{w}_t)}{\sqrt{\sum_{t=k+1}^n (v_t - \bar{v}_t)^2} \sqrt{\sum_{t=k+1}^n (w_t - \bar{w}_t)^2}}$$

Donde  $\bar{v}_t$  y  $\bar{w}_t$  son los promedios de las series  $v_t$  y  $w_t$ . El número marca los componentes de la serie.

Una vez creadas las funciones de ACF y PACF, serán utilizadas para detectar que modelo ARIMA se ajusta mejor a la serie temporal. La estructura general para detectar el modelo:

Modelo	Función ACF	Función PACF
AR(p)	Decrecimiento rápido exponencial atenuado u ondas sinusoidales	Solo los p primeros coeficientes son significativos. El resto se anulan
MA(q)	Solo los q primeros coeficientes son significativos. El resto se anulan	Decrecimiento rápido exponencial atenuado u ondas sinusoidales
ARMA(p,q)	Comportamiento irregular en los retardos (con picos).Decrecimiento retardos posteriores a q	Decrece exponencialmente u ondas sinusoidales. No decrece para rápidamente

Tabla 3.1.1. Comportamiento gráfico de un modelo ARMA

Un valor es considerado casi cero cuando su módulo es inferior a  $2/\sqrt{n}$ . A parte de la información reflejada en la tabla, hay una serie de gráficas (APENDICE) que permiten identificar un modelo ARIMA dependiendo del comportamiento de las correlaciones en la ACF y PACF



### 3. Tercera Fase: Estimación de los parámetros

En esta fase se cuantifican los parámetros de las formula. Los métodos utilizados para la estimación pueden ser ML, CLS o ULS.

### 4. Cuarta Fase: Diagnósis del modelo

Un modelo es adecuado y sus predicciones son buenas cuando los residuos se comportan como un ruido blanco. La identificación de los residuos se realiza con las siguientes metodologías:

#### Representación Gráfica

Representación gráfica de las funciones de la ACF y PACF para los residuos en el que ninguno de los datos debe de ser significativo (deben de ser muy pequeños) y estar dentro de las bandas de confianza mencionadas en el apartado anterior  $(-2/\sqrt{n} \cdot 2/\sqrt{n})$ .

#### Test de Ljung-Box

El test de encarga de contrastar si los m primeros periodos de los residuos del modelo tienen correlación entre ellos. El estadístico se define de la siguiente forma:

$$LB = n(n + 2) \sum_{h=1}^m \frac{\rho_h^2}{n - r}$$

Donde  $\rho_h$  es el coeficiente de autocorrelación de los residuos, n es el número de componentes de la serie  $X_t$  y r es el número de parámetros estimados.

### 5. Quinta Fase: Predicción

Una vez completadas las fases anteriores correctamente, se realiza el ejercicio de predicción.

En el caso de no obtener buenos resultados, se comenzaría de nuevo el proceso y se estimaría un nuevo modelo.



### 3.2 Modelos ARIMA con variables de Intervención

Aplicación de modelos ARIMA con variables de intervención, debido a que la serie temporal a predecir puede verse afectada por fenómenos externos, tales como festivos, vacaciones o huelgas, por ejemplo.

Estos fenómenos intervienen en el comportamiento de la serie, por lo que se debe evaluar su efecto mediante la introducción en el modelo ARIMA de variables artificiales binarias. Por lo tanto, se incorporan como variables Dummy en forma de impulsos temporales y escalones. Si no se incorporan dichas variables se pueden producir sesgos en las estimaciones, así como un elevado error en las predicciones.

En el análisis realizado en este proyecto, se puede reconocer dos grandes grupos.

**VARIABLES LABORALIDAD:** Son variables de tipo impulso o de tipo escalón que recogen periodos vacacionales que afectan directamente a la demanda de Gas.

**VARIABLES CALENDARIO:** Marcan la estacionalidad en el comportamiento de la demanda, ya sea mensual o diaria.

### 3.3 Modelos ARMA con variables externas (ARMAX)

Además de incorporar valores autorregresivos y de media móvil, se pueden incorporar variables externas en forma de regresores. Son variables que provienen de una fuente externa distinta a la de la serie de predicción.

Dichos modelos presentan los siguientes parámetros: ARIMA (p,q,n), siendo p la parte autorregresiva, q la parte media móvil y n las variables externas. La nueva ecuación quedaría de la siguiente forma:

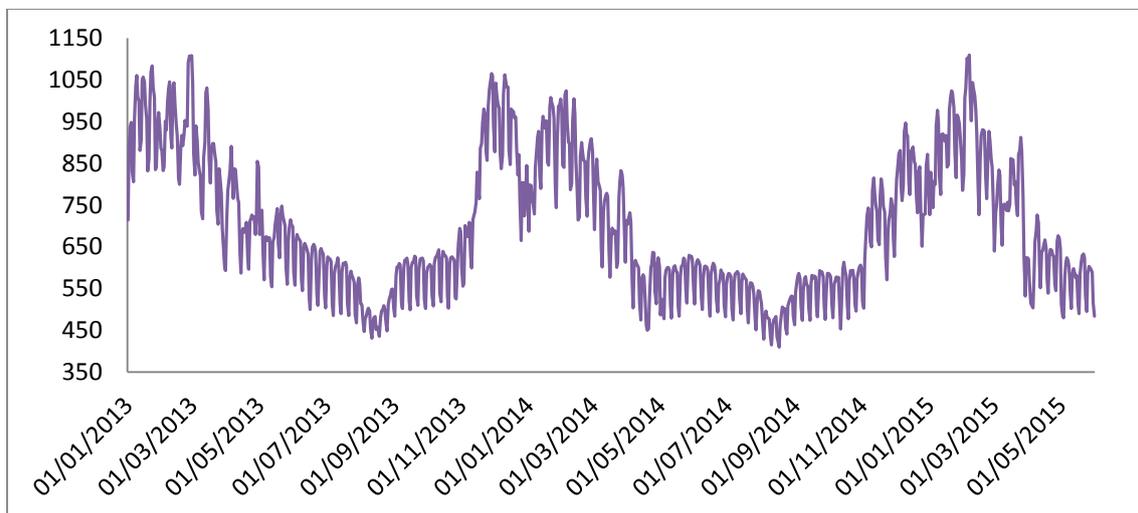
$$X_t = \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + \beta_p X_{t-p} + e_t - \alpha_1 e_{t-1} - \alpha_2 e_{t-2} - \dots - \alpha_q e_{t-q} + Y_{1,t} + \dots + Y_{n,t}$$

## 4 Estudio de la Base de Datos

### 4.1 Análisis de la Serie Temporal de Demanda de Gas

#### 4.1.1 Análisis Gráfico

Para la realización del estudio se considera la serie de Demanda de Gas Convencional a nivel Nacional. La serie temporal tiene frecuencia diaria, siendo el intervalo de tiempo del 01 de Enero de 2013 al 31 de Mayo de 2015. La Gráfica 4.1.1.1. muestra la evolución histórica de la serie, mientras que la Gráfica 4.1.1.2 muestra la serie desestacionalizada.



Gráfica 4.1.1.1 Evolución demanda diaria de gas

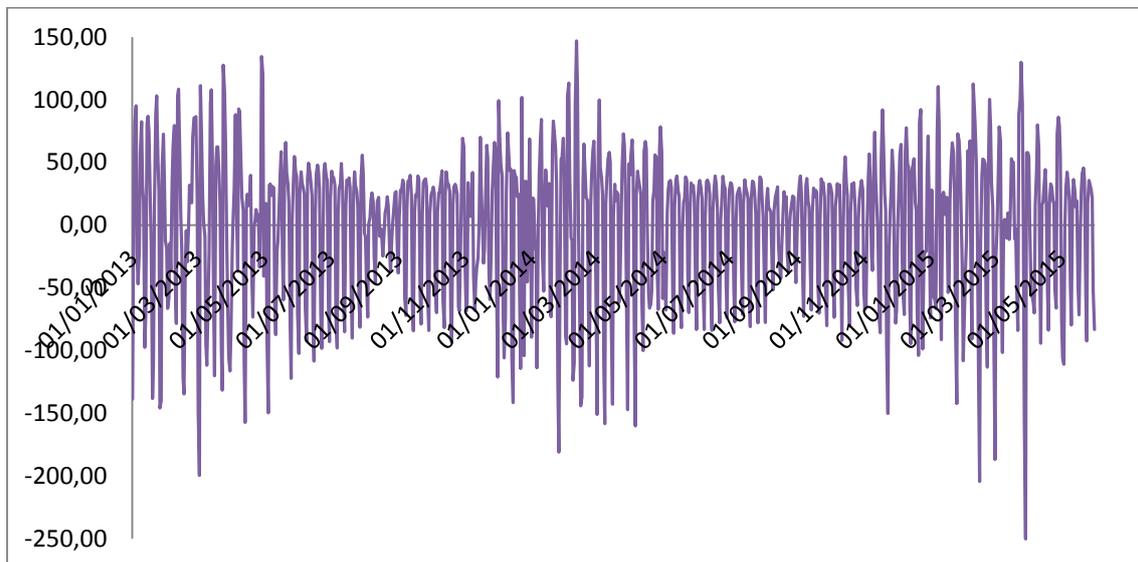
A simple vista se pueden visualizar ciertos componentes estacionales. A nivel mensual se aprecia como Agosto es el mes con menor Demanda, y Febrero, el de más Demanda. A nivel diario, los picos más bajos corresponden a los Domingos. En la apartado de Estacionalidad diaria se estudiará el comportamiento del consumo para los siete días de la semana y como se produce un descenso de este los fines de semana.

Se realiza la gráfica aplicando la siguiente fórmula, restando el valor en el momento  $t$  menos el promedio de Demanda semanal:

$$Y_t = X_t - \overline{X_{semana}}$$



Con el fin de desestacionalizar la serie temporal y ver un movimiento más preciso de la Demanda de Gas



Gráfica 4.1.1.2 Evolución demanda diaria de gas desestacionalizada

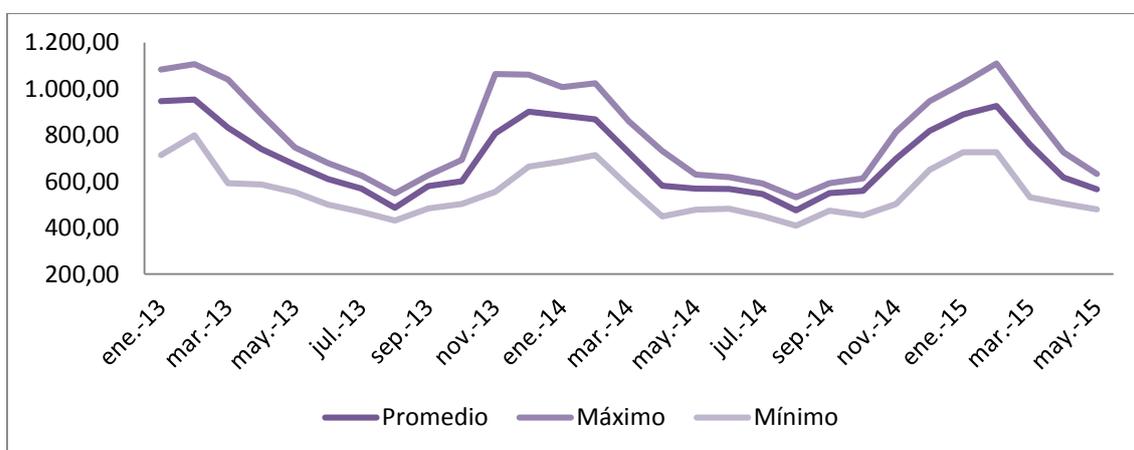
Se aprecia una mayor variabilidad de los datos en la época invernal, mientras que en Verano se produce un comportamiento más constante. Los datos negativos, principalmente reflejan los datos del fin de semana.

En la Tabla 4.1.1.1. se aprecia la evolución mensual de la Demanda de Gas, así como un estudio de valores promedio, máximos y mínimos por mes.

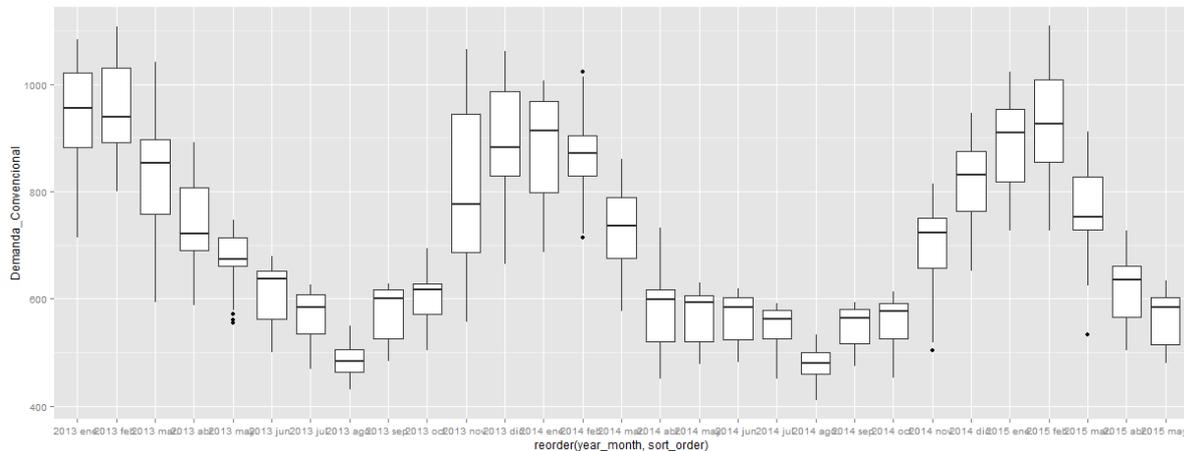


DATOS DEMANDA CONVENCIONAL GAS					
Año	Mes	Suma	Promedio	Máximo	Mínimo
2013	1	29.362	947,16	1.084	714
2013	2	26.716	954,14	1.108	800
2013	3	25.813	832,68	1.041	593
2013	4	22.204	740,13	891	587
2013	5	20.869	673,19	748	554
2013	6	18.303	610,10	680	500
2013	7	17.671	570,03	626	468
2013	8	15.086	486,65	549	431
2013	9	17.394	579,80	628	484
2013	10	18.637	601,19	694	503
2013	11	24.228	807,60	1.065	556
2013	12	27.971	902,29	1.062	665
2014	1	27.458	885,74	1.008	687
2014	2	24.340	869,29	1.024	714
2014	3	22.496	725,68	860	577
2014	4	17.473	582,43	732	450
2014	5	17.655	569,52	630	478
2014	6	17.025	567,50	619	482
2014	7	16.919	545,77	591	451
2014	8	14.754	475,94	532	410
2014	9	16.513	550,43	593	474
2014	10	17.350	559,68	613	453
2014	11	20.992	699,73	815	503
2014	12	25.413	819,77	947	651
2015	1	27.557	888,94	1.024	727
2015	2	25.938	926,36	1.110	727
2015	3	23.544	759,48	912	532
2015	4	18.529	617,63	726	504
2015	5	17.565	566,61	633	480
<b>Total</b>		<b>615.775</b>	<b>698,95</b>	<b>1.110</b>	<b>410</b>

Tabla 4.1.1.1. Evolución mensual de la demanda de gas



Gráfica 4.1.1.3 Evolución promedio, máximo, y mínima demanda de gas



Gráfica 4.1.1.4. Boxplot del comportamiento mensual

En la Gráfica 4.1.1.3. y el Boxplot de la Gráfica 4.1.1.4. se confirma en comportamiento estacional identificado en la gráfica de la serie desestacionalizada. Se puede ver claramente la estacionalidad de la serie, pero también se aprecia una variabilidad mucho mayor en los datos de Invierno, en comparación con los datos de Verano.

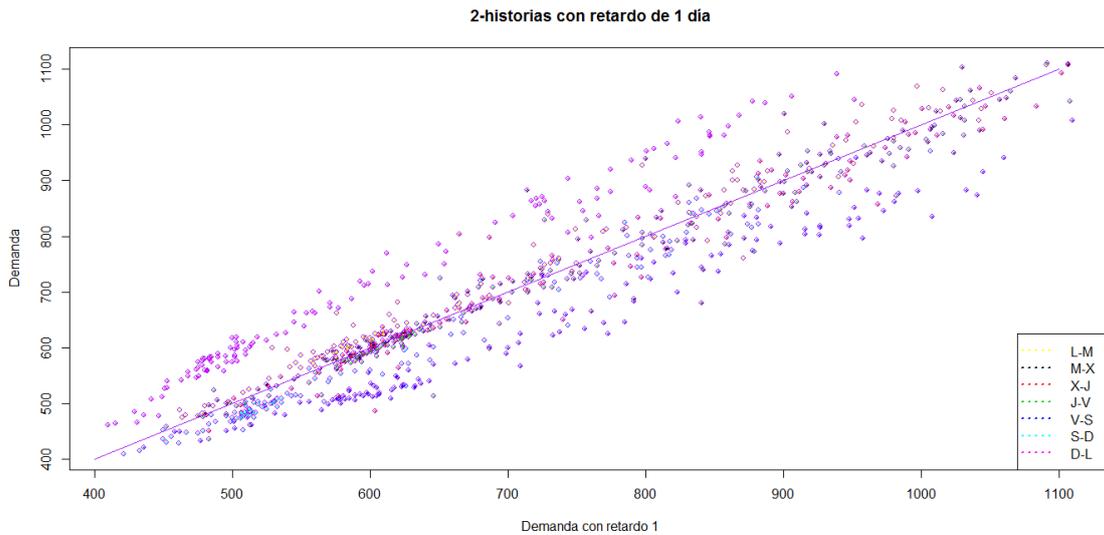
En el apartado destinado a los datos de temperaturas, se verá el efecto de estas en los distintos años. De esta forma se puede comprobar si la introducción de temperaturas ayudará a controlar la volatilidad de las predicciones para la época Invernal.

### **Estudio de los movimientos diarios de demanda**

Mediante el siguiente estudio gráfico se pretende identificar los movimientos de la Demanda de Gas entre los siete días de la semana, entendiéndose como evolucionan los datos día a día, identificando la relación existente entre ellos.

Se pretende comparar lunes con martes, martes con miércoles, miércoles con jueves, jueves con viernes, viernes con sábados, sábados con domingos y domingos con lunes, siendo:

*L: lunes, M: martes, X: miércoles, J:jueves, V:viernes, S:sábados y D:domingos*

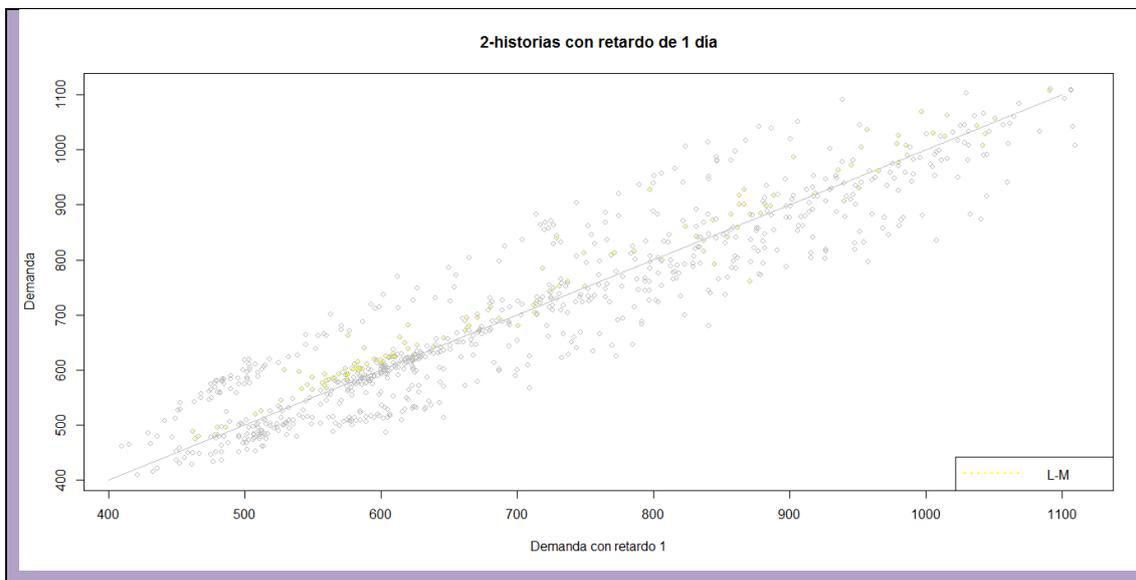


Gráfica 4.1.1.5. Representación de los movimientos diarios de Demanda

El eje X representa los datos en el momento t-1 y el eje Y en el momento t. Se producen tres comportamientos distintos representados mediante tres líneas rectas que contienen la siguiente información:

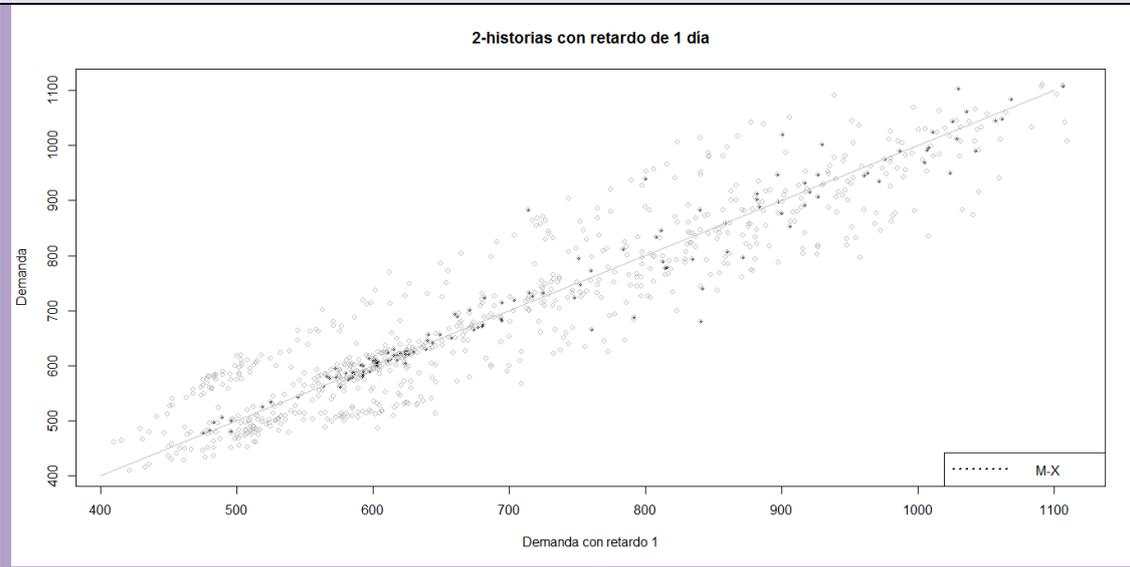
- **1º Recta (Superior):** Representa el ascenso de Demanda del día t-1 al t.
- **2º Recta (Central):** Representa los días en lo que no se produce una desviación elevada de Demanda en el paso del día t-1 al t.
- **3º Recta (Inferior):** Representa el descenso de Demanda del día t-1 al t.

A continuación se detalla uno por uno los siete movimientos semanales, identificando que comportamiento cumple cada uno:

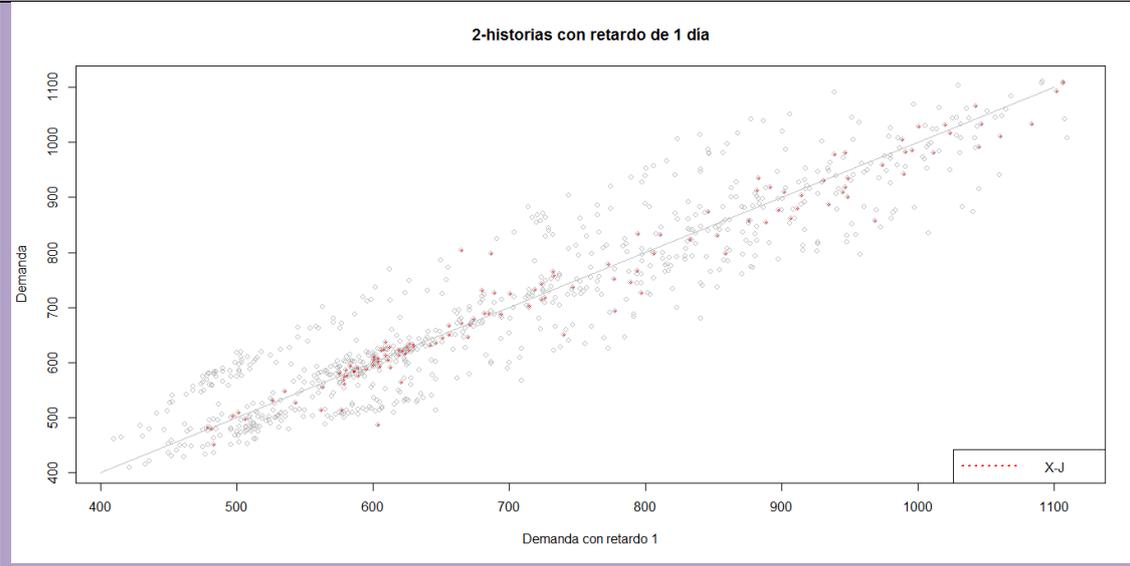




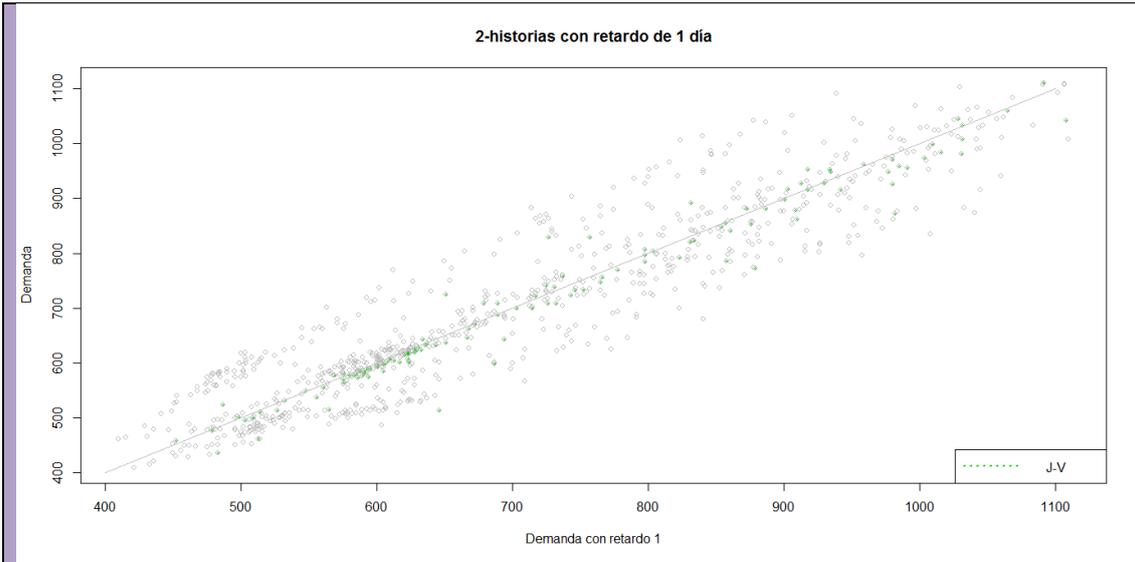
**Lunes y Martes (Línea Central):** Se encuentra dentro del grupo de comportamiento estándar en el que el dato para  $t$  y  $t-1$  es muy similar. Como se aprecia sus datos están un poco por encima de la línea por lo que el dato del Martes es algo superior al de los Lunes.



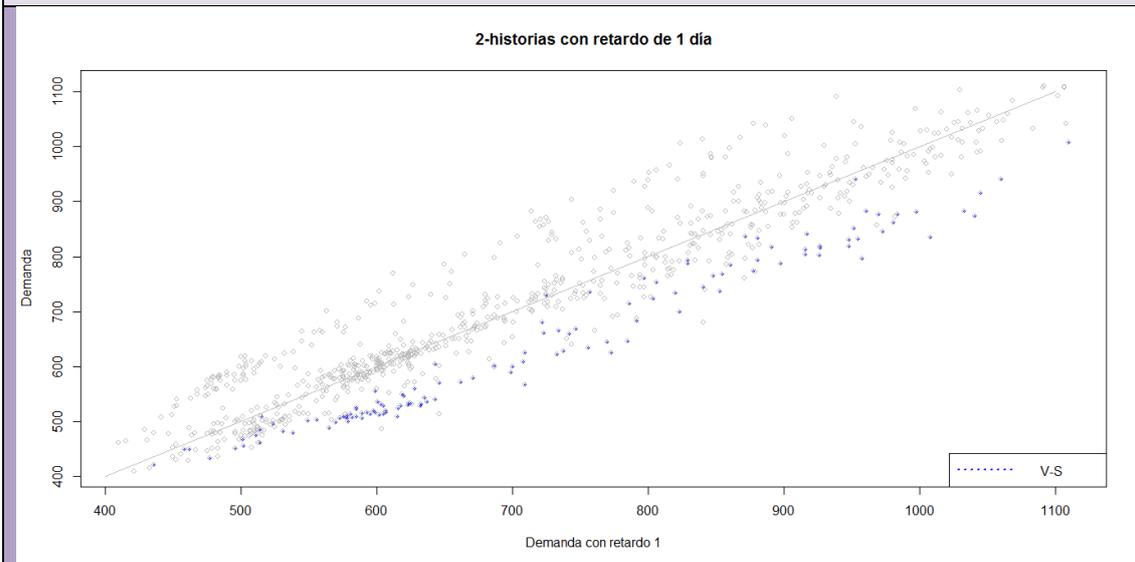
**Martes y Miércoles (Línea Central):** Se encuentra dentro del grupo de comportamiento estándar en el que el dato para  $t$  y  $t-1$  es muy similar. Se aprecia como los datos están sobre la línea por lo que en general los datos para Martes y Miércoles son muy similares



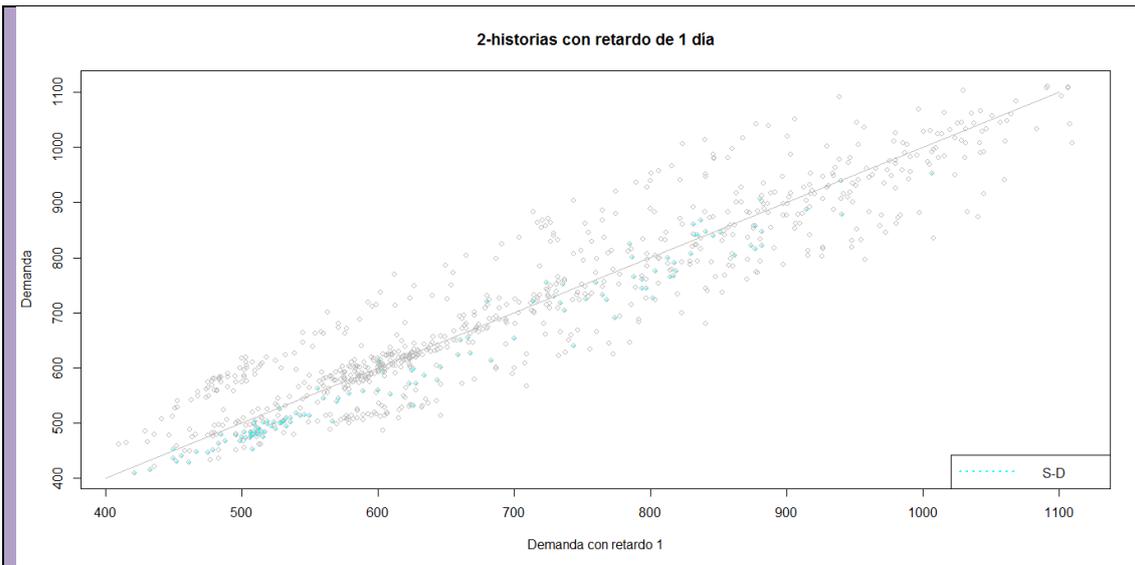
**Miércoles y Jueves (Línea Central):** Se encuentra dentro del grupo de comportamiento estándar en el que el dato para  $t$  y  $t-1$  es muy similar. Se aprecia como los datos están sobre la línea por lo que en general los datos para Miércoles y Jueves son muy similares



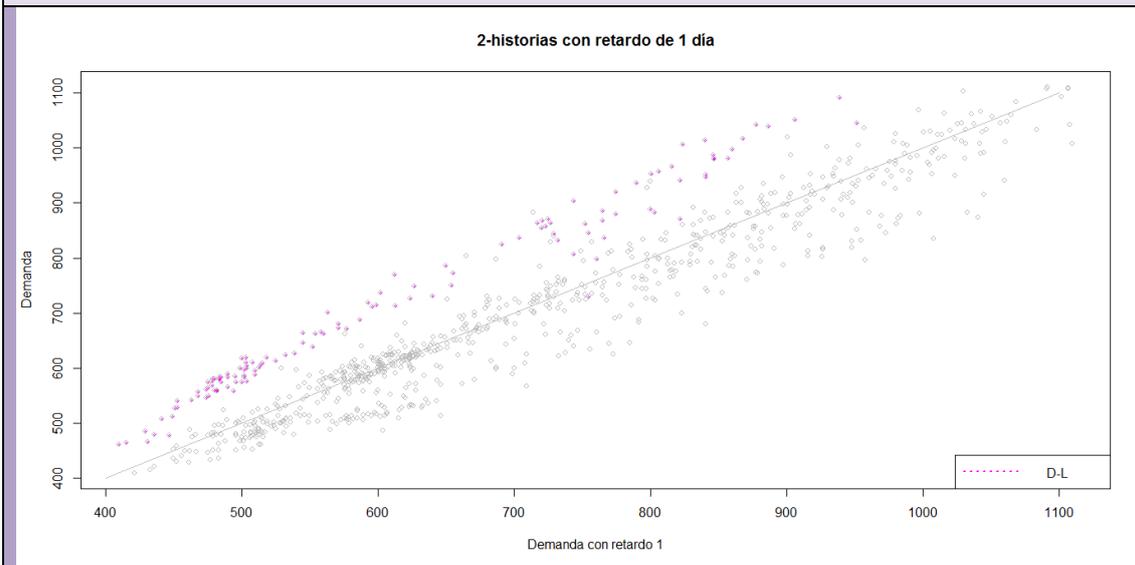
**Jueves y Viernes (Línea Central):** Se encuentra dentro del grupo de comportamiento estándar en el que el dato para  $t$  y  $t-1$  es muy similar. Como se aprecia sus datos están un poco por debajo de la línea por lo que el dato del Jueves es algo superior al de los Viernes.



**Viernes y Sábados (Línea Inferior):** Refleja un claro cambio en el valor de la Demanda entre estos dos días. Como se aprecia sus datos están claramente por debajo de la línea central por lo que el valor de los Sábados es inferior al de los Viernes (comienzo del fin de semana).



**Sábados y Domingos (Línea Central):** Se encuentra dentro del grupo de comportamiento estándar en el que el dato para  $t$  y  $t-1$  es muy similar. Como se aprecia sus datos están un poco por debajo de la línea por lo que el dato del Domingo es algo superior al de los Sábados.



**Domingo y Lunes (Línea Superior):** Refleja un claro cambio en el valor de la Demanda entre estos dos días. Como se aprecia sus datos están claramente por encima de la línea central por lo que el valor de los Lunes es inferior al de los Domingos (comienzo de la semana laborable).

Tabla 4.1.1.2. Representación día a día de la comparación de demanda entre  $t$  y  $t-1$



A parte de estudiar la evolución de los datos de demanda durante los días de la semana y de diferenciar la estacionalidad semanal que presenta la serie temporal, también permite confirmar los datos observados en las Gráficas 4.1.1.3. y 4.1.1.4., en los que se aprecia una mayor variabilidad de los datos en la época Invernal (coincide con los meses de más demanda).

Comparando los datos en el momento  $t$  y  $t-1$  se ve como a mayor valor de la demanda, mayor diferencia de un día a otro (mayor variabilidad) y por tanto valores más lejanos a la línea central.

## 4.1.2 Análisis descriptivo y de comportamiento

### Estadística Descriptiva

A continuación se exponen algunos estadísticos del histórico de consumos. La mediana del Consumo de Gas diario (un estadístico más robusto que la media, y que elimina el problema de datos atípicos) es de 650 Gwh. Dichos valores suelen aparecer en el periodo de Primavera.

La serie presenta un mínimo consumo durante los 881 días de 410 Gwh el día 17 de Agosto de 2014 (coincidiendo con la época de Verano, temperaturas altas y mayor movimiento vacacional) y un máximo consumo de 1.110 el 6 de Febrero de 2015 (coincidiendo con la época más fría del año).

Estadísticos	Valores
Media	698,95
Mediana	650
Moda	624
Desviación estándar	167,265
Varianza de la muestra	27922,423
Curtosis	-0,791
Coefficiente de asimetría	0,51
Rango	700
Mínimo	410
Máximo	1110
Suma	615775
Cuenta	881

Tabla 4.1.2.1 Estadísticos descriptivos de la serie



**Estacionalidad Diaria**

Día de la semana	Promedio Dem. Convencional
Lunes	716,00
Martes	736,67
Miercoles	736,82
Jueves	730,77
Viernes	720,91
Sabado	638,90
Domingo	612,71
<b>TOTAL</b>	<b>698,95</b>

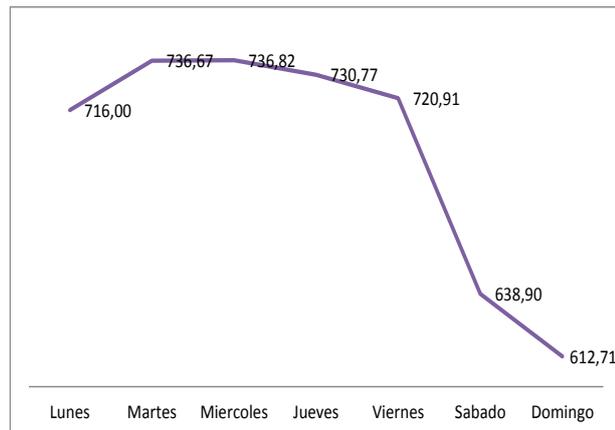


Tabla 4.1.2.2. Promedio Diario

Gráfico 4.1.2.1 Representación Promedio Diario

Los resultados muestran los Consumos medios de Gas por día de la semana. Se aprecia como los días centrales son los de más demanda de gas (martes, miércoles y jueves) y como los fines de semana baja claramente el consumo.

**Estacionalidad Mensual**

Mes	Promedio Dem. Convencional
Enero	907,28
Febrero	916,60
Marzo	772,61
Abril	646,73
Mayo	603,11
Junio	588,80
Julio	557,90
Agosto	481,29
Septiembre	565,12
Octubre	580,44
Noviembre	753,67
Diciembre	861,03
<b>TOTAL</b>	<b>698,95</b>

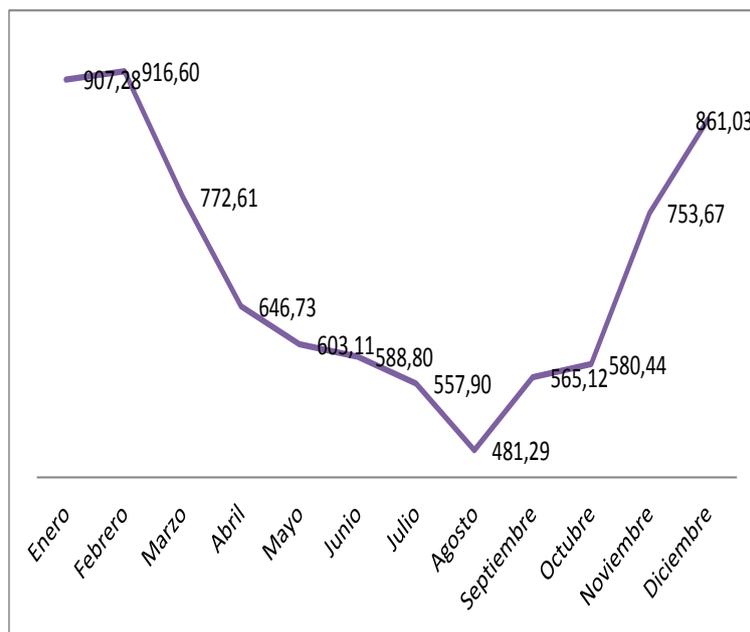


Tabla 4.1.2.3. Promedio Mensual

Gráfica 4.1.2.2. Representación Promedio Mensual

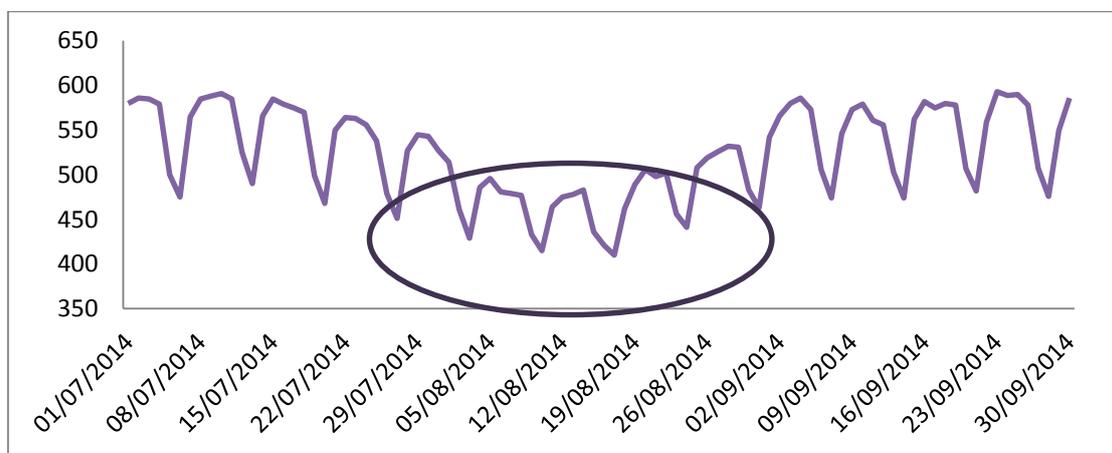
Los resultados muestran como los meses de Invierno son los que más demanda de Gas presentan (especialmente Febrero) y los de Verano los que menos (especialmente Agosto). Se puede apreciar una clara tendencia ascendente en el periodo Septiembre – Febrero y un descenso en el periodo Marzo – Agosto.



### 4.1.3 Detección de Datos Atípicos

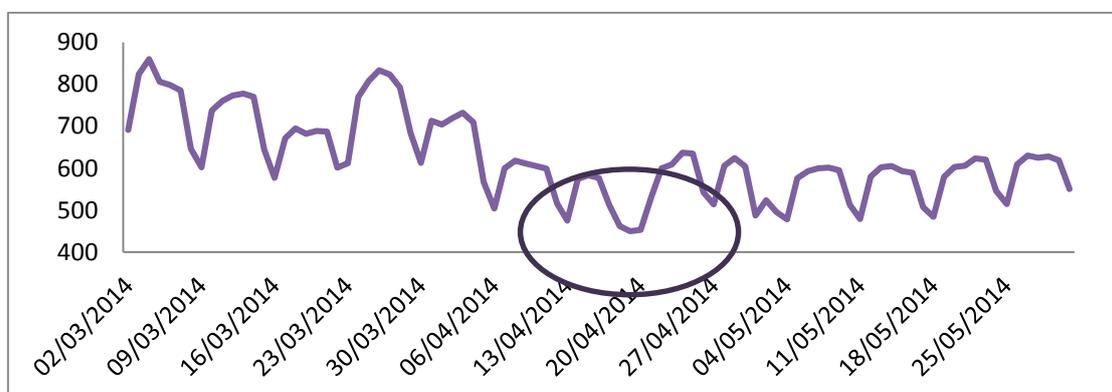
Se escoge el periodo temporal de 2014 para identificar los datos atípicos que se pueden encontrar en la serie temporal de Demanda de Gas.

- Datos atípicos de tipo escalón: Rompe en un periodo de tiempo el comportamiento de la serie. Puede tratarse de comportamientos estacionales. Como ejemplo se utiliza la Gráfica 4.1.3.1 que muestra el comportamiento vacacional del mes de Agosto de 2014.



Gráfica 4.1.3.1. Detección del comportamiento de la demanda de gas en Agosto

- Datos atípicos de tipo pulso: Rompe de forma puntual el comportamiento de la serie. Puede tratarse de un momento vacacional. Como ejemplo se utiliza la Gráfica 4.1.3.2 que muestra el comportamiento vacacional de Semana Santa para el año 2014.



Gráfica 4.1.3.2. Detección del comportamiento de la demanda de gas en Semana Santa



Los tipos de datos atípicos comentados en este punto se tratarán introduciendo variables Dummy estacionales y en forma de pulso para que el modelo pueda captar este comportamiento a lo largo del histórico diario de la serie.

## 4.2 Análisis variables exógenas

En el ejercicio de predicción de la Demanda de Gas se utilizarán datos pasados de la misma, así como variables explicativas con el fin de mejorar las predicciones.

Las variables exógenas se pueden dividir en dos grandes grupos:

- Variables Dummy. Pueden ser de dos tipos:
  - Variables Dummy de laboralidad: Son los pulsos que toman valor 0 y 1 que permiten controlar los datos atípicos de tipo pulso comentados en el análisis descriptivo de la serie.
  - Variables Dummy de estacionalidad: Son los pulsos que toman valor 0 y 1 que permiten controlar los datos atípicos de tipo escalón comentados en el análisis descriptivo de la serie.
- Variables temperaturas: Visualizar si es cierto el fuerte efecto de las temperaturas sobre la demanda de Gas.

El Esquema resume los tipos de variables externas que se van a utilizar de cara a la realización de la predicción.

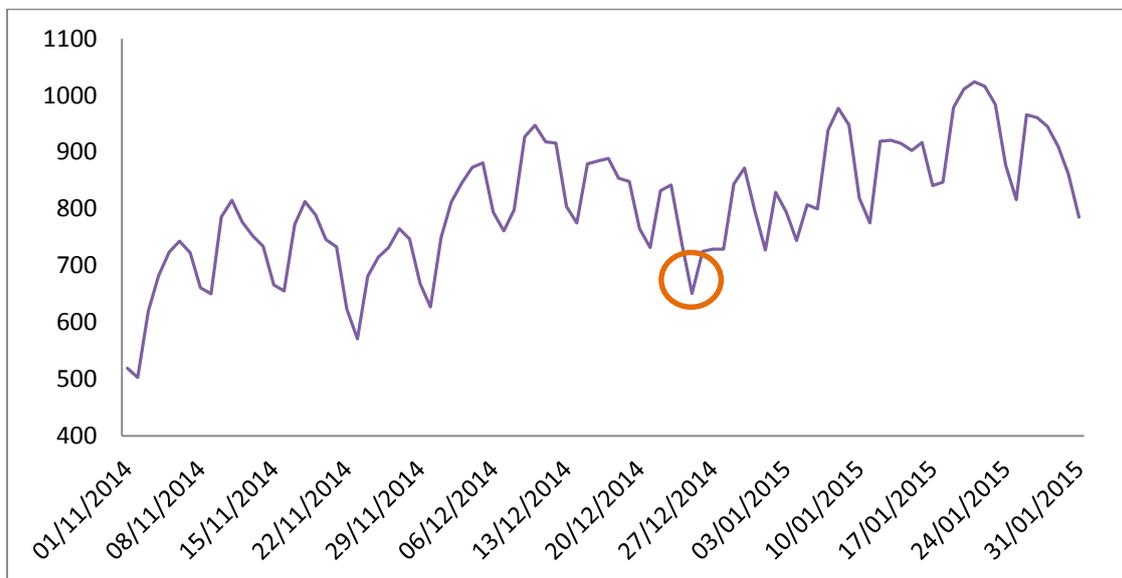


Figura 4.2.1 Esquema Variables Exógenas

### 4.2.1 Variables Dummy Laboralidad

VARIABLES que recogen diferencias en la demanda de gas agregada debido a diferencias en días laborables y a festivos. Dichas diferencias provocan un mayor o menor consumo de gas.

En la Gráfica 4.2.1.1 se aprecia el efecto de Día 25 de Diciembre (Navidad) y como afecta a la serie temporal produciendo una fuerte caída de demanda de gas.



Gráfica 4.2.1.1. Demanda de gas el día 25 de Diciembre de 2014

Se detalla que festivos se han incluido en la base de datos:

1. Festivos Nacionales: Se ha creado una variable Dummy que toma valor uno cada vez que hay un festivo nacional en el Calendario. También se han creado variables Dummies individuales para cada uno de los festivos nacionales.
2. Festivos Autonómicos: Se ha creado una variable Dummy para cada una de las Comunidades Autónomas, tomando valor uno aquellos días que hay fiesta en la Autonomía

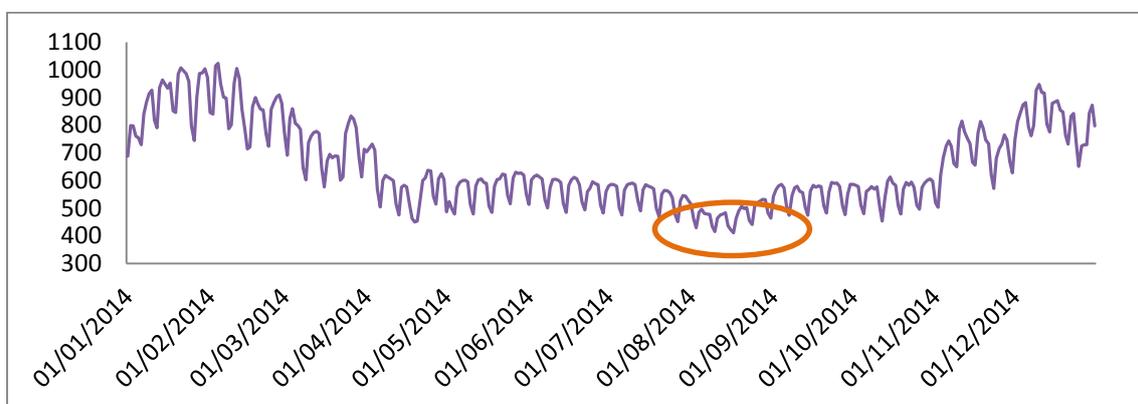


3. Fiestas de Semana Santa: Creación de una variable que toma valor uno para el periodo de Semana Santa, así como variables Dummy individuales para cada uno de los días.
4. Periodo Navidades: Variable Dummy que toma valor uno en el periodo vacacional de Navidades.
5. Post Festivo Nacional: Variable Dummy que toma valor uno siempre que le día anterior haya habido una fiesta nacional.
6. Pre Festivo Nacional: Variable Dummy que toma valor uno siempre que al día siguiente vaya a haber una fiesta nacional.
7. Puentes: Variable Dummy que toma valor uno siempre que se cree un puente con un Festivo Nacional. Como ejemplo, si un festivo nacional cae en jueves, la variable Dummy “Puentes” tomará valor 1 para el jueves, viernes, sábado y domingo.

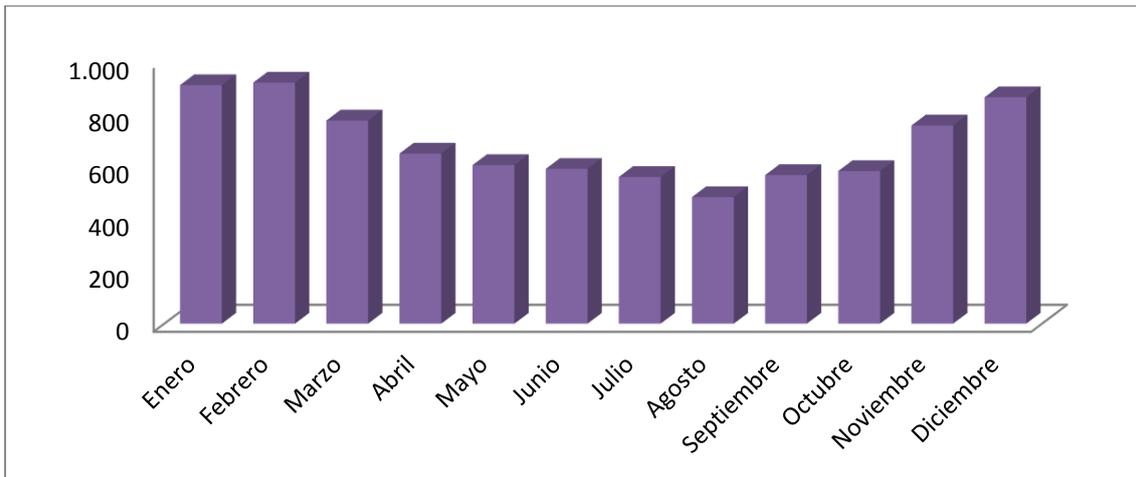
#### 4.2.2 Variables Dummy Estacionales

Variables que identifican los movimientos cíclicos de la serie temporal, pudiendo identificar, por ejemplo, como la demanda de gas es inferior los fines de semana en comparación con los días laborables.

En la Gráfica 4.2.2.1. y la Gráfica 4.2.2.2. se identifica como a lo largo de la serie histórica, el mes de Agosto es cuando menos consumo de gas se realiza, especialmente en la primera mitad del mes (coincidiendo con una de los mayores periodos vacacionales del año).



Gráfica 4.2.2.1 Detección demanda de gas en Agosto 2014



Gráfica 4.2.2.2 Demanda de gas promedio por meses

Se detalla que variables estacionales se han incluido en la base de datos:

1. Variables de Lunes a Domingo: Siete variables Dummy, cada una de ellas para uno de los días de la semana, tomando valor uno cada vez que la fecha se corresponda con el día de la semana de la variable.
2. Variables de Enero a Diciembre: Doce variables Dummy, cada una de ellas para uno de los meses, tomando valor uno cada vez que el mes de la fecha corresponda con el mes de la variable.
3. Fines de Semana para los distintos meses: Variables Dummy de sábado y domingo que tomarán valores uno dependiendo del mes de la fecha.

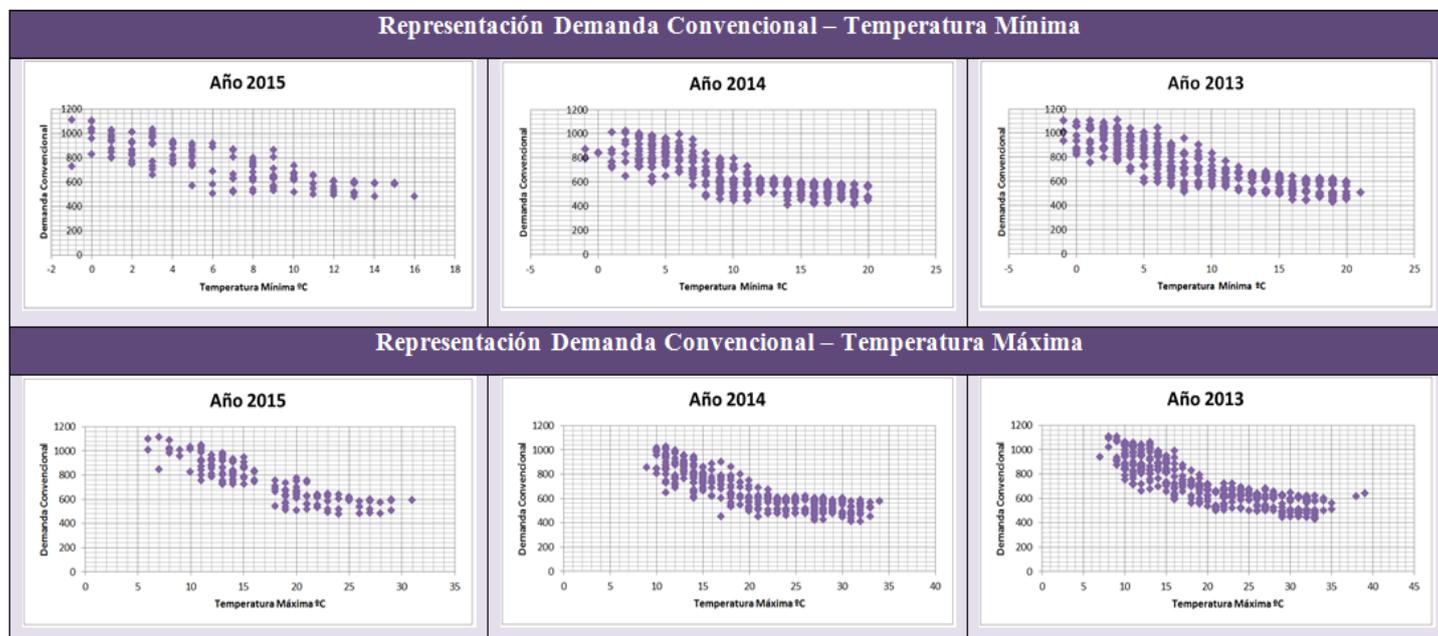
### 4.2.3 Variables Temperaturas

Los requerimientos de Gas de un país se encuentran fuertemente vinculados a factores climatológicos; los cuales, en muchos casos, explican los movimientos de la demanda de Gas.

En estudios anteriores también se ha optado por incluir en los modelos otras variables climatológicas (por ejemplo, la luminosidad, humedad, precipitaciones, etc), pero en este caso se optará solo por el uso de las temperaturas.



Se han obtenido datos de temperaturas mínimas y máximas a nivel nacional con frecuencia diario. Se exponen en las siguientes gráficas la relación entre el consumo de demanda de gas y las temperaturas por año de estudio.



Gráfica 4.2.3.1 Relación Demanda y Temperaturas

Visualizando las gráficas se obtiene el resultado de que la relación entre las temperaturas y la Demanda de Gas no es lineal, ya que en todos los casos se observa una cierta curvatura. La Tabla 4.2.3.1. muestra las temperaturas medias, así como el promedio de Demanda de Gas por mes:

Mes	Promedio Dem. Convencional	Promedio de tmin	Promedio de tmax
Enero	907,28	3,65	12,15
Febrero	916,60	3,49	11,98
Marzo	772,61	5,71	15,82
Abril	646,73	8,57	19,49
Mayo	603,11	10,46	22,22
Junio	588,80	14,27	26,57
Julio	557,90	17,92	30,40
Agosto	481,29	17,79	30,16
Septiembre	565,12	15,88	27,32
Octubre	580,44	12,65	23,05
Noviembre	753,67	7,28	15,42
Diciembre	861,03	3,02	12,16
<b>TOTAL</b>	<b>698,95</b>	<b>9,46</b>	<b>19,89</b>

Tabla 4.2.3.1. Relación del promedio de consumos con el de temperaturas máximas y mínimas



El promedio de Demanda de Gas es más elevado para aquellos meses que presentan menores temperaturas mínimas y máximas. Estos meses son aquellos que presentan mayor variabilidad en la Demanda de Gas (como se ha estudiado en el apartado 4.1.), por lo que se espera un mayor efecto de las temperaturas en las predicciones y por lo tanto que ayuden a mejorar los resultados.

Por lo tanto, a la vista de los resultados obtenidos, se va a optar por introducir las temperaturas transformadas mediante una regresión cuadrática. Dicha transformación se realiza con el fin de aumentar el efecto de las temperaturas en el modelo. Las nuevas funciones de Demanda de Gas en función de las temperaturas son las siguientes:

- **Función cuadrática de la Demanda en función de la temperatura mínima.**

$$\text{Nueva Demanda Gas}(tmin)_t = \beta_0 + \beta_1 tmin_t + \beta_2 tmin_t^2$$

- **Función cuadrática de la Demanda en función de la temperatura máxima.**

$$\text{Nueva Demanda Gas}(tmax)_t = \beta_0 + \beta_1 tmax_t + \beta_2 tmax_t^2$$

De esta forma, en el modelo no se introducirán las variables de temperatura mínima y máxima, si no las dos nuevas variables creadas. Esas variables son datos de Demanda de Gas para un momento t en función de las temperaturas de ese mismo momento t.

Las Tablas 4.2.3.2. y 4.2.3.3. muestran los resultados de las regresiones.

Parameter Estimates						
Variable	Etiqueta	DF	Parameter Estimate	Standard Error	Valor t	Pr >  t
Intercept	Intercept	1	1443.03015	22.71958	63.51	<.0001
tmax	tmax	1	-58.43204	2.37184	-24.64	<.0001
tmax_2		1	0.93631	0.05661	16.54	<.0001

Tabla 4.2.3.2. Regresión Demanda de Gas – Temperatura Máxima

Parameter Estimates						
Variable	Etiqueta	DF	Parameter Estimate	Standard Error	Valor t	Pr >  t



<b>Intercept</b>	Intercept	<b>1</b>	994.67778	8.78734	113.19	<.0001
<b>tmin</b>	tmin	<b>1</b>	-44.27777	2.04853	-21.61	<.0001
<b>tmin_2</b>		<b>1</b>	1.01258	0.10009	10.12	<.0001

Tabla 4.2.3.3. Regresión Demanda de Gas – Temperatura Mínima

Los resultados muestran como las variables de temperatura son significativas en ambas regresiones. Mediante esta metodología (regresión de la Demanda de Gas con respecto a las temperaturas) se espera obtener una predicción más ajustada.

## 5 Metodología Box-Jenkins

Se procede a realizar la predicción de Demanda de Gas y para ello se va a utilizar la metodología Box-Jenkins. El siguiente esquema reproduce todos los pasos que se van a seguir hasta obtener un resultado final:

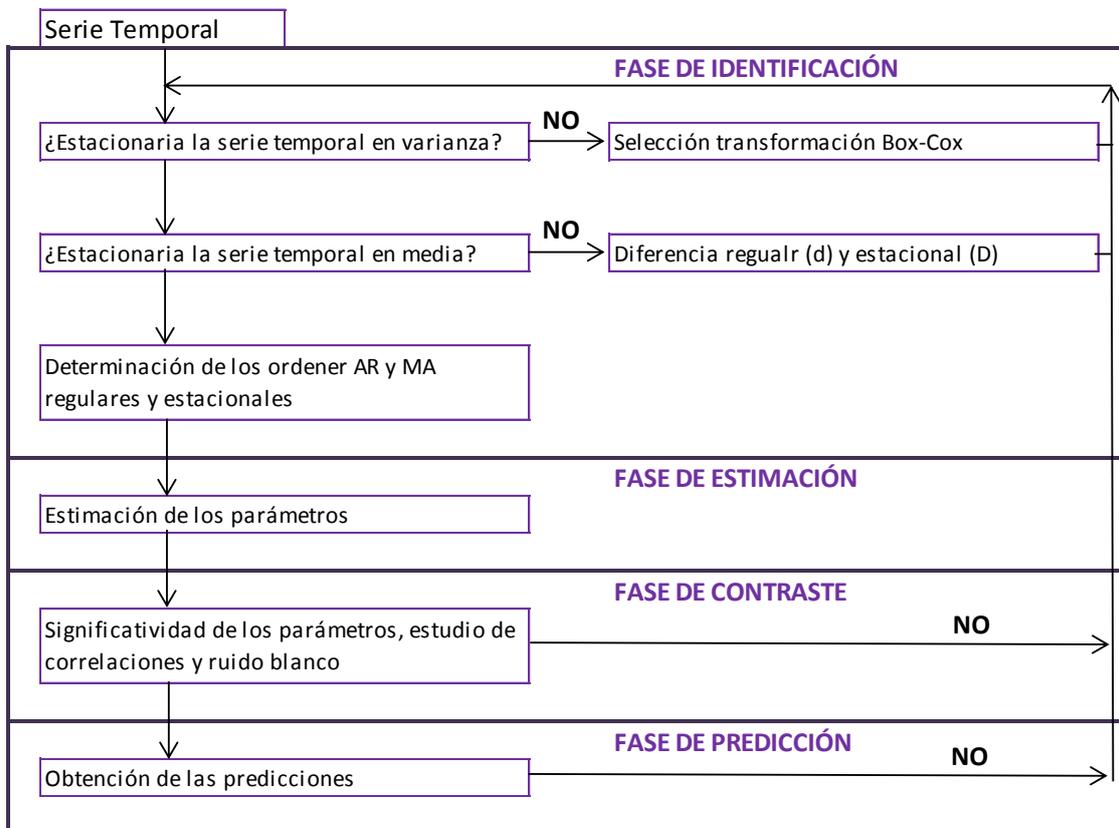


Figura 5.1. Esquema proceso metodología Box-Jenkins



## 5.1 Estacionariedad en Varianza

El primer paso en el proceso de predicción con Metodología Box-Jenkins es el de comprobar la estacionariedad en varianza, o lo que es lo mismo, contrastar si hay Heterocedasticidad en la serie temporal de estudio.

La palabra Heterocedasticidad surge de la unión de dos palabras. La palabra “hetero” que significa diferente, y de la palabra “cedasticidad” que significa dispersión. Es la existencia de una varianza no constante en las perturbaciones aleatorias de un modelo econométrico.

En un modelo de regresión lineal, uno de los supuestos para poder estimarlo, es que las perturbaciones del modelo tienen que ser homocedásticas, o lo que es lo mismo, la dispersión de los datos ha de ser constante a lo largo del tiempo.

La presencia de heterocedasticidad producirá estimadores sesgado y por lo tanto, una predicción incorrecta de la serie de Demanda de Gas.

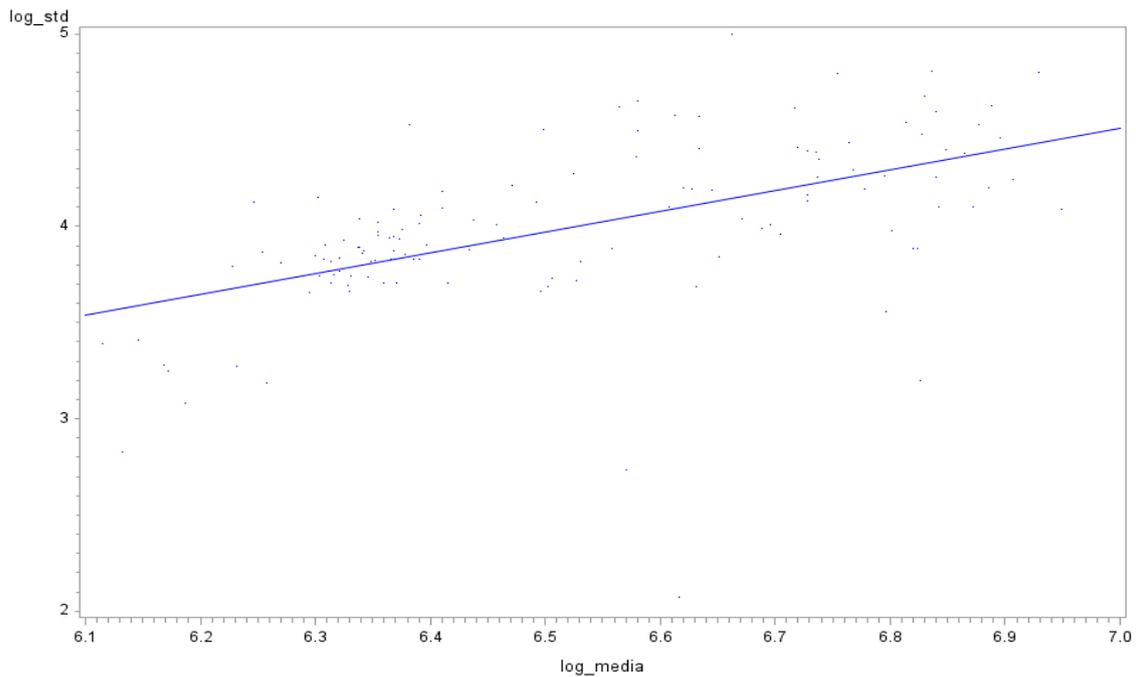
Para ello se utiliza un gráfica que muestre una medida de centralidad, por ejemplo la media de la Demanda Semanal, y una medida de dispersión, por ejemplo la desviación típica de la Demanda Semanal. Se obtiene la siguiente representación de la pendiente del modelo de regresión:

$$\log(Std) = \beta_0 + \beta_1 \log(media) + \varepsilon$$

Mediante dicha ecuación se determina la variabilidad de la serie temporal. La Tabla 5.1.1. y el Gráfico 5.1.1. muestran los resultados obtenidos:

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	Valor t	Pr >  t
Intercept	1	-3.01910	0.96499	-3.13	0.0022
log_media	1	1.07554	0.14780	7.28	<.0001

Tabla 5.1.1. Regresión para detección de no Estacionariedad en Varianza



Gráfica 5.1.1. Gráfica de la regresión entre la Desv. Típica y la Media

Se observa en los resultados que el parámetro del logaritmo de la media es significativo y muy próximo a 1, por lo que se recomienda transformar la serie antes de realizar la predicción.

## 5.2 Estacionariedad en Media

La aproximación a los modelos estocásticos con procedimientos ARIMA está restringida a aquellos que cumplan, al menos de forma débil, la condición de estacionariedad.

La no estacionariedad en media implica la presencia de “tendencia determinista” en la serie temporal. En muchos casos, es suficiente observar el gráfico de la serie para ver si su valor medio se mantiene constante, o por el contrario, si crece o decrece en el tiempo.

Visualizando la Gráfica 4.1.1.1., vemos como no se aprecia una tendencia fuerte en la serie. Igualmente para la detección de estacionariedad en media de forma más precisa se va a utilizar el test de Dickey-Fuller. Dicho test es un contraste de raíces unitarias que



permite evaluar la necesidad de diferenciar la serie temporal para obtener estacionariedad en media.

El contraste es el siguiente:

$$H_0 = |\phi| = 1 \text{ (raiz unitaria)}$$

$$H_1 = |\phi| < 1 \text{ (estacionario)}$$

Los resultados obtenidos con la serie original son los siguientes:

Pruebas aumentadas de la raíz unidad de Dickey-Fuller							
Tipo	Retardos	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-0.1374	0.6517	-0.35	0.5606		
	1	-0.2000	0.6374	-0.44	0.5224		
	7	-0.1382	0.6515	-0.66	0.4310		
Single Mean	0	-60.0828	0.0018	-5.52	<.0001	15.26	0.0010
	1	-85.2313	0.0018	-6.50	<.0001	21.16	0.0010
	7	-23.5633	0.0049	-3.49	0.0088	6.24	0.0053
Tendencia	0	-61.9093	0.0008	-5.62	<.0001	15.82	0.0010
	1	-87.5488	0.0008	-6.59	<.0001	21.71	0.0010
	7	-23.7408	0.0313	-3.47	0.0436	6.10	0.0585

Tabla 5.2.1. Test de Dickey-Fuller para la serie original en logaritmos

Los resultados muestran el estudio de la estacionariedad en media para los retardos 0, 1 y 7. En este caso nos fijamos en el retarde 0 y 7:

- **Estacionariedad en la Tendencia:** El estadístico es -5,62 para el retardo 0, mientras que para el retardo 7 es de -3,47, con p-valores muy pequeños en ambos casos, por lo que no se aprecia tendencia en la serie temporal.
- **Estacionariedad en Media:** El estadístico es -5,52 para el retardo 0, mientras que para el retardo 7 es de -3,49, con p-valores muy pequeños en ambos casos, por lo que se centra en un valor constante de la media
- **Estacionariedad en Media 0:** El estadístico es -0,35 para el retardo 0, mientras que para el retardo 7 es de -0,66. En este caso, el p-valor es de 0,56 y 0,52 respectivamente, por lo que la media de los valores de la serie temporal no se centra en torno al 0.



Debido a los resultados para Zero Mean, se puede optar por diferenciar la serie. Se muestran los resultados tanto con una diferencia regular (dif=1) en la Tabla 5.2.2., como para una diferencia estacional (dif=7) en la Tabla 5.2.3.

Pruebas aumentadas de la raíz unidad de Dickey-Fuller							
Tipo	Retardos	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-764.994	0.0001	-26.09	<.0001		
	1	-1798.65	0.0001	-30.00	<.0001		
	7	-1337.93	0.0001	-10.95	<.0001		
Single Mean	0	-765.021	0.0001	-26.08	<.0001	339.97	0.0010
	1	-1798.80	0.0001	-29.98	<.0001	449.44	0.0010
	7	-1337.25	0.0001	-10.96	<.0001	60.04	0.0010
Tendencia	0	-764.994	0.0001	-26.06	<.0001	339.58	0.0010
	1	-1798.81	0.0001	-29.96	<.0001	448.93	0.0010
	7	-1352.63	0.0001	-10.95	<.0001	60.00	0.0010

Tabla 5.2.2. Test de Dickey-Fuller para la diferencia regular del logaritmo de la serie

Pruebas aumentadas de la raíz unidad de Dickey-Fuller							
Tipo	Retardos	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-136.575	0.0001	-8.84	<.0001		
	1	-238.764	0.0001	-10.98	<.0001		
	7	-232.979	0.0001	-8.09	<.0001		
Single Mean	0	-136.850	0.0001	-8.85	<.0001	39.16	0.0010
	1	-239.155	0.0001	-10.98	<.0001	60.34	0.0010
	7	-234.998	0.0001	-8.11	<.0001	32.89	0.0010
Tendencia	0	-136.906	0.0001	-8.85	<.0001	39.20	0.0010
	1	-239.200	0.0001	-10.98	<.0001	60.30	0.0010
	7	-235.749	0.0001	-8.12	<.0001	32.97	0.0010

Tabla 5.2.3. Test de Dickey-Fuller para la diferencia estacional del logaritmo de la serie

En ambos casos se cumplen todos los criterios, por lo que la serie temporal es estacionaria en media.

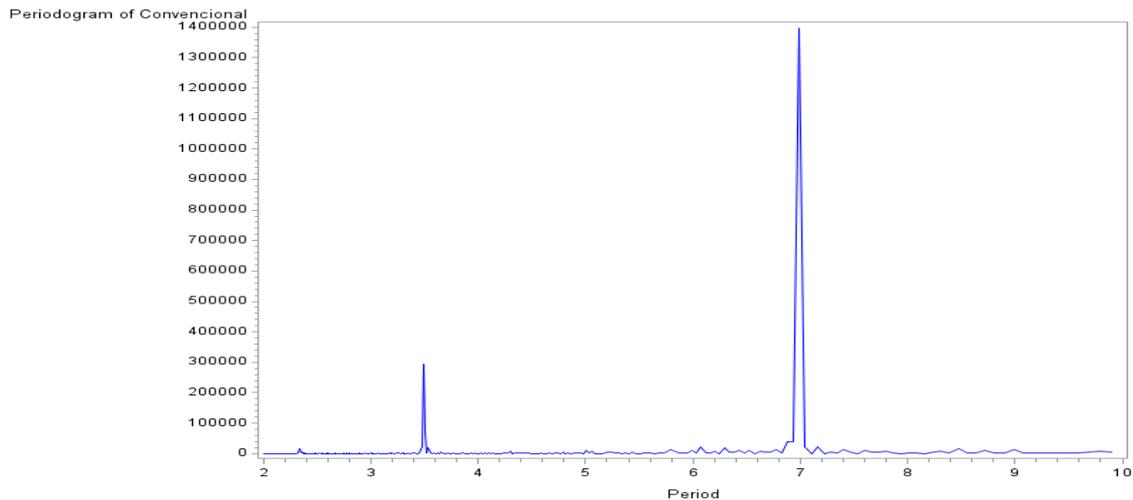
### 5.3 Detección de Estacionalidad

En apartados anteriores se ha comprobado como la serie temporal presenta distintos comportamientos en diferentes periodos temporales, especialmente a lo largo de los diferentes días de la semana.



En este caso, se va a utilizar como herramienta para la detección de la estacionalidad, el Periodograma.

Consiste en ajustar la serie temporal a través de una combinación de sinusoides que se mueven a distinta frecuencia y analizar la influencia que estas frecuencias tienen sobre la evolución de la serie.



Gráfica 5.3.1.Periodograma de la serie temporal de Demanda

El Periodograma marca una clara estacionalidad semanal en la serie temporal (claro repunte en el 7). La estacionalidad también se puede detectar mediante la función de autocorrelación y la función de autocorrelación parcial de la serie temporal (se podrá visualizar un puntos posteriores del trabajo).

## 5.4 Selección del mejor modelo ARIMA

A continuación se detallan, paso a paso, las modificaciones realizadas en la parte autorregresiva y media móvil hasta llegar al resultado óptima.



En primer lugar se grafica la función de autocorrelación (ACF) y autocorrelación parcial (PACF) de la serie en logaritmos para identificar que comportamiento presenta y de esta forma poder asignar un primer modelo.

Autocorrelaciones						
Retardo	Covarianza	Correlación	-1	9	8	7
0	0.055247	1.00000	*****			0
1	0.051353	0.92951	*****			0.033691
2	0.046478	0.84128	*****			0.055646
3	0.044552	0.80641	*****			0.068580
4	0.043806	0.79292	*****			0.078610
5	0.044224	0.80049	*****			0.087217
6	0.047296	0.85608	*****			0.095192
7	0.049628	0.89830	*****			0.103562
8	0.045870	0.83028	*****			0.112058
9	0.041523	0.75159	*****			0.118836
10	0.040060	0.72511	*****			0.124114
11	0.039626	0.71725	*****			0.128833
12	0.040253	0.72860	*****			0.133289
13	0.043512	0.78760	*****			0.137735
14	0.046032	0.83320	*****			0.142756
15	0.042555	0.77027	*****			0.148173
16	0.038474	0.69640	*****			0.152650
17	0.037154	0.67251	*****			0.156215
18	0.036777	0.66569	*****			0.159467
19	0.037335	0.67579	*****			0.162591
20	0.040469	0.73251	*****			0.165748
21	0.042887	0.77628	*****			0.169383
22	0.039328	0.71185	*****			0.173374
23	0.035114	0.63558	*****			0.176661
24	0.033605	0.60828	*****			0.179237

Tabla 5.4.1 Autocorrelación de la serie

Autocorrelaciones parciales						
Retardo	Correlación	-1	9	8	7	6
1	0.92951	*****				
2	-0.16698	***				
3	0.37109	*****				
4	0.02105	.				
5	0.31233	*****				
6	0.47353	*****				
7	0.10010	**				
8	-0.62332	*****				
9	0.20822	****				
10	0.01613	.				
11	0.06000	*				
12	0.09367	**				
13	0.19873	****				
14	0.01676	.				
15	-0.29799	*****				
16	0.10172	**				
17	-0.03766	*				
18	0.04521	*				
19	-0.00632	.				
20	0.12268	**				
21	-0.00497	.				
22	-0.22758	*****				
23	0.02880	*				
24	-0.05507	*				

Tabla 5.4.2. Autocorrelación parcial de la serie

La gráfica de la ACF muestra como la serie temporal no es estacionaria en media, ya que va descendiendo a 0 de forma lenta, mientras que debería hacer rápidamente.

Por lo visto en los resultados, lo mejor es comenzar con un modelo AR (1).

### Modelo AR(1) + Constante

Los resultados muestran que tanto el parámetro del AR(1) como la variable constante son estadísticamente significativos.

Estimación por mínimos cuadrados condicional					
Parámetro	Estimador	Error estándar	Valor t	Approx Pr >  t	Retardo
MU	6.52646	0.03813	171.16	<.0001	0
AR1,1	0.93178	0.01236	75.39	<.0001	1

Tabla 5.4.3. Estimación parámetros

Se aprecia como la correlación entre ambas parámetros es baja.



Correlaciones de los estimadores de parámetro		
Parámetro	MU	ARI,1
MU	1.000	0.015
ARI,1	0.015	1.000

Tabla 5.4.4. Correlación parámetros

Todavía no se ha obtenido ruido blanco en las correlaciones de los residuos.

Comprobación de los residuales de autocorrelación										
Para retardo	Chi-cuadrado	DF	Pr > ChiSq	Autocorrelaciones						
6	242.93	5	<.0001	0.160	-0.335	-0.120	-0.118	-0.306	0.122	
12	1011.10	11	<.0001	0.791	0.108	-0.329	-0.103	-0.109	-0.307	
18	1681.12	17	<.0001	0.118	0.777	0.106	-0.317	-0.095	-0.093	
24	2423.70	23	<.0001	-0.300	0.115	0.773	0.111	-0.312	-0.099	
30	3113.85	29	<.0001	-0.092	-0.300	0.085	0.740	0.111	-0.302	
36	3751.48	35	<.0001	-0.085	-0.100	-0.302	0.092	0.751	0.111	
42	4475.40	41	<.0001	-0.323	-0.106	-0.105	-0.302	0.102	0.744	
48	4690.76	47	<.0001	0.116	-0.305	-0.096	-0.109	-0.311	0.083	

Tabla 5.4.5. Correlación de los residuos

Representación de la ACF y PACF:

Trazado de residuales de autocorrelación					
Retardo	Covarianza	Correlación	-198765432101234567891	Error std	
0	0.0074172	1.00000	*****	0	
1	0.0011833	0.15954	.***	0.033691	
2	-0.0024844	-0.33495	*****	0.034538	
3	-0.0008879	-0.11971	**.	0.038047	
4	-0.0008754	-0.11802	**.	0.038472	
5	-0.0022664	-0.30556	*****	0.038881	
6	0.00090300	0.12174	**.	0.041517	
7	0.0058667	0.79097	*****	0.041920	
8	0.00079743	0.10751	**.	0.056370	
9	-0.0024412	-0.32913	*****	0.056602	
10	-0.0007657	-0.10323	**.	0.058734	
11	-0.0008065	-0.10874	**.	0.058940	
12	-0.0022758	-0.30683	*****	0.059167	
13	0.00087826	0.11841	**.	0.060947	
14	0.0057604	0.77663	*****	0.061207	
15	0.00078635	0.10602	**.	0.071523	
16	-0.0023494	-0.31675	*****	0.071701	
17	-0.0007062	-0.09522	**.	0.073273	
18	-0.0006888	-0.09286	**.	0.073413	
19	-0.0022222	-0.29961	*****	0.073546	
20	0.00085108	0.11474	**.	0.074919	
21	0.0057363	0.77339	*****	0.075118	
22	0.00082383	0.11107	**.	0.083669	
23	-0.0023177	-0.31247	*****	0.083836	
24	-0.0007362	-0.09926	**.	0.085148	

Tabla 5.4.6. Autocorrelación de los residuos

Autocorrelaciones parciales												
Retardo	Correlación	-198765432101234567891										
1	0.15954	.***										
2	-0.36982	*****										
3	0.01485	.										
4	-0.25982	*****										
5	-0.34299	*****										
6	0.13722	.***										
7	0.69725	*****										
8	-0.18883	*****										
9	0.03864	.*										
10	-0.00844	.										
11	-0.03373	.*										
12	-0.11087	**.										
13	0.09192	.***										
14	0.35065	*****										
15	-0.07905	**.										
16	0.07222	.*										
17	-0.01238	.										
18	0.04930	.*										
19	-0.07048	.*										
20	0.06733	.*										
21	0.25539	*****										
22	-0.02391	.										
23	0.05538	.*										
24	-0.02978	.*										

Tabla 5.4.7. Autocorrelación Parcial de los Residuos

El parámetro del AR(1) está cerca de 1, por lo que se podría realizar una diferencia regular, pero de momento se continua con la misma ejecución.

El gráfico de la ACF corresponde con un modelo AR(7). Se aprecia estructura multiplicativa.



### Modelo SAR(1)x(1)<sub>7</sub> + Constante

El estimador de todos los parámetros es significativo. El estimador de la parte AR(1) ha descendido su valor por la introducción del autorregresivo estacional.

Estimación por mínimos cuadrados condicional					
Parámetro	Estimador	Error estándar	Valor t	Approx Pr >  t	Retardo
MU	6.53725	0.03571	183.06	<.0001	0
AR1,1	0.86790	0.01702	50.99	<.0001	1
AR2,1	0.80762	0.02039	39.61	<.0001	7

Tabla 5.4.8. Estimación parámetros

Las correlaciones entre los tres parámetros continúan siendo muy bajas.

Correlaciones de los estimadores de parámetro			
Parámetro	MU	AR1,1	AR2,1
MU	1.000	0.040	0.013
AR1,1	0.040	1.000	-0.159
AR2,1	0.013	-0.159	1.000

Tabla 5.4.9. Correlación parámetros

No hay ruido blanco en la correlación de los residuos.

Comprobación de los residuales de autocorrelación									
Para retardo	Chi-cuadrado	DF	Pr > ChiSq	Autocorrelaciones					
6	61.05	4	<.0001	0.241	-0.070	-0.021	-0.005	0.054	-0.049
12	166.59	10	<.0001	-0.325	-0.094	-0.046	0.010	-0.019	-0.042
18	171.97	16	<.0001	0.020	0.064	0.006	-0.008	0.024	0.029
24	199.89	22	<.0001	-0.007	0.075	0.148	0.029	-0.038	-0.031
30	205.49	28	<.0001	0.020	-0.016	-0.058	-0.012	0.013	0.043
36	223.56	34	<.0001	0.061	-0.014	-0.030	0.008	0.121	0.015
42	260.31	40	<.0001	-0.105	-0.071	-0.009	-0.006	0.076	0.134
48	268.36	46	<.0001	0.057	0.023	-0.003	-0.036	-0.053	-0.027

Tabla 5.4.10. Correlación de los residuos



Trazado de residuales de autocorrelación					
Retardo	Covarianza	Correlación	-1	9	Error std
0	0.0027005	1.00000			0
1	0.00065187	0.24139	*****		0.033691
2	-0.0001889	-0.06995	*.		0.035600
3	-0.0000578	-0.02140	. .		0.035756
4	-0.0000125	-0.00463	. .		0.035770
5	0.00014647	0.05424	. .*		0.035771
6	-0.0001333	-0.04936	*.		0.035864
7	-0.0008764	-0.32454	***** .		0.035941
8	-0.0002551	-0.09446	** .		0.039126
9	-0.0001238	-0.04583	*.		0.039384
10	0.00002582	0.00956	. .		0.039445
11	-0.0000518	-0.01917	. .		0.039447
12	-0.0001130	-0.04185	*.		0.039458
13	0.00005475	0.02027	. .		0.039508
14	0.00017172	0.06359	. .*		0.039520
15	0.00001716	0.00636	. .		0.039636
16	-0.0000213	-0.00790	. .		0.039637
17	0.00006401	0.02370	. .		0.039639
18	0.00007928	0.02936	. .*		0.039655
19	-0.0000185	-0.00684	. .		0.039680
20	0.00020142	0.07459	. .*		0.039681
21	0.00040100	0.14849	. ***		0.039840
22	0.00007708	0.02854	. .*		0.040463
23	-0.0001017	-0.03767	*.		0.040486
24	-0.0000841	-0.03116	*.		0.040526

Tabla 5.4.11. Autocorrelación de los residuos

Retardo	Correlación	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9		
1	0.24139																						
2	-0.13616																						
3	0.03316																						
4	-0.01823																						
5	0.06440																						
6	-0.08958																						
7	-0.30163																						
8	0.06611																						
9	-0.11514																						
10	0.05260																						
11	-0.06599																						
12	0.01551																						
13	0.00485																						
14	-0.05287																						
15	-0.00126																						
16	-0.05975																						
17	0.06782																						
18	-0.03425																						
19	-0.01215																						
20	0.11168																						
21	0.12486																						
22	-0.03384																						
23	-0.03682																						
24	0.01765																						

Tabla 5.4.12. Autocorrelación Parcial de los residuos

A la vista de la representación de la PACF, se necesita introducir un parámetro de media móvil de orden 7, ya que claramente el lag 7 es significativo.

**Modelo SARIMA(1,0,0)x(1,0,1)<sub>7</sub> + Constante**

La inclusión de la parte media móvil estacional no afecta a la significatividad del resto de parámetros. Es destacable que el estimador del AR(7) toma un valor muy cercano a uno.

Estimación por mínimos cuadrados condicional					
Parámetro	Estimador	Error estándar	Valor t	Approx Pr >  t	Retardo
MU	6.56444	0.04181	156.99	<.0001	0
MA1,1	0.80572	0.02258	35.68	<.0001	7
AR1,1	0.95654	0.01053	90.86	<.0001	1
AR2,1	0.99826	0.0040605	245.85	<.0001	7

Tabla 5.4.13. Estimación parámetros

Sigue sin haber correlaciones altas. Tanto el autorregresivo regular como el estacional presentan una correlación muy similar con la media móvil.

Correlaciones de los estimadores de parámetro				
Parámetro	MU	MA1,1	AR1,1	AR2,1
MU	1.000	-0.012	-0.014	-0.001
MA1,1	-0.012	1.000	0.342	0.332
AR1,1	-0.014	0.342	1.000	0.075
AR2,1	-0.001	0.332	0.075	1.000

Tabla 5.4.14. Correlación parámetros



No hay ruido blanco en la correlación de los residuos.

Comprobación de los residuales de autocorrelación									
Para retardo	Chi-cuadrado	DF	Pr > ChiSq	Autocorrelaciones					
6	62.62	3	<.0001	0.197	-0.123	-0.092	-0.076	0.029	0.039
12	72.76	9	<.0001	0.020	-0.042	-0.090	-0.007	-0.026	0.018
18	79.78	15	<.0001	0.047	0.004	-0.039	-0.034	0.030	0.045
24	84.53	21	<.0001	0.042	0.046	0.031	-0.011	-0.014	0.013
30	100.99	27	<.0001	0.054	0.042	-0.073	-0.079	0.002	0.043
36	106.72	33	<.0001	0.061	0.006	0.021	-0.026	0.026	0.027
42	112.96	39	<.0001	-0.052	-0.037	-0.014	0.018	0.035	0.031
48	120.38	45	<.0001	0.060	0.037	0.020	-0.026	-0.025	-0.036

Tabla 5.4.15. Correlación de los residuos

Representación de la ACF y de la PACF.

Trazado de residuales de autocorrelación					
Retardo	Covarianza	Correlación	-198765432101234567891	Error std	
0	0.0019163	1.00000	*****	0	
1	0.00057819	0.19736	.****	0.033691	
2	-0.0002366	-.12347	**.	0.034979	
3	-0.0001755	-.09157	**.	0.035470	
4	-0.0001462	-.07629	**.	0.035737	
5	0.00005605	0.02925	.*	0.035921	
6	0.00007450	0.03888	.*	0.035948	
7	0.00003823	0.01995	. .	0.035996	
8	-0.0000804	-.04195	* .	0.036009	
9	-0.0001726	-.09008	** .	0.036064	
10	-0.0000139	-.00726	. .	0.036319	
11	-0.0000507	-.02648	* .	0.036320	
12	0.00003521	0.01837	. .	0.036342	
13	0.00008995	0.04694	.*	0.036353	
14	7.31492E-6	0.00382	. .	0.036421	
15	-0.0000745	-.03888	* .	0.036422	
16	-0.0000653	-.03406	* .	0.036469	
17	0.00005717	0.02983	.*	0.036505	
18	0.00008647	0.04512	.*	0.036533	
19	0.00007992	0.04171	.*	0.036596	
20	0.00008732	0.04557	.*	0.036650	
21	0.00005895	0.03076	.*	0.036714	
22	-0.0000213	-.01112	. .	0.036743	
23	-0.0000276	-.01439	. .	0.036747	
24	0.00002437	0.01272	. .	0.036754	

Tabla 5.4.16. Autocorrelación de los residuos

Autocorrelaciones parciales											
Retardo	Correlación	-198765432101234567891									
1	0.19736	****									
2	-0.16900	** .									
3	-0.03185	* .									
4	-0.07434	* .									
5	0.04654	.*									
6	-0.00133	. .									
7	0.01506	. .									
8	-0.04868	* .									
9	-0.06389	* .									
10	0.01804	. .									
11	-0.05848	* .									
12	0.02628	.*									
13	0.01872	. .									
14	-0.00171	. .									
15	-0.03414	* .									
16	-0.01091	. .									
17	0.03125	.*									
18	0.01653	. .									
19	0.03116	.*									
20	0.03690	.*									
21	0.04158	.*									
22	-0.00572	. .									
23	0.00423	. .									
24	0.01626	. .									

Tabla 5.4.17. Autocorrelación Parcial de los residuos

El modelo presentado es inestable ya que se viola la condición de estacionariedad. Tal y como se ha visto en el estudio de la serie. La ACF de la serie en logaritmos también nos ha mostrado que la serie no es estacionaria sin diferenciar.

El polinomio AR (7) presenta una raíz unitaria (su estimador toma un valor muy cercano a 1) que justifica reemplazarlo por una diferencia estacional.

### Modelo SARIMA (1,0,0)x(0,1,1)<sub>7</sub> +Constante

Mediante la modificación de los parámetros, desaparece la parte AR(7), al introducir la diferencia estacional se aprecia como la variable constante deja de ser significativa.



Estimación por mínimos cuadrados condicional					
Parámetro	Estimador	Error estándar	Valor t	Approx Pr >  t	Retardo
MU	0.0096894	0.0074540	1.30	0.1940	0
MA1,1	0.67215	0.02640	25.46	<.0001	7
AR1,1	0.92891	0.01385	67.08	<.0001	1

Tabla 5.4.18. Estimación parámetros

Se mantienen bajos niveles de correlación entre los parámetros

Correlaciones de los estimadores de parámetro			
Parámetro	MU	MA1,1	AR1,1
MU	1.000	0.022	0.299
MA1,1	0.022	1.000	0.300
AR1,1	0.299	0.300	1.000

Tabla 5.4.19. Correlación parámetros

Todavía no se presenta ruido blando en las correlaciones de los residuos.

Comprobación de los residuales de autocorrelación									
Para retardo	Chi-cuadrado	DF	Pr > ChiSq	Autocorrelaciones					
6	29.18	4	<.0001	0.138	-0.074	-0.053	-0.058	0.047	-0.018
12	36.82	10	<.0001	0.005	-0.073	-0.045	0.010	-0.020	0.028
18	44.43	16	0.0002	0.014	0.003	-0.059	0.006	0.045	0.052
24	48.47	22	0.0009	0.042	0.028	0.026	-0.028	0.016	0.015
30	67.92	28	<.0001	0.074	0.034	-0.067	-0.077	-0.008	0.065
36	70.97	34	0.0002	0.046	0.016	0.016	-0.012	0.023	-0.007
42	77.84	40	0.0003	-0.032	-0.043	-0.003	0.009	0.043	0.051
48	84.60	46	0.0005	0.048	0.059	0.009	-0.015	-0.028	-0.023

Tabla 5.4.20. Correlación de los residuos

Representación de la ACF y de la PACF

Trazado de residuales de autocorrelación					
Retardo	Covarianza	Correlación	-198765432101234567891	Error std	
0	0.0021455	1.00000	*****	0	
1	0.00029591	0.13792	***	0.033826	
2	-0.0001596	-0.07438	*	0.034463	
3	-0.0001135	-0.05292	*	0.034646	
4	-0.0001234	-0.05753	*	0.034738	
5	0.00010153	0.04732	*	0.034847	
6	-0.0000393	-0.1831	.	0.034921	
7	0.00001101	0.00513	.	0.034932	
8	-0.0001559	-0.07266	*	0.034933	
9	-0.0000972	-0.04531	*	0.035105	
10	0.00002137	0.00996	.	0.035172	
11	-0.0000418	-0.01950	.	0.035175	
12	0.00006055	0.02822	*	0.035188	
13	0.00003045	0.01419	.	0.035213	
14	5.39757E-6	0.00252	.	0.035220	
15	-0.0001271	-0.05924	*	0.035220	
16	0.00001189	0.00554	.	0.035334	
17	0.00009623	0.04485	*	0.035335	
18	0.00011263	0.05249	*	0.035400	
19	0.00008939	0.04166	*	0.035489	
20	0.00006016	0.02804	*	0.035545	
21	0.00005679	0.02647	*	0.035570	
22	-0.0000605	-0.02821	*	0.035593	
23	0.00003424	0.01596	.	0.035618	
24	0.00003274	0.01526	.	0.035627	

Tabla 5.4.21. Autocorrelación de los residuos

Autocorrelaciones parciales									
Retardo	Correlación	-198765432101234567891							
1	0.13792	.	***						
2	-0.09522	**	.						
3	-0.02937	*	.						
4	-0.05394	*	.						
5	0.05839	*	.						
6	-0.04597	*	.						
7	0.01998	.	.						
8	-0.08435	**	.						
9	-0.01614	.	.						
10	-0.00035	.	.						
11	-0.02703	*	.						
12	0.02316	.	.						
13	0.00817	.	.						
14	0.00112	.	.						
15	-0.06343	*	.						
16	0.02833	*	.						
17	0.02186	.	.						
18	0.04632	*	.						
19	0.02348	.	.						
20	0.04119	*	.						
21	0.02768	*	.						
22	-0.02446	.	.						
23	0.02612	*	.						
24	0.00764	.	.						

Tabla 5.4.22. Autocorrelación Parcial de los residuos



A la vista de los resultados, se decide eliminar la constante del análisis. Del mismo modo estudiando la ACF y la PACF se aprecia necesaria la introducción de un nuevo parámetro de media móvil. MA(1).

**Modelo SARIMA (1,0,1)x(0,1,1)<sub>7</sub>**

En la estimación por mínimos cuadrados ya no se encuentra la constante, y se aprecia como los tres parámetros son significativos.

Estimación por mínimos cuadrados condicional					
Parámetro	Estimador	Error estándar	Valor t	Approx Pr >  t	Retardo
MA1,1	-0.20154	0.03652	-5.52	<.0001	1
MA2,1	0.62873	0.02781	22.61	<.0001	7
AR1,1	0.88101	0.01839	47.90	<.0001	1

Tabla 5.4.23. Estimación parámetros

No existe una correlación alta entre los parámetros.

Correlaciones de los estimadores de parámetro			
Parámetro	MA1,1	MA2,1	AR1,1
MA1,1	1.000	0.149	0.417
MA2,1	0.149	1.000	0.316
AR1,1	0.417	0.316	1.000

Tabla 5.4.24. Correlación parámetros

En este caso, y mediante la introducción del último parámetro MA(1), si hay presencia de ruido blando en los residuos.

Comprobación de los residuales de autocorrelación									
Para retardo	Chi-cuadrado	DF	Pr > ChiSq	Autocorrelaciones					
6	8.52	3	0.0363	-0.006	-0.029	-0.000	-0.038	0.085	-0.012
12	13.69	9	0.1339	0.004	-0.047	-0.022	0.035	-0.017	0.040
18	19.90	15	0.1757	0.020	-0.000	-0.045	0.019	0.043	0.047
24	23.53	21	0.3162	0.035	0.029	0.024	-0.024	0.028	0.003
30	39.08	27	0.0623	0.071	0.034	-0.052	-0.068	0.002	0.061
36	41.53	33	0.1464	0.036	0.006	0.020	-0.015	0.027	-0.003
42	46.30	39	0.1964	-0.024	-0.038	0.004	0.003	0.039	0.040
48	50.52	45	0.2647	0.034	0.052	0.002	-0.010	-0.024	-0.001

Tabla 5.4.25. Correlación de los residuos



Representación de la ACF y de la PACF

Trazado de residuales de autocorrelación				
Retardo	Covarianza	Correlación	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1	Error std
0	0.0020883	1.00000	*****	0
1	-0.0000133	-0.00635	. .	0.033826
2	-0.0000614	-.02940	* .	0.033827
3	-4.1699E-7	-0.00020	. .	0.033856
4	-0.0000796	-.03810	* .	0.033856
5	0.00017708	0.08480	. .**	0.033905
6	-0.0000242	-.01157	. .	0.034147
7	8.42932E-6	0.00404	. .	0.034151
8	-0.0000982	-.04703	* .	0.034152
9	-0.0000467	-.02238	. .	0.034226
10	0.00007299	0.03495	. .*	0.034243
11	-0.0000362	-.01736	. .	0.034283
12	0.00008289	0.03969	. .*	0.034294
13	0.00004146	0.01985	. .	0.034346
14	-4.2789E-7	-0.00020	. .	0.034359
15	-0.0000947	-.04536	* .	0.034359
16	0.00004066	0.01947	. .	0.034428
17	0.00009052	0.04334	. .*	0.034440
18	0.00009907	0.04744	. .*	0.034503
19	0.00007403	0.03545	. .*	0.034577
20	0.00005993	0.02870	. .*	0.034619
21	0.00005076	0.02431	. .	0.034646
22	-0.0000499	-.02392	. .	0.034665
23	0.00005892	0.02821	. .*	0.034684
24	6.21407E-6	0.00298	. .	0.034711

Tabla 5.4.26. Autocorrelación de los residuos

Autocorrelaciones parciales				
Retardo	Correlación	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1		
1	-0.00635	. .		
2	-0.02944	* .		
3	-0.00058	. .		
4	-0.03900	* .		
5	0.08447	. .**		
6	-0.01317	. .		
7	0.00919	. .		
8	-0.04986	* .		
9	-0.01580	. .		
10	0.02394	. .		
11	-0.01564	. .		
12	0.03713	. .*		
13	0.02581	. .*		
14	0.00678	. .		
15	-0.05116	* .		
16	0.02393	. .		
17	0.03375	. .*		
18	0.05046	. .*		
19	0.03411	. .*		
20	0.04413	. .*		
21	0.02974	. .*		
22	-0.02580	* .		
23	0.02066	. .		
24	-0.00133	. .		

Tabla 5.4.27. Autocorrelación parcial de los residuos

Se ha creado un modelo con diferencia estacional, al que se ha incorporado una parte regular autorregresiva y media móvil y una parte estacional media móvil.

Se ha conseguido llegar a un modelo que presentan ruido blanco en los residuos y que por lo tanto es válido para hacer la predicción.

### 5.5 Inclusión de variables Dummy de laboralidad o calendario

Seleccionado los parámetros del modelo ARIMA, se procede a la inclusión de variables externas Dummies de laboralidad.

El modelo desarrollado contiene una parte media móvil estacional por lo que se opta por descartar las variables Dummies estacionales y el estudio se centrará en la incorporación de variables de laboralidad.



Durante este proceso se han ido incluyendo variables externas, eliminando aquellas que no fuesen significativas en la predicción de la Demanda de Gas. El resultado final es el siguiente:

Parámetro	Estimador	Error Aprox Estándar	Valor t	Pr >  t	Retardo	Variable
MA1,1	-0.31128	0.03461	-8.99	<.0001	1	log_Convencional
MA2,1	0.81414	0.02245	36.27	<.0001	7	log_Convencional
AR1,1	0.96023	0.01056	90.96	<.0001	1	log_Convencional
NUM1	-0.26232	0.01204	-21.79	<.0001	0	Año_nuevo
NUM2	-0.08615	0.01091	-7.89	<.0001	0	Epifania
NUM3	-0.15609	0.01068	-14.62	<.0001	0	Trabajo
NUM4	-0.08072	0.01297	-6.22	<.0001	0	Asuncion_virgen
NUM5	-0.09090	0.01778	-5.11	<.0001	0	Santos
NUM6	-0.08529	0.01778	-4.80	<.0001	0	Constitucion
NUM7	-0.12162	0.01825	-6.66	<.0001	0	Inmaculada
NUM8	-0.23482	0.01766	-13.30	<.0001	0	Navidad
NUM9	-0.13443	0.01715	-7.84	<.0001	0	Fiesta_24dic
NUM10	-0.13199	0.01455	-9.07	<.0001	0	Fiesta_31dic
NUM11	-0.11141	0.01585	-7.03	<.0001	0	Jueves_Santo
NUM12	-0.18282	0.02338	-7.82	<.0001	0	Viernes_Santo
NUM13	-0.12023	0.02539	-4.73	<.0001	0	Sabado_Santo
NUM14	-0.09013	0.02380	-3.79	0.0002	0	Domingo_Santo
NUM15	-0.11412	0.01996	-5.72	<.0001	0	Lunes_Pascua
NUM16	-0.03037	0.01371	-2.22	0.0270	0	escalonNavidad
NUM17	-0.02214	0.0064865	-3.41	0.0007	0	Pos_festivo
NUM18	-0.06293	0.0097090	-6.48	<.0001	0	Puente

Tabla 5.5.1. Estimación de los parámetros ARIMA + Var. Dummy

Los resultados muestran la introducción de los tres parámetros del modelo ARIMA, así como la inclusión de otras 18 variables exógenas.



A la vista de los resultados, se obtienen las siguientes conclusiones:

- Se aprecia como el P-valor es muy pequeño en todos los casos por lo que son significativas.
- El estimador de todas las variables exógenas es negativo. Esto es debido a que al tratarse de festivos, esos días la Demanda de Gas es inferior a la normal, por lo que los resultados tienen sentido.
- Finalmente han entrado en el modelo los festivos nacionales en forma de pulsos individuales.
- Especial efecto en la época de Navidades y en Semana Santa.
- Se introduce en el modelo la variable “Pos-Festivo”, esperando que la Demanda de Gas baja del día anterior (debido al festivo nacional) afecte a la del siguiente día, así como la variable “Puente” que detecta un consumo bajo durante la unión de un festivo nacional con un fin de semana.

A continuación se introducen las dos variables de temperatura transformadas, respetando las variables exógenas que ya se han introducido.

## 5.6 Inclusión de variables de Temperatura

Los resultados de la regresión introduciendo las temperaturas son los siguientes:

Parámetro	Estimador	Error Aprox Estándar	Valor t	Pr >  t	Retardo	Variable
<b>MA1,1</b>	-0.12314	0.03680	-3.35	0.0009	1	log_Convencional
<b>MA2,1</b>	0.73757	0.02566	28.75	<.0001	7	log_Convencional
<b>AR1,1</b>	0.94890	0.01207	78.62	<.0001	1	log_Convencional
<b>NUM1</b>	-0.24870	0.01091	-22.81	<.0001	0	Anyo_nuevo
<b>NUM2</b>	-0.09605	0.01017	-9.45	<.0001	0	Epifania
<b>NUM3</b>	-0.15608	0.0098012	-15.92	<.0001	0	Trabajo
<b>NUM4</b>	-0.08442	0.01195	-7.06	<.0001	0	Asuncion_virgen
<b>NUM5</b>	-0.11124	0.01656	-6.72	<.0001	0	Santos
<b>NUM6</b>	-0.09722	0.01653	-5.88	<.0001	0	Constitucion
<b>NUM7</b>	-0.10779	0.01684	-6.40	<.0001	0	Inmaculada
<b>NUM8</b>	-0.23469	0.01485	-15.81	<.0001	0	Navidad
<b>NUM9</b>	-0.11683	0.01450	-8.06	<.0001	0	Fiesta_24dic
<b>NUM10</b>	-0.12098	0.01326	-9.12	<.0001	0	Fiesta_31dic
<b>NUM11</b>	-0.11783	0.01316	-8.95	<.0001	0	Jueves_Santo



<b>NUM12</b>	-0.17928	0.01787	-10.03	<.0001	0	Viernes_Santo
<b>NUM13</b>	-0.11644	0.01899	-6.13	<.0001	0	Sabado_Santo
<b>NUM14</b>	-0.09934	0.01802	-5.51	<.0001	0	Domingo_Santo
<b>NUM15</b>	-0.11556	0.01680	-6.88	<.0001	0	Lunes_Pascua
<b>NUM16</b>	-0.04097	0.01152	-3.55	0.0004	0	escalonNavidad
<b>NUM17</b>	-0.02283	0.0057296	-3.98	<.0001	0	Pos_festivo
<b>NUM18</b>	-0.06565	0.0088159	-7.45	<.0001	0	Puente
<b>NUM19</b>	0.000267	0.00001863	14.38	<.0001	0	consumo_tmax_pred
<b>NUM20</b>	0.000178	0.00001712	10.44	<.0001	0	consumo_tmin_pred

Tabla 5.6.1. Estimación parámetros ARIMA + Var. Dummy + Temperaturas

A la vista de los resultados se obtienen las siguientes conclusiones:

- Tanto la variable relacionada a la temperatura máxima, como la de la temperatura mínima son significativas.
- El parámetro de las dos variables es positivo por lo que a movimientos de temperatura se espera un aumento de la Demanda de Gas.
- Todos los parámetros introducidos en los pasos anteriores continúan siendo significativos.

Una vez obtenido los modelos se procede a comparar los resultados para los tres escenarios posibles:

1. Solo con metodología ARIMA.
2. Metodología ARIMA más introducción de variables Dummies externas.
3. Metodología ARIMA, introducción de variables Dummies externas y temperaturas máximas y mínimas.

Para ellos se procede a validar los tres modelos en el periodo del 01 de Enero de 2015 al 31 de Mayo de 2015, para 10 periodos temporales (predicción 1 día vista, 2 días vista, ... , 10 días vista).

Se espera que según se vayan introduciendo variables externas, se vayan obteniendo mejores resultados. Las variables Dummies disminuirán los errores en los periodos vacaciones mientras que las temperaturas se espera que reduzcan más el error en la época Invernal (donde hay una mayor variabilidad de Demanda de Gas).



## 5.7 Resultados de Predicción

Los tres modelos expuestos anteriores se entrenan en el periodo del 01 de Enero de 2013 a 31 de Diciembre de 2015, y se validan sobre el periodo 01 de Enero de 2015 a 31 de Mayo de 2015.

Se realizan las predicciones a diez horizontes temporales (desde el propio día, horizonte un, hasta diez días vista), con el fin de ver el comportamiento de las predicciones según se alejan en el tiempo con respecto a los datos reales.

Se calcula para cada día de predicción el Error Absoluto en Porcentaje (EAP) mediante la siguiente ecuación:

$$EAP = \frac{|Valor_{Real,t} - Valor_{Predicho,t}|}{|Valor_{Real,t}|} * 100\%$$

Este indicador muestra para cada día la desviación porcentual que se ha producido con respecto al valor real, ya sea por haber predicho un valor inferior o superior al real.

Para visualizar las desviaciones por horizontes temporales, se calculan diferentes estadísticos, destacando el Error Absoluto Medio en Porcentaje (EAMP) y el Error Absoluto Mediano en Porcentaje (EAMeP).

$$EAMP = \frac{\sum_{t=1}^n \frac{|Valor_{Real,t} - Valor_{Predicho,t}|}{|Valor_{Real,t}|}}{n} * 100\%$$

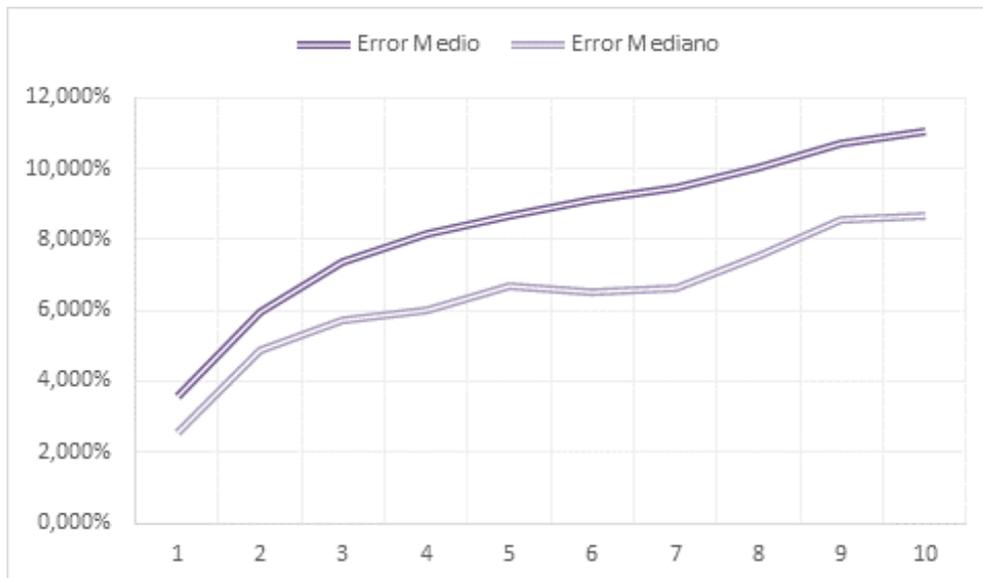
$$EAMeP = Median \left( \frac{|Valor_{Real,t} - Valor_{Predicho,t}|}{|Valor_{Real,t}|} \right)$$



### 5.7.1 Resultado solo con modelización ARIMA

Horizonte temporal	Error Medio	Error Mediano	Error Máximo	Desviación Típica
1	3,594%	2,554%	22,253%	3,488%
2	5,970%	4,862%	28,174%	5,366%
3	7,353%	5,734%	43,150%	6,667%
4	8,170%	5,981%	43,389%	7,634%
5	8,688%	6,686%	42,778%	8,348%
6	9,120%	6,533%	45,267%	8,994%
7	9,482%	6,652%	53,705%	9,509%
8	10,031%	7,552%	68,136%	10,212%
9	10,732%	8,531%	68,428%	10,467%
10	11,019%	8,668%	64,262%	10,726%

Tabla 5.7.1.1. Errores pred. ARIMA



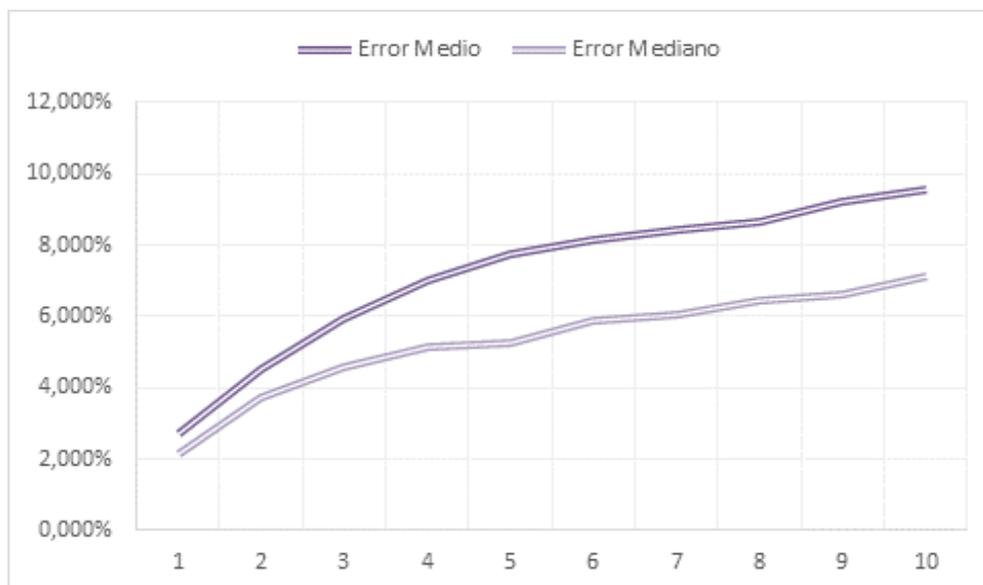
Gráfica 5.7.1.1. Representación errores.ARIMA



### 5.7.2 Resultados incluyendo variables Dummy

Horizonte temporal	Error Medio	Error Mediano	Error Máximo	Desviación Típica
1	2,717%	2,167%	11,888%	2,401%
2	4,540%	3,736%	24,820%	4,025%
3	5,946%	4,606%	40,004%	5,525%
4	7,009%	5,165%	44,281%	6,579%
5	7,734%	5,237%	42,800%	7,425%
6	8,143%	5,869%	45,288%	8,057%
7	8,411%	6,041%	48,010%	8,458%
8	8,639%	6,425%	44,274%	8,768%
9	9,217%	6,594%	46,342%	8,629%
10	9,578%	7,116%	40,475%	8,693%

Tabla 5.7.2.1. Errores pred. ARIMA + Dummy



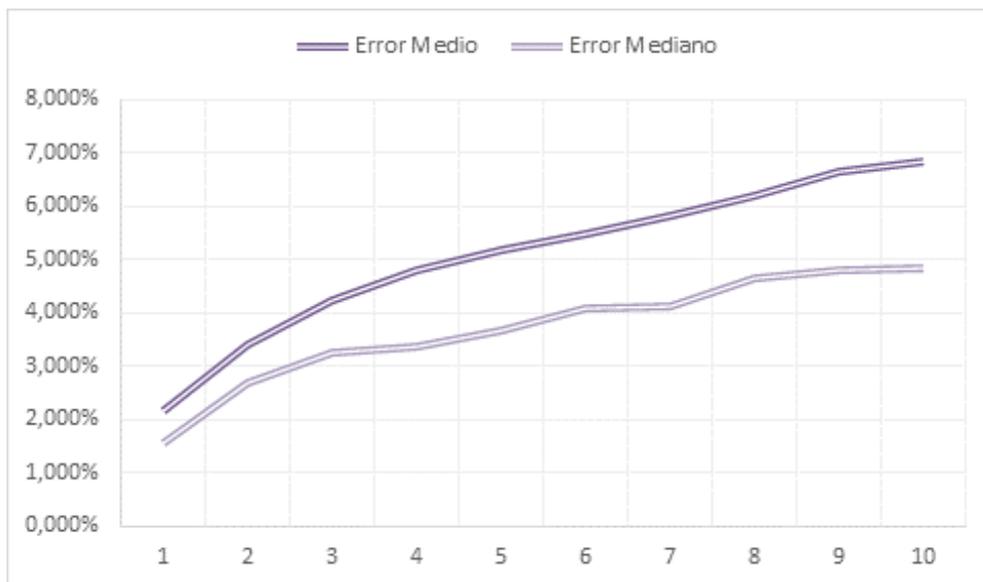
Gráfica 5.7.2.1. Representación errores ARIMA+Dummy



### 5.7.3 Resultados incluyendo variables Dummy y Temperaturas.

Horizonte temporal	Error Medio	Error Mediano	Error Máximo	Desviación Típica
1	2,161%	1,564%	12,289%	2,207%
2	3,407%	2,676%	25,500%	3,255%
3	4,236%	3,269%	32,003%	4,145%
4	4,811%	3,364%	32,153%	4,901%
5	5,185%	3,671%	31,331%	5,515%
6	5,474%	4,076%	35,403%	5,946%
7	5,812%	4,137%	35,530%	6,183%
8	6,191%	4,661%	33,165%	6,345%
9	6,655%	4,788%	33,324%	6,282%
10	6,826%	4,834%	31,844%	6,443%

Tabla 5.7.3.1. Errores pred ARIMA+Dummy+Tem

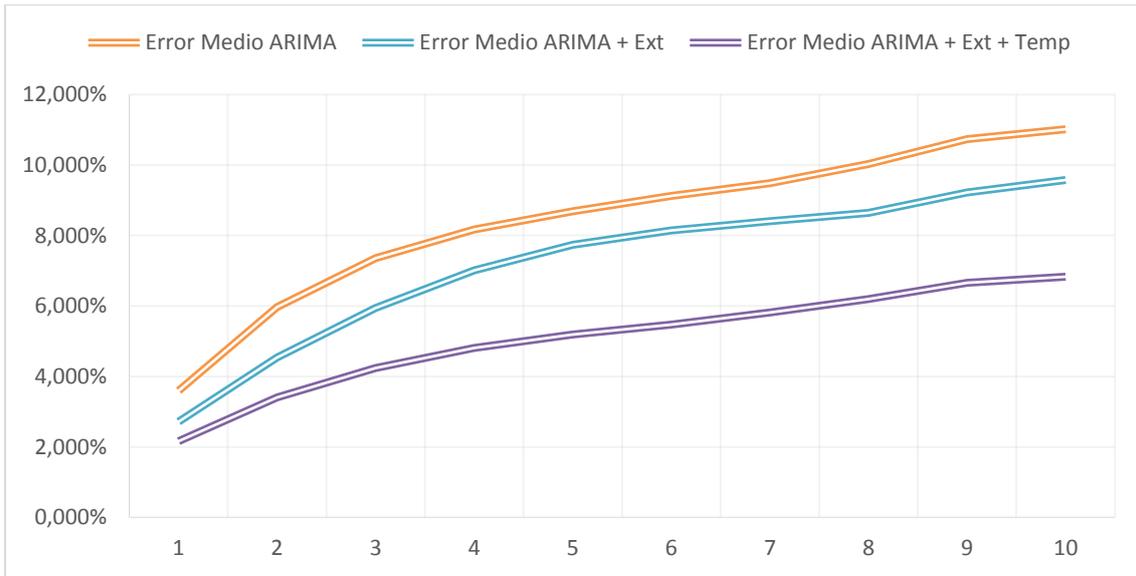


Gráfica 5.7.3.1. Representación errores ARIMA+Dummy+Temp

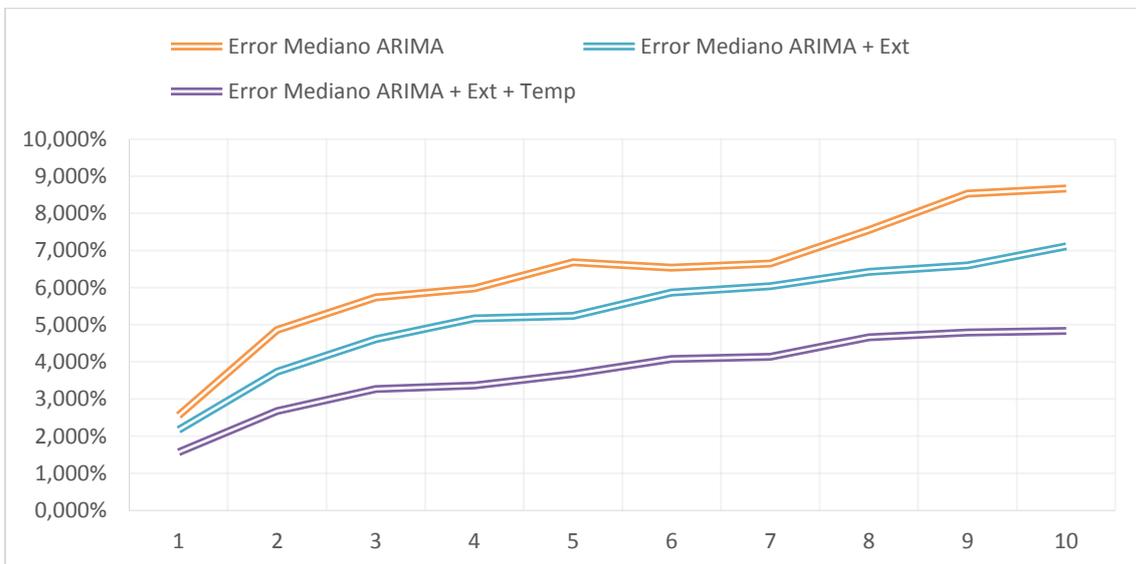


Mediante la inclusión de variables externas se obtiene un mejor resultado.

La Gráfica 5.7.3.2. y 5.7.3.3. muestran los resultados para el EAMP y el EAMeP de los tres modelos.



Gráfica 5.7.3.2. Comparación errores tres modelos



Gráfica 5.7.3.3. Representación error mediano tres modelos



Ambas gráficas reflejan como el Error Porcentual desciende claramente al introducir variables exógenas al modelo. Se produce un mayor descenso mediante la inclusión de las temperaturas.

En los resultados se ha observado cómo según aumenta el horizonte temporal (predicción a más días vista), tanto el error máximo como la desviación en los errores de predicción también aumenta. La inclusión de variables exógenas Dummy y de temperatura también permite controlar la las grandes desviaciones en los errores de predicción, disminuyendo en todos los horizontes el error máximo y la desviación de las predicciones.

Una de las limitaciones de estudio detalladas en el Apartado 1.5. es que los datos de Temperatura son reales en los horizontes de predicción. En un escenario real se espera que el Error Porcentual sea algo más alta que el mostrado en los resultados.

Se procede a estudiar fechas concretas (periodos festivos) para ver el efecto de introducir dichas variables en la predicción final.

### 5.7.4. Resultados de la predicción en periodos festivos

Se detallan los errores para cada uno de los tres modelos, en los festivos que se encuentran dentro del periodo de validación.

Fecha	Horizonte	EAMP Arima	EAMP Arima + Dummy	EAMP Arima+Dummy+Temp
<b>Año Nuevo</b>				
01/01/2015	1	2,284%	7,42%	7,66%
	2	6,691%	10,30%	9,24%
	3	7,366%	9,54%	9,06%
	4	5,877%	8,11%	5,90%
	5	3,982%	12,08%	10,12%
	6	1,373%	16,09%	11,87%
	7	6,854%	12,47%	11,20%
	8	2,742%	12,17%	11,54%
	9	15,726%	12,76%	10,13%
	10	19,849%	14,91%	10,27%
<b>Epifanía del Señor</b>				



06/01/2015	1	4,255%	5,11%	4,28%
	2	9,279%	4,37%	6,85%
	3	12,716%	6,67%	8,18%
	4	11,891%	8,22%	8,37%
	5	5,251%	11,57%	7,05%
	6	6,654%	3,51%	0,81%
	7	8,906%	1,04%	0,41%
	8	9,572%	1,88%	0,19%
	9	7,964%	3,57%	3,62%
	10	5,887%	1,05%	1,44%
<b>Semana Santa</b>				
02/04/2015 - 06/04/2015	1	6,698%	2,58%	3,05%
	2	10,078%	3,32%	3,59%
	3	12,996%	4,13%	4,05%
	4	14,449%	3,43%	4,14%
	5	20,155%	4,98%	5,49%
	6	25,845%	8,07%	8,36%
	7	32,044%	14,51%	14,19%
	8	40,841%	20,82%	18,28%
	9	48,116%	28,58%	24,18%
	10	52,152%	34,42%	27,48%
<b>Fiesta de Trabajo</b>				
01/05/2015	1	22,253%	3,02%	3,52%
	2	25,925%	7,52%	6,67%
	3	28,191%	10,25%	6,25%
	4	30,379%	12,37%	9,35%
	5	26,791%	8,45%	4,49%
	6	24,010%	4,39%	3,37%
	7	24,975%	5,03%	2,10%
	8	21,862%	2,52%	0,60%
	9	23,970%	5,03%	4,71%
	10	26,841%	8,13%	6,15%

Tabla 5.7.4.1. Errores días festivos

- **Año Nuevo:** Los resultados muestran como la inclusión de las variables Dummy no afecta positivamente a los resultados y aumenta los errores en prácticamente todos los horizontes.

Este comportamiento es debido al bajo consumo del resto de días y también a la inclusión en el modelo de una variable Dummy para el día 31 de Diciembre.



- **Epifanía del Señor:** Para el 06 de Enero si se produce un importante descenso de los errores pese a no incluirse su variable ficticia en el modelo (esto es debido a que no hay una diferencia significativa en la Demanda de Gas entre este día y los días que le rodean).

Si está incluida la variable Pos Festivo (afecta directamente al día 07 de Enero) y que ha producido el descenso en el error del día 06.

La introducción de las temperaturas también produce otro efecto positivo en el descenso del error. Al tratarse de la época Invernal, la Demanda de Gas es más sensible a los movimientos de las temperaturas por lo que estas han hecho mejorar los resultados.

- **Semana Santa:** La introducción de las variables Dummy para los días de Jueves Santo a Lunes de Pascua produce un claro descenso en el error de predicción.

El modelo con temperaturas presenta resultados muy similares ya que en ese periodo temporal del año ya no hay un efecto tan fuerte de las temperaturas sobre la Demanda de Gas.

- **Fiesta del Trabajo:** Al igual que en Semana Santa la inclusión de una variable ficticia para controlar el cambio de Demanda de Gas del día 01 de Mayo produce un claro descenso del error de predicción para todos los horizontes temporales.



## 6 Conclusiones

Como resultado final de este estudio se han obtenido las siguientes conclusiones:

- Comportamiento de la Demanda de Gas, sin una tendencia clara, obteniendo los valores más altos para el mes de Febrero (con un máximo de 1110 Gwh, el 06/02/2015) y los valores más bajos el mes de Agosto (con un mínimo de 410 Gwh, el 17/08/2014).
- Estacionalidad diaria de la serie temporal. Clara estacionalidad diaria, obteniéndose una Demanda Media similar para los días de diario, y un consumo inferior para los fines de semana (especialmente los Domingos).
- Estacionalidad mensual de la serie temporal. Se obtienen los datos más altos de Demanda en Febrero y los más bajos en Agosto. Se produce un descenso de la Demanda de los meses de Febrero a Agosto y un aumento de Septiembre a Diciembre.
- Relación no lineal entre las temperaturas máximas y mínimas con la Demanda de Gas.
- Claro descenso de la Demanda de Gas en festivos, lo cual exige la necesidad de introducir variables Dummy de intervención para controlar dichos movimientos en la Demanda.
- Se obtiene un modelo SARIMA  $(1,0,1) \times (0,1,1)_7$ , demostrando el efecto sobre la Demanda del dato del día interior, así como un comportamiento estacional diario.
- Necesidad de introducir variables exógenas en modelo, tales como variables Dummy o de temperaturas, con el fin de mejorar los resultados de predicción.
- Desarrollo de un modelo aplicable a la Demanda Convencional de Gas de cualquier empresa energética, e imprescindible para anticiparse de la forma más precisa al dato de Demanda, y de esta forma minimizar las sanciones.



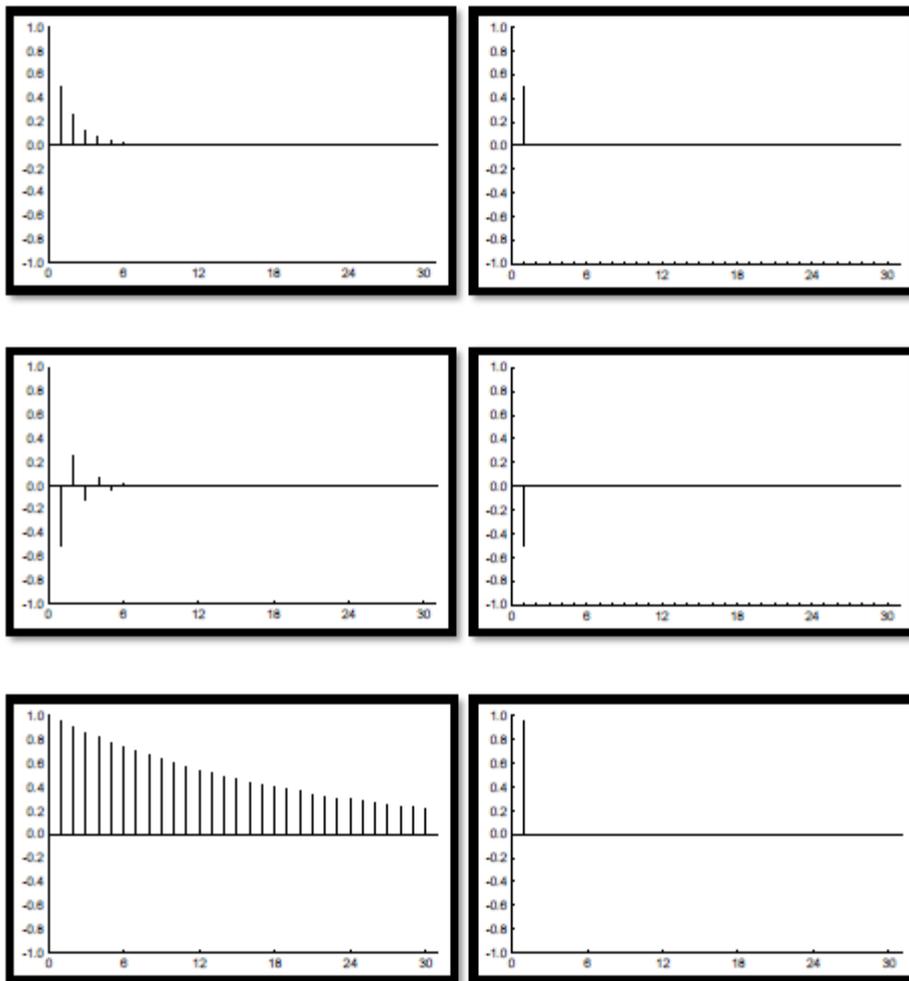
## 7 Anexo

A continuación se exponen una serie de combinaciones de resultados para la Función de autocorrelación y la Función de autocorrelación parcial para identificar con que modelo se debe ajustar una serie temporal

Ejemplo de visualización para modelos AR(1)

ACF

PACF

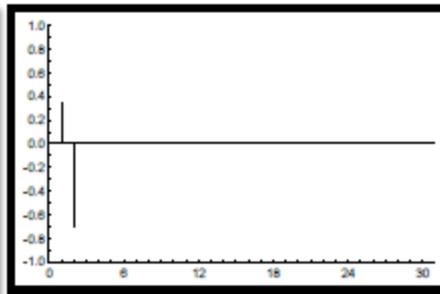
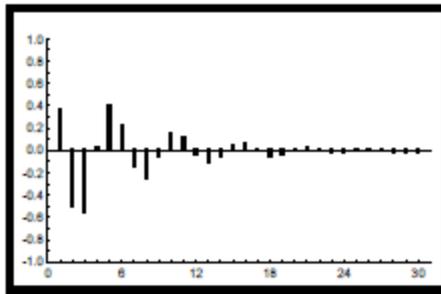
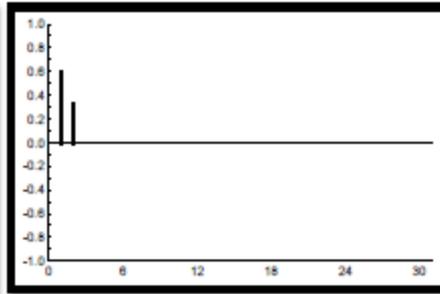
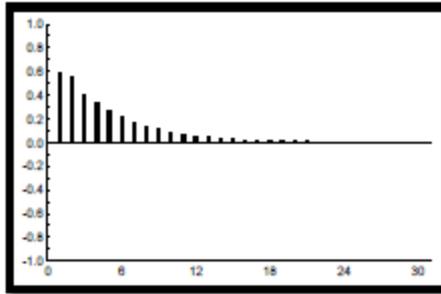


Ejemplo de visualización para modelos AR(2).



ACF

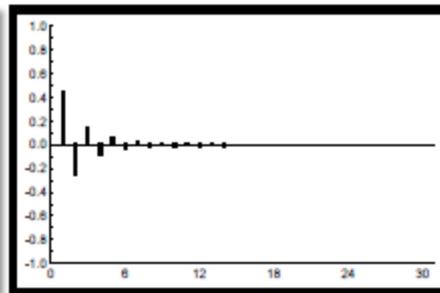
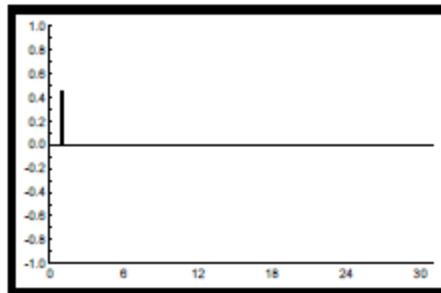
PACF



Ejemplo de visualización para un modelo MA(1).

ACF

PACF

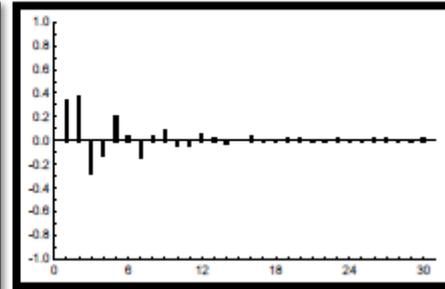
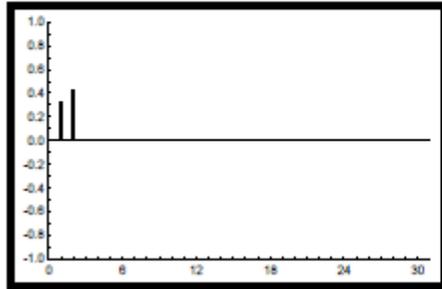




Ejemplo de visualización para un modelo MA(2).

ACF

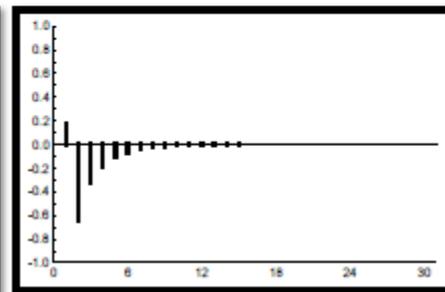
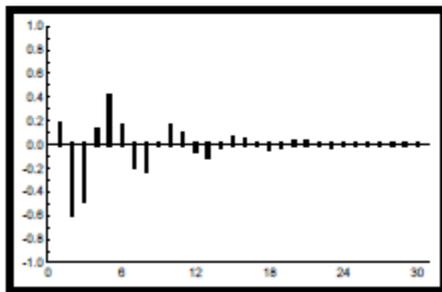
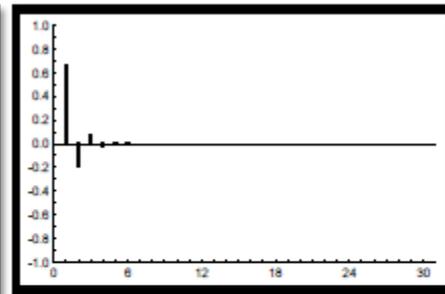
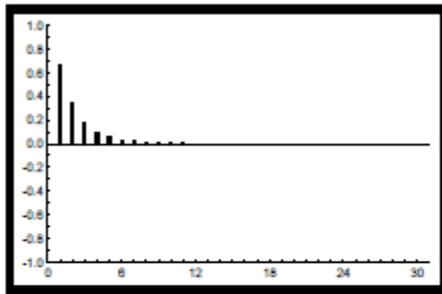
PACF



Ejemplo de visualización para un modelo ARMA(1,1)

ACF

PACF





## 8 Bibliografía

- **Enagas (2014).** *Informe del Sistema Gasista Español.*
- **CNMC (2014).** *Informe de Supervisión del Mercado Minorista del Gas Natural en España.*
- **BOE-A-2014-13057.** *Orden IET/2355/2014, de 12 de diciembre, por la que se establece la retribución de las actividades reguladas del sector gasista para el segundo período de 2014*
- **Fernández Fernández, Santiago; Córdoba Largo, Alejandro; Cordero Sánchez, José María (2002).** *Estadística Descriptiva.*
- **SAS.** *SAS/ETS 12.1 User's Guide.*
- **Box and Jenkins (1970).** *Time Series Analysis: Forecasting and Control.*
- **Hernández Alonso, José (2014).** *Análisis de Series Temporales Económicas II.*
- **Cancelo JR, A Espasa; R Grafe (2008).** *Forecasting the electricity load from one day to one week ahead for the Spanish system operator.*
- **Chikobvu, Delson; Sigauke, Caston (2012).** *Regression-SARIMA modelling of daily peak electricity demand in South Africa.* Journal of Energy in Sothern Africa Vol 23 No 3
- **Climent Diranzo, Francisco J y Otros (2003).** *Incidencia de la Climatología en el Consumo de Gas y Electricidad en España.*

### Otras fuentes consultadas

- Gestor Técnico del Sistema Gasista español ([www.Enagas.com](http://www.Enagas.com))
- Comisión Nacional de los Mercados y las Competencias ([www.cnmc.es](http://www.cnmc.es))
- Página Oficial de SAS ([support.sas.com](http://support.sas.com))
- Manual R para la realización de Gráficas Box Plot (<https://stat.ethz.ch/R-manual/R-devel/library/graphics/html/boxplot.html>)