Índice de documentos

DOCUMENTO I. MEMORIA

Parte I. Memoria	pág. 8 a 88	81 páginas

DOCUMENTO II. PLIEGO DE CONDICIONES

1. Generales y económicas	pág. 3 a 4	2 páginas
2. Técnicas y particulares	pág. 5 a 6	2 páginas

DOCUMENTO III. PRESUPUESTO

1. Mediciones	pág. 3 a 4	2 páginas
2. Precios unitarios	pág. 5 a 6	2 páginas
3. Sumas parciales	pág. 7 a 8	2 páginas
4. Presupuesto general	pág. 9	1 páginas

Autorizada la entrega del proyecto del alumno:

Manuel Gómez de la Calle

LOS DIRECTORES DEL PROYECTO

Dra. Alicia Mateo González

Fdo.: ..... Fecha: ..... / ..... / .....

Prof. Dr. Eugenio Francisco Sánchez Úbeda

Fdo.: ..... Fecha: ..... / ..... / .....

V^{O} B^{O} DEL COORDINADOR DE PROYECTOS

Prof. Dr. Álvaro Sánchez Miralles

Fdo.: ..... Fecha: ..... / ..... / .....

UNIVERSIDAD PONTIFICIA COMILLAS

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA (ICAI)

INGENIERO INDUSTRIAL

PROYECTO FIN DE CARRERA

MODELO DE PREVISIÓN DE DEMANDA DE ELECTRICIDAD DE LARGO PLAZO

AUTOR: Manuel Gómez de la Calle

MADRID, Septiembre de 2010

Resumen

Con la entrada en vigor de la Tarifa de Último Recurso (TUR) en julio del año pasado el Mercado Eléctrico Español queda totalmente liberalizado. Supone el final de un largo proceso que se inició en 1997 con la aprobación de la Ley 54/1997. Con la aparición de esta nueva tarifa, cualquier consumidor con independencia de sus características de consumo, pueden contratar libremente a su suministrador.

Este último paso en la liberalización del mercado eléctrico español ha tenido un gran impacto en las comercializadoras, que han visto notablemente aumentado el número de clientes a los que suministrar electricidad, con el consiguiente crecimiento de la incertidumbre a la hora de gestionarles la energía. Para minimizar el riesgo de precio las comercializadoras pueden firmar contratos de compra/venta de energía a medio y largo plazo en distintos mercados. Antes de acudir a los mercados de medio y largo plazo (físicos o financieros) es necesario conocer, con la menor incertidumbre posible, el consumo de su cartera de clientes, y así poder ajustar las cantidades de energía para la que desea cubrir su riesgo de precio. En este punto es cuando surge la necesidad de disponer de un buen modelo de previsión de energía eléctrica a medio plazo.

Existen multitud de factores que influyen en la demanda de la energía eléctrica, pero la mayoría de los autores coinciden en que son variables determinantes la laboralidad y las temperaturas, por lo que es indispensable modelar de la mejor manera su impacto sobre el consumo de energía eléctrica. El impacto de la temperatura está directamente relacionada con la necesidad de disponer de equipos de aire acondicionado o calefacción para mantener una temperatura de bienestar, por lo que, como se verá, será una variable claramente estacional en la que se podrán identificar los periodos de verano e invierno sin dificultad.

Por otro lado, la variable laboralidad nos ayudará a explicar el consumo de energía eléctrica en días laborales donde la actividad industrial es muy alta, y distinguirlos de los días festivos o vacaciones, donde como también se verá la demanda cae notablemente por el descenso de dicha actividad industrial.

En este proyecto se va a proponer una nueva metodología para realizar la previsión de la demanda de energía eléctrica en el medio-largo plazo para clientes industriales. Para ello se propondrá un modelo estadístico y ajustará una muestra de clientes. La novedad que se presenta en este proyecto es el uso de técnicas de *clustering* combinadas con modelos de regresión lineal. El proceso para la construcción del modelo constará de dos etapas, que muestran de manera esquemática en la Figura 1.

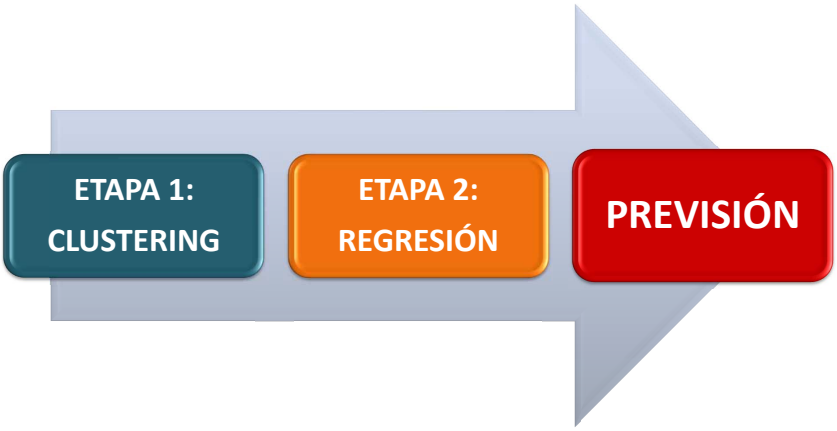

Figura 1. Esquema construcción Modelo.

La primera etapa abarca todo el trabajo de caracterización y clasificación de los clientes. Cada cliente se identificará de manera inequívoca según su forma de consumir durante ciertas horas del día, a lo largo de la semana y a lo largo de los diferentes meses. Atendiendo a estas características, se va a utilizar la red neuronal conocida como "Mapa de Kohonen" para clasificar a los clientes industriales de forma automática. El objetivo es agrupar aquellos clientes con características de consumo similares. Una vez clasificados los clientes se obtendrán series agregadas de consumo por cada grupo, de manera que el problema de estudiar la serie de consumo total de los clientes quedaría dividido en el estudio de tantas series agregadas como grupos se hayan identificado.

La segunda etapa consiste en construir los modelos de regresión lineal para ajustar las series agregadas de los grupos construidos en la etapa anterior. Como variables explicativas se utilizarán la temperatura y la laboralidad en cada grupo. Para poder utilizar adecuadamente la información aportada por estas variables será necesario construir, también para cada grupo, la variable de temperatura y laboralidad agregada, cuyo proceso se definirá con detalle y en él será importante la comunidad autónoma o la localización del cliente para su determinación.

La formación de los distintos grupos, así como de las diferentes series temporales de cada uno, permite el uso de modelos de regresión específicos para cada grupo de clientes

con características similares de consumo. A parte de la distinción de los grupos, se van a realizar modelos específicos para cada tipo de día de la semana, puesto que dicha variable será determinante. El uso de estos modelos específicos supone una mejora frente al uso de un único modelo para ajustar toda la demanda de energía, puesto que el modelo siempre buscará una solución de compromiso para formas muy distintas de consumir.

El modelo desarrollado en este proyecto se ha aplicado a una muestra de clientes industriales de la cartera de ENDESA durante el periodo comprendido entre enero de 2003 y diciembre de 2009, poniendo especial interés al periodo de crisis económica vivido en los últimos meses en España. El resultado ha sido un modelo que ajusta la demanda de energía eléctrica con un error MAPE de aproximadamente el 4 %, nivel similar a los referenciados en la literatura.

Abstract

WHEN the TUR ("Tarifa de Último Recurso" in Spanish, last resort rate) came into effect last year in July, the Spanish Electric Market was completely liberalized. This is the end of a long process which started in 1997 with the approval of the Law 54/1997. This new rate involves that any electric consumer, regardless of their electric consumption characteristics, can contract freely his electric supplier.

This last step in the Spanish electricity market liberalization has had a major impact on the electric.retailers, since the number of customers has increased significantly. This fact involves that there is greater uncertainty to manage their electric energy. To minimize the price risk, the electric retailers can sign medium and long term purchase and sell energy contracts in different markets. Before resort to these medium and long term markets (physical and financial markets), it is necessary to know, with the lower uncertainly as possible, the clients list energy consumption. In this way, they can adjust the energy quantities for which they can cover their price risk. Is in this point when the need of having a good medium term forecasting electric energy model appears.

There are several factors that have an impact on the electric energy demand, but most of the authors agree about that the temperature and the working day are determinants. For that reason, is essential to model in the best way the impact of these two variables on the electricity consumption. The effect of the temperature is directly related with the necessity of having air conditioning or heating machines, for keep a comfort temperature. This is a seasonal variable in which can identify the different periods of summer and winter without any difficulty.

On the other hand, the working day variable will help to explain the electric consumption on working days, when the industrial activity is really high, and distinguish it from holidays, when the electric consumption decreases remarkably by the fall of this activity.

This proyect will propose a new methodology for perform the medium term electric energy consumption forecasting for a industrial client list. A statistic model will be proposed and adjusted for a industrial clients sample. The innovation in the forecasting field that it is shown in this proyect is the use of clustering techniques combine with linear regression models. The process for the construction of this model will consist of two stages, which appear in a schematic way on the figure 1.

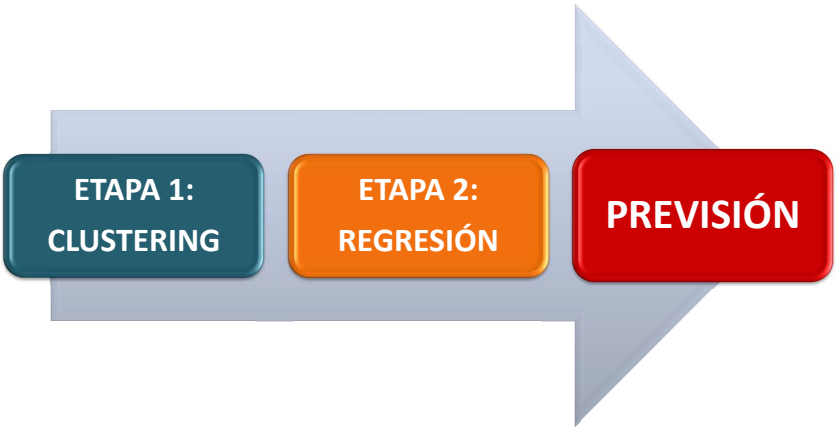

Figure 1. Construction scheme model.

The first stage covers all the characterization and clustering of the client list sample. Each client will be identified according the way of using the electric energy along certain hours of the day, along the week and along the different months. According to these characteristics, a neuronal network known as Kohonen Maps will be use to clust the industrial client sample in an automatic way. The objective is to group those clients with similar consumption characteristics. The next step is to calculate the aggregated consumption series for each group. Now, the problem of studying the total consumption serie it is divided into the study of as many aggregated consumption series as groups are formed with the neural network.

The second and last step consist of constructing the linear regression models for fitting the groups aggregated consumption series of the previous step. Working day and temperature will be used as explanatory variables in each group. It is necessary to construct, for each group, the aggregated working day and aggregated temperature variable, for using in the correct way the information of those variables. This process will be defined in detail and it will be very important the autonomous regions and the client location.

The different groups made and the aggregated consumption time series allow the use of specific linear regression models for each client group with similar consumption

characteristics. Apart from the distinction of the groups, specific models for each type of the different days of the week will be made, due to this variable will be determinant. Using these specific model will be a improvement against the use of a single model to fit all the electric energy consumption, due to this last model always will look for a intermediate solution for very different consumption ways.

The developed model of this proyect will be applied to a industrial clients sample of ENDESA during the period between January 2003 and December 2009, with a special interest in the economic crisis period in last months in Spain. The result is a model that fit the total energy consumption with a MAPE of approximately 4 %, similar level to those referenced in the literature.

A mi familia
por todo lo que les debo.

"El futuro tiene muchos nombres.
Para los débiles es lo inalcanzable.
Para los temerosos, lo desconocido.
Para los valientes es la oportunidad."
VICTOR HUGO

Agradecimientos

Al finalizar un trabajo tan importante como el Proyecto Fin de Carrera, que supone la finalización de los estudios universitarios, es inevitable llenarse de sentimientos egocentristas, que tienden a abarcar todos los méritos y felicitaciones. Pero un análisis objetivo de los hechos revela que todo el trabajo realizado sería imposible sin la ayuda y participación de las personas e instituciones que han facilitado los medios para que el proyecto llegue a un feliz término. Por ello, es un placer utilizar este espacio para ser justo y consecuente con ellas, expresándoles mis agradecimientos.

En primer lugar me gustaría agradecer la inconmensurable ayuda y el apoyo que me ha brindado mi familia a lo largo de la carrera, y de manera especial, durante la realización de este proyecto. Me gustaría destacar a mis padres, Yolanda y J. Manuel, mis primeros y mejores maestros, que creyeron en mí y me han guiado a lo largo de mi vida con su ejemplo de dedicación y esfuerzo. Muchas gracias.

Aprovecho también la ocasión para hacer mención a mis compañeros, que me han acompañado durante estos maravillosos años de universidad, haciendo que el tiempo transcurrido se convierta en una experiencia única e irrepetible. Nunca olvidaré las experiencias vividas, los consejos y las ayudas en los momentos de duda.

Me gustaría agradecer la inestimable ayuda brindada por mis directores de proyecto, Alicia y Eugenio, por estar siempre pendientes del desarrollo del proyecto y que me han ayudado a solucionar todos los problemas que han aparecido a lo largo de la realización del mismo.

También quisiera agradecer a los profesores de la universidad su contribución en mi formación académica y que han ayudado a que este proyecto se convierta en una realidad.

Por último, me gustaría agradecer a ENDESA, que me brindó una beca para realización de este proyecto, la confianza depositada en mi persona para su desarrollo.

DOCUMENTO I

MEMORIA

Índice

I. Memoria **8**

1. Introducción **9**

1.1. Motivación 9

1.2. Objetivos 17

1.3. Recursos 18

1.3.1. Hardware 18

1.3.2. Software 18

2. Estado del arte **20**

2.1. Historia 20

2.2. Horizontes temporales de previsión 21

2.2.1. Corto Plazo 22

2.2.1.1. Factores influyentes corto plazo 22

2.2.2. Medio Plazo 22

2.2.2.1. Factores influyentes medio plazo 22

2.2.3. Largo Plazo 22

2.2.3.1. Factores influyentes largo plazo 23

2.3. Métodos de previsión 23

2.3.1. Métodos cualitativos 23

2.3.2. Métodos cuantitativos 23

2.3.2.1. Regresión Lineal 23

2.3.2.2. Modelos ARIMA 24

2.3.2.3. Modelos de Descomposición 24

2.3.2.4. Modelos de redes neuronales 25

3. Descripción del modelo de previsión **26**

3.1. Introducción 26

3.2. Variables 29

3.2.1. Datos consumo 29

3.2.2. Clientes Industriales 29

3.2.2.1. Comunidades Autónomas 30

3.2.2.2. Observatorios 30

3.2.3. Variables exógenas 31
3.2.3.1. Temperatura 31
3.2.3.2. Festividades 31
3.3. Etapa 1: *Clustering* 31
3.3.1. Perfiles de consumo 31
3.3.1.1. Perfil diario 34
3.3.1.2. Perfil semanal 45
3.3.1.3. Perfil anual 51
3.3.2. Mapas autoorganizados de Kohonen 57
3.3.3. *Clustering* 60
3.3.4. Cálculo Series Agregadas 64
3.3.4.1. Series Consumo Agregadas 66
3.3.4.2. Series Festividades Agregadas 67
3.3.4.3. Series Temperatura Agregadas 68
3.3.4.4. Series Número de Clientes 70
3.4. Etapa 2: Modelos de regresión 70
3.4.1. Series Consumo Anual 70
3.4.2. Variables indicadoras calendario 72
3.4.2.1. Tipo de mes 72
3.4.2.2. Navidad 73
3.4.2.3. Semana Santa 73
3.4.2.4. Periodos 74
3.4.3. Modelos de regresión 74

4. Resultados **77**
4.1. Introducción 77
4.2. Tiempos Ejecución 78
4.3. Ajuste de los modelos de regresión 78
4.4. Regresión Serie Consumo Total 79
4.5. Comparación Modelos Literatura 81

5. Conclusiones **82**

6. Futuros desarrollos **84**

Bibliografía **86**

Índice de figuras

1. Cruce de las curvas agregadas de compra y venta 15
2. Evolución de la demanda de energía eléctrica peninsular. 16
3. Esquema modelo previsión 28
4. Ejemplo perfil diario 1. 35
5. Perfil diario 5 junio de 2009. 36
6. Ejemplo perfil diario 2. 37
7. Ejemplo perfil diario 3. 38
8. Ejemplo perfil diario 4 39
9. KSOM 5x5 perfil diario. 40
10. Grupo 8 perfil diario. 41
11. Componentes Principales perfil diario. 42
12. Matriz U perfil diario. 43
13. Ejemplo perfil semanal 1. 46
14. Ejemplo perfil semanal 2. 47
15. Ejemplo perfil semanal 3. 47
16. Componentes Principales perfil semanal. 48
17. KSOM 5x5 perfil semanal. 49
18. Matriz U perfil semanal. 50
19. Ejemplo perfil anual 1 52
20. Ejemplo perfil anual 2 52
21. Ejemplo perfil anual 3 53
22. KSOM 5x5 perfil anual. 54
23. Componentes Principales perfil anual. 55
24. Matriz U perfil anual. 55
25. Estructura mapa de Kohonen 58
26. Pantalla de Inicio KSOM 60
27. Mapa de KSOM perfil_consumo 62
28. Mapa componentes principales 63
29. Serie Consumo Grupo 3 66
30. Serie Consumo Grupo 7 66
31. Serie Temperatura Grupo 3 69

32. Serie Consumo Grupo 3 VS Serie Consumo Normalizada 71
33. Serie Consumo Grupo 7 VS Serie Consumo Normalizada 72
34. Regresión Grupo 3. 75
35. Regresión Grupo 7. 76
36. Serie Real Consumo VS Serie Estimada. Residuos 80

Índice de tablas

1. Calendario de elegibilidad para los consumidores de electricidad 13
2. Fechas Inicio y Fin estaciones 33
3. Resumen Grupos KSOM 64
4. Fechas Inicio y Fin Semana Santa 73
5. MAPE's separado por años. 81
6. Comparación errores literatura. 81

Acrónimos

ARIMA	*Autoregressive Integrated Moving Average*
CC.AA.	Comunidades Autónomas
CNAE	Clasificación Nacional de Actividades Económicas
CUR	Comercializadora de Último Recurso
ENDESA	Empresa Nacional de Electricidad, Sociedad Anónima
GPL	*General Public License*
ICAI	Instituto Católico de Artes e Industrias
INM	Instituto Nacional de Meteorología
KSOM	*Kohonen Self-Organized Maps*
LSE	Ley del Sector Eléctrico
MAPE	*Mean Average Percentage Error*
MRC	*Matlab Component Runtime*
OMEL	Operador del Mercado de la Electricidad
OMIP	Operador del Mercado Ibérico a Plazo
OS	Operador del Sistema
PFC	Proyecto Fin de Carrera
PIB	Producto Interior Bruto
REE	Red Eléctrica Española
RMSE	*Root Mean Square Error*
RNA	Red Neuronal Artificial
SEE	Sistema Eléctrico Español
SQL	*Structured Query Language*
SS	Semana Santa
TUR	Tarifa de Último Recurso

PARTE I

MEMORIA

Capítulo 1

Introducción

"Paréceme, Sancho, que no hay refrán que no sea verdadero,
porque todos son sentencias sacadas de la misma experiencia,
madre de las ciencias todas."
MIGUEL DE CERVANTES

LA **"previsión"** de una magnitud consiste en el *cálculo anticipado que se hace de un hecho que va a suceder, a partir de unas determinadas señales o indicios*. En el caso concreto de este trabajo, se va a desarrollar un modelo de previsión de la demanda de energía eléctrica a medio plazo. Es importante recalcar que se pretende abordar el problema desde el punto de vista de una comercializadora. Por tanto, este proyecto debe contribuir a determinar la energía que una cartera de clientes de una comercializadora consumirá en el horizonte del medio plazo, que se establecerá en al menos un año.

1.1. Motivación

Entre los numerables motivos por lo que es importante realizar previsiones de largo plazo de electricidad, se puede destacar la importancia de disponer de una buena previsión, que permita actualizar y mantener al parque de generación adecuado en cada momento. Dado que la instalación de nuevas centrales es un proceso muy lento y que requiere una fuerte inversión, se necesita disponer de estimaciones que garanticen la necesidad de estas centrales con un tiempo de anticipación adecuado, de lo contrario se corre el riesgo de no ser capaces de garantizar el suministro de energía eléctrica a los futuros consumidores. En este sentido, la previsión de la demanda de medio-largo plazo es una pieza clave para la sociedad actual, en tanto que contribuye a garantizar el suministro de medio-largo plazo, y por tanto, a garantizar el funcionamiento de la economía moderna y el estado del bienestar.

En este sentido, desde un punto de vista financiero, las inversiones en el sector eléctrico suele contribuir de manera muy importante en el PIB de los países desarrollados según [23]. Por tanto, las previsiones de largo plazo de la demanda se convierten en indicadores relevantes a considerar en los estudios, no sólo para conocer la evolución del sector eléctrico, sino también para la evolución de la economía general de un país.

Son varios los agentes que integran el sector eléctrico: generadores, distribuidores, comercializadores, reguladores, Ministerios, Operadores Técnicos del Sistema (OS), etc. Para todos ellos, la previsión de la demanda de medio-largo es fundamental, aunque cada uno de ellos tiene objetivos distintos, con horizontes temporales diferentes.

Desde el punto de vista de medio plazo, el agente regulador, el Ministerio o los Operadores del Sistema, deberán utilizar una buena previsión de demanda de energía eléctrica, entre otras muchas cosas, para trasladar la señal económica adecuada a los generadores, con el objetivo de que éstos realicen las inversiones que garanticen el suministro eléctrico. En el medio plazo, los generadores deberán aprovisionarse de la materia prima necesaria para poder operar sus centrales de manera óptima y programar las operaciones de mantenimiento de los equipos necesarias para el correcto funcionamiento de las mismas. Los distribuidores deberán establecer y mejorar sus redes de abastecimiento, adecuándolas a la demanda prevista. Y por último, las comercializadoras deben conocer también el consumo que tendrán sus clientes en el medio plazo, con el objetivo de conocer su posición y optimizar el riesgo de su cartera de clientes entre otros, motivación principal de este proyecto.

Las comercializadoras han sido pieza clave del sector eléctrico desde el comienzo de la liberalización del sector, puesto que durante dicho proceso ha dado lugar a cambios importantes para todos los participantes del mercado. Éstos han sido especialmente intensos para las comercializadoras, que a día de hoy gestionan toda la demanda del Sistema Eléctrico Español.

El proceso de liberalización del sector eléctrico español comenzó en 1997 con la aprobación de la Ley 54/1997 el 27 de noviembre, la Ley del Sector Eléctrico (LSE), que constituye el paso definitivo para liberalizar el Sector Eléctrico Español. De esta Ley merece destacar el establecimiento de los siguientes tramos de mercado:

- Mercado mayorista, Pool[1] centralizado o sistema de subasta, denominado Mercado Diario, donde la energía es ofertada y demandada diariamente por los generadores y los consumidores, con entrega para el día siguiente, disponiendo de un mecanismo de ajuste para las posibles modificaciones (Mercado Intradiario). Se crea la figura del **Operador del Mercado**, entidad encargada de gestionar los aspectos económicos del mercado. A pesar de la plena libertad de entrada, no aparecen nuevos participantes.

- Mercado Minorista, con libertad de elección de suministrador (de aplicación progresiva). Se contempla la posibilidad de llevar a cabo transacciones bilaterales entre oferentes y demandantes de energía eléctrica al margen del mercado diario para su liquidación por entrega de energía eléctrica o por diferencias. Aparece de la figura del comercializador.

- Regulación y acceso garantizado de terceros a las redes de transporte y distribución.

En 1998 entró en funcionamiento el mercado liberalizado de la energía, convirtiéndose en el primer mercado libre y organizado al contado de energía eléctrica de los países de la Unión Monetaria Europea. Sin embargo, sólo los usuarios cualificados[2] pudieron beneficiarse en un primer momento de la apertura del mercado eléctrico. En este mercado, los generadores compiten para vender su producción y las comercializadoras y distribuidores compran la energía para venderla posteriormente a los clientes cualificados y a los clientes a tarifa, respectivamente.

A partir del 1 de enero de 2003, la legislación española reconoce al consumidor de energía el derecho de elegir la forma en la que contrata el suministro eléctrico. En este sentido, el consumidor peninsular podrá optar por permanecer a tarifa integral y seguir contratando la electricidad con las empresas distribuidoras de las zonas, o pasar al mercado liberalizado. Para consumidores conectados en alta tensión, estaba previsto que en 2007 desaparecerá el mercado a tarifa, aunque finalmente se alargó hasta julio 2008.

Finalmente, el mercado eléctrico español ha quedado totalmente liberalizado con la entrada en vigor de la Tarifa de Último Recurso (TUR) el pasado 1 de julio de 2009.

[1] El Pool es un mercado de compra y venta de energía. El Operador de Mercado (OMEL) recibe ofertas de los generadores para cada una de las horas y cubre la demanda mediante un sistema de pujas, adjudicando los paquetes de electricidad hasta cubrir la demanda total del día, siguiendo un rango de precio ascendente, es decir, de la oferta más barata a la más cara. Así, la energía vendida es igual a la energía comprada. Sin embargo, el precio final que se paga es único y corresponde a la oferta más elevada que se haya tenido que aceptar para cubrir la demanda

[2] Durante esta época, se consideraba consumidor cualificado a todo aquel que tuviera un consumo anual superior a 15GWh.

El objetivo de la TUR es garantizar que el suministro de energía eléctrica siga siendo un servicio básico y asequible para todos. Esto ha supuesto un punto de inflexión para las comercializadoras. Los cambios más significativos que acarrea esta nueva tarifa se resumen en los siguientes apartados:

- Las tarifas integrales de energía eléctrica quedan extinguidas. Esto supone la desaparición de las tarifas reguladas en baja tensión para aproximadamente 25 millones de clientes en España. Dentro de este marco, aumentan las oportunidades de negocio para las comercializadoras, ya que el número de potenciales clientes se ve incrementado ante las nuevas posibilidades que brinda la liberación del mercado. Las comercializadoras se enfrentan ante un aumento significativo de los clientes a los que ha de suministrar energía.

- Los consumidores suministrados por un distribuidor que no hayan optado por elegir empresa comercializadora pasarán a ser suministrados por un comercializador de último recurso (CUR). Dicho comercializador sucederá a la empresa distribuidora con los derechos y obligaciones establecidos en el artículo 45 de la Ley 54/1997, de 27 de noviembre de 1997, del Sector Eléctrico Español. Las empresas de distribución de electricidad dejarán por tanto de prestar servicios de comercialización, para dedicarse en exclusiva a construir, mantener y operar las redes, así como a leer los contadores y facilitar el servicio de averías 24 horas.

- Sólo podrán acogerse a tarifas de último recurso los consumidores de energía eléctrica conectados en baja tensión cuya potencia contratada sea inferior o igual a 10kW. Con la aparición de la TUR, todos los consumidores deben de contratar la energía eléctrica en el mercado liberalizado, sin embargo, para aquellos que no hayan formalizado contrato de su suministro eléctrico en el mercado libre, la TUR es una tarifa de transición, que garantiza su suministro eléctrico, aunque no tengan contrato en el mercado libre.

En conclusión, el sector eléctrico se ha liberalizado en diferentes etapas, hasta permitir que todos los usuarios, independientemente de su potencia contratada, participen del mercado liberalizado. Esto ha supuesto un punto de inflexión para los agentes que operan en el sector eléctrico, en especial a las comercializadoras.

En la Tabla 1 se resume el proceso seguido en la liberalización del sector eléctrico español. Muestra la evolución de los consumos mínimos exigidos para adquirir la condición de consumidor cualificado dentro del sector y poder participar el mercado liberalizado. Los datos han sido obtenidos de la Comisión Nacional de Energía.

Marco Jurídico	Fecha Efecto	Consumo Anual	Apertura Mercado
Ley 54/1997	01/01/1998	Superior a 15GWh	26 %
			700 suministros
RD 2820/1998	01/01/1999	Superior a 5GWh	33.40 %
			2.300 suministros
-	01/01/1998	Superior a 3GWh	37 %
			3.800 suministros
-	01/01/1998	Superior a 2GWh	39.60 %
			5.600 suministros
-	01/01/1998	Superior a 1GWh	43.40 %
			10.000 suministros
RD-L 6/2000	01/01/2003	Consumidores de AT	52.30 %
		(V>1000V)	65.000 suministros
RD 485/2009	01/07/2009	Todos los consumidores	100 %
			25.000.000 suministros

Tabla 1. Calendario de elegibilidad para los consumidores de electricidad

Desde el 1 de julio de 2009, existen por tanto en el mercado dos tipos fundamentes de comercializadoras:

- **Comercializadoras Mercado Libre**: operan en el mercado liberalizado, ofreciendo distintos precios a los clientes de media y baja tensión. Su función es suministrar energía eléctrica a los consumidores que estén en el mercado libre, es decir, a aquellos que han elegido libremente su comercializadora y pactado unas condiciones de contrato. La factura que el cliente paga a su empresa comercializadora es un precio pactado libremente con el cliente que incluye la tarifa de acceso por usar las redes eléctricas de la empresa distribuidora y el precio por la energía consumida según el contrato firmado con ella.

- **CUR**: únicas habilitadas para suministrar a los clientes con la tarifa de último recurso. Su función es suministrar energía eléctrica a los consumidores que estén acogidos a la tarifa de último recurso. En particular, en ENDESA se ha creado una nueva comercializadora de último recurso, que recibe el nombre de ENDESA ENERGÍA XXI.

La tarifa de último recurso que el cliente pagará a su CUR está formado por los términos de la Ecuación 1:

$$TUR = Coste_energia + Tarifa_acceso + Costes_comercializacion \quad (1)$$

Cada uno de los términos de la expresión anterior se explican a continuación:

- **Coste de la energía**: fijado con base a los Mercados a Plazo (CESUR y OMIP) y a otros costes de la energía, como los Costes de Ajuste del Sistema y los Pagos por Capacidad. Se revisan como mínimo cada 6 meses.
- **Tarifa de acceso**: posibilidad de revisión semestral de los peajes, para garantizar la aditividad de la TUR.
- **Costes de comercialización**: costes reconocidos a la actividad de comercialización.

Como se ha visto anteriormente, el proceso de liberalización ha llegado a un punto muy importante recientemente, y con un fuerte impacto para los grupos con comercializadoras CUR, que han visto como su cartera de clientes ha aumentado considerablemente, pasando a gestionar millones de clientes. Este aumento de la cartera de clientes también trae consigo un crecimiento de la incertidumbre a la hora de gestionarla.

Gestionar a los clientes supone comprarles la electricidad en los distintos horizontes temporales (corto, medio y largo plazo) y facturarla. En el corto plazo, se acude a los mercados diarios e intradiarios, donde se realiza la compra de la energía eléctrica, que se rige bajo la libre concurrencia de ofertas de compra y venta. Los mercados se convocan según un horario preestablecido al que los agentes de las comercializadoras acuden para comprar la energía que necesitan a un precio dado. El precio de la energía se obtiene por la intersección de las curvas agregadas de oferta y demanda (Figura 1).

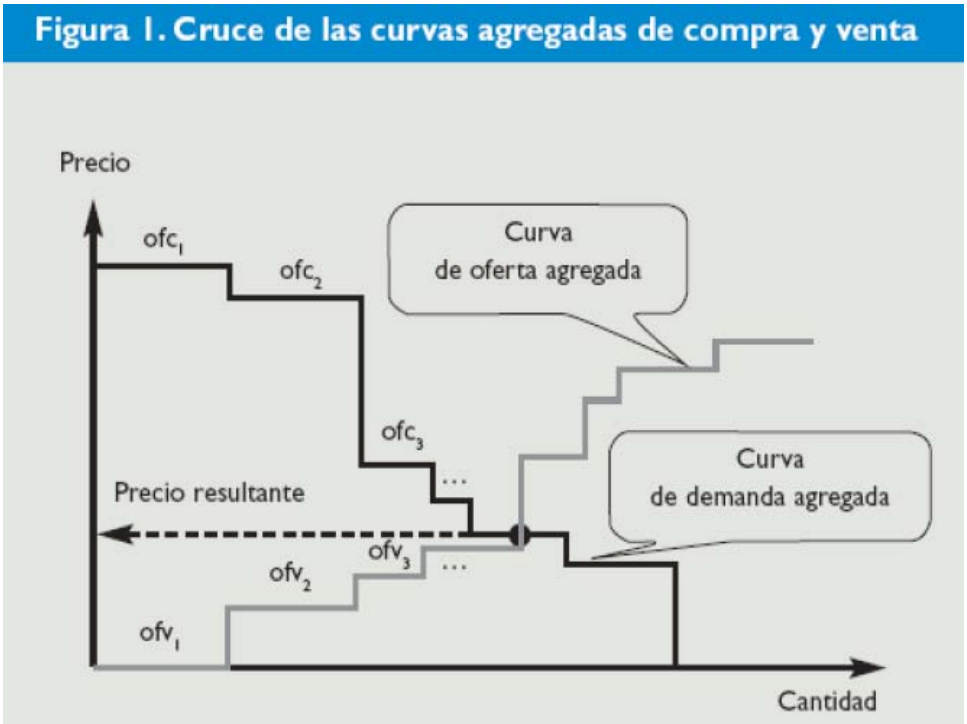

Figura 1. Cruce de las curvas agregadas de compra y venta

Este precio es el resultado de las estrategias propias de los agentes, el comportamiento de la demanda de la cartera de clientes y de los recursos del sistema. A parte de la incertidumbre generada por el comportamiento de los clientes, también existe la incertidumbre generada por la variabilidad de los precios de la materia prima, por lo que existe una alta volatilidad en los precios finales de estos mercados. Con el objetivo de minimizar este riesgo de precio, los agentes pueden acudir a firmar contratos de compra/venta de energía a medio y largo plazo en distintos mercados.

Básicamente existen dos tipos, los mercados físicos (OTC) y los financieros. Para poder acudir con garantías a un mercado de largo plazo, una empresa comercializadora debe de conocer con la menor incertidumbre posible el consumo de su cartera de clientes, y así poder ajustar las cantidades de energía para la que desea cubrir su riesgo de precio. En este sentido debe desarrollar modelos de previsión de electricidad de medio-largo plazo para su cartera.

Se puede decir que la cartera de una comercializadora está formada básicamente por grandes consumidores (fábricas, cogeneradores, etc) y clientes domésticos, generalmente conectados a baja tensión. Como se deduce de la breve exposición sobre los periodos de liberalización del mercado, las comercializadoras han tenido continuamente que adaptarse a los cambios y a la nueva legislación. Esto ha dado lugar a que su cartera de clientes haya fluctuado continuamente, tanto en clientes de alta tensión, como en consumidores domésticos.

Desde el punto de vista de la previsión, los modelos, algoritmos, metodología y herramientas utilizadas para la previsión son totalmente distintos, según la topología de los clientes, debido a las características muy distintas de los mismos. Mientras que de los consumidores de alta tensión se dispone de valores horario de consumo y medidas en tiempo real, en los consumidores de baja tensión las lecturas son bimestrales y dependientes de unos perfiles horarios estándares publicados por Red Eléctrica Española (REE). En el caso concreto de este proyecto, el estudio y desarrollo del modelo se centrará en una muestra de los clientes de alta tensión de ENDESA.

Los modelos de previsión de la energía eléctrica han ido evolucionando y adaptándose a las distintas situaciones de mercado y regulaciones[3]. Como ejemplo, a continuación se muestran gráficas obtenidas con los datos disponibles en la página de Red Eléctrica Española (www.ree.es), en los que se muestra la evolución de la demanda desde el año 2004. Un ejemplo claro, con un impacto importante de este hecho, es el relacionado con la crisis económica actual, que ha dado lugar a caídas muy relevantes de la demanda y cambios de comportamiento en la misma.

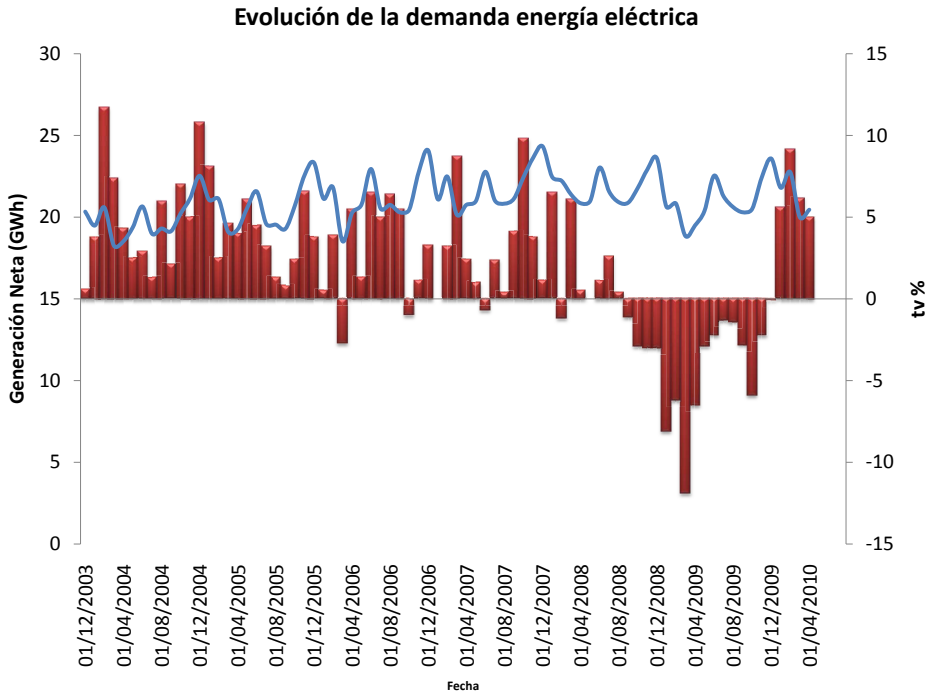

Figura 2. Evolución de la demanda de energía eléctrica peninsular.

En la Figura 2 se puede apreciar este efecto. La variación mensual del consumo de energía en tv %[4] es una serie con una ligera tendencia decreciente, que oscila entre

[3]En la bibliografía se realiza un breve repaso por la historia de los modelos de previsión.
[4]Tasa de variación respecto al mismo período del año anterior.

el 0 y el 10 %. Pero a partir de una fecha concreta y durante un periodo de tiempo, la variación cae abruptamente, para después continuar con su variación habitual.

Durante este periodo, las variaciones son siempre negativas, por lo que se produce un descenso en el consumo de energía, comparado con los periodos anteriores. Es la primera vez que se observa una situación en la que se producen caídas de demanda de esta magnitud. Este tipo de situaciones son claves y deben de ser consideradas de forma aislada en cualquier modelo de previsión.

Por esta razón, es necesario volver a diseñar los modelos para tener en cuenta este fenómeno, que se une a las variables que influyen de manera notoria en el consumo, como son el efecto del calendario y la temperatura. Un buen modelo debe contemplar toda la problemática que influye en la demanda de energía, para así conseguir un buen ajuste y llegar a hacer una predicción lo más cercana a la realidad.

1.2. Objetivos

El objetivo principal de este proyecto es investigar nuevos modelos de largo plazo que permitan realizar previsiones de la demanda eléctrica de una cartera de clientes industriales. Se va a estudiar, desarrollar e implementar una herramienta informática que estime la demanda de estos clientes en el medio-largo plazo. Dicha herramienta estará basada en modelos estadísticos que ajusten el comportamiento de los clientes, a partir de ciertas variables explicativas y deberá ser capaz de realizar una previsión de la demanda de energía a partir del pronóstico de las variables explicativas elegidas.

Como primer paso, se revisarán de forma general los modelos existentes para la previsión de la energía eléctrica, distinguiendo entre los modelos de corto, medio y largo plazo y haciendo especial hincapié en los modelos de medio-largo plazo, ya que es el horizonte temporal objeto de este proyecto.

Una vez revisados los modelos actuales, el segundo paso es organizar la información histórica disponible. Este punto es crítico para la correcta realización del proyecto, ya que es la base sobre la que se apoya el modelo estadístico. Se va a diseñar una base de datos en la que se organice la información sobre:

- **Clientes:** se almacenará toda la información relativa a los clientes.
- **Calendario:** se almacenarán las fechas que abarcan el periodo de estudio con resolución horaria y la información relevante de cara a su influencia en la demanda de energía eléctrica (festividades, verano, invierno...).

- **Consumos:** se dispondrán un conjunto de tablas donde se podrán consultar los consumos de cada cliente en un instante determinado.

A continuación se definirá y ajustará un modelo de previsión para consumidores industriales en que se combinen técnicas de *clustering* y modelos lineales. El modelo estará dividido en dos etapas:

Etapa 1: ***Clustering***: Mediante técnicas de *clustering* se agruparán los clientes según tengan características similares de consumo.

Etapa 2: Regresión: Se probarán distintos modelos estadísticos de regresión lineal con el fin de seleccionar aquel que presente el mínimo error en términos estadísticos. Para ello, se llevarán a cabo distintos análisis estadísticos para verificar la validez o no de las variables explicativas propuestas.

Una vez ajustado el modelo de previsión, es necesario comparar los resultados obtenidos con los datos originales, para poder ponderar la validez o no del modelo elegido. Se va a calcular el error cometido en el ajuste de los datos utilizando el modelo propuesto. De esta manera, se podrá comparar con los modelos existentes encontrados en la literatura y valorar la bondad del ajuste efectuado. Así pues se podrán sacar conclusiones acerca del modelo utilizado, analizando sus virtudes y mejorando las debilidades del mismo.

1.3. Recursos

Los principales recursos utilizados para la realización del proyecto se pueden dividir en dos grupos: *hardware* y *software*:

1.3.1. Hardware

Este proyecto es de carácter puramente informático, por lo que el único *hardware* necesario para su realización es un ordenador que cumpla con los requerimientos mínimos que exige el software utilizado.

1.3.2. Software

- **Microsoft Access y MySQL**
  Bases de datos utilizadas para almacenar los series históricas de consumo de electricidad.

- **Matlab**
  Software matemático utilizado para manipular los datos de las bases de datos.

Se han utilizado las funciones de la *toolbox database* para la conexión a las distintas bases de datos. También se ha utilizado para realizar distintas funciones para realizar operaciones con los datos, visualización de las series y análisis estadísticos.

- **Idat**
  Herramienta de análisis de datos desarrollada por el Instituto de Investigación Tecnológica del ICAI, capaz de crear distintos modelos de predicción de series temporales, así como modelos de decisión y mapas autoorganizativos de Kohonen.

- **Distribución de LaTeX**
  Software de distribución librea para el procesamiento, la maquetación y la redacción final de la memoria.

- **Paquete Microsoft Office 2007**
  Uso de los distintos programas como la hoja de cálculo de Excel o el programa de presentaciones PowerPoint.

Capítulo 2

Estado del arte

"En los momentos de crisis, sólo la imaginación es más importante que el conocimiento."
ALBERT EINSTEIN

A lo largo de la historia, las sociedades han ido evolucionando en torno a la tecnología y sus avances. El consumo de energía no ha sido inmune y la demanda de energía eléctrica ha ido aumentando con el paso del tiempo, fruto del creciente nivel de vida de las sociedades industrializadas, en las que cada vez más se depende de la energía eléctrica. El hecho de que las necesidades de energía no han sido siempre las mismas, ni la situación del sector eléctrico se ha visto inalterada con el paso de los años[5], ha repercutido en el campo de la previsión de la energía eléctrica. La evolución de las técnicas de previsión y los medios usados han estado determinados principalmente por el marco económico.

2.1. Historia

En los años 50 y 60 reinaba una situación de total estabilidad. Los precios estables de los combustibles, los bajos costes del capital, economías de escala aplicadas a las plantas de generación, una previsible tendencia demográfica y la inexistencia de la ley de la oferta y la demanda aplicada al sector eran algunas de sus principales características. Ante un escenario tan invariable, sencillas técnicas de predicción, como una simple extrapolación de la tendencia, eran suficientes para estimar los futuros picos de carga.

A partir de los años 70, los métodos para la predicción de la demanda de energía eléctrica han sufrido una profunda evolución. La primera crisis del petróleo en 1973[6]

[5]Es completamente distinto afrontar la previsión de la demanda de energía eléctrica en un mercado totalmente monopolizado y regulado, que en una situación liberalizada.

[6]La OPEP decidió dejar de exportar petróleo a los países que habían apoyado a Israel en la guerra del *Yom Kippur*, provocando un fuerte efecto inflacionista.

provocó una recesión económica mundial en los años 1974 y 1975 (en algunos países, como España, se prolongó más tiempo). Se pasó de una situación anterior de estabilidad a un escenario, donde la volatilidad de los precios de la energía y la inesperada elasticidad de la demanda de los consumidores a los crecientes precios de la energía eléctrica eran protagonistas. La planificación de la energía eléctrica adquirió un gran protagonismo dentro de las empresas del sector eléctrico. Estas enormes incertidumbres hicieron necesario el desarrollo y aplicación de técnicas mucho más sofisticadas.

No es casualidad que durante este periodo, se produjera un importante avance de la teoría estadística en lo referente a las técnicas de análisis de series temporales. El gran salto en el estudio de las series temporales se produjo en 1976, con la publicación del libro *Time Series Analysis: Forecasting and Control* de los autores G.E.P. Box y G.M. Jenkins. Su aparición dio nombre a la *metodología Box-Jenkins*. Esta metodología ha demostrado ya grandes resultados en su aplicación a los diferentes campos del análisis económico, como por ejemplo el sector energético.

A partir de 1990 se extendió el proceso de la desregularización, lo que supuso un punto de inflexión, convirtiendo la energía eléctrica en una mercancía más dentro de un mercado gobernado por la ley de la oferta y la demanda, donde los consumidores se lanzan a la búsqueda de los productos más competitivos, justo lo contrario que ocurría durante los años 50 y 60.

Los errores en la predicción del consumo de energía pueden acarrear una carencia o superabundancia de energía, lo que se traduce en costes extras que añadir a los precios de la energía. Estos errores no son admisibles, ya que lastran los precios en un mercado muy competitivo. Por ello, las empresas empiezan a invertir en el desarrollo e investigación de diversas técnicas avanzadas de predicción de electricidad, para obtener la ventaja competitiva de conocer el comportamiento del consumo de electricidad en el futuro con el mínimo error posible.

2.2. Horizontes temporales de previsión

El primer paso antes de construir un modelo es elegir el horizonte temporal para el cual se pretende realizar previsión de la demanda de energía eléctrica. La mayoría de autores coinciden en dividir los horizontes en tres diferentes grupos, pero no se ponen de acuerdo en la duración de cada intervalo. Se va a seguir la clasificación propuesta en [24] y las características de [3].

2.2.1. Corto Plazo

Aquí se pueden distinguir a su vez dos subgrupos: el **muy corto plazo**, que abarca los próximos 30 minutos en tiempo real a partir del momento en que se efectúa la predicción; y el **corto plazo**, que se extiende desde una hora hasta una semana en el futuro.El objetivo principal que se intenta alcanzar es la planificación de las operaciones tácticas y para realizar las decisiones del día a día (operaciones en los mercados diarios).

2.2.1.1. Factores influyentes corto plazo

La demanda de energía eléctrica tiene una naturaleza variable con el tiempo y afectada por multitud de factores (sociales, meteorológicos, económicos ...). Si el horizonte temporal es pequeño, las condiciones sociales y económicas no tienen influencia en las series. Por otra parte, el tiempo tiene una influencia fundamental en las series. Los patrones de consumo durante el fin de semana, vacaciones y días festivos difieren de los patrones durante los días ordinarios. El máximo valor de consumo coincide cerca del extremo máximo o mínimo de temperatura, dependiendo de si se usa la energía para calefacción o refrigeración. Habitualmente, la curva de demanda horaria sigue el perfil de temperaturas diarias. Las variables a incluir en este grupo de modelos son la hora del día, la temperatura horaria, la humedad horaria, la velocidad del viento horaria, etc.

2.2.2. Medio Plazo

Se entiende por medio plazo la previsión que abarca un periodo desde un mes hasta un año a partir del momento en que se estima la futura demanda. Esta previsión es necesaria para establecer el calendario de mantenimiento de las plantas generadoras y del sistema de transmisión, así como el aprovisionamiento de combustible en las centrales térmicas.

2.2.2.1. Factores influyentes medio plazo

Al medio plazo le influyen más factores a parte de los meramente climatológicos o los derivados del calendario. Las condiciones socio-económicas juegan un papel muy importante en este escenario, tales como la renta per cápita, los tipos de interés, etc. así como variables demográficas y tecnológicas.

2.2.3. Largo Plazo

El largo plazo suele abarcar la predicción desde uno a diez años en el futuro, aunque es variable dependiendo de los autores que se consulten. Es una previsión de

vital importancia para planificar la expansión del sistema generador y el sistema de transporte, y en la planificación de la tarifa. Una correcta modelización y predicción del consumo a largo plazo puede anticipar la capacidad de generación óptima y de ahí inferir la combinación de potencia generadora más conveniente con la que se debe contar del parque generador.

2.2.3.1. Factores influyentes largo plazo

A medida que crece el horizonte temporal, es necesario incluir nuevas variables explicativas. Hay que añadir a las citadas en el medio plazo, el precio de la electricidad, el Producto Interior Bruto (PIB) así como los indicadores económicos más relevantes.

2.3. Métodos de previsión

Como se ha visto anteriormente, las técnicas para la previsión de la demanda de energía eléctrica a evolucionado de manera muy significativa a partir de 1970. Estos métodos, muy influenciados por el horizonte temporal a considerar, se pueden agrupar de la siguiente manera:

2.3.1. Métodos cualitativos

Se emplean para aquellas situaciones en las que el pasado no proporciona información directa sobre el fenómeno considerado. Son métodos subjetivos que se basan en estimaciones y opiniones, razón por la cual es difícil cuantificar la precisión del pronóstico.

2.3.2. Métodos cuantitativos

Se basan en la idea de que se pueden usar los datos históricos para inferior los futuros. El objetivo es extraer toda la información posible contenida en los datos y, en base al patrón de conducta seguido, realizar estimaciones.

2.3.2.1. Regresión Lineal

Es un método matemático que modeliza la relación entre la variable dependiente $y(t)$ (demanda de electricidad) y las variables independientes $x_i(t)$ (día de la semana, mes del año,temperatura ...) y un término aleatorio $\varepsilon(t)$. Se puede expresar según la ecuación Ecuación 2:

$$y(t) = a_0 + a_1 x_1(t) + ... + a_n x_n(t) + \varepsilon(t) \tag{2}$$

Donde $a_0, a_1, ..., a_n$ son los coeficientes de la regresión. Estos coeficientes suelen ser estimados por el método de minimos cuadrados.

Este modelo es el usado por [19] para predecir la demanda de energía eléctrica, una vez que la serie temporal ha sido corregida del efecto de la laboralidad. Las variables independientes utilizadas en este modelo son la temperatura[7], un indicador de actividad económica[8] y una variable del número de viviendas[9].

2.3.2.2. Modelos ARIMA

Metodología formalizada por Box y Jenkins en 1976, según se comentó anteriormente. Este enfoque se basa en el hecho de que la serie temporal que se trata de predecir es generada por un proceso estocástico, cuya naturaleza puede ser caracterizada mediante un modelo. Para ello, se necesita una serie con un elevado número de observaciones.

Este tipo de modelos se han convertido en la opción por defecto para realizar predicciones en el corto plazo.

Esta es la metodología adoptada por [20]. Los autores de [4, 6] lo utilizan para comparar con métodos más modernos de previsión.

2.3.2.3. Modelos de Descomposición

Los métodos clásicos de descomposición (Holt-Winters) consideran que la serie temporal se pude descomponer en los siguientes componentes:

- **Tendencia**: componente que recoge la tendencia de la serie a largo plazo.
- **Factor cíclico**: curvas de largo periodo asociadas con ciclos económicos.
- **Estacionalidad**: movimientos ascendentes y descendentes de las series que se repiten periódicamente (diario, semanal, anual ...).
- **Componente irregular**: fruto de factores aleatorios que inciden en de forma aislada y no permanente en una serie. Es la parte que no se puede explicar. También se denomina residuo.

[7] tratando la no linealidad de la respuesta al consumo eléctrico usando un modelo de regresión de doble umbral

[8] variable sintética obtenida tras la agrupación en sectores económicos

[9] indica la tendencia creciente a largo plazo en el consumo

Las series son el resultado de la integración de estos componentes, bien de modo aditivo (las fluctuaciones no se ven afectadas por la tendencia) o de modo multiplicativo (las fluctuaciones varían con la tendencia).

Este tipo de modelo se plantea en [22] para la previsión del consumo de gas. En [12] se propone una nueva forma de utilizar los modelos de descomposición. En lugar de utilizar los elementos anteriormente comentados, aplica un algoritmo que va descomponiendo en varias componentes la serie hasta encontrar la tendencia.

2.3.2.4. Modelos de redes neuronales

Estos modelos se caracterizan por ser capaces de modelar automáticamente relaciones no lineales entre las variables, pero tienen el inconveniente de ser cajas negras, al ser muy complejas de interpretar. Han sido aplicadas extensamente en la práctica, existiendo mucha documentación en la comunidad científica.

Capítulo 3

Descripción del modelo de previsión

"Si buscas resultados distintos, no hagas siempre lo mismo."
ALBERT EINSTEIN

DURANTE este capítulo se va a presentar el modelo propuesto en este trabajo para cumplir los objetivos del Apartado 1.2 y ajustarlo para una cartera de clientes industriales, que como ya se ha adelantado en el Capítulo 1, está basado en dos etapas fundamentales: (1) *Clustering* y (2) Ajuste de modelos de regresión.

3.1. Introducción

Son múltiples los factores que influyen en el consumo de energía eléctrica, pero toda la literatura consultada coincide en que existen dos variables claves, que son de uso obligatorio en cualquier modelo de previsión: la temperatura y la laboralidad. Sin embargo, a pesar de que todos los modelos de previsión deben considerar dichas variables, el impacto de las mismas puede ser radicalmente distinto dependiendo de la topología de clientes que se estén modelando.

Como es natural, la sensibilidad a la temperatura no es la misma para un gran cliente industrial que para un modesto consumidor doméstico. De igual manera, el impacto que un día festivo del calendario tiene sobre los consumidores industriales, es muy diferente al de un consumidor doméstico. Este hecho que se entiende de forma clara entre dos tipos de clientes tan dispares, también hay que considerarlo entre distintos tipos de clientes industriales. Más aún, perteneciendo al mismo sector de actividad, el impacto no sólo de la temperatura o la laboralidad, sino de otros factores externos, puede ser muy distinto entre dos clientes industriales.

Por tanto, si se plantea un modelo de previsión de todos los clientes agregados, serían necesarios un número elevadísimo de variables explicativas para poder captar las diferentes casuísticas del perfil de consumo de éstos, respetando el impacto que cada variable pueda tener en cada uno de ellos.

En el lado opuesto se encontraría la opción de ajustar un modelo de demanda de forma individual para cada uno de los clientes. Sin embargo, esto puede resultar muy costoso, e incluso ineficiente, puesto que supondría en muchos casos desarrollar miles de modelos de previsión.

El modelo que se describe en este documento, se centra en el consumo de los clientes industriales y propone una solución intermedia, basada en una primera etapa, muy importante y fundamental para el desarrollo del mismo, de clasificación de los clientes atendiendo a sus principales características de consumo; y una segunda etapa, donde se procederá al ajuste de los modelos de regresión para cada una de las series agregadas de consumo, fruto de la agrupación de la etapa previa.

Con el objetivo de describir de forma ordenada y estructurada cada una de las partes que componen el modelo propuesto, en la Figura 3 se presenta un esquema global del mismo. El modelo se divide en tres partes claramente diferenciadas:

1. **Variables**. En este primer bloque se explicarán los datos disponibles para el ajuste del modelo. Por un lado se detallan los datos directamente relacionados con los clientes, esto es: consumo, Comunidades Autónomas (CCAA), observatorio meteorológico al que pertenecen, etc. Por otro lado se describirán las variables que especifican las festividades y las medidas de temperatura.

2. ***Clustering***. En esta parte, se realizará una agrupación de clientes utilizando técnicas de *clustering*, atendiendo a sus perfiles de consumo, variable que se definirá más adelante. Tras realizar la agrupación, se va a proceder al cálculo de las series agregadas necesarias para el ajuste de los modelos. Para cada grupo se calculará las serie temporal de consumos, de temperatura y de festividades agregadas, además de calcular la evolución del número de clientes durante el periodo de estudio.

3. **Modelos de regresión**. El último paso es ajustar un modelo de regresión lineal para cada una de las series agregadas de consumo calculada en cada grupo. Estos modelos utilizarán las variables calculadas en el apartado anterior (número de clientes, temperatura, festividades), además de las variables indicadoras que hacen referencia al calendario (mes del año, Semana Santa, Navidad, etc) para ajustar

cada una de las series agregadas de consumo. Por último, se obtendrá la previsión de la demanda de energía de la cartera de clientes de la comercializadora, sumando las estimaciones de cada una de las series temporales de demanda de cada uno de los grupos.

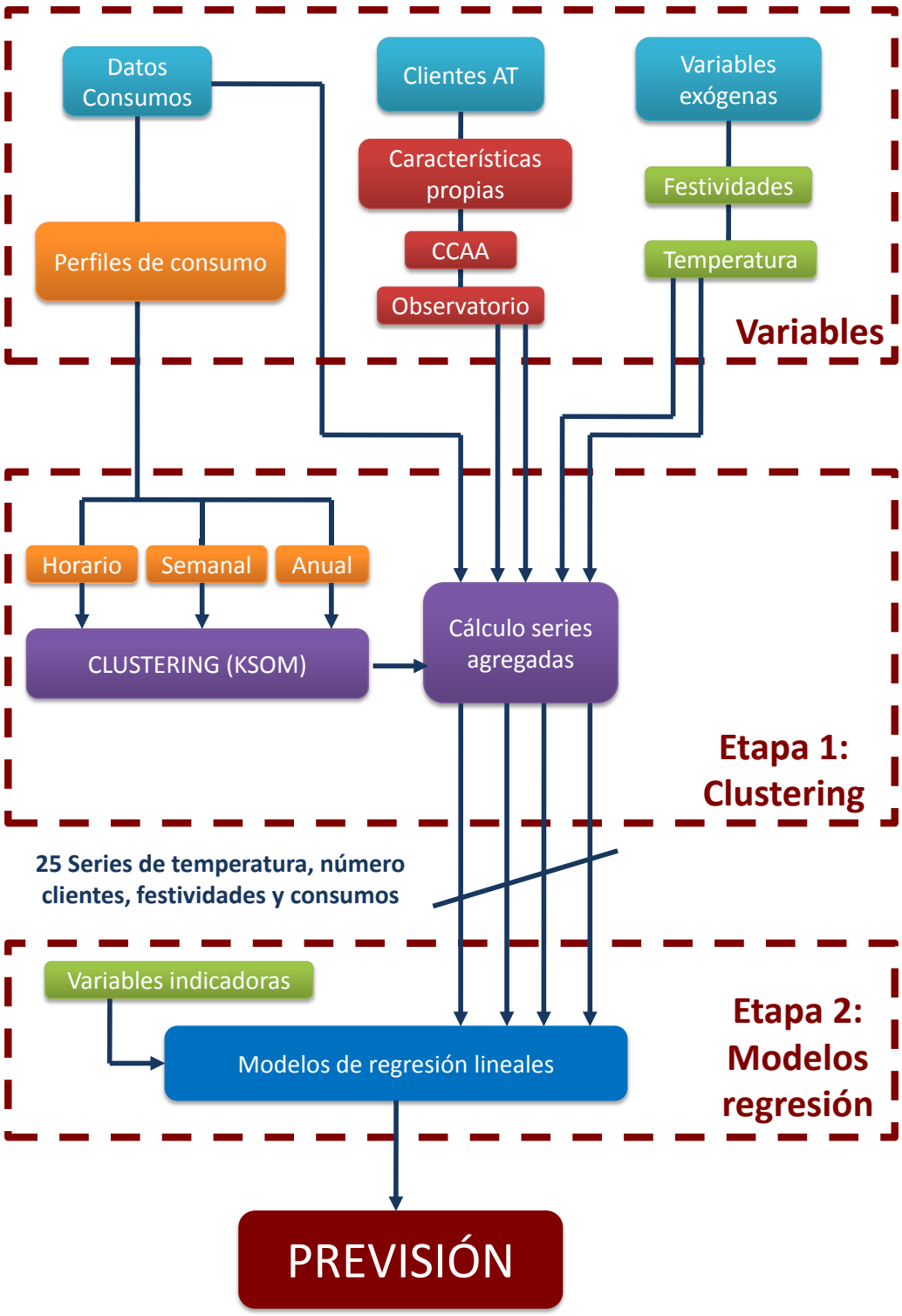

Figura 3. Esquema modelo previsión

3.2. Variables

El primer módulo que aparece en la Figura 3 corresponde con las variables disponibles que se van a utilizar en el modelo de regresión. Estas variables son la fuente de donde se sacará toda la información necesaria para realizar las dos etapas del modelo: *clustering* y modelos de regresión.

3.2.1. Datos consumo

Como ya se ha comentado en la introducción, se dispone de una base de datos con los consumos de los clientes. Los clientes de alta tensión cuentan con medidas en tiempo real y con detalle horario. Por esta razón, los consumos disponibles en las bases de datos tienen resolución horaria[10] para cada uno de los clientes.

Usando estos datos horarios de consumo, se van a proceder al cálculo de los perfiles de consumo (el procedimiento empleado para su obtención se detallará más adelante). Estos perfiles de consumo son un punto clave en el desarrollo del proyecto, ya que permiten identificar a cada cliente según la manera de consumir la energía eléctrica. De esta forma, se podrán clasificar a los clientes en la etapa 1 de *clustering*.

3.2.2. Clientes Industriales

Para obtener una predicción razonable utilizando un modelo explicativo es imprescindible identificar para cada cliente el conjunto de variables explicativas clave que puedan explicar el consumo de energía. Según se ha comentado, las variables principales que influyen de forma más determinante son la laboralidad y la temperatura. Sin embargo, es necesario también considerar otro conjunto de variables que nos ayuden a construir mejores modelos de previsión.

En este sentido, las variables explicativas utilizadas se pueden dividir en dos grupos: variables explicativas directas y perfiles de consumo. Se van a denominar como variables explicativas directas aquellas que no requieren ninguna operación de cálculo para su obtención, como por ejemplo la distribuidora a la pertenecen, la comunidad autónoma donde están localizados, el volumen de energía consumida a lo largo del año (consumo anual) o el sector económico al que pertenecen (Clasificación Nacional de Actividades Económicas CNAE[11]). Dentro de este grupo resultan especialmente importantes dos

[10] Debido a los cambios de hora producidos en el calendario, cada año hay dos días especiales, uno de 23 horas y otro de 25 horas.

[11] La Clasificación Nacional de Actividades Económicas de España permite la clasificación y agrupación de las unidades productoras según la actividad que ejercen de cara a la elaboración de estadísticas.

características, las comunidades autónomas y los observatorios, por su importancia en la construcción de las series temporales de temperatura y festividades.

El segundo grupo está relacionado con la forma de consumir que presentan los clientes. Se van a calcular los perfiles de consumo que presentan los clientes a lo largo del día, de la semana y del año. Estas variables tendrán gran importancia en la primera etapa de *clustering*.

3.2.2.1. Comunidades Autónomas

De las características más importantes que se disponen de los clientes se encuentra la comunidad autónoma a la que pertenecen. La mayor parte de los clientes de la muestra escogida se encuentran en dos comunidades CCAA1 y CCAA2, mientras que el resto de la muestra quedan repartidos entre las restantes comunidades autónomas.

Las comunidades autónomas son una característica muy importante de los clientes a la hora de calcular las series temporales de las festividades. Cada comunidad autónoma tiene festividades propias además de las festividades nacionales, que son comunes a todas ellas. Por lo tanto, a la hora de construir las series temporales de las festividades de cada grupo (que se detallará más adelante), será importante conocer la distribución de clientes de ese grupo según las diferentes comunidades autónomas.

3.2.2.2. Observatorios

Cada cliente tiene asociado uno de los 48 observatorios que se han utilizado en este modelo, atendiendo a la proximidad geográfica. Nótese que hay más observatorios que comunidades autónomas, puesto que hay comunidades muy extensas, donde se pueden encontrar distintas climatologías, que pueden ser muy dispares entre sí. Existen comunidades autónomas, como Andalucía, donde varía mucho la climatología, ya que existen zonas costeras con temperaturas cálidas, mientras que en las zonas cercanas a la montaña tienen temperaturas más frías.

La importancia de los observatorios radica en la construcción de las series temporales de temperatura de cada grupo. No pesa lo mismo un observatorio que otro dentro de un grupo a la hora de construir las series de temperaturas, ya que el número de clientes del grupo pertenecientes a cada observatorio es distinto. De esta manera, será necesario conocer la evolución del número de clientes dentro de cada grupo que pertenecen a los diferentes observatorios.

3.2.3. Variables exógenas

El último grupo de datos que se van a utilizar es el que se ha llamado en la Figura 3 variables exógenas. En los modelos de previsión de la demanda de energía, los factores externos más influyentes son la temperatura y las festividades.

3.2.3.1. Temperatura

Como se ha comentado anteriormente, cada cliente tiene asociado un observatorio. Estos observatorios suministran datos de temperatura diarios acerca de la temperatura mínima y máxima, tanto real como prevista. De todos estos datos, se va a calcular una temperatura media diaria, utilizando las máximas y las mínimas reales. Cada cliente está influenciado por la serie temporal del observatorio que tiene asociado. Estas series de temperatura se utilizarán para calcular las series de temperatura agregadas de cada grupo, cuyo cálculo se detallará más adelante.

3.2.3.2. Festividades

Existen dos tipos de festividades con las que se va a trabajar: nacionales y autonómicas[12]. En primer lugar están las festividades nacionales, que son comunes a todas las comunidades autónomas. En el segundo grupo se encuentran las festividades autonómicas. Estas festividades solamente son comunes a las ciudades y provincias pertenecientes a la comunidad autónoma, fruto de su herencia histórica y cultural. Usando estas fechas se van a construir las series temporales indicadoras de las festividades agregadas de cada grupo. El detalle del cálculo se realizará en el apartado siguiente.

3.3. Etapa 1: *Clustering*

Una vez que se han explicado los distintos datos y variables disponibles para la construcción del modelo, el siguiente paso es realizar la agrupación de la muestra de clientes en distintos grupos, con similares características. Para realizar el *clustering*, se van a utilizar los "Mapas autoorganizativos de Kohonen". Una vez realizado la agrupación de la cartera de clientes, se calcularán las distintas series agregadas necesarias para ajustar los modelos de regresión en la siguiente etapa.

3.3.1. Perfiles de consumo

Anteriormente se han analizado las variables que más importancia tienen para la construcción del modelo. Sin embargo, se ha comprobado en este proyecto que

[12]Existen más festividades, como por ejemplo las festividades locales, pero no han sido consideradas.

utilizando únicamente las características de los clientes resultan insuficientes para realizar una clasificación satisfactoria de los clientes, por lo que se ha recurrido a la construcción de unas nuevas variables: los **perfiles de consumo**.

Para construir un modelo predictivo con discriminación horaria interesa conocer tanto el volumen anual del consumo de energía eléctrica como su evolución temporal. Esta importancia radica en que dos clientes pueden tener consumos anuales similares, pero la forma en que consumen esa electricidad a lo largo del año, de la semana o incluso a lo largo de las horas del día, puede ser totalmente distinta. Por ejemplo, dos clientes pueden tener el mismo valor de consumo diario, pero su forma de consumo puede ser muy dispar, consumiendo uno la electricidad durante las horas de la mañana y la tarde, mientras que el otro cliente, debido a la naturaleza de su actividad consume la energía eléctrica durante las horas nocturnas. Este fenómeno se puede reproducir de igual manera ampliando el periodo de observación (semana, mes, año).

Cada cliente tiene una forma habitual de consumir a lo largo del día, de la semana o del año. Estas formas de consumo son características propias de cada cliente y pueden utilizarse para identificarlos dentro de la extensa cartera de la comercializadora. Conociendo las distintas formas de consumo, se va a realizar una clasificación utilizando estas características para formar los diferentes grupos.

El uso de los perfiles de consumo para realizar la clasificación presenta una ventaja importante de cara al ajuste de los modelos de regresión. De esta manera, en cada grupo del *clustering* se encontrarán los clientes que consumen de forma parecida, a pesar de las diferencias en los volúmenes de energía.

Debido a la gran influencia que tiene la temperatura en el consumo de energía, se va a prestar atención a las estaciones del año a la hora de construir los perfiles. Durante estos periodos del año, las condiciones climáticas se mantienten dentro de un cierto rango. Las fechas de inicio y fin de las estaciones se recogen en la siguiente Tabla 2. Los datos se han obtenido de la web del Instituto Geográfico Nacional.

Año	Estación	Fecha Inicio
2003	Primavera	21 marzo
	Verano	21 junio
	Otoño	23 septiembre
	Invierno	22 diciembre
2004	Primavera	20 marzo
	Verano	21 junio
	Otoño	22 septiembre
	Invierno	21 diciembre
2005	Primavera	20 marzo
	Verano	21 junio
	Otoño	22 septiembre
	Invierno	21 diciembre
2006	Primavera	20 marzo
	Verano	21 junio
	Otoño	23 septiembre
	Invierno	22 diciembre
2007	Primavera	21 marzo
	Verano	21 junio
	Otoño	23 septiembre
	Invierno	22 diciembre
2008	Primavera	20 marzo
	Verano	21 junio
	Otoño	22 septiembre
	Invierno	21 diciembre
2009	Primavera	20 marzo
	Verano	21 junio
	Otoño	22 septiembre
	Invierno	21 diciembre

Tabla 2. Fechas Inicio y Fin estaciones

Se va a considerar que el año sólo se compone de dos estaciones dependiendo de la temperatura. El "Invierno" quedará constituido por las estaciones de otoño e invierno, puesto que son aquellos en los que las temperaturas son más bajas, mientras que en primavera y verano se pueden encontrar las temperaturas más altas, por lo que constituirán la estación del "Verano".

3.3.1.1. Perfil diario

Se define el perfil diario[13] como *"la mediana del consumo de la hora i normalizada con el consumo total del día"*. Para cada hora del día se calcula el consumo normalizado, dividiendo el consumo de esa hora entre el consumo total de ese día. Después se calcula como valor típico de cada hora la mediana de los valores normalizados para todo el conjunto de datos de estudio. Esta definición se concreta en la Ecuación 3 para el invierno y en la Ecuación 4 para el verano.

$$h_i^{inv} = mediana(H_i^{inv}) \qquad i = 1, 2, \cdots, 24 \tag{3}$$
$$\text{donde} \qquad H_i^{inv} = \frac{hora_i^{inv}}{\sum_{j=1}^{24} hora_j^{inv}}$$

$$h_i^{ver} = mediana(H_i^{ver}) \qquad \forall i = 1, 2, \cdots, 24 \tag{4}$$
$$\text{donde} \qquad H_i^{ver} = \frac{hora_i^{ver}}{\sum_{j=1}^{24} hora_j^{ver}}$$

De esta forma, cada cliente queda caracterizado por 48 valores de consumo normalizados (24 valores para los periodos de invierno y 24 valores para los periodos de verano). Estos valores constituyen el **perfil diario** de cada cliente, es decir, la forma más habitual de consumo de un cliente a lo largo del día. Esta nueva variable queda definida en la Ecuación 5, combinando en un vector los valores obtenidos tras aplicar la Ecuación 3 y la Ecuación 4.

$$Perfil_horario_j = (h_1^{inv}, h_2^{inv} ... h_{24}^{inv}, h_1^{ver}, h_2^{ver} ... h_{24}^{ver}) \tag{5}$$
$$\forall j \in Clientes$$

A continuación se van a comentar varios perfiles de consumo diario obtenidos tras la representación de la variable definida en la Ecuación 5. Para poder apreciar las posibles diferencias entre los meses más calurosos y los más fríos, se van a representar los valores de invierno y verano por separado.

[13] Se ha elegido el miércoles siguiendo lo propuesto en [19] como día de referencia, por ser típicamente el más homogéneo de la semana, por ser el día medio del ciclo semanal y ser el de mayor consumo

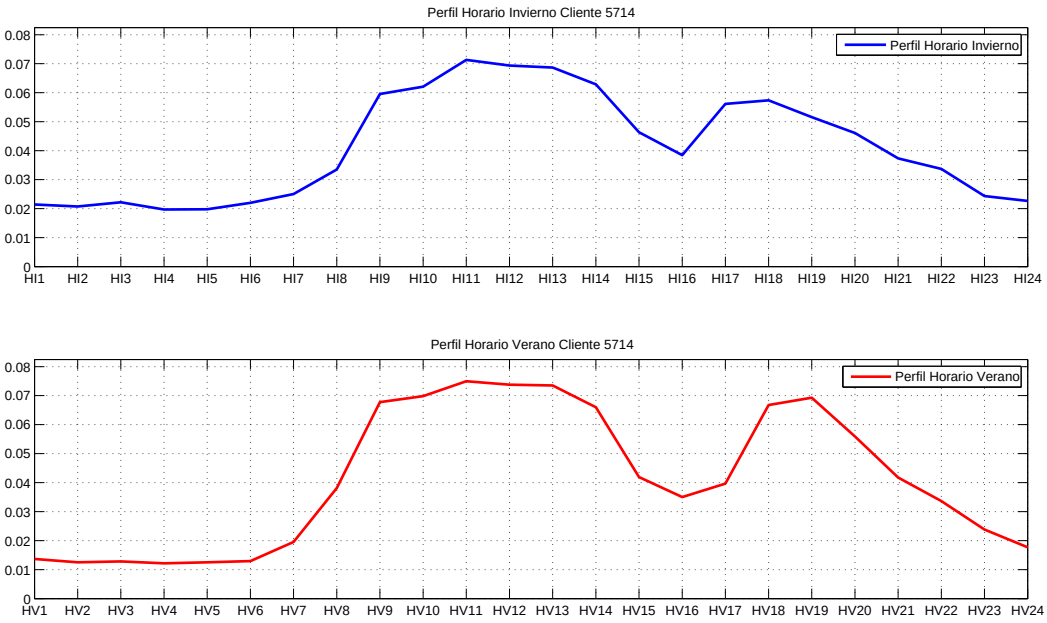

Figura 4. Ejemplo perfil diario 1.

El primer perfil que aparece con cierta frecuencia entre los clientes es el representado en la Figura 4 y muestra una de las formas más típicas de consumo de energía eléctrica a lo largo del día. El consumo se mantiene en valores mínimos durante la noche, mientras que durante las horas diurnas, el consumo es mucho mayor, especialmente durante la mañana (de 9h a 14h) y durante la tarde (de 15h a 20h).

Este comportamiento en la forma de consumir parece lo más lógico que se cabría esperar *a priori*: la mayor parte de la energía se consume durante las horas en que la gente está en horario laboral, las fábricas comienzan a funcionar a plena carga y los comercios abren sus puertas al público. Existe un valle en el consumo aproximadamente alrededor de las 15h, que se corresponde aproximadamente con la hora de comer, tiempo durante el cual la gente deja sus puestos de trabajo. Nótese que el perfil de verano es ligeramente diferente del de invierno, produciéndose un incremento del consumo de energía durante las horas de tarde-noche del verano, posiblemente debido al mayor uso de los equipos de refrigeración.

Este perfil es prácticamente idéntico a la forma en que consume España a lo largo del día. En la Figura 5 se muestra el consumo de energía de España a lo largo de las 24 horas del día 5 de junio de 2009. Este gráfico puede consultarse a través de la web de REE, donde se muestra la evolución del consumo horario de España en tiempo real.

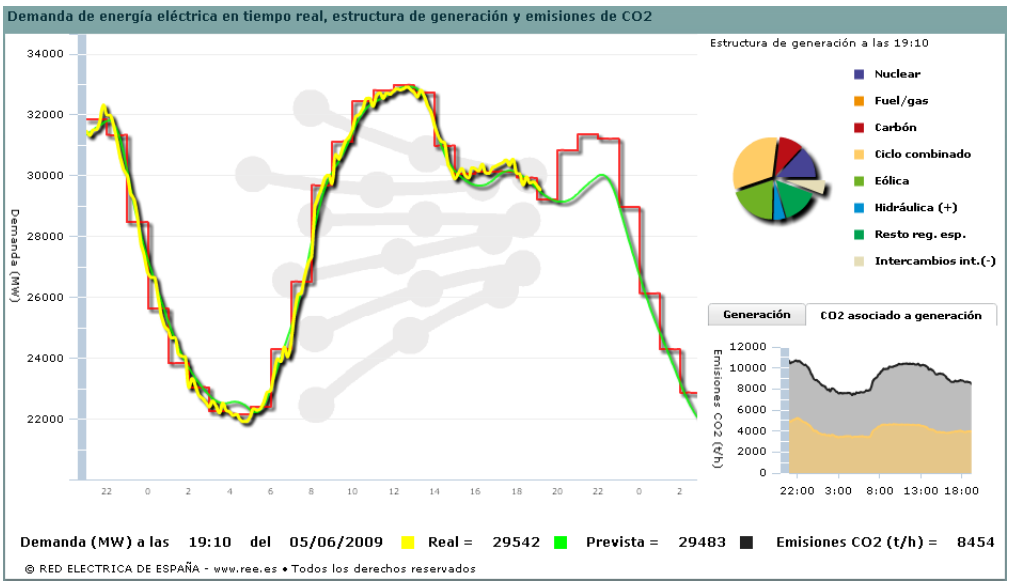

Figura 5. Perfil diario 5 junio de 2009.

Otro perfil que aparece con bastante frecuencia es el mostrado en la Figura 6. Es muy similar al comentado anteriormente, se sigue manteniendo el consumo al mínimo durante las horas nocturnas, al contrario que durante las hora diurnas, que se consume la mayor parte de la energía.

La principal diferencia existente con el perfil anterior es que no existe el valle alrededor de las 15h, sino que se consume de forma prácticamente constante entre las 9h y las 20h, observándose una pequeña disminución a partir de las 16h.

Este tipo de cliente no reduce su consumo en ningún momento a lo largo del día, por lo que se puede concluir que no se deja de trabajar durante todo el horario laboral. Los perfiles de invierno y verano son prácticamente idénticos, sin encontrar diferencias notables entre los consumos para la calefacción o la refrigeración.

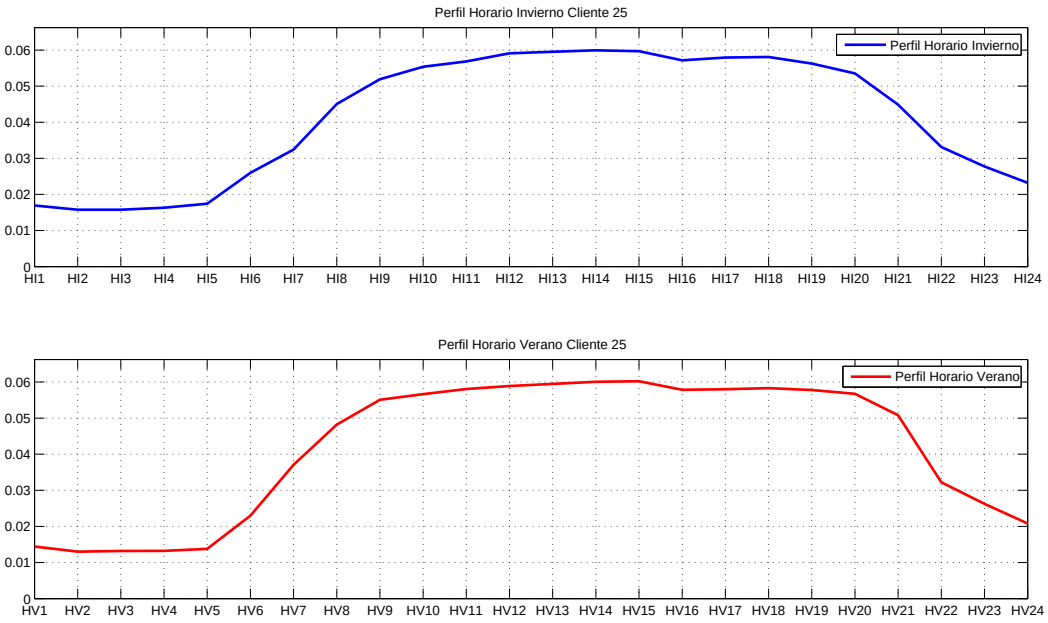

Figura 6. Ejemplo perfil diario 2.

En el extremo opuesto, podemos encontrar perfiles como el de la Figura 7. El consumo nocturno sigue situándose en valores mínimos, pero ahora el valle que se produce alrededor de las 15h es mucho más pronunciado, dividiendo claramente la forma del consumo entre las horas de la mañana y las de la tarde.

Se puede observar como los consumos durante la noche son prácticamente nulos, por lo que podemos estar hablando de clientes cuya actividad nocturna es prácticamente nula, posiblemente oficinas o fábricas que solo necesitan de la energía eléctrica fuera de la jornada laboral para mantener con un mínimo de energía los equipos.

Este comportamiento es justo al contrario que, por ejemplo, empresas de alimentación, que necesitan mantener cierto consumo energético para alimentar los grandes frigoríficos que almacenan alimentos. Al igual que en el perfil anterior, las diferencias entre invierno y verano son mínimas, pudiéndose confundir ambos perfiles si se superponen.

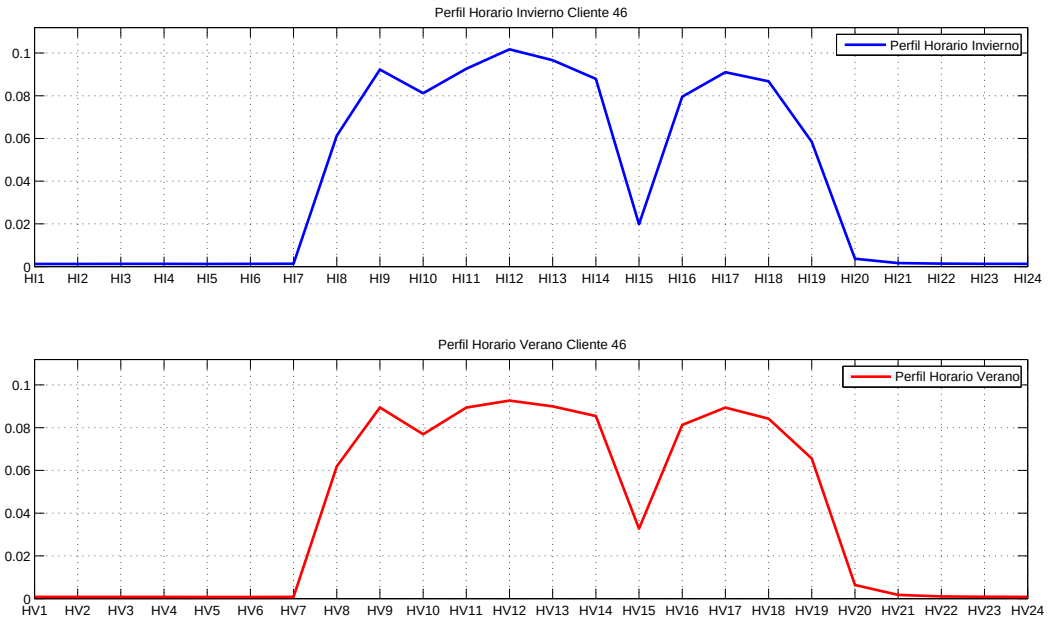

Figura 7. Ejemplo perfil diario 3.

Por último, también se pueden encontrar perfiles como el representado en Figura 8. Al contrario que los anteriormente comentados, este es un consumo prácticamente constante a lo largo de todo el día, sin distinción entre la mañana o la tarde, ni descansos para comer, ni menor consumo durante las horas nocturnas.

Estos perfiles corresponden a clientes industriales, posiblemente grandes fábricas, que no paran su producción a lo largo del día, debido a los altos costes de arranque y parada de sus equipos, y que organizan a su plantilla en turnos, para que siempre pueda haber gente trabajando de manera continuada y sin interrupciones. Al igual que los perfiles anteriores, no existen grandes diferencias entre los perfiles de invierno y verano.

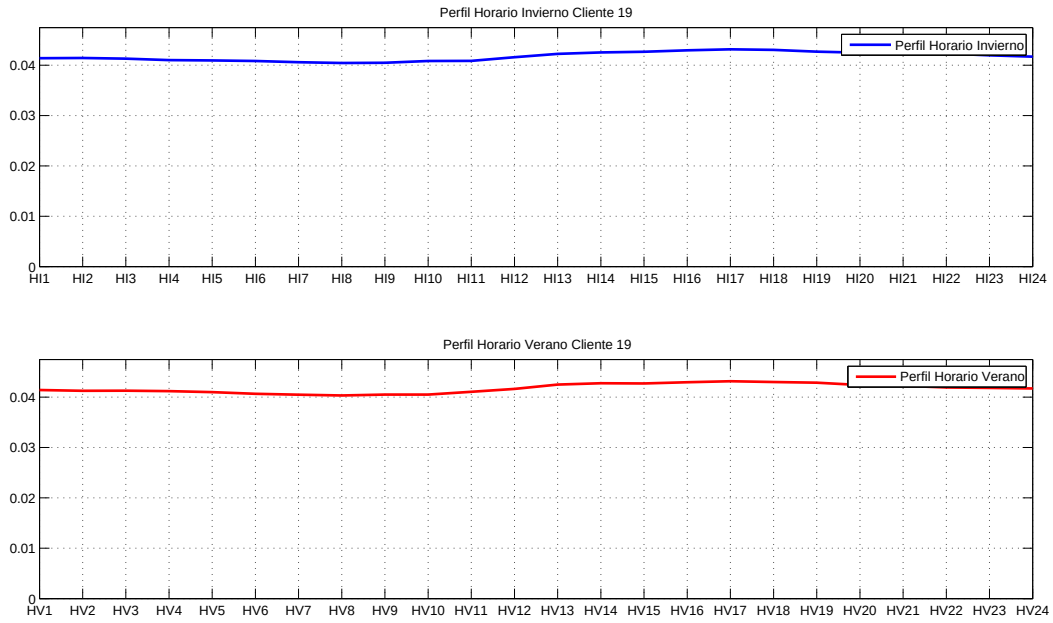

Figura 8. Ejemplo perfil diario 4

Una vez generados los perfiles horarios, se está en disposición de analizar la información contenida en esta nueva variable. Se espera que los clientes que presenten formas de consumir horarias parecidas en la realidad, tengan perfiles horarios, según la definición de Ecuación 5, también similares. Estos perfiles horarios se han utilizado para clasificar a los clientes atendiendo a la forma de consumir a lo largo de un día típico de invierno y verano. Para ello se ha utilizado una técnica automática de *clustering* denominada "Mapas autoorganizativos de Kohonen". Estos mapas son redes neuronales que agrupan los datos de entrada en *clusters* o grupos con las mismas características.

Tras introducir los datos y ajustar el modelo del "Kohonen Self-Organized Map" KSOM para 25 grupos, el resultado es el que se muestra en la Figura 9. En ella, se muestran 2 tipos de gráficas: unas correspondientes a los grupos que se han formado y otra correspondiente a la representación de los datos en el espacio de las dos primeras componentes principales. A continuación se van a explicar estos dos tipos de gráficas, detallando los aspectos más importantes de cada una de ellas.

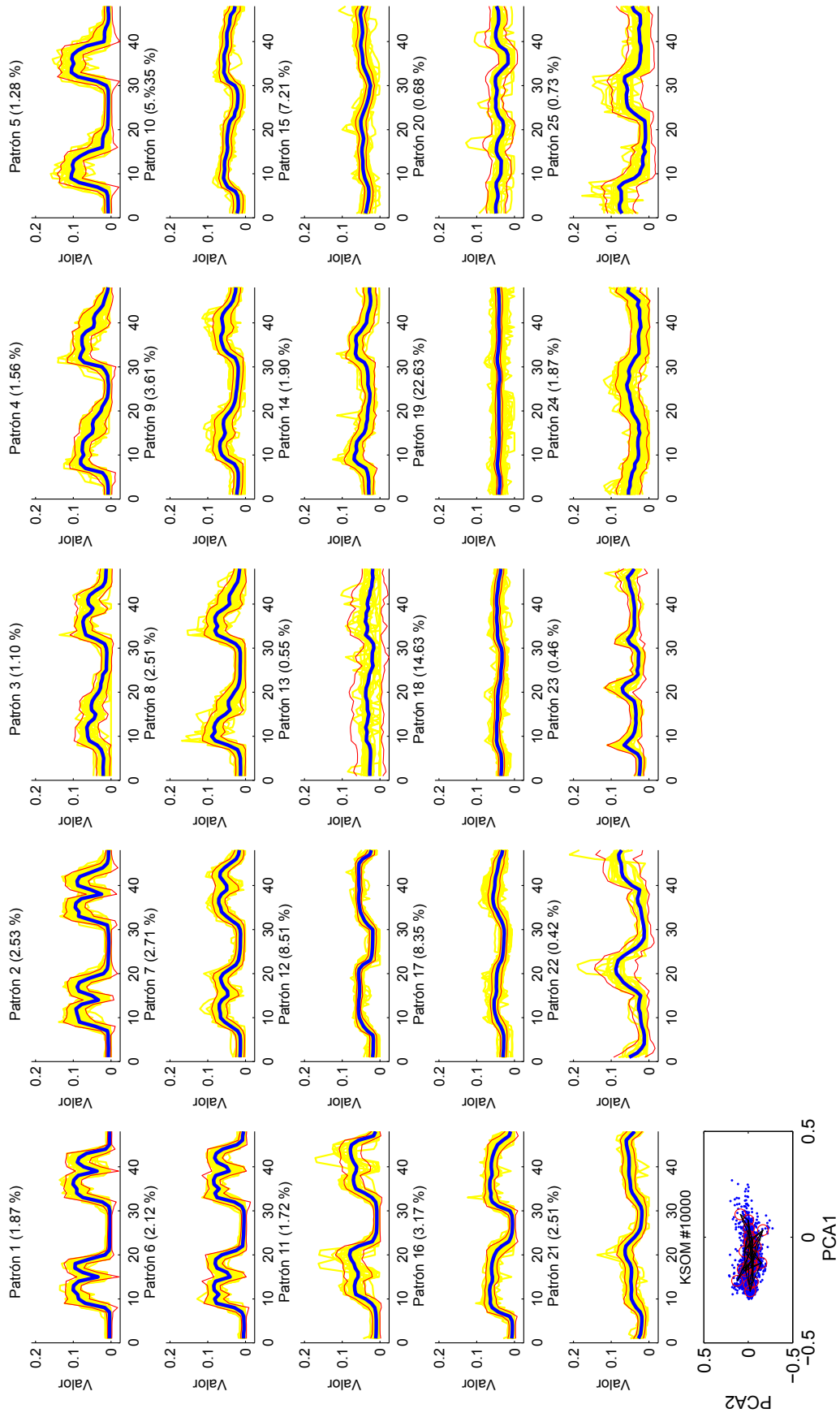

Figura 9. KSOM 5x5 perfil diario.

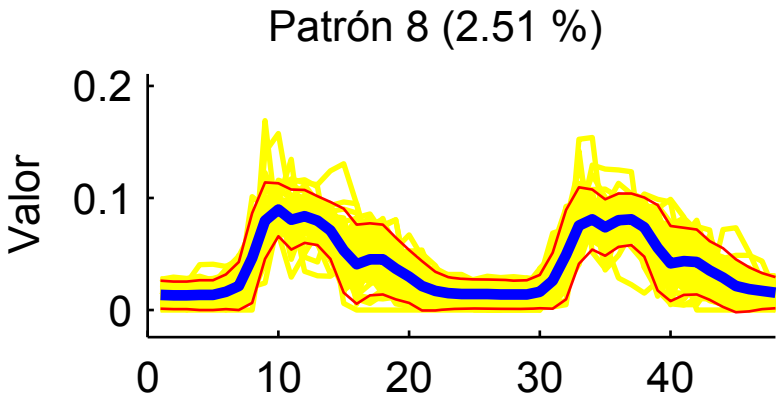

Figura 10. Grupo 8 perfil diario.

En la Figura 10 se muestra uno de estos grupos formados. En los ejes de la gráfica, se colocan las variables en el eje de abscisas, mientras que el valor que éstas toman se colocan en el eje de ordenadas. Según la Ecuación 5, los primeros 24 valores corresponden con las medianas de consumo normalizadas de invierno y los siguientes son las 24 horas normalizadas de verano. En el título del gráfico, se identifica el número de grupo que es, así como los integrantes que lo componen. En este caso concreto, se está analizando el grupo 8, que tiene el 2.51 % de los clientes.

En esta gráfica, se pueden apreciar diferentes trazos, diferenciados por colores. Cada uno de estos colores se corresponden con:

- **Datos**: se representan con líneas de color amarillo. Son los valores de los distintos clientes que pertenecen a un grupo. En grupos muy numerosos es difícil distinguir las distintas lineas, ya que todo parece una superficie.
- **Representante**: se representa con la línea de color azul. El representante se calcula a partir de los valores de los datos que pertenecen al grupo. En este caso, el representante de cada grupo es el perfil medio que identifica a todos los clientes pertenecientes a él.
- **Límites**: se corresponden con las líneas rojas. Estas líneas representan la dispersión existente en un grupo en cada una de sus variables. Los grupos cuyos integrantes sean más parecidos entre sí, tendrán las líneas rojas muy cercanas a la línea azul, es decir, al representante del grupo.

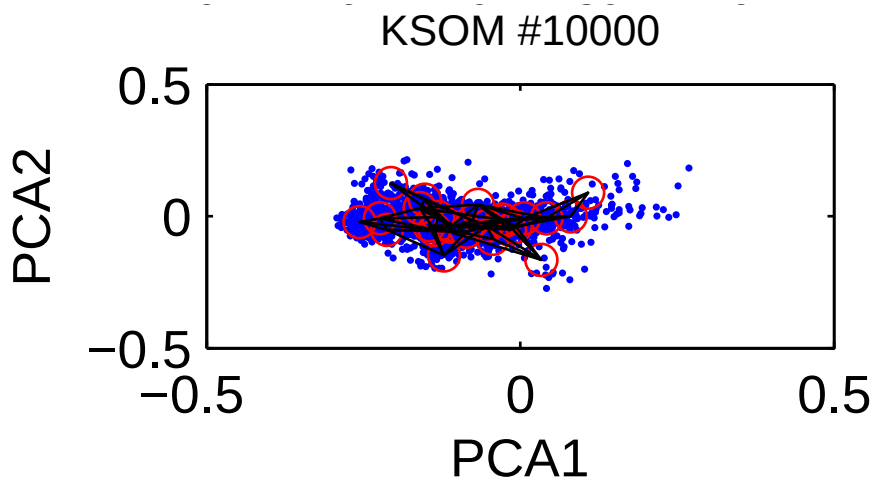

Figura 11. Componentes Principales perfil diario.

La Figura 11 es la gráfica de componentes principales[14] del KSOM. Los clientes, cada uno con 48 valores del perfil horario, quedan representados por un punto de color azul, que es su proyección en el espacio de las dos primeras componentes principales. Por tanto, la nube de puntos de color azul representan la disposición de los clientes en el espacio de componentes principales.

En esta figura también se pueden apreciar diferentes círculos de color rojo repartidos por la nube de puntos. Estos círculos son la proyección de los representantes de cada grupo en el espacio de componentes principales. De manera intuitiva, se puede considerar que el KSOM está bien ajustando cuando los círculos rojos se reparten en el espacio de las componentes principales, cubriendo la mayoría de la nube de puntos, que representan a los clientes.

Otra de las visualizaciones de que disponen los KSOM, es la representación de la matriz U. Esta visualización permite hacerse una idea de las diferencias entre los componentes de un mismo grupo, así como las diferencias existentes con el resto de los grupos. Normalmente los KSOM utilizan neuronas hexagonales, como es este caso. En la Figura 12 se muestra la matriz U perteneciente al mapa de la Figura 9.

[14]El Análisis de Componentes Principales es una técnica estadística de síntesis de la información, o reducción de la dimensión (número de variables). Es decir, ante un conjunto de datos con muchas variables, el objetivo será reducirlas a un menor número perdiendo la menor cantidad de información posible.

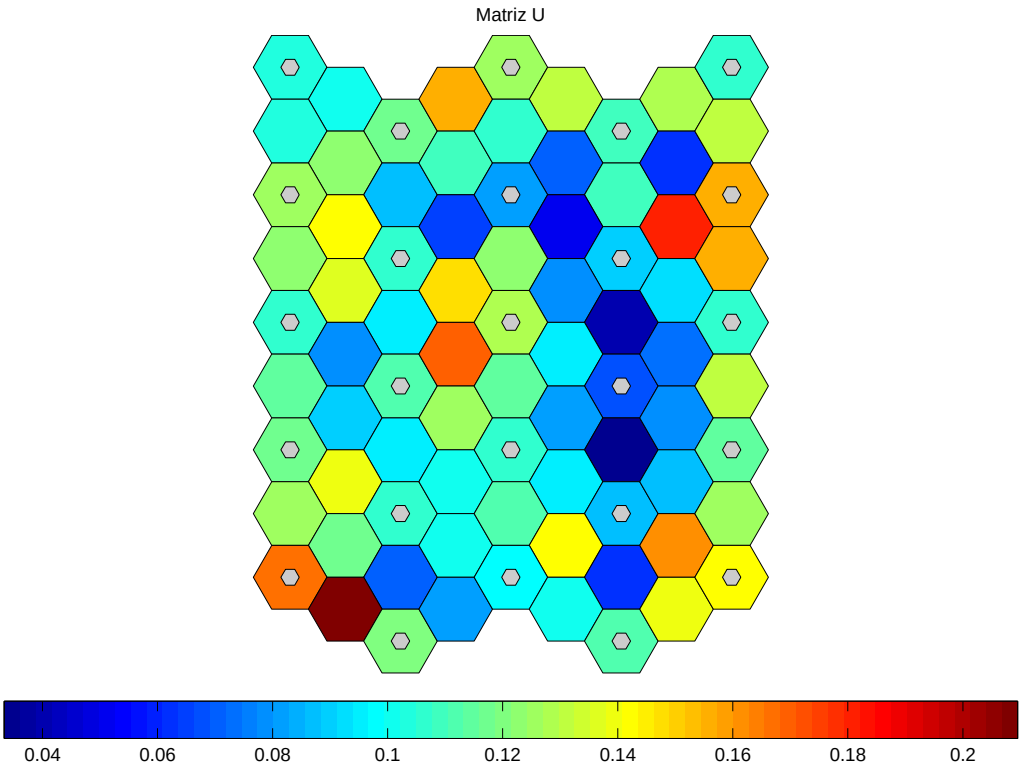

Figura 12. Matriz U perfil diario.

Cada uno de los 25 grupos formados en el ajuste del KSOM son neuronas. Estas neuronas están repartidas en el espacio, formando una rejilla de dimensión $5x5$. Cada neurona no se encuentra de forma aislada en la rejilla, sino que está rodeada a su vez de otras 6 neuronas (esto es debido al uso de neuronas hexagonales), que constituyen la vecindad de la neurona. La matriz U utiliza un código de colores para representar las distancias existentes en el espacio de las componentes principales entre los integrantes de un mismo grupo y entre los representantes de diferentes grupos.

En la Figura 12 los 25 grupos están representados por neuronas hexagonales, numerados de izquierda a derecha, y de arriba a abajo. Para diferenciarlas de las distancias entre neuronas, que también se han representando como hexágonos, se ha trazado en su interior un pequeño hexágono de color gris. Utilizando el código degradado de colores de la parte inferior de la Figura 12, se puede intuir cuáles son los grupos más diferentes entre sí o aquellos que están formados por un grupo más homogéneo de clientes.

Por ejemplo, el grupo 14 tiene una tonalidad azulada, que según el código de colores se corresponde con un valor de 0.06 aproximadamente. Esto significa que la media de distancias entre los integrantes del grupo es de 0.06 en el gráfico de las componentes principales. Volviendo a la Figura 9, se observa como la líneas rojas están bastante

cerca de la línea azul, que es el representante, por lo que la varianza del grupo es pequeña. No ocurre lo mismo con el grupo 13, que presenta un tono amarillento, lo que se corresponde con un valor de 0.13 aproximadamente. Esto quiere decir que sus integrantes están más distantes que en el grupo 14, afirmación que se confirma tras observar la Figura 9, donde los trazos rojos están mucho más alejados del representante.

Siguiendo con el grupo 13, la tonalidad de la neurona que indica la distancia entre el grupo 13 y el grupo 12 es de color anaranjado, lo que se traduce en un valor de 0.17. Esto quiere decir que la distancia entre estos representantes en el espacio de componentes principales es grande, lo que se traduce en una diferencia entre la forma de los perfiles de los representantes de ambos grupos.

Volviendo a la Figura 9, se observa una gran diversidad en los representantes de los grupos formados. Algunos de los grupos formados se asemejan a los perfiles comentados anteriormente. Los grupos 1, 2, 6 y 7 se asemejan al perfil de la Figura 7. En algunos caso, como en el grupo 6, aparece un pequeño valle en las horas de la mañana que no se observa en el resto de grupos. En cambio, en el grupo 7, la caída en el consumo que se produce alrededor de las 15h no es tan notoria como en el resto de los grupos.

También se encuentran grupos cuyos representantes tienen una forma parecida a la de la Figura 6. Los grupos 11, 12 y 16 presentan una forma similar, consumiendo prácticamente de manera plana entre las primeras horas de la mañana y las primeras de la noche. La diferencia principal entre estos grupos es la hora de principio y la de fin de consumo. El grupo 12 comienza a consumir alrededor de las 8h y termina sobre las 22h, mientras que el grupo 11 comienza a consumir sobre las 10h y deja de consumir alrededor de las 21h.

En cuanto al perfil de la Figura 8, también aparece en la forma de los representantes de los grupos 15, 17, 18 y 19, que son lo más numerosos. En estos grupos se encuentran los clientes que tienen una forma más o menos plana de consumir a lo largo de todo el día.

Los grupos 3, 4 y 8 tienen un representante que se asemeja a la forma del perfil de la Figura 4. Estos grupos tienen mayores consumos por la mañana que por la tarde, con una hora en la que el consumo desciende de manera notoria, aunque en el grupo 4 se produce de manera muy sutil.

Los grupos anteriormente citados constituyen la mayor parte de los clientes analizados, aunque existen otros clientes con formas distintas de consumo. Por ejemplo,

la forma de consumo del representante de los grupos 22 y 25 muestra que los clientes pertenecientes consumen la mayor parte de la energía por la noche, situando el consumo entre las 22h y las 8h aproximadamente.

3.3.1.2. Perfil semanal

Se define el perfil semanal como *"la mediana del consumo del dia i normalizada con el consumo total de la semana"*. Es decir, primero se normalizan todos los días de las semanas entre el consumo total de la semana correspondiente, para después realizar la mediana de estos días normalizados. Esta definición se concreta en la Ecuación 6 para el invierno y en la Ecuación 7 para el verano.

$$d_i^{inv} = mediana(D_i^{inv}) \qquad i = 1, 2 \cdots, 7 \tag{6}$$
$$\text{donde} \qquad D_i^{inv} = \frac{dia_i^{inv}}{\sum_{i=1}^{7} dia_i^{inv}}$$

$$d_i^{ver} = mediana(D_i^{ver}) \qquad i = 1, 2 \cdots 7 \tag{7}$$
$$\text{donde} \qquad D_i^{ver} = \frac{dia_i^{ver}}{\sum_{i=1}^{7} dia_i^{ver}}$$

De esta forma, cada cliente queda caracterizado por 14 valores de consumo semanal normalizados (7 valores para los periodos de invierno y 7 valores para los periodos de verano). Estos valores constituyen el **perfil semanal** de cada cliente, es decir, la forma más habitual de consumo de un cliente cualquiera a lo largo de la semana. Esta nueva variable queda definida en la Ecuación 8.

$$Perfil_semanal_j = (d_1^{inv}, d_2^{inv} ... d_7^{inv}, d_1^{ver}, d_2^{ver} ... d_7^{ver}) \tag{8}$$
$$\forall j \in Clientes$$

Se va a proceder de manera similar a la usada para la obtención de los perfiles horarios. Utilizando la Ecuación 6 y la Ecuación 7 se van a calcular los valores de la variable definida en la Ecuación 8.

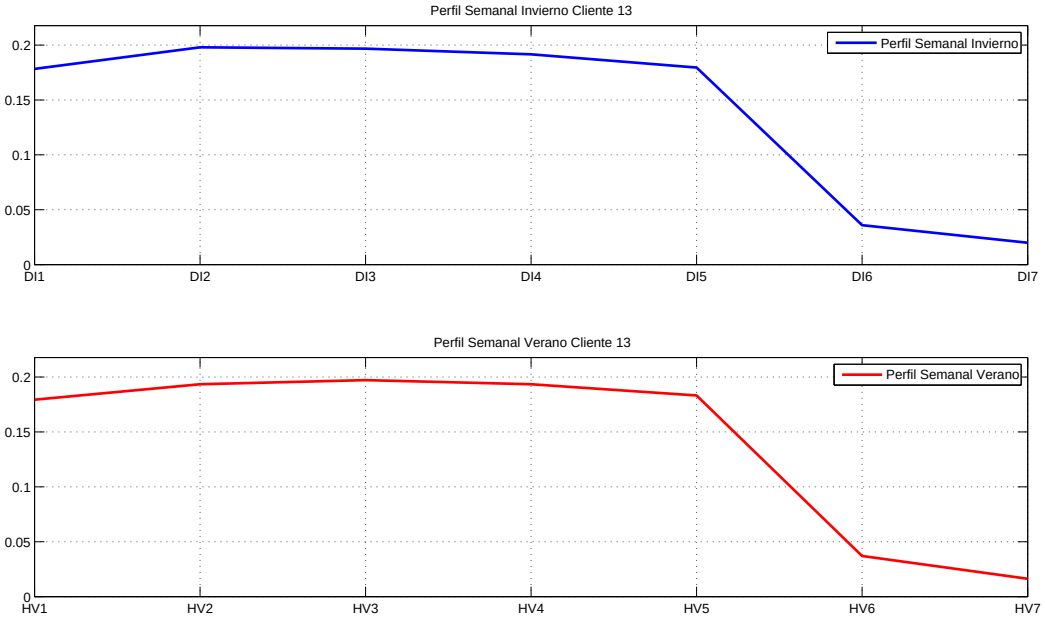

Figura 13. Ejemplo perfil semanal 1.

En la Figura 13 aparece uno de los perfiles más habituales entre los clientes. Como se puede apreciar, el consumo se mantiene prácticamente constante durante toda la semana, entendiendo por semana de lunes a viernes, y el fin de semana disminuye de manera drástica, manteniéndose en valores mínimos.

Este efecto se denomina *"Working day effect"*. La mayoría de las empresas funcionan durante los días laborables de la semana, dejando los fines de semana libres para el descanso y el ocio. Durante los días laborables las empresas funcionan a pleno rendimiento, utilizando todos los recursos de que disponen para desarrollar su actividad adecuadamente. En cambio, los fines de semana que la empresa se encuentra cerrada, el consumo de energía se utiliza para mantener en funcionamiento aquellos equipos y dispositivos que sean imprescindibles y no puedan ser desconectados. Al igual de lo que se había observado en los perfiles horarios, las diferencias de consumo entre invierno y verano no son apreciables.

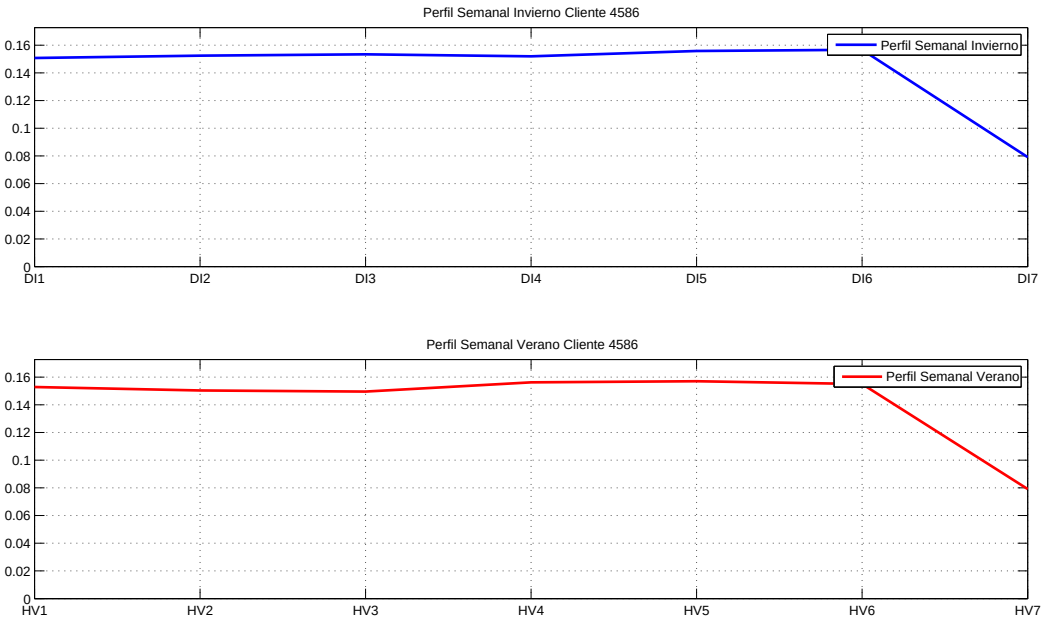

Figura 14. Ejemplo perfil semanal 2.

Otro perfil bastante común es el representado en la Figura 14. Los clientes que presentan este tipo de perfil, consumen prácticamente la misma cantidad de energía desde el lunes hasta el sábado, produciéndose un pronunciado descenso en el consumo el domingo. Aparentemente, las formas de consumo no se ven afectadas por los periodos de verano o invierno, puesto que prácticamente son iguales

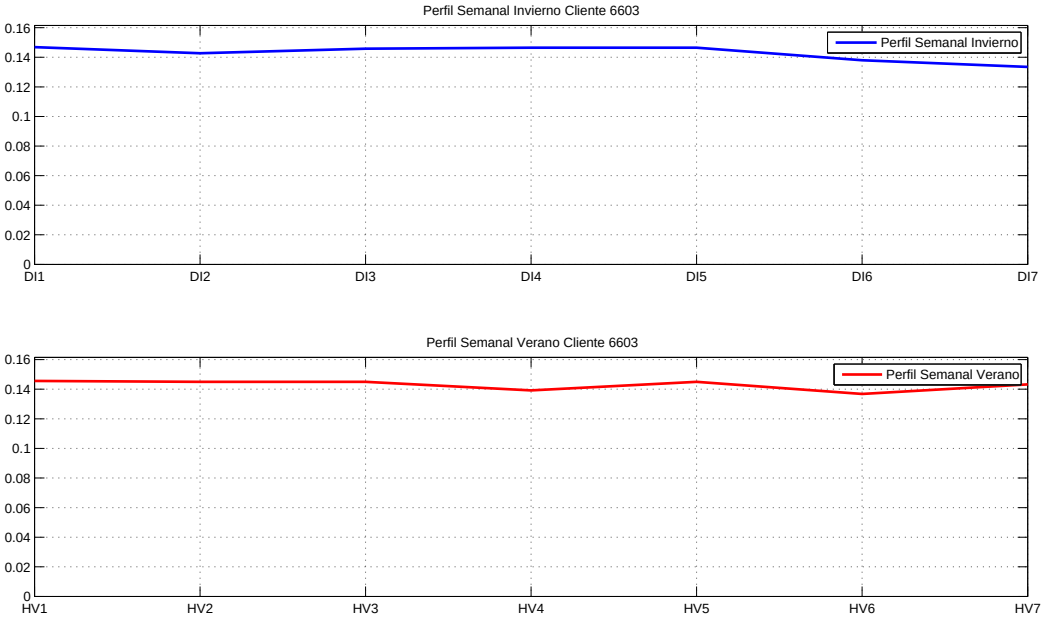

Figura 15. Ejemplo perfil semanal 3.

Por último comentar el perfil de la Figura 15, que se encuentra muy relacionado con el perfil horario de la Figura 8. Aquí se puede apreciar como el consumo se mantiene

constante a lo largo de la semana. No descansan los fines de semana, lo que hace pensar que ocurre algo parecido a la Figura 8, donde tampoco dejaban de consumir en ningún momento del día. Este comportamiento es muy común en grandes empresas industriales en las que no se para la producción salvo causas excepcionales, ya que los costes de arranque son muy altos. Como ocurría en el resto de perfiles, no hay diferencias apreciables entre los periodos de verano e invierno.

Al igual que se realizó con los perfiles horarios, también se ha construido un mapa de Kohonen. Se pretende aprovechar las ventajas de visualización de que disponen para mostrar las distintas formas de consumo semanal que aparecen en la muestra de clientes. Los resultados se muestran en la Figura 17.

Observando la Figura 17 se pueden sacar varias conclusiones. La primera conclusión es que el número de grupos realizados es claramente excesivo. Existen grupos como el 1 y el 2 que tienen representantes con formas muy similares, al igual que los grupos 11 y 12, los grupos 17, 18 y 19, o los grupos 9 y 14. Este comportamiento se puede confirmar observando la proyección en las componentes principales de los clientes en la Figura 16.

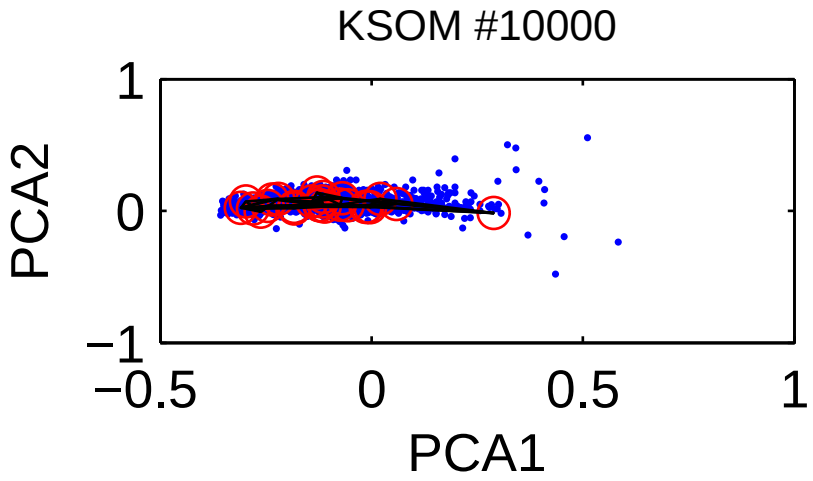

Figura 16. Componentes Principales perfil semanal.

Todos los clientes se encuentran concentrados en una franja situada a lo largo del eje de abcisas, a excepción de algunos *outlayers* que se alejan de los valores habituales. De hecho, se puede observar como la mayoría de los representantes se encuentran concentrados en esa franja.

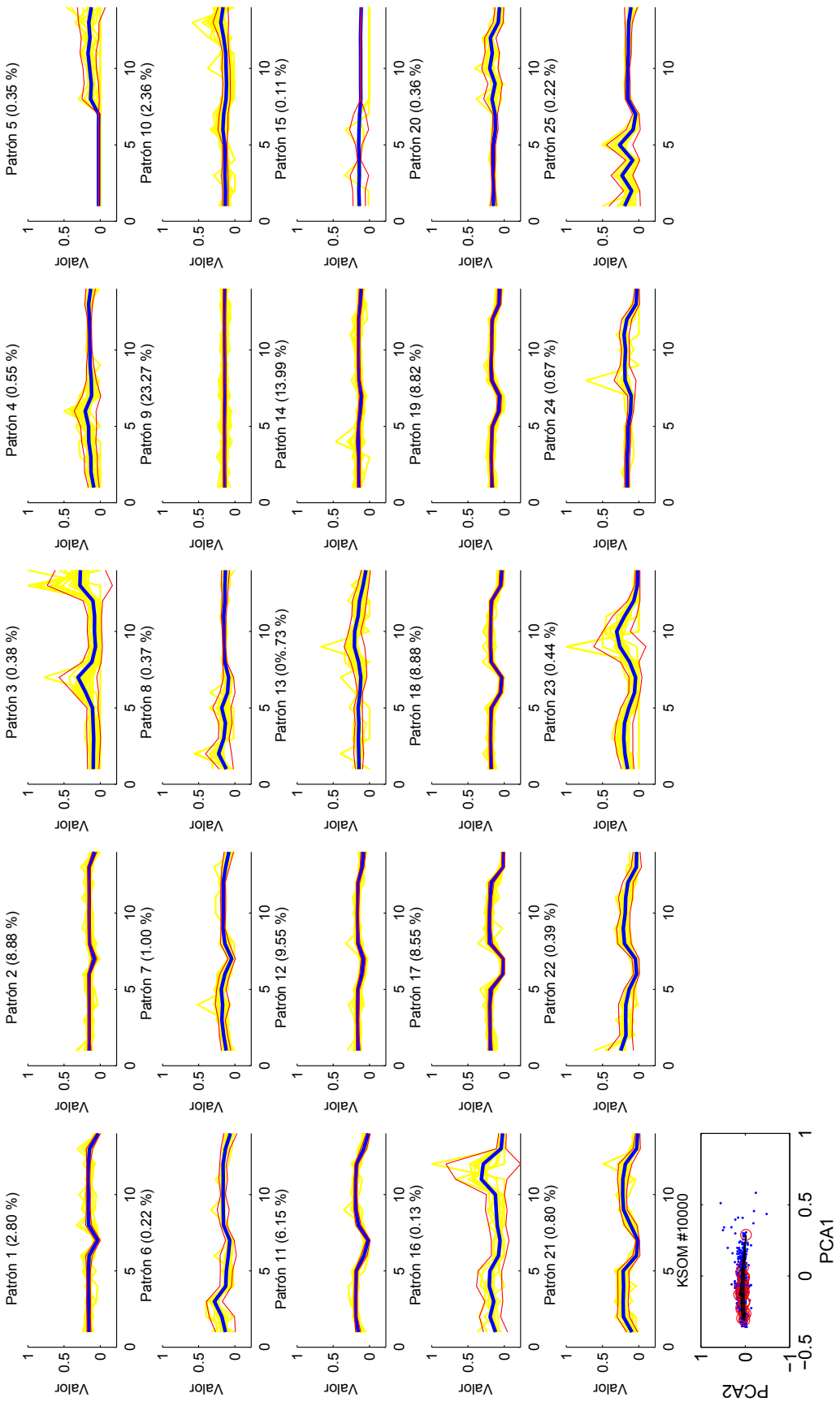

Figura 17. KSOM 5x5 perfil semanal.

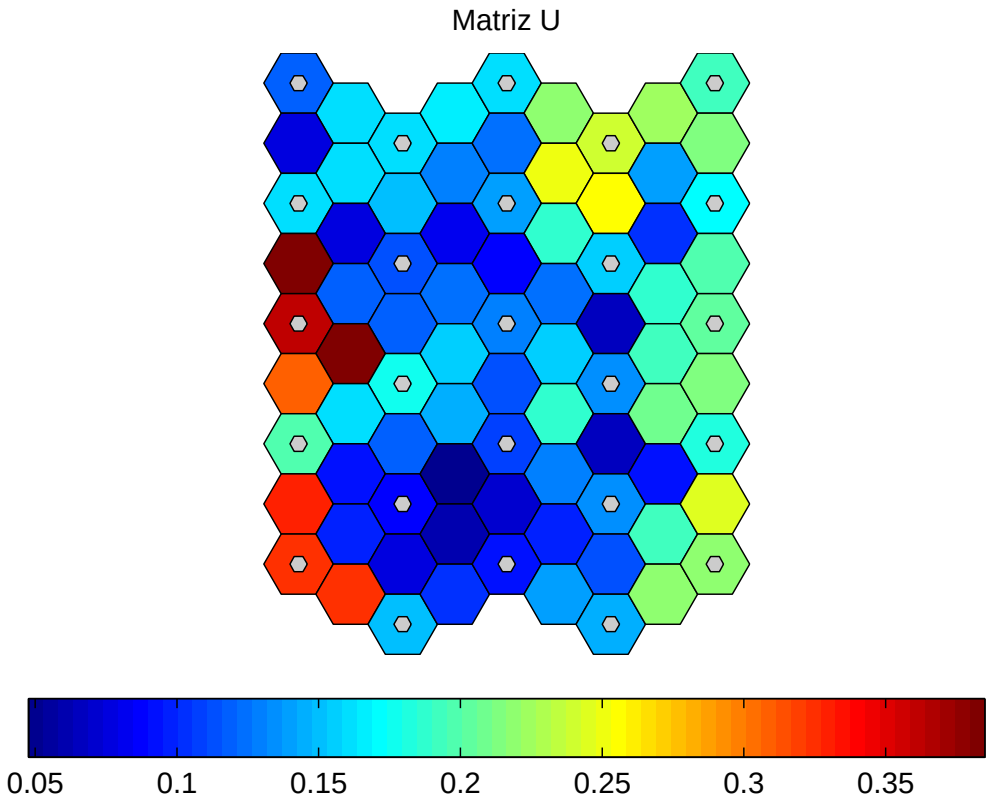

Figura 18. Matriz U perfil semanal.

Analizando la matriz U de la Figura 18 se confirma nuevamente el hecho de que el número de grupos es excesivo. Existen 3 zonas que presentan una tonalidad azul oscura, que se corresponden con los grupos citados anteriormente. Esto quiere decir que los grupos se encuentran muy próximos y apenas existen diferencias entre ellos.

El otro hecho que se puede concluir tras observar la Figura 17 es que los clientes se hayan concentrados en una serie de grupos. Entre los grupos 1, 2, 9, 11, 12, 14, 17, 18 y 19 se concentran cerca del 90 % de los datos utilizados para construir el mapa.

El resto de grupos recoge las formas más extrañas de consumo de los clientes. Por ejemplo, los grupos 5 y 20 contienen clientes que tan sólo consumen durante las semanas de verano; justo al contrario que el grupo 25, que presenta los mayores consumos durante la época del invierno.

Tras analizar con detenimiento el mapa de KSOM del perfil semanal, se puede concluir que entre los clientes no existe demasiada variedad a la hora de consumir a lo largo de la semana. La mayoría de los clientes pueden agruparse en tres grandes grupos, que coinciden con los ejemplos expuestos anteriormente:

- : **Grupo 1**: formado por los grupos 1 y 2. Se correspondería con el ejemplo de perfil 2, donde los clientes consumen constante de lunes a sábado, con una caída en el consumo durante el domingo.
- : **Grupo 2**: formado por los grupos 11, 12, 17, 18 y 19. Se correspondería con el ejemplo de perfil 1, donde los clientes consumen constante de lunes a viernes, con una caída en el consumo durante el fin de semana.
- : **Grupo 3**: formado por los grupos 9 y 14. Se correspondería con el ejemplo de perfil 3, donde los clientes consumen de manera constante a lo largo de toda la semana.

3.3.1.3. Perfil anual

Se define el perfil anual como *"la mediana del consumo del mes i normalizada con el consumo total del año"*. Es decir, primero se normalizan todos los meses de los años entre el consumo total del año correspondiente, para después realizar la mediana de estos meses normalizados. Esta definición se concreta en la Ecuación 9.

$$m_i = mediana(M_i) \qquad i = 1, 2 \cdots, 12 \tag{9}$$
$$\text{donde} \qquad M_i = \frac{mes_i}{\sum_{j=1}^{12} mes_j}$$

Lógicamente, en la obtención de estos perfiles no se realiza la diferenciación entre los perfiles de invierno y verano, y tan sólo se van a obtener 12 valores de consumo anual normalizado. Al contrario de los perfiles horarios y semanales, donde existían suficientes datos para calcular las medianas, de algunos clientes no se pudo calcular estos perfiles por falta de valores[15].

De esta forma, cada cliente queda caracterizado por 12 valores de consumo anual normalizados. Estos valores constituyen el **perfil anual** de cada cliente, es decir, la forma más habitual de consumo de un cliente a lo largo del año. Esta nueva variable queda definida en la Ecuación 10.

$$Perfil_anual_j = (m_1, m_2 ... m_{12}) \tag{10}$$
$$\forall j \in Clientes$$

[15]No todos los clientes llevan consumiendo energía durante al menos un año completo.

De manera análoga a la usada con los perfiles horario y semanal, se va a aplicar la Ecuación 9 para calcular los valores de la variable $perfil_anual$ de cada cliente. A continuación se muestran ejemplos de representar esta variable para algunos clientes.

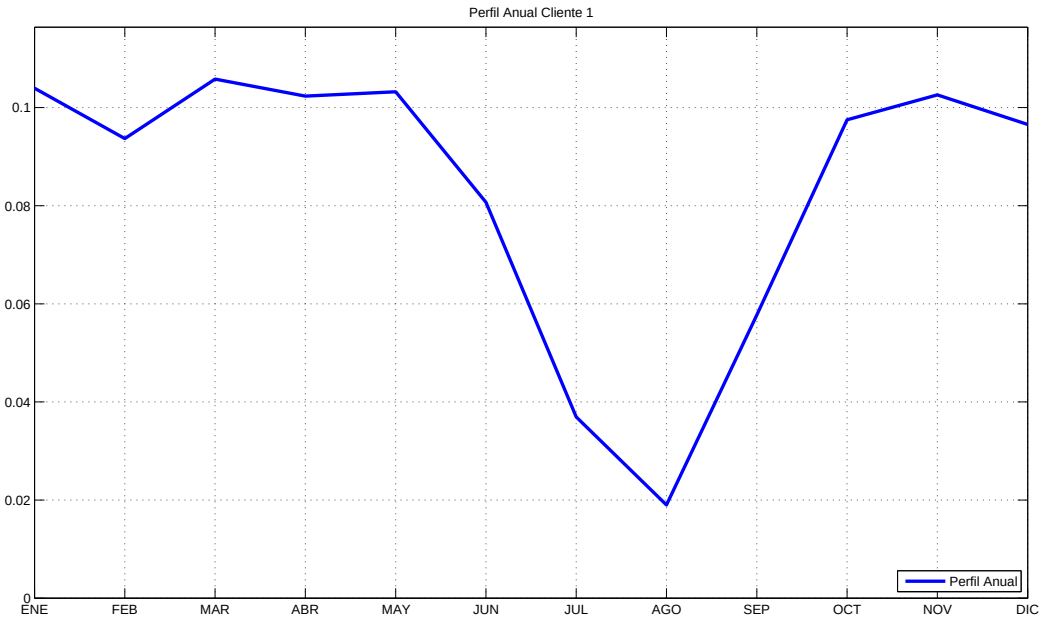

Figura 19. Ejemplo perfil anual 1

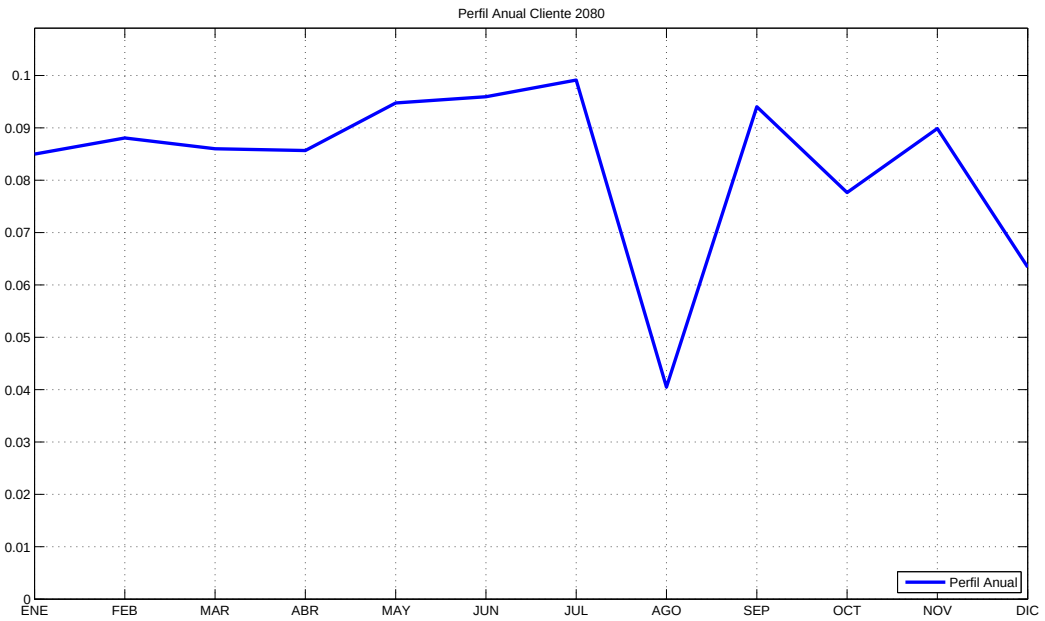

Figura 20. Ejemplo perfil anual 2

En las Figura 19 y Figura 20 se puede observar la forma del consumo habitual durante el año, con un fuerte descenso del consumo durante el mes de agosto[16]. Esto puede ser debido a que la mayoría de las personas escoge este mes para sus vacaciones de verano.

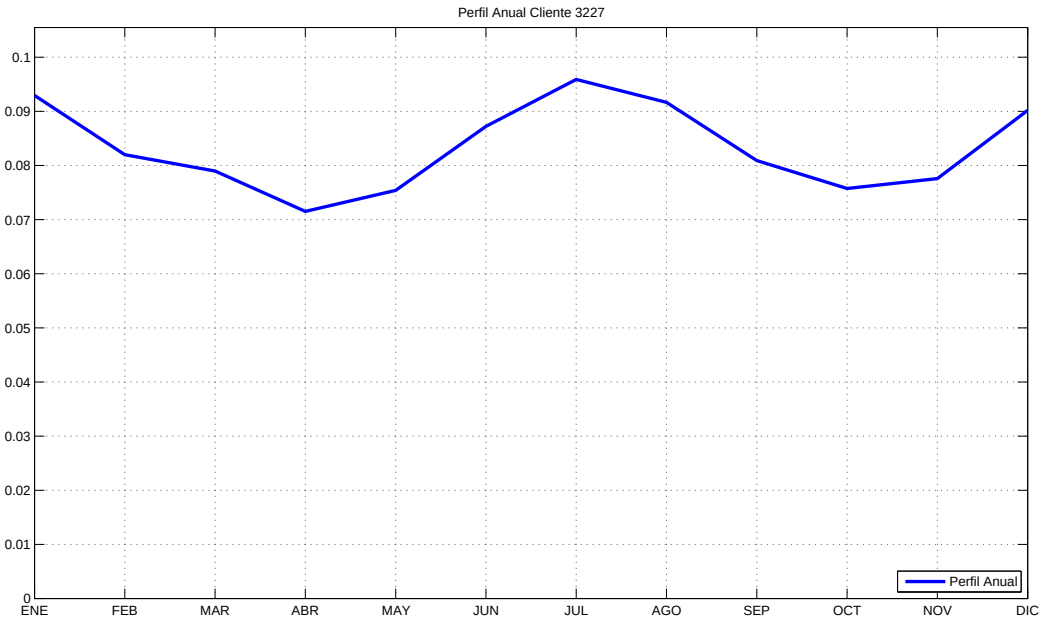

Figura 21. Ejemplo perfil anual 3

Justamente en la Figura 21 se produce un fenómeno contrario, donde la mayor parte del consumo se concentra en los meses de verano (junio, julio y agosto), produciéndose una fuerte disminución en los meses de abril y de octubre. Posiblemente se trate de un establecimiento hotelero situado en la playa, que requiere mayor energía eléctrica en los días con más clientes, que suelen localizarse en el mes de agosto.

Siguiendo con la metodología utilizada anteriormente con los perfiles horario y semanal, se va a usar un KSOM para clasificar los diferentes perfiles anules que presentan los clientes. La pantalla resultante tras llevar a cabo el ajuste del mapa se muestra en la Figura 22.

En un primer vistazo del mapa, se observan más formas de consumo a parte de las expuestas en los ejemplos. Esto quiere decir, que el perfil semanal es mucho más variado de lo que se esperaba. Para confirmar esta variedad en las formas de consumo, se va a recurrir al resto de visualizaciones, el gráfico de las componentes principales y la matriz U.

[16]Esta característica será tratada de forma especial en el modelo de regresión

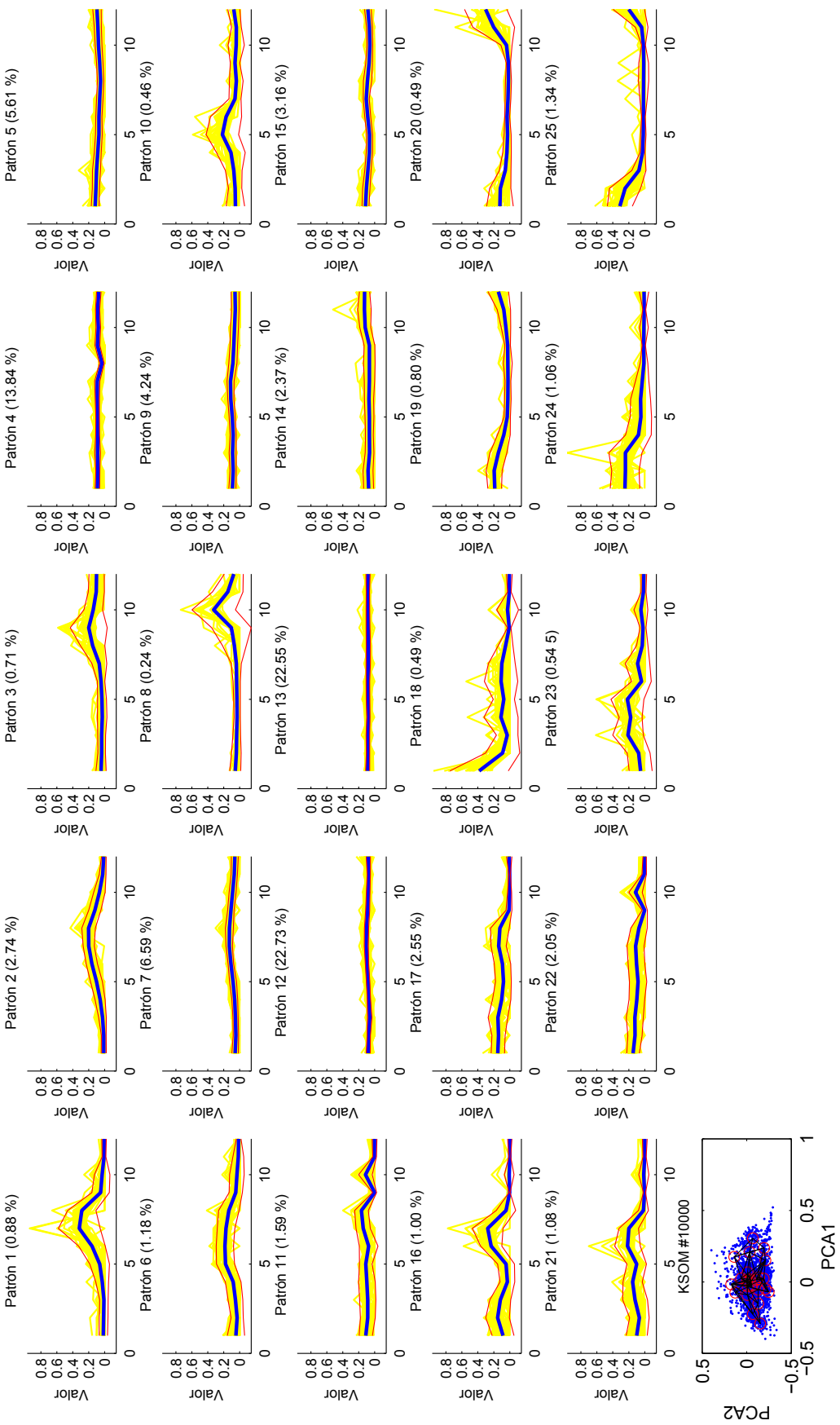

Figura 22. KSOM 5x5 perfil anual.

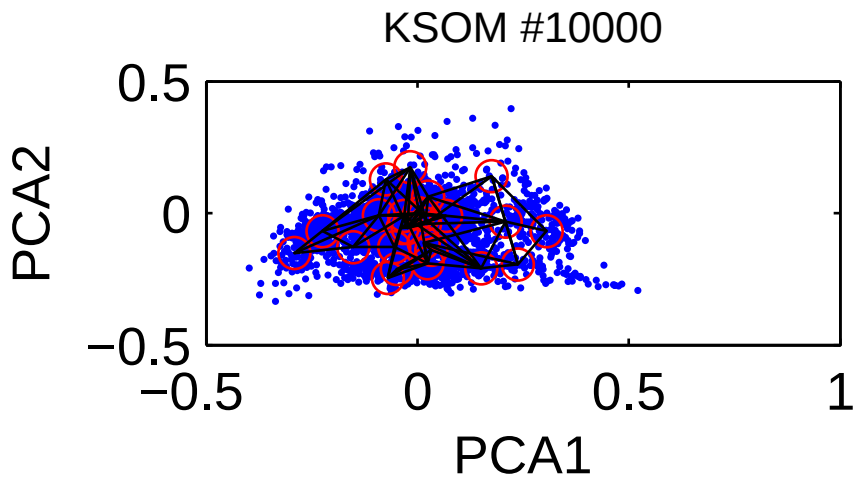

Figura 23. Componentes Principales perfil anual.

Las componentes principales de este KSOM, representadas en la Figura 23, muestra que los clientes se hayan muy repartidos en el gráfico, bastante más repartidos que en la Figura 16, que se encontraban muy concentrados. De igual manera, los representantes se reparten en el espacio de componentes principales, intentando abarcar a la mayoría de los clientes.

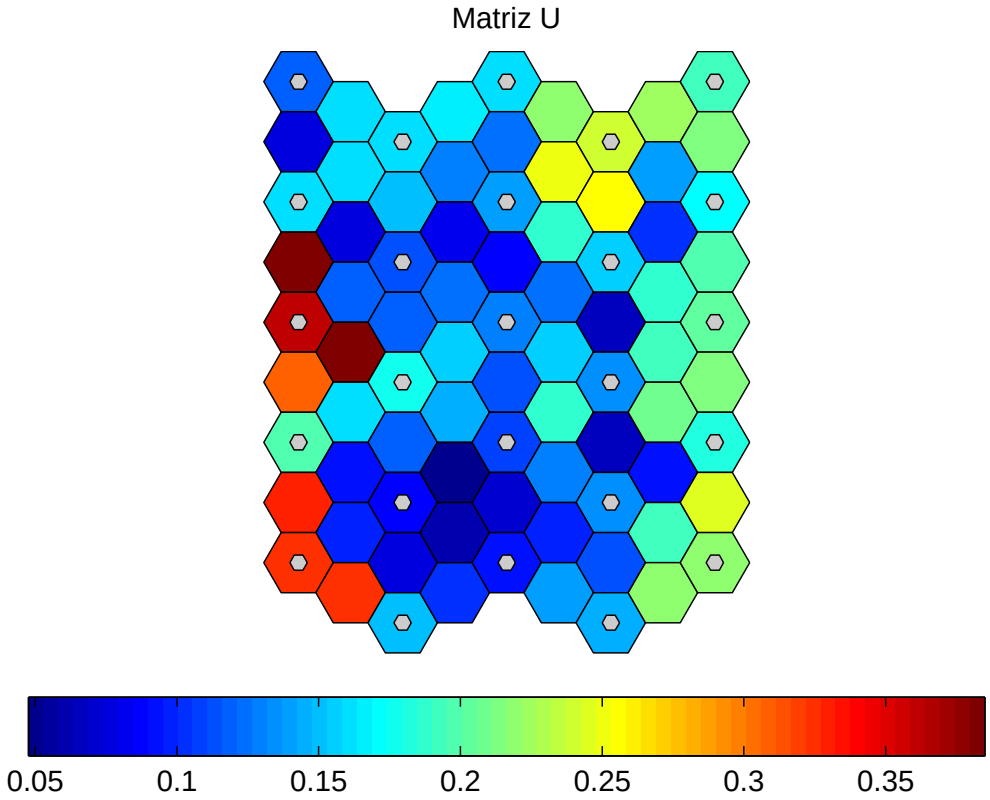

Figura 24. Matriz U perfil anual.

Por último, queda por analizar la matriz U correspondiente al perfil anual. Esta matriz está representada en la Figura 24. Existe una gran variedad en los grupos

formados, con una matriz heterogénea en cuanto a tonalidades. Esto confirma que la forma de consumir de los clientes a lo largo del año es mucho más variada que, por ejemplo, la que presentan a lo largo de la semana.

Una vez que se han caracterizado cada cliente de forma unívoca con 48 valores de consumo horario, 14 valores de consumo semanal y 12 valores de consumo anual, se va a utilizar esta información, que describe la forma habitual en que consume durante el día tipo de invierno y verano, la semana durante invierno y verano y durante el año, para clasificar a los clientes y formar grupos.

Antes de introducir los datos y ajustar un KSOM, hay que pararse a analizar las variables disponibles, teniendo en cuenta el horizonte de previsión para el que se pretende realizar el modelo. Se han agrupado los clientes en 25 *clusters* de acuerdo a la forma de consumir horaria, por otra parte se han agrupado en otros 25 *clusters* atendiendo a su forma de consumir a lo largo de la semana y finalmente, 25 grupos con la forma de consumir anual. Intentar clasificar a los clientes a partir de estos grupos realizados implica un gran número de posibilidades de combinación, por lo que se ha optado por crear una nueva variable, utilizando las ya disponibles.

En total, se disponen de 74 variables, de las cuáles, más de la mitad se corresponden con el consumo horario. En un modelo de previsión a medio plazo, no tiene sentido que tenga más peso las variables correspondientes al muy corto plazo (horas), que las que realmente están relacionadas con el medio plazo (semanas y meses).

Se va crear una nueva variable, llamada **perfil_consumo**, que agrupe las características que se han utilizado para construir el KSOM. La definición de esta nueva variable se concreta en la Ecuación 11:

$$Perfil_consumo_j = (perfil_diario^{✠}, perfil_semanal, perfil_anual) \quad (11)$$

$$\text{donde} \qquad perfil_diario^{✠} = (h_i^{inv}, h_i^{ver})\ \forall i \in 2, 4, 9, 11, 15, 18, 20$$
$$\forall j \in Clientes$$

Puesto que el perfil horario disponía de demasiados valores, se ha optado por reducir su influencia en la nueva variable, escogiendo las horas que son más representativas y ocurren los cambios más notables. Las 2h y 4h se han escogido para tener en cuenta a aquellos clientes que consumen más energía durante las horas nocturnas. Las 9h y 20h

se corresponden con la subida y bajada de los consumos de la mayoría de los clientes, mientras que en las 11h y las 18h se producen los picos en el consumo. Durante las 15h se produce el valle más importante del día.

Con esta nueva variable, se agruparán a los distintos clientes atendiendo a su forma de consumir. Si se agregan los clientes que siguen un perfil de consumo similar, la serie resultante será más fácil de estimar y se podrán aplicar modelos específicos a cada uno de los grupos. Esta agrupación se va a realizar utilizando los ya conocidos KSOM, pero antes de comentar el mapa resultante, se va a explicar en profundidad el funcionamiento de este algoritmo.

3.3.2. Mapas autoorganizados de Kohonen

Para llevar a cabo la agrupación de la cartera de clientes, se va a utilizar un algoritmo llamado "Mapa Autoorganizativo de Kohonen" (Kohonen Self Organising Map). Es un modelo de red neuronal presentado por T. Kohonen en el libro *Self-Organizing Maps* en 1989. La idea principal que Kohonen pretendía lograr con este trabajo, según afirmó en 1995 se resume en la siguiente grase: *"I just wanted an algorithm that would effectively map similar patterns (pattern vectors close to each other in the input signal space) onto contiguous locations in the output space"*. Según continua diciendo Kohonen, son los modelos más realistas en cuanto a plausibilidad biológica de las redes neuronales, dado que el resultado es una representación con orden espacial, circunstancia que se da en el cerebro y que no es tenida en cuenta por otros modelos de RNA. La información acerca de este algoritmo se ha obtenido de [25].

Se puede concluir que el KSOM es un algoritmo cuyo objetivo es agrupar un conjunto de datos con una serie de características, descubriendo la estructura subyacente de los mismos. La arquitectura de los mapas ha demostrado ser aplicable a una diversidad de dominios de datos, en especial a aquellos en los que se posee un gran volumen de información y numerosos atributos[17] Los principales usos de este algoritmo son los siguientes:

1. ***Clustering***: agrupar el conjunto de datos de entrada en diferentes *clusters*, con características similares entre los integrantes del *cluster*, pero diferentes del resto de agrupaciones.

2. **Visualización**: este agrupamiento, como se realiza de una forma ordenada, permite visualizarlo y descubrir características nuevas, o relaciones que no se

[17]Esta característica los convierte en una herramienta muy potente para analizar las características de la cartera de clientes de una comercializadora.

habían previsto de antemano. También permite visualizar la evolución temporal de un conjunto de datos.

3. **Clasificación**: una vez calibrado el mapa, o asignada algún tipo de etiqueta a cada *cluster*, se puede usar para clasificar datos desconocidos que tengan las mismas características que los datos utilizados para crear el KSOM, evaluando el modelo.

Son algoritmos no supervisados, es decir, no necesitan de un supervisor externo que juzgue los procesos del aprendizaje. Esto es debido a que en los KSOM no se dispone de ninguna salida objetivo hacia la cual debería de tender la red neuronal. Además, estos algoritmos sólo manejan patrones de entrada.

También se denominan algoritmos competitivos, porque durante el entrenamiento se entrena una sola neurona de cada vez. Las neuronas compiten unas con otras para activarse cuando se presenta a la red un patrón de entrada. Se pretende que sólo una de las neuronas resulte vencedora ante el patrón de entrada, desactivando el resto de neuronas de la capa de salida.

La estructura de los KSOM está compuesta por dos capas de neuronas. Se puede observar de manera gráfica en la Figura 25:

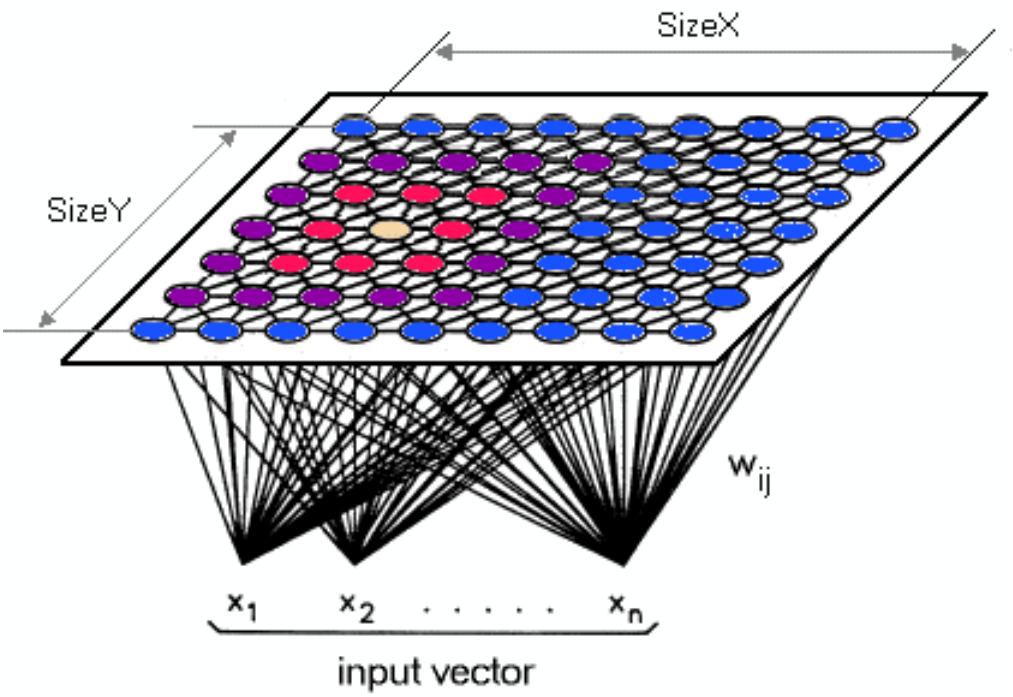

Figura 25. Estructura mapa de Kohonen

- **Capa de entrada**: Está formada por m neuronas, una por cada variable de entrada. Se encarga de recibir y transmitir a la capa de salida la información procedente del exterior.

- **capa de salida**: Está compuesta por $n = X\text{x}Y$ neuronas, que son el número de *clusters* que va a construir el mapa después de procesar la información. Las neuronas de la capa de salida se organizan en forma de mapa bidimensional.

La información se propaga desde la capa de entrada hacia la capa de salida. Cada neurona $i \in m$ de la capa de entrada está conectada con cada neurona de la capa $j \in n$ mediante un peso w_{ij}.

Entre las neuronas de la capa de salida existe cierta relación, ya que cada una de ellas tiene cierta influencia sobre sus vecinas. Esto se consigue a través de un proceso de competición entre neuronas y la aplicación de una función de vecindad, que produce la topología del mapa (las más frecuentes son la rectangular o la hexagonal). Esta topología permanece fija al inicio del algoritmo. Cada neurona de la capa de salida j tiene asociada una vecindad N_j, que está constituida por las neuronas adyacentes a ésta.

Siguiendo con la nomenclatura anterior, se va a resumir la terminología usada para explicar el modelo de *clustering*.

- $X = x_0, x_1, x_2, ..., x_{m-1}$ representa un conjunto de N entradas y para el cual cada x_i tiene m dimensiones o características.

- $m =$ número de neuronas de entrada. Forma la capa de entrada.

- $n =$ número de neuronas de salida o número de *clusters*. Forma la capa de salida.

- $W_j =$ vector $[w0_j, w1_j, ..., w(m-1)_j]^T$ que corresponde a la neurona de salida j, donde $0 <= j <= n-1$. Este vector contiene los coeficientes de las m neuronas de entrada hasta la neurona de salida j.

Cuando se presenta el vector de entrada x_i a la red, el nodo de salida j queda asignado a $d_{ij} = (x_i - W_j)^T(x_i - W_j)$. El vector W_j contiene los coeficientes de los m nodos de entrada hasta el nodo de salida j.

El objetivo global es mover los coeficientes hacía los centros de los *clusters* mediante la actualización de los mismos por cada una de las entradas. Los pasos necesarios para conseguirlo son los siguientes:

1. Inicializar los coeficientes W_j de una forma aleatoria para todos los nodos j y asignar el tamaño de la "vecindad" (EN) a $n/2$. Asignar todas las actualizaciones de pesos ΔW_j a cero. El parámetro de aprendizaje $Lrate$ se inicializa a un valor entre 0 y 1.

2. Para cada entrada x_i, se selecciona el nodo de entrada $j*$, $0 <= j <= n-1$, tal que la distancia d_{ij} sea mínima. Después se actualiza el vector W_j utilizando la siguiente regla:

$$W_j = W_j + Lrate * \Delta W_j$$
$$\Delta W_j = \Delta W_j + (x_i - W_j)$$

   donde j incluye el nodo de salida $j*$ y a sus EN vecinos hacia la izquierda y la derecha. Se repite el paso número 2 hasta que no existan cambios en los coeficientes.

3. Comprobar si EN=0. Si es así, el algoritmo ha acabado. En cada caso contrario, se reduce el valor de EN en una unidad y se vuelve al paso número 2.

3.3.3. *Clustering*

El algoritmo comentado anteriormente se encuentra implantado en una herramienta informática llamada IDAT. Este herramienta permite usar multitud de modelos: unos supervisados (regresión lineal, bisagras, árboles ...) y otros no supervisados (PRBFN, Neural Gas y KSOM). En este PFC se va a utilizar el KSOM, cuya pantalla de inicio es la que aparece en la Figura 26.

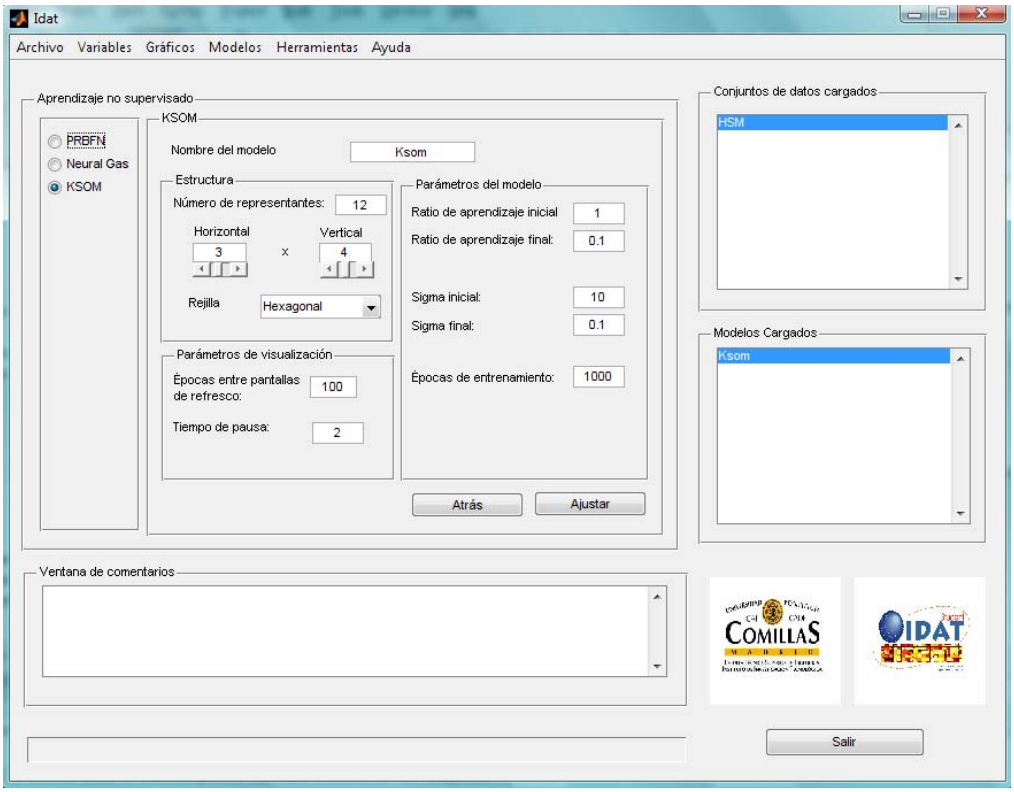

Figura 26. Pantalla de Inicio KSOM

En esta pantalla se pueden configurar los principales parámetros del KSOM:

- **Estructura**: permite configurar el número de neuronas que van a formar la capa de salida, es decir, el número de *clusters* que se quieren realizar. Para el estudio que se va a realizar, se han elegido formar 25 grupos, es decir un mapa de 5x5. Además, se indica la relación de vecindad que va a existir entre las neuronas, en este caso, se ha elegido una rejilla hexagonal.
- **Ratios de aprendizaje**: estas opciones se han mantenido los valores por defecto.
- **Épocas de entrenamiento**: define el número de iteraciones que debe realizar el modelo antes de terminar el algoritmo. Dada la cantidad de datos, se ha ajustado este parámetro a 10000 iteraciones, para que el modelo quede lo más ajustado posible.

Una de las ventajas del uso de este tipo de algoritmos, como ya se ha comentado anteriormente, son sus grandes posibilidades a la hora de representar el resultado obtenido.

Se va a construir un modelo que agrupe a los clientes según sus características (perfiles de consumo). El KSOM se va a crear utilizando una muestra representativa dentro de la cartera de clientes. Los clientes que no se han incluido para construir el modelo tendrán que ser evaluados para el modelo construido, es decir, comparar las características de las que sí se dispone (perfiles horarios y semanales) para ver cual es *cluster* al que pertenecen. Este proceso se realiza calculando la distancia a cada uno de los representantes y eligiendo aquel que la minimice. Al finalizar el algoritmo de KSOM con la configuración anterior, el resultado es el que se muestra en la Figura 27.

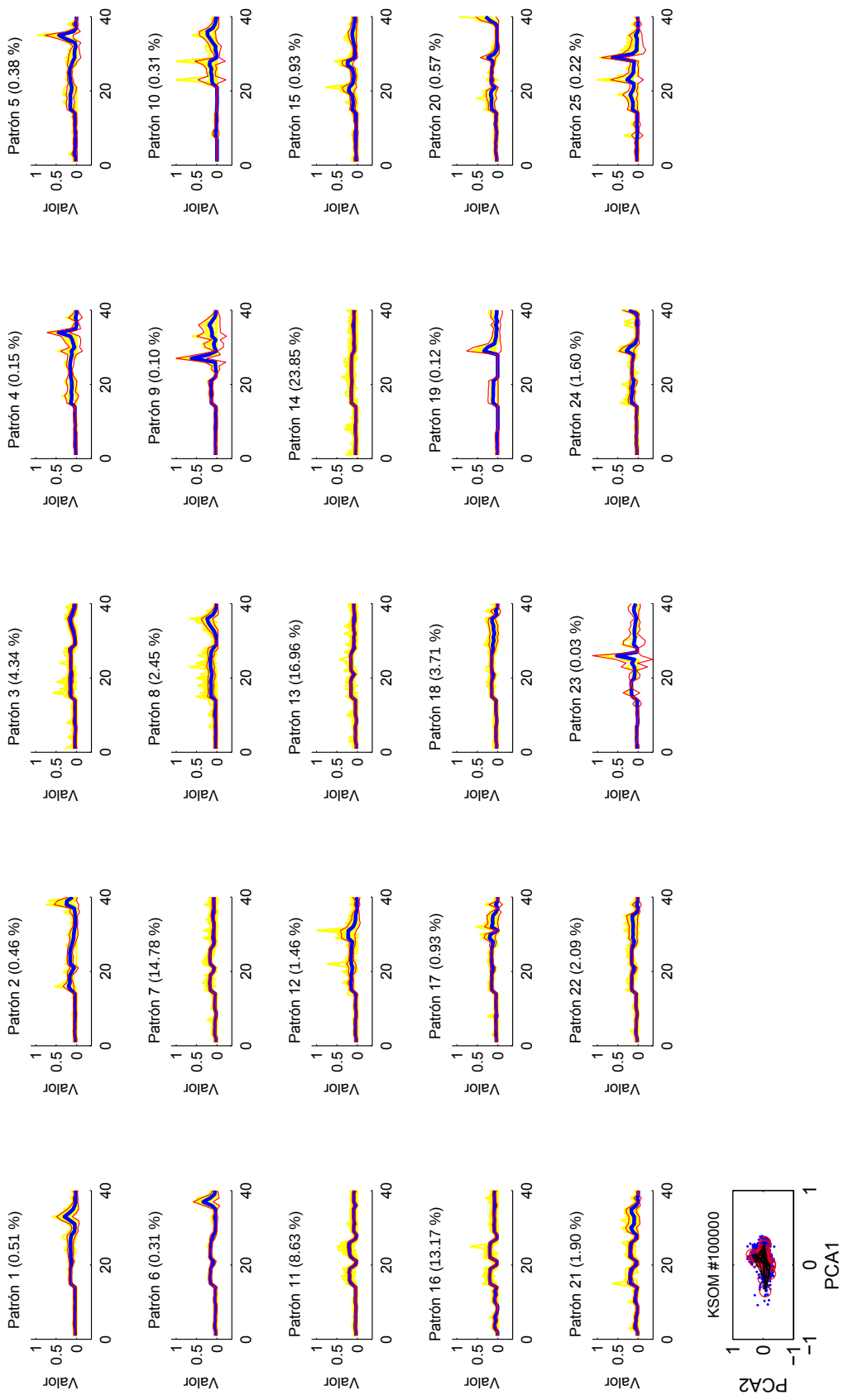

Figura 27. Mapa de KSOM perfil_consumo

Al representar la nueva variable, es más difícil localizar las distintas formas de consumo analizadas anteriormente entre los distintos representantes, debido al uso de tan sólo 14 horas de las 48 disponibles para el perfil horario (7 horas de las 24 horas para cada perfil de verano y de invierno). En la Figura 27 se muestran los 25 grupos que se han creado, así como la proyección de los clientes en el gráfico de las componentes principales. Este gráfico se muestra ampliado en la Figura 28.

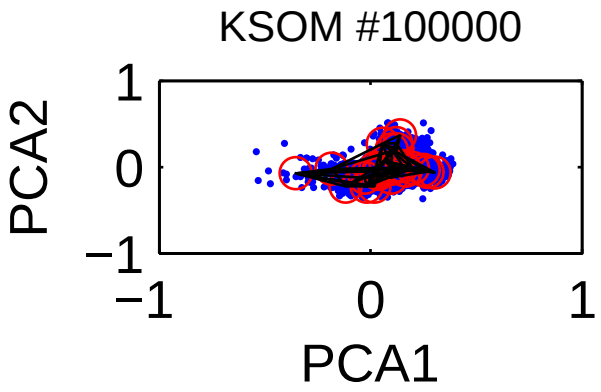

Figura 28. Mapa componentes principales

En este gráfico, se puede apreciar que los clientes se hayan repartidos por todo el espacio de las componentes principales, sin seguir un patrón determinado, como por ejemplo el observado en la Figura 16. A pesar de que los clientes se encuentran bastante dispersos, los representantes también se encuentran razonablemente distribuidos sobre este espacio, abarcando a la mayoría de ellos, salvo alguna excepción. Podemos concluir que los grupos formados agrupan de forma correcta los datos de entrada.

Evaluando este modelo para el resto de clientes, se ha conseguido agrupar toda la muestra en 25 grupos. Estos grupos se han construido automáticamente agrupando a aquellos clientes que presentan mayores similitudes en sus formas de consumir diariamente, semanalmente y anualmente. El resultado de esta agrupación se resume en la Tabla 3.

Grupo	Porcentaje	Grupo	Porcentaje
1	0.51 %	14	23.85 %
2	0.46 %	15	0.93 %
3	4.34 %	16	13.17 %
4	0.15 %	17	0.93 %
5	0.38 %	18	3.71 %
6	0.31 %	19	0.12 %
7	14.78 %	20	0.57 %
8	2.45 %	21	1.90 %
9	0.10 %	22	2.09 %
10	0.31 %	23	0.03 %
11	8.63 %	24	1.60 %
12	1.46 %	25	0.22 %
13	16.96 %		

Tabla 3. Resumen Grupos KSOM

A la vista de los resultados, los grupos formados no son uniformes en cuanto al número de integrantes, puesto que existen tres grupos que abarcan casi el 60 % de los clientes. Este resultado no tiene porque ser anómalo, puesto que se busca agrupar en función de la forma de consumir y ésta no tiene porque ser homogénea en los clientes de la muestra. Además, estos grupos formados no tienen en cuenta el volumen de energía consumida, por lo que sí podrían ser homogéneos teniendo en cuenta el orden de magnitud de las series agregadas de consumo.

3.3.4. Cálculo Series Agregadas

Una vez que se tienen agrupados los clientes, el siguiente paso a realizar es calcular las series agregadas necesarias para introducir en los modelos de regresión. Tal y como se muestra en la Figura 3, este es el último módulo de la etapa 1, y de aquí saldrán las cuatro series que se detallan a continuación.

Para el cálculo de las series se han implementado distintas funciones en Matlab, que realizan las consultas a las bases de datos y todas las operaciones necesarias. La estructura fundamental de estas funciones se detalla en el Código 1.

```
function serie_temporal=SeriesAgregadas(grupo)

% Función que realiza las consultas a las bases de datos para calcular las series
% agregadas.
% Esta función recibe como argumento el número del grupo y devuelve un cell con la
% fecha y los datos de la serie.

N= ;
%número de bases de datos a consultar.
fecha = [año,mes];                      %Fecha de comienzo de las bases de datos
serie_temporal = [];     %Variable para almacenar las fechas y los consumos.

for i=1:N
    % Nombre de la base de datos y de las tablas a consultar
    nombre_BD = sprintf('...%d...%d', fecha(1), fecha(2));
    nombre_tabla = sprintf('...%d...%d', fecha(1), fecha(2));

    % Construcción de la consulta de SQL
    sql = sprintf('...%s...%d...%d...%d', nombre_tabla, fecha(1), fecha(2), grupo );

    % Conexión a la base de datos
    conn = database(nombre_BS, 'usuario', 'contraseña');

    % Ejecutar la consulta
    curs = exec(conn, sql);
    curs = fetch(curs);

    % Almacenar los datos en la variable
    serie_temporal = [serie_temporal; curs.Data];

    % Cerrar la conexión con la base de datos
    close(curs);
    close(conn);

    %Avanzar el vector de la fecha
    fecha(2)=fecha(2)+1;

    if fecha(2)==13
        fecha(2)=1;
        fecha(1)=fecha(1)+1;
    end
end
```

Código 1. Estructura funciones Series Agregadas

Sobre la estructura del Código 1 se han realizado modificaciones para ajustar el código a las necesidades que requiere el cálculo de cada una de las series, variando la sintaxis del código en SQL para realizar la consulta necesaria en cada uno de los casos.

3.3.4.1. Series Consumo Agregadas

Las series de consumo agregadas se calculan como "*la suma de todos los valores de consumo diario de los clientes que pertenezcan al mismo grupo*". Para realizar esta operación, la estructura del Código 1 no debe cambiarse, ya que lo único necesario es modificar la sintaxis de la consulta en SQL para realizar una *query* para las series de consumo agregado.

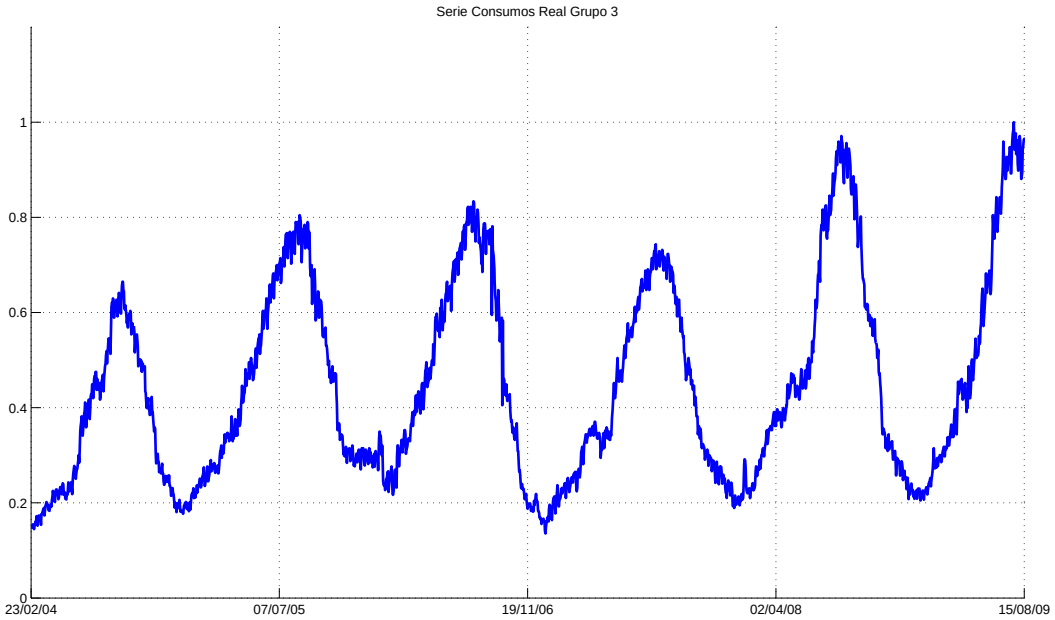

Figura 29. Serie Consumo Grupo 3

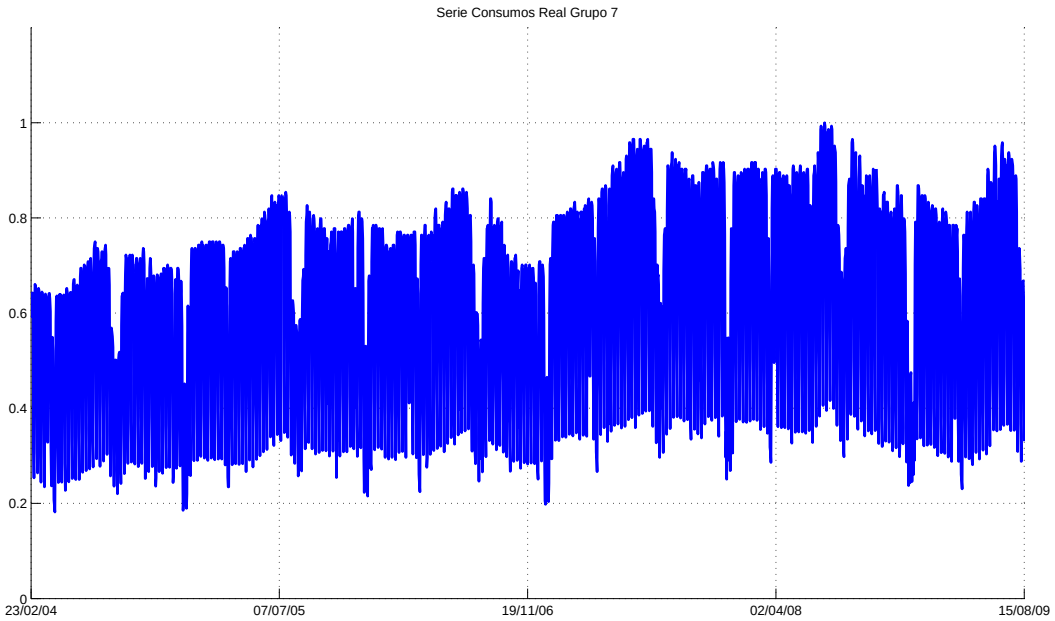

Figura 30. Serie Consumo Grupo 7

Después de realizar este proceso, las series resultantes tienen formas diversas. La serie del grupo 3 y 7 se representan algunos ejemplos de series de consumo agregadas tal y como se muestran en la Figura 29 y la Figura 30 respectivamente.

Como se puede apreciar en la Figura 30 serie existe una componente de tendencia creciente en el consumo de energía, mientras que la Figura 29 se pueden distinguir varias partes: una que mantiene constante su consumo, otra en la que se ven incrementado y finalmente una disminución en el consumo de energía.

Para modelar el efecto de la tendencia se van a normalizar las series utilizando las series de consumo anual de cada grupo. Las series de las figuras anteriores son las series sin normalizar, con la componente de la tendencia. Tras normalizarlas, se eliminará esta componente. Este proceso se detallará más adelante

3.3.4.2. Series Festividades Agregadas

Los festivos son fechas importantes a tener en cuenta a la hora de modelar las variables influyentes en el consumo de energía. En los días festivos, la actividad en las empresas baja notablemente. Por ello, es necesario construir una serie temporal que indique cuándo se producen estos momentos, para que el modelo predictivo pueda tenerlo en cuenta.

Es importante distinguir entre las festividades nacionales y las autonómicas. Se dispone de una serie binaria que indica cuándo es festivo nacional, y de 19 series binarias que indican las festividades autonómicas. La idea es construir una serie de festividades agregadas para cada grupo, combinando la serie de festividades nacionales con las de festividades automáticas.

La serie de festividades agregadas para cada grupo se define siguiendo la Ecuación 12.

$$festividad_agregada_i = festividad_nacional_i + \sum_{j=1}^{19} \alpha_{ij} * festividad_{ij} \quad (12)$$

$$\forall i \in dias\ periodo\ estudio$$

$$\text{donde} \qquad \alpha_{ij} = \frac{numero_clientes_j}{n\ total\ clientes}$$

$$\forall j \in CCAA$$

Esta nueva serie de festividades agregadas combina el efecto de las festividades nacionales con el efecto de las festividades autonómicas. El peso que tiene cada comunidad autónoma en la serie final se calcula dividiendo el número de clientes de la comunidad autónoma, entre el número total de clientes de todas las comunidades autónomas. Por lo tanto, se necesita una función que devuelva las series temporales del número de clientes por comunidad autónoma de cada grupo.

Al igual que para el cálculo de las series de consumo agregadas, se va a utilizar la estructura del Código 1, pero modificando la consulta para que el resultado que devuelva la función sea un *cell array* con 20 columnas, la primera para indicar la fecha y las 19 restantes para las series del número de clientes por comunidad autónoma.

3.3.4.3. Series Temperatura Agregadas

La temperatura es un factor muy influyente en el consumo de energía. El Instituto Nacional de Meteorología (INM) cuenta con numerosos observatorios distribuidos por la geografía española, para disponer de datos representativos de todas las zonas geográficas.

A cada cliente se le asocia un observatorio concreto, de tal manera que la temperatura que registre ese observatorio será la que caracterice su forma de consumo. Manejar una serie de temperaturas para cada cliente es complicado y costoso, más aun teniendo en cuenta que muchas de ellas se repiten, por lo que se van a construir las series de temperaturas agregadas para cada grupo.

La idea es similar a la expuesta en el aparatado anterior para el cálculo de las festividades. En este caso, consiste en sumar las series de temperatura, utilizando un factor de ponderación para cada una, obteniéndose una nueva variable: temperatura agregada. En la Ecuación 13

$$temperatura_agregada_i = \sum_{j=1}^{N_obs} \beta_{ij} * temperatura_{ij} \quad (13)$$

$$\forall i \in dias\ periodo\ estudio$$

$$\text{donde} \qquad \beta_{ij} = \frac{numero_clientes_j}{n\ total\ clientes}$$

$$\forall j \in Observatorios$$

La estructura del Código 1 sigue siendo válida, pero modificando la consulta para que el resultado que devuelva la función sea un *cell array* con una columna para las fechas y las restantes columnas con las series del número de clientes por cada unos de los observatorios.

En la Figura 31 se muestra un ejemplo de la serie de temperatura del grupo 3.

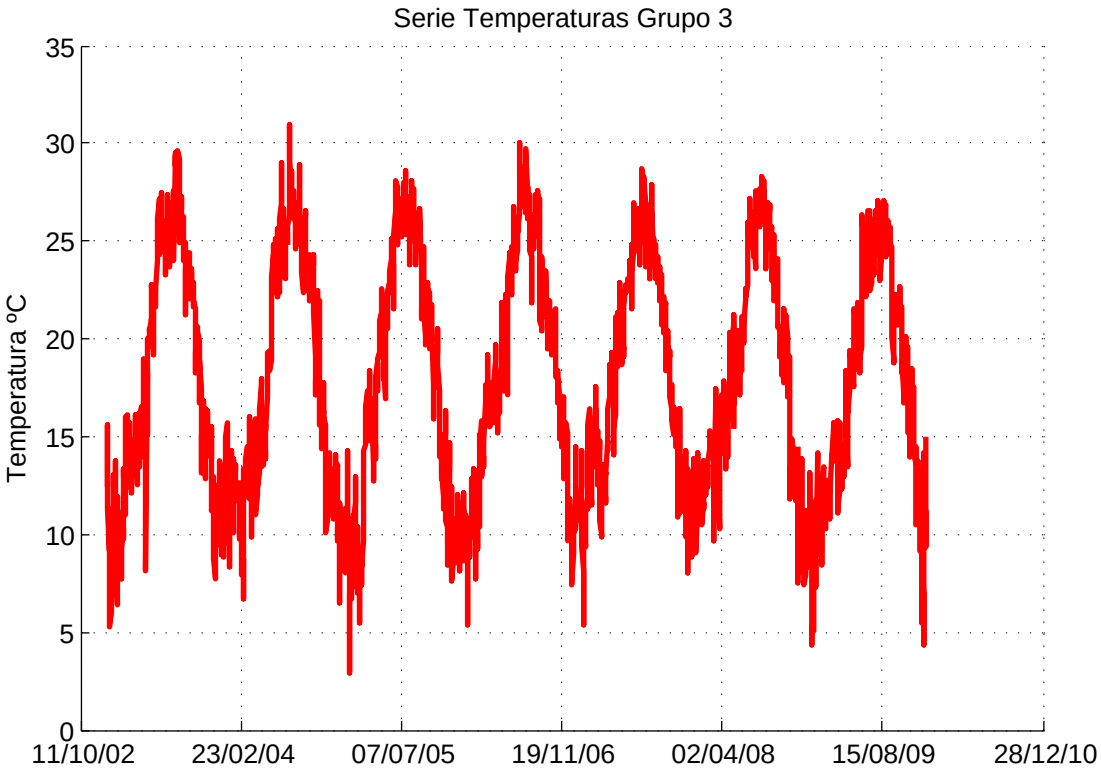

Figura 31. Serie Temperatura Grupo 3

Al observar esta serie temporal, se puede apreciar la estacionalidad de la temperatura. El mismo comportamiento se repite a lo largo de los años, comenzando el año con los valores mínimos de la temperatura (invierno), ir aumentando a medida que pasa el tiempo hasta alcanzar un máximo en julio (verano) y volver a descender hasta los

valores mínimos otra vez. A simple vista, la serie no muestra ninguna tendencia, ya que la serie se mantiene más o menos entre los mismo valores, sin aumentar ni disminuir.

3.3.4.4. Series Número de Clientes

El número de clientes es una variable muy importante a tener en cuenta en los modelos de medio plazo, puesto que va a determinar de forma importante el consumo final de la cartera. Por lo tanto, es una variable importante a tener en cuenta en nuestro modelo.

Anteriormente se han obtenido series con el número de clientes por comunidad autónoma y por observatorio para cada grupo. Se puede aprovechar cualquiera de estas dos matrices para obtener la serie de número de clientes por grupo, puesto que sumando para cada día el número de clientes en cada observatorio o en cada comunidad autónoma se obtendrán el número de clientes total que existen en ese día para ese grupo. De esta forma, se utilizar los resultados anteriores y no es necesaria ninguna consulta a las bases de datos.

3.4. Etapa 2: Modelos de regresión

Se va a utilizar un modelo de descomposición para realizar previsión de las series agregadas de consumo. Estos modelos separa la serie temporal en distintas componentes: tendencia, factor cíclico, estacionalidad y componente irregular. La tendencia se va a modelar utilizando las series de consumo anual. Una vez que se ha eliminado esta componente de las series temporales, se van a ajustar modelos de regresión lineal para modelar el factor cíclico y la estacionalidad, utilizando las series de temperatura, festividades agregadas y las variables indicadoras del calendario.

3.4.1. Series Consumo Anual

En la base de datos de los clientes se tiene almacenado un campo con el consumo anual que han tenido a lo largo de un año. Estos consumos dependen de la potencia contratada y suelen variar a lo largo del periodo de estudio. La serie de consumo anual, que va a ser la que se tome como base para normalizar las series de consumo agregadas, se construye sumando el consumo anual de los clientes que consumen cada día y que pertenezcan a un grupo determinado. Para obtenerlas, se va a utilizar la estructura del Código 1.

Esta serie es una forma de cuantificar la entrada, o salida, de clientes del grupo, que es la principal causante de que las series agregadas de consumo tengan cierta tendencia.

De esta forma, cuando un cliente aparezca por primera vez en un grupo, el valor de la serie de consumo de ese grupo aumentará, pero también lo hará el valor de la serie de consumo anual.

Como ya se ha comentado, el ajuste de los modelos de regresión se va a realizar para las series normalizadas de consumo. Esta nueva variable se calcula según la Ecuación 14

$$serie_normalizada_consumo_i = \frac{serie_consumo_i}{serie_consumo_anual_i} \qquad (14)$$
$$\forall i \in GRUPOS$$

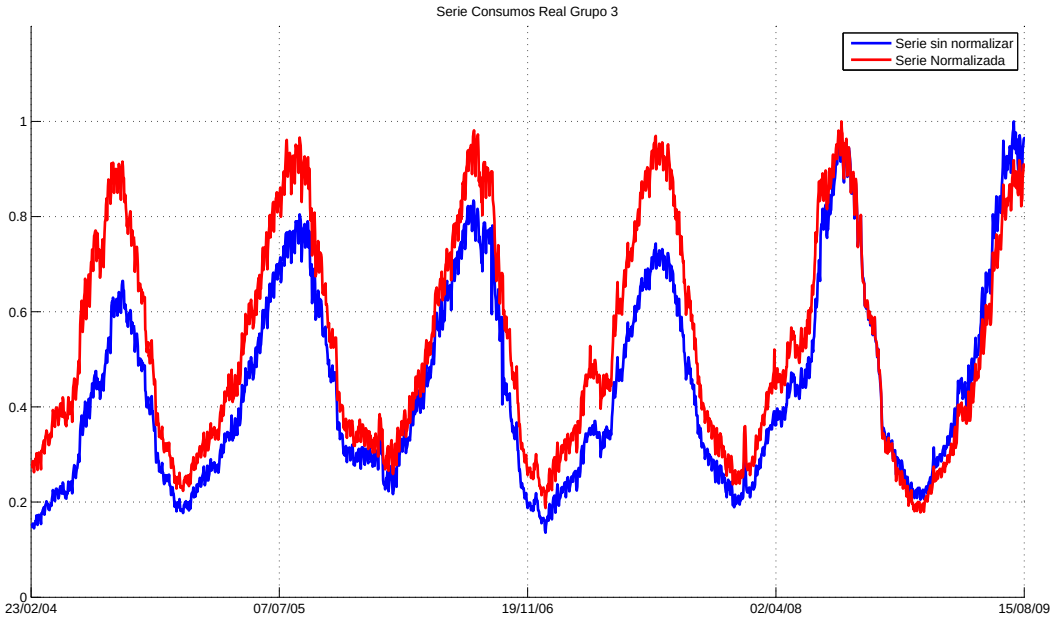

Figura 32. Serie Consumo Grupo 3 VS Serie Consumo Normalizada

Siguiendo con el ejemplo de los grupos 3 y 7, se van a comparar las series sin normalizar y las series normalizadas del consumo. Los resultados se muestran en la Figura 32 y la Figura 33, respectivamente. A simple vista se puede observar como en ambas figuras, la tendencia desaparece, sobre todo al principio del periodo. En la Figura 32 se comprueba que los ciclos ahora tienen aproximadamente la misma amplitud, y en la Figura 33, se corrigen los escalones de consumo del principio para dejarlo aproximadamente al mismo nivel que el resto de la serie.

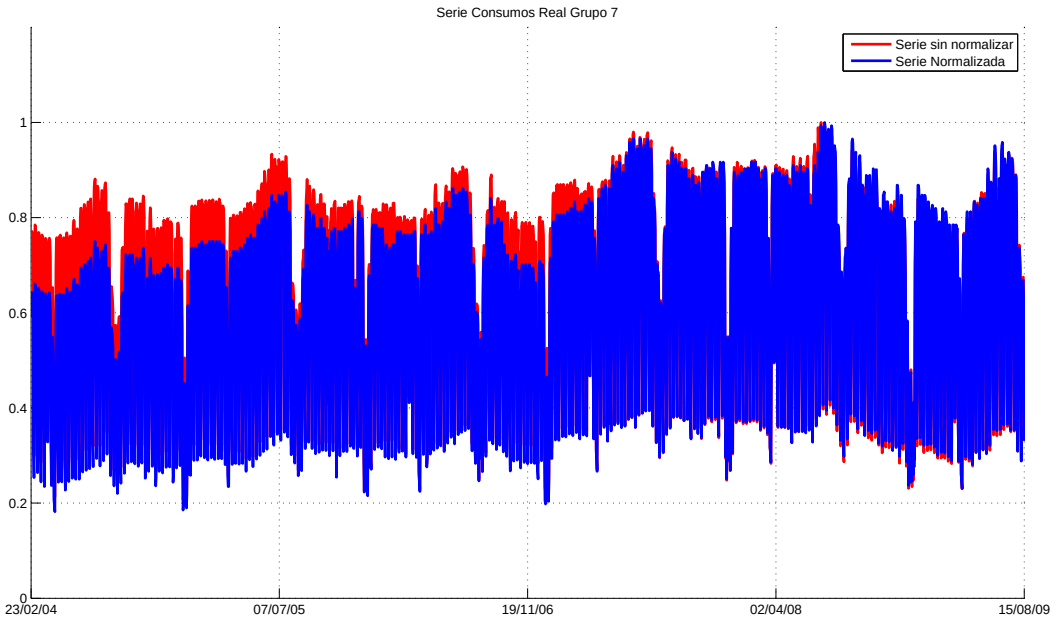

Figura 33. Serie Consumo Grupo 7 VS Serie Consumo Normalizada

3.4.2. Variables indicadoras calendario

Existen ciertos periodos, como las Navidades o la Semana Santa, que tienen un claro impacto en el consumo de energía eléctrica, por lo que deben modelarse correctamente.

El uso de variables indicadoras, o *dummy*, permite incluir este tipo de variables categóricas independientes como parte del modelo de regresión. En nuestro caso, las variables categóricas sólo tienen dos categorías, puesto que se van a referir a la existencia o no de un evento importante en una fecha concreta. Una variable indicadora X_d se define de la siguiente manera:

$$X_d = 1 \text{ si la observación corresponde a la categoría 1.}$$
$$X_d = 0 \text{ si la observación corresponde a la categoría 2.}$$

Las distintas variables indicadoras que se van a utilizar en el modelo se explican a continuación.

3.4.2.1. Tipo de mes

Esta variable indicadora se utiliza para diferenciar los distintos meses que hay en el calendario. Como se ha visto en la Figura 22, el consumo varía bastante en función del mes. Por esta razón, se han construido 12 variables, una por cada mes del año, que indican cuando una fecha está contenida en dicho mes. Formalmente se puede definir como:

$M_i = 1$ si la fecha corresponde al mes i.
$M_i = 0$ si la fecha no corresponde al mes i.
$i = 1, 2 \cdots, 12;\ i \neq 8$

En la definición anterior no se ha incluido el mes de agosto, ya que va a tener un tratamiento especial. Debido a que es un mes que altera bastante el consumo por ser el mes central del verano, se va a dividir esta variable en cuatro variables, una por cada una de las semanas que tiene agosto, para que exista más flexibilidad del modelo a la hora de ajustar los consumos.

3.4.2.2. Navidad

Al igual que las variables anteriores, la Navidad es un acontecimiento que tiene una gran influencia en el consumo, ya que éste tiende a bajar en ese periodo. Al igual que agosto, la Navidad también se ha dividido, pero en dos partes:

$Nav_1 = 1$ si la fecha $\in$ [20 diciembre-31 de diciembre].
$Nav_1 = 0$ en caso contrario i.
$Nav_2 = 1$ si la fecha $\in$ [1 de enero-6 de enero].
$Nav_2 = 0$ en caso contrario i.

3.4.2.3. Semana Santa

Es un periodo en el que también se ve alterado el consumo de energía debido a las festividades que se celebran. Los periodos que abarca la Semana Santa de cada año se recogen en la Tabla 4.

Año	Fecha Inicio	Fecha Fin
2003	13 abril	20 abril
2004	4 abril	11 abril
2005	20 marzo	27 marzo
2006	9 abril	16 abril
2007	1 abril	8 abril
2008	16 marzo	23 marzo
2009	5 abril	12 abril

Tabla 4. Fechas Inicio y Fin Semana Santa

La variable correspondiente a la Semana Santa quedaría de la siguiente forma:

$SS = 1$ si la fecha $\in$ [periodo Tabla 4].
$SS = 0$ En caso contrario.

3.4.2.4. Periodos

Por último, se va a dividir el horizonte temporal en tres tramos: 2003-2006, 2007-2008, 2009. Esta distinción viene a modelar los distintos regímenes de comportamiento de la demanda a lo largo del tiempo. El efecto de la crisis, ha tenido una importante repercusión en el consumo de energía, hecho que se observa en las series donde, aunque aumentó el número de clientes, los consumos no lo hicieron con la misma celeridad. Con todas estas variables, se va a proceder a ajustar los diferentes modelos, analizar las series temporales que construyen los modelos de regresión cuantificando sus errores y comparar la serie real con la previsión realizada.

$P1 = 1$ si el año $\in$ [2003-2006].
$P1 = 0$ En caso contrario.
$P2 = 1$ si el año $\in$ [2007-2008].
$P2 = 0$ En caso contrario.
$P3 = 1$ si el año $\in$ [2009].
$P3 = 0$ En caso contrario.

3.4.3. Modelos de regresión

Una vez que se han obtenido todas las series necesarias para aplicar los modelos de regresión, el siguiente paso es ajustar los modelos de regresión lineal. Como se vio en los perfiles semanales, el consumo de energía no es el mismo a lo largo de la semana, por lo que se ha decidido ajustar un modelo para cada tipo de día. En resumen, se van a confeccionar $25\ grupos\ x\ 7\ dias\ =\ 175\ modelos$ de regresión lineal independientes, que después sumados van a ajustar la demanda total de energía. En forma de ecuación, cada uno de los modelos tiene la forma que se muestra en la Ecuación 15.

$$d_{ij} = \alpha_0 + \alpha_1 * NumClient_{ij} + \alpha_2 * Temp_{ij} + \alpha_3 * Fest_{ij} + \vec{\alpha}_4 * Cal_{ij} \quad (15)$$

1. i: índice que indica el grupo, con $i = 1, 2, \cdots, 25$.
2. j: índice que indica el tipo de día, con $j \in \{L, M, X, J, V, S, D\}$
3. $NumClient_{ij}$: serie del número de clientes del grupo i para el tipo de día j.
4. $Temp_{ij}$: serie de la temperatura agregada del grupo i para el tipo de día j.
5. $Fest_{ij}$: serie de las festividades agregadas del grupo i para el tipo de día j.
6. Cal_{ij}: conjunto de variables indicadoras o *dummy* que se usan para explicar el calendario.
7. $\vec{\alpha}_k$: coeficientes de la regresión.

Tras realizar el ajuste del modelo para cada una de las series, se obtienen los coeficientes de regresión. Evaluando estos modelos se obtienen las previsiones de las series de consumo normalizadas. En las figuras siguientes se muestran ejemplos de algunas series normalizadas de los grupos 3 y 7 y sus ajustes.

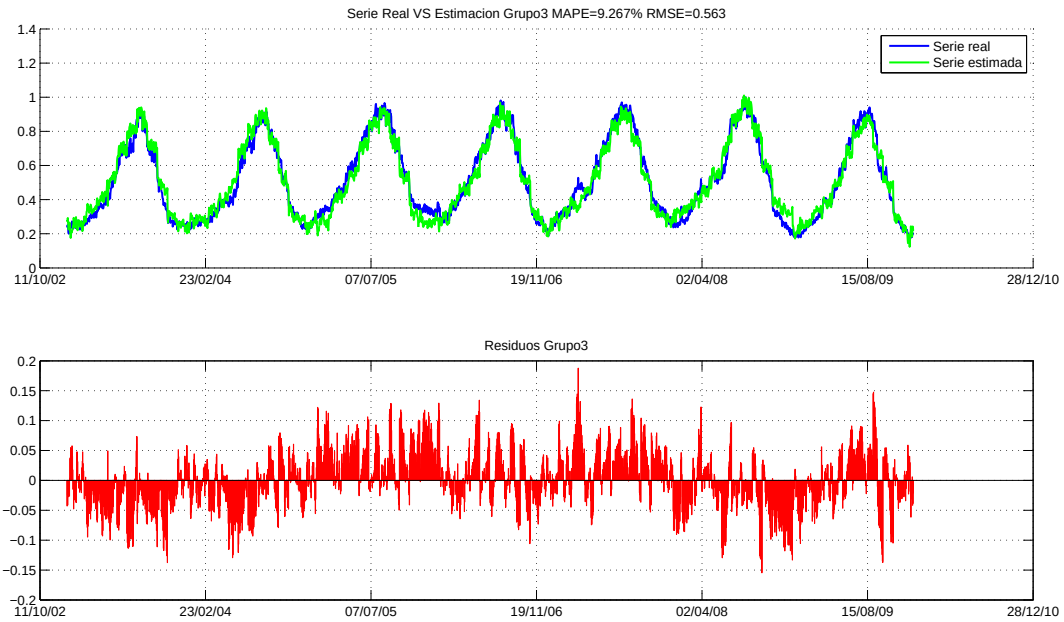

Figura 34. Regresión Grupo 3.

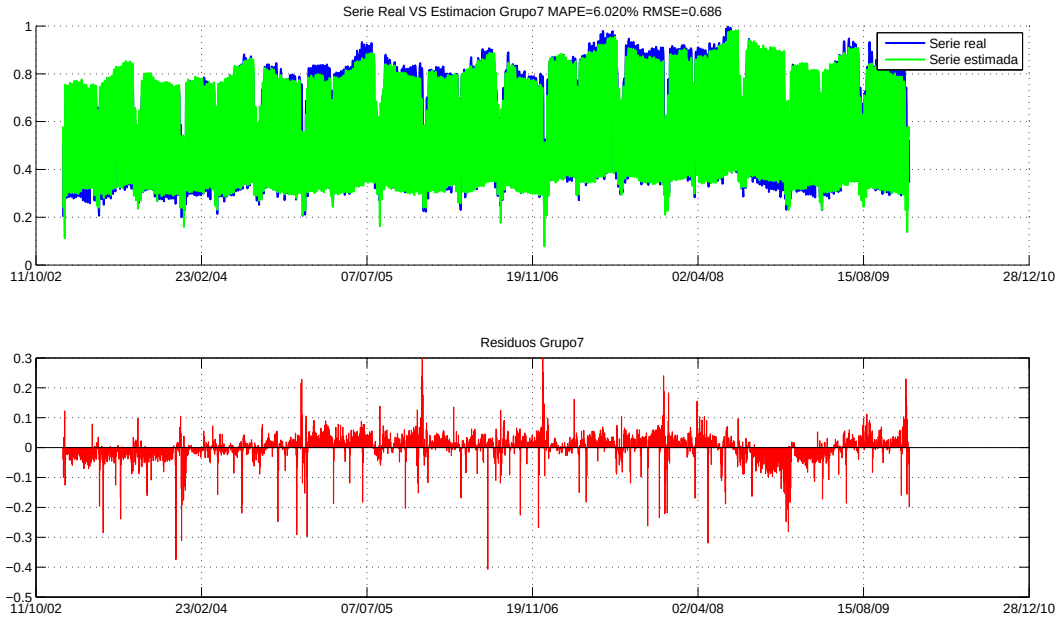

Figura 35. Regresión Grupo 7.

La gráfica de los residuos es el resultado de restar la serie real de consumo menos la estimación de la regresión. El MAPE y el RMSE se definen en la Ecuación 16 y la Ecuación 17.

Capítulo 4

Resultados

"No hay más que un modo de dar una vez en el clavo,
y es dar cientos de veces en la herradura."
MIGUEL DE UNAMUNO

En este capítulo se presenta una aplicación real del modelo propuesto a los datos de consumo de una comercializadora. Se ha llevado a cabo el ajuste del modelo para el consumo de energía de una muestra de la cartera de clientes industriales de ENDESA.

4.1. Introducción

Se ha utilizado una muestra de clientes industriales de la cartera de ENDESA para ajustar, usando el modelo propuesto, su consumo durante el periodo comprendido entre el 1 de enero de 2003 hasta el 31 de diciembre de 2009 inclusive. Por cada uno de los 84 meses que componen el periodo de estudio se dispone de una base de datos que contienen los 24 datos de consumo horario de los clientes.

Las distintas funciones programadas en Matlab, así como la herramienta Idat han sido ejecutadas en un ordenador con las siguientes características:

- **Procesador**: Intel®Core™Duo CPU P8600 **@**2.40GHz.
- **Memoria RAM**: 4GB.
- **Tipo de Sistema Operativo**: Sistema operativo 32bits.
- **Sistema Operativo**: Windows Vista™Home Premium ©2007 Service Pack 2.

4.2. Tiempos Ejecución

Dado el volumen de datos manejado, tanto la preparación de la información como el ajuste de los diferentes submodelos requiere tiempos que no son despreciables. Es importante recordar que no era objetivo de este trabajo desarrollar una herramienta de ajuste optimizada en velocidad. Tanto las consultas como las distintas funciones desarrolladas en Matlab se pueden acelerar. De acuerdo con el tiempo de ejecución de las funciones, la construcción del modelo se puede dividir en tres partes:

Parte 1: Cálculo de los perfiles de consumo

Durante esta primera parte se calculan los valores de las variables definidas en el capítulo anterior. Esto supone calcular 48 valores de consumo diario, 14 valores de consumo semanal y 12 valores de consumo anual para cada uno de los clientes de la muestra. Este cálculo requiere un tiempo de ejecución de aproximadamente 24 horas (conviene recordar que el cálculo de los perfiles sólo se realiza una vez, a no ser que se elija otra muestra).

Parte 2: Ajuste del KSOM

Una vez obtenidos los valores que caracterizan a los clientes, el siguiente paso consiste en agruparlos automáticamente utilizando un modelo KSOM. El ajuste de un mapa de Kohonen de 10000 con 40 variables cada uno en 25 grupos tras 10000 iteraciones se lleva a cabo en aproximadamente 10 minutos. El tiempo de ajuste del mapa es muy sensible al número de iteraciones y al número de grupos que se definan.

Parte 3: Cálculo Series Agregadas

Con los grupos definidos, sólo resta el cálculo de las series agregadas. Como se explicó en el capítulo 3, serán necesarias las series de consumo agregadas y las series con el número de clientes por comunidad y por observatorio para cada grupo. La función que calcula las series de consumo agregadas requiere 5 horas para completar todos los grupos, mientras que cada una de las otras dos funciones requieren aproximadamente 8 horas, debido a que manejan mayor número de datos y las consultas que realizan son de mayor complejidad.

4.3. Ajuste de los modelos de regresión

Para cuantificar los errores cometidos por los modelos, se han calculado dos parámetros de error muy utilizados en el campo de las series temporales: el MAPE (“Mean Absolute Percentage Error”) y el RMSE (“Root Mean Square Error”). La notación empleada para definir los errores es la siguiente:

- Y_i: Demanda Real de energía.
- $\hat{Y}_i$: Demanda estimada de energía, resultante del modelo propuesto.

El MAPE es una estimación del error porcentual que el modelo comete. Se calcula como el promedio de las diferencias absolutas entre los valores estimados y los reales. Se expresa como porcentaje de los valores reales. Se define matemáticamente en la Ecuación 16.

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|Y_i - \hat{Y}_i|}{Y_i} * 100\,\% \tag{16}$$

Por otro lado, el RMSE, conocido como error estándar, es la raíz cuadrada de la suma de varianzas. En la Ecuación 17 se presenta la fórmula de este parámetro.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2} \tag{17}$$

Estos parámetros son los indicadores de la exactitud de los modelos construidos, y como tales se han empleado para guiar el proceso iterativo de mejora del modelo durante la fase de diseño. Estos errores se pueden medir en distintos subconjuntos de datos. Es muy habitual medir el error en el conjunto de aprendizaje y en un conjunto de validación, para controlar el posible sobreaprendizaje. Por otro lado, también se puede medir el MAPE y el RMSE en los distintos submodelos que constituyen el modelo global para identificar las fuentes de mayor error y así introducir mejoras más efectivas.

4.4. Regresión Serie Consumo Total

Todos los análisis realizados anteriormente sobre los clientes, búsqueda de nuevas variables, aplicación de técnicas de *clustering*, etc., iba encaminado a conseguir realizar la previsión de la demanda de energía eléctrica mediante un modelo de descomposición basado en modelos sencillos de regresión, intentando separar las series de consumo de los clientes que puedan presentar características comunes. El resultado de todo el trabajo se muestra en la Figura 36.

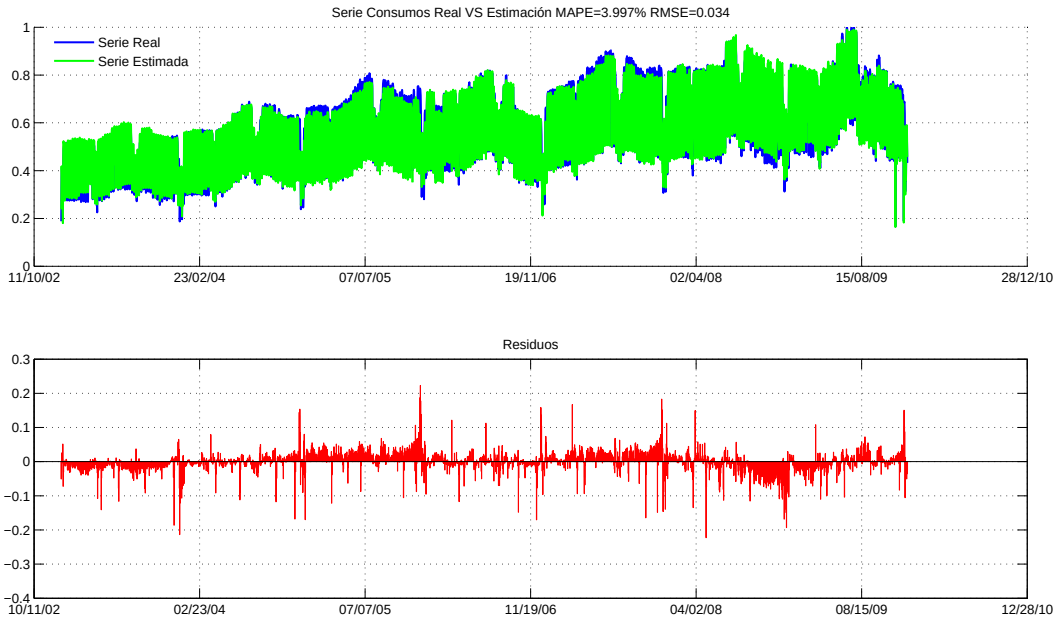

Figura 36. Serie Real Consumo VS Serie Estimada. Residuos

El resultado obtenido es un modelo que ajusta el consumo real agregado de la muestra de clientes con un **MAPE=4.0 %** y un **RMSE=0.034**.

Un análisis de los residuos[18], permite observar como existen momentos puntuales en que el modelo se equivoca, por exceso o por defecto. Estos días en que el modelo no es capaz de estimar con exactitud la demanda de energía puede ser debido entre otros motivos a que sean días festivos a nivel local, que no quedaron recogidos en la serie de festividad (se han cogido sólo las festividades autonómicas) y por ello el modelo no consigue distinguirlos de los días normales.

En el año 2008 se pueden observar unos residuos de valor negativo. A lo largo de este año, en el cual comenzó la crisis económica, el modelo calcula un valor de energía por encima del que realmente se consumió. Sin embargo, durante el año 2009, los residuos del modelo son de menor valor y presentan un carácter más aleatorio.

También puede observarse cómo los residuos son diferentes de un año para otro. A lo largo del mismo año, el residuo se mantiene en valores positivos o negativos, ya sea porque el modelo estima por defecto o por exceso. En la Tabla 5 se presenta los MAPE's de los distintos años del periodo de estudio.

[18]Diferencia entre la medida real y la medida prevista

2003	2004	2005	2006	2007	2008	2009
5.09 %	3.06 %	4.88 %	3.10 %	3.89 %	4.67 %	3.30 %

Tabla 5. MAPE's separado por años.

El MAPE a lo largo de los años es bastante similar, variando entre el 3 % y el 5 %. Como puede verse, el mayor error se comete en el primer año del periodo de estudio, probablemente debido a la peor calidad de los datos. A continuación Se observa un error más alto coincidiendo con el inicio de la crisis económica en el año 2008. A continuación, es en el año 2009 donde el error se mantiene en torno al 3 %, a pesar de la situación económica existente. Este hecho muestra que el modelo ha conseguido ajustar adecuadamente este periodo.

4.5. Comparación Modelos Literatura

Para poder evaluar la calidad de los resultados obtenidos, se han comparado los errores calculados del modelo propuesto con los errores que se han publicado en algunos artículos científicos. Nótese que tanto las series de demanda como el horizonte de predicción difiere de los utilizados en este trabajo. Por tanto, esta comparativa es meramente cualitativa. Estos se recogen en la Tabla 6.

Trabajo	Periodo estudio	Error periodo
[16]	1996	4.76 %
	1997	4.65 %
[20]	enero 1993 - abril 1998	$\pm$1.00 % - $\pm$4.00 %
[19]	agosto 1995 - diciembre 1999	1.00 %
[4]	1987 - 1992	GR-SC 18.3 % ANN 3.63 %

Tabla 6. Comparación errores literatura.

El modelo presentado se encuentra entre medias de los modelos de la literatura, con un error de aproximadamente un 4 %, similar al que obtienen en [16] o en [4], pero ajustando un periodo de estudio mayor y con una coyuntura económica distinta.

Capítulo 5

Conclusiones

"Una experiencia nunca es fracaso, pues siempre viene a demostrar algo"
THOMAS ALVA EDISON

En el presente proyecto se ha desarrollado un modelo de previsión de la demanda de energía eléctrica a medio plazo y se ha aplicado para ajustar el consumo de energía de los clientes industriales de una comercializadora. En un campo tan extenso y con multitud de alternativas como es el de los modelos de previsión, se ha presentado una forma totalmente novedosa, que combina las técnicas de *clustering* con modelos de regresión lineal.

Los resultados obtenidos con la metodología propuesta muestran un error MAPE del 4 % al ajustar los datos de consumo históricos desde 2003 hasta 2009. Tras revisar los resultados publicados en los artículos consultados, se deduce que el modelo propuesto puede ser válido para su uso en la predicción de la demanda de energía eléctrica. Por esta razón, puede convertirse en una alternativa a los métodos comentados en el Capítulo 2.

Entre las virtudes de esta nueva metodología se encuentra la flexibilidad del modelo. La idea de agrupar los clientes en grupos atendiendo las características que definen a cada uno, permite realizar diferentes agrupaciones en función de los rasgos que se elijan. De esta manera, es posible obtener diferentes agrupaciones, para elegir entre la más homogénea y que permita una previsión con el mínimo error.

Otra virtud de esta metodología es su versatilidad. En el estado del arte se han mencionado otros tipos de metodologías para la previsión de series temporales, como los modelos ARIMA o las redes neuronales. Una vez que los grupos están definidos y calculadas las distintas series temporales, es sencillo cambiar los modelos de regresión por otros métodos que se consideren más adecuados, pudiendo encontrar una metodología que mejor se ajuste a las particularidades de las series en concreto.

El desarrollo de una nueva metodología para la previsión de la demanda de energía eléctrica conlleva ir explorando distintas variantes que pueden añadirse al modelo propuesto, actualizándose y mejorando las debilidades encontradas.

Capítulo 6

Futuros desarrollos

"Nunca he encontrado una persona tan ignorante
que no se pueda aprender algo de ella."
ALBERT EINSTEIN

Los distintos ámbitos científicos están siempre en continua evolución y mejora, y el campo de los modelos de predicción no podía ser una excepción, ofreciendo un amplio y variado abanico de posibilidades a la hora de realizar previsiones, no sólo aplicadas al campo de la demanda de energía eléctrica, sino también a otras variables de carácter económico.

Se han detectado posibles mejoras que sería importante estudiar de cara a la optimización del modelo de previsión. En primer lugar, habría que repasar la programación utilizada para realizar tanto las consultas a las bases de datos como las funciones de cálculo y ajuste de modelos, para mejorar los tiempos de ejecución. Como ya se ha comentado, el tiempo nada despreciable de cálculo que requieren las distintas funciones y consultas es un *handicap*. Para ello MATLAB cuenta con la herramienta "profile", la cual, entre otras cosas, calcula el tiempo de ejecución de cada función, y muestra que variables puedes descartarse del código para acelerarlo.

También se podría estudiar otra forma de modelar la tendencia de las series agregadas de consumo. El uso de los consumos anuales para modelar la tendencia podría no resultar del todo satisfactorio, sobre todo de cara a realizar previsiones. Por lo que podría ser interesante intentar asociar la tendencia a otras variables que normalicen las series de manera aceptable y de las que se puedan obtener previsiones con facilidad de una fuente externa. Las mejores candidatas pueden ser variables económicas, aunque no hay que descartar el uso de alguna variable sintética.

Siguiendo con la mentalidad fijada en la previsión de la demanda de electricidad, sería muy interesante desarrollar un mecanismo automático para clasificar a los clientes entrantes en los grupos existentes. La idea sería introducir las características disponibles del nuevo cliente en un sistema, para que las evalúe y determine el grupo al que pertenece.

Por último, sería conveniente añadir alguna variable derivada de la temperatura que recoja las inercias térmicas, distinguiendo entre los periodos de invierno y de verano.

Bibliografía

[1] **Cha, J., Park, J., Choi, J., Jung, Y. and Yun, Y.**, *Determination of A Deterministic Reliability Criterion for Composite Power System Expansion Planning*. Power & Energy Society General Meeting IEEE, pp 1-6, Jun 09.

[2] **Ahsan, Q. and Uddin, M.**, *A Probabilistic Approach of Electrical Energy Forecasting*. Conferente Proceedings of the IEEE Instrumentation and Measurement Technology Conference, pp 1070-1074, May 2005.

[3] **Al-Alawi, S.M. and Islam S.M.**, *Principles of electricity demand forecasting. Part 1 Methodologies*. Power Engineering Journal, Vol.10, pp 139-143, Jun 1996.

[4] **Al-Alawi, S.M. and Islam S.M.**, *Principles of electricity demand forecasting. Part 2 Applications*. Power Engineering Journal, Vol.10, pp 91-95, April 1997.

[5] **Padmakumari, K., Mohandas, K.P. and Thiruvengadam, S.**, *Application Of Fuzzy System Theory In Land Use Based Long Term Distribution Load Forecasting*. International Conference on Energy Management and Power Delivery, Proceedings of EMPD IEEE, Vol.1, pp 311-316, Mar 1998

[6] **Teixeira, M.A. and Zaverucha, G.**, *Fuzzy Multi-Hidden Markov Predictor in Electric Load Forecasting*. International Joint Conference on Neural Networks, Proceedings IJCNN IEEE, Vol.3, pp 1758-1763, August 2005.

[7] **Da-peng, L. and Dong-hai, M.**, *The forecasting of China's Energy Consumption* International Conference on Management Science & Engineering, Proceedings ICMSE IEEE, pp 1557-1562, September 2008

[8] **Meng, M. and Shang, W.**, *Research on Annual Electric Power Consumption Forecasting Based on Partial Least-squares Regression*. International Seminar on Business and Information Management, vol1, pp 125-127, 2008.

[9] **Batmunkh, S. and Batjargal, Ts.**, *Modeling Approach of Energy Demand Based on Macroeconomic Indices* The 1st International Forum on Strategic Technology, pp 383-384, October 2006.

[10] **Broehl, J.H.**, *An End-Use Approach to Demand Forecasting* IEEE Transactions on Power Appartus and Systems, vol. Pas-100, No. 6 June 1981.

[11] **Choi, J., Thomas, R., Wang, Z., El-Keib, A. and Billinton, R.**, *A Study on Probabilistic Optimal Reliability Criterion Determination in Composite Power System Expansion Planning*. IEEE Power Engineering Society General Meeting, vol.2 pp1277-1284, June 2005.

[12] **Mahmoud, M.O.M and F. Jaidane-Saidane, M.**, *Long Term Multi-Scale Analysis of the Daily Peak Load Based on the Empirical Mode Decomposition* IEEE Bucharest Power Tech Conference, pp 1-6, June/July 2009.

[13] **Asbury, C.E.**, *Weather Load Model for Electric Demand and Energy Forecasting*. IEEE Transactions on Power Appartus and Systems, vol. Pas-94, No. 4, July/August 1975.

[14] **Dong-xiao, N., Yan-chang, L. and Qing, Z.**, *Research of Residul Error-Participle Swarm Optimization Gray Model Based on Markov in Load Forecasting*. IEEE International Conference on Grey Systems and Intelligent Services, November 2007.

[15] **Badran, I., El-Zayyat, H. and Halasa, G.**, *Short-Term and Medium-Term Load Forecasting for Jordan's Power System*. American Journal of Applied Sciences 5 (7): 763-768, July 2008.

[16] **Alturki, F.A. and ABdennour, A.B.**, *Medium to Long-Term Peak Load Forecasting for Riyadh City Using Artificial Neural Networks*. Journal of King Saud University. Engineering Sciences. Vol 12, No. 2 (2000/1420)

[17] **Simon, G., Lendasse, A., Cottrell, M. and Verleysen M.**, *Long-Term Time Series Forecasting Using Self-Organizing Maps: the Double Vector Quantization Method*. ANNPR'2003 proceedings. Artificial Neural Networks in Pattern Recognition Florence (Italy), IAPR-TC3, pp 8-14, September 2003.

[18] **Lendasse, A., Lee, J., Werz, V. and Verleysen M.**, *Time Series Forecasting using CCA and Kohonen Maps - Application to Electricity Consumption*. ESANN'2000 proceedings. European Symposium on Artificial Neural Networks Brugues (Belgium), pp 329-334, April 2000.

[19] **Moral Carcedo, J. y Vicéns Otero, J.**, *Un Modelo de Previsión de la Demanda de Energía Eléctrica: THOR II*. Instituto L. R. Klein. Centro Gauss. U.A.M. D.T. No. 4 Noviembre 2003

[20] **Valor, E., Climent, F.J., Meneu, V. y Caselles, V.**, *El Modelo Español de Consumo Sectorial de Electricidad*. Revista Española de Física, Vol. 16, No. 3, pp 24-29, 2002.

[21] **Moral Carcedo, J. y Vicéns Otero, J.**, *Modelos STR para el Análisis de la Respuesta no Lineal de la Demanda de Electricidad a la Temperatura*. Instituto L. R. Klein. Centro Gauss. U.A.M. abril 2004.

[22] **Sánchez-Úbeda, E.F. y Berzosa, A.**, *Modeling and Forecasting Industrial End-Use Natural Gas Consumption*. Energy Economics 29 (2007) 710-742.

[23] **Labys, W.C.**, *Modelling Mineral and Energy Markets*. Kluwer Academic Publishers.

[24] **Gutiérrez Moya, E.** *Tesis Doctoral: La demanda de residencial de energía eléctrica en la Comunidad Autónoma de Andalucía: un análisis cuantitativo. Universidad de Sevilla 2003* www.cervantesvirtual.com

[25] **Nettleton, D. F.**, *Técnicas para el análisis de datos clínicos*. Ediciones Díaz de Santos 334 págs..

DOCUMENTO II

PLIEGO DE CONDICIONES

Índice

1. Condiciones generales y económicas **3**

1.1. Condiciones generales 3

1.2. Condiciones económicas 4

2. Condiciones técnicas y particulares **5**

2.1. Equipo informático 5

2.2. Condiciones *Software* 6

2.3. Normas de seguridad e higiene 6

Capítulo 1

Condiciones generales y económicas

El presente Pliego de Condiciones tiene por objeto definir fijar las condiciones técnicas y económicas, así como las condiciones generales y contractuales que han de regir en la ejecución del modelo de previsión de la energía eléctrica, detallada en el presente proyecto. Las condiciones y cláusulas que se establecen en este documento son de obligado cumplimiento por las partes contratantes.

1.1. Condiciones generales

I. Las partes contratantes del presente proyecto se comprometen, desde la firma del mismo, a llevar a cabo todas las disposiciones que en él se estipulen.

II. Las reclamaciones o discrepancias en lo concerniente al cumplimiento de lo pactado por cualquiera de las partes, serán tramitadas por las autoridades legales competentes dispuestas para este fin, en caso de no llegar a un acuerdo acuerdo mutuo. El dictamen o sentencia será acatado por ambas partes.

III. La firma del contrato implica que el suministrador se compromete a facilitar toda la información necesaria para la instalación y buen funcionamiento del sistema, siempre que se requieran sus conocimientos para este fin.

IV. De igual manera, el comprador está obligado a detallar al fabricante todas las características propias del sistema en que se va a implantar, con el fin de facilitar su instalación, quedando el proveedor exento de cualquier responsabilidad sobre las incidencias que surjan por el incumplimiento de dicha obligación.

V. Los retrasos imputables al suministrador, se penalizarán con una indemnización del 1 % del valor estipulado por semana de retraso.

VI. A partir de la entrega del sistema, existirá un plazo de garantía por validez de un año. Dicha garantía quedará invalidada si se demuestra que el sistema ha sido alterado.

VII. Agotado el plazo de garantía, quedará a criterio del suministrador atender las peticiones del cliente y fijar la compensación por la reparación del sistema.

1.2. Condiciones económicas

I. Los precios indicados en este proyecto son inamovibles y exentos de revisión bajo ningún concepto, siempre y cuando sean aceptados dentro del periodo de validez del presupuesto, que se fija hasta marzo de 2010.

II. El pago del proyecto se realizará de la forma que sigue:

- 30 % En el momento de la firma del contrato.
- 70 % A la entrega del proyecto.

III. El pago del proyecto se realizará al contado mediante un cheque nominativo o bien una transferencia bancaria a la cuenta correspondiente. En ningún caso se aceptarán letras de cambio.

IV. Durante el plazo de garantía estipulado, la totalidad de los gastos generados por las reparaciones y sus derivados, correrán por cuenta del suministrador.

V. Fuera del plazo de garantía, los costes de las reparaciones necesarias serán fijados exclusivamente por el suministrador del sistema.

Capítulo 2

Condiciones técnicas y particulares

En la utilización del proyecto de aplicación informática se derivan una serie de condiciones que limitan de alguna forma la utilización del mismo. Estas condiciones resultaron como consecuencia de llevar a la práctica la realización del proyecto.

2.1. Equipo informático

I. El equipo informático donde se va a utilizar el sistema deberá de cumplir con las homologaciones dispuestas por la reglamentación europea y española vigentes a fecha de septiembre de 2010.

II. La instalación del equipo deberá llevarse a cabo en un lugar que cumpla los niveles de humedad y temperatura dispuestos por el fabricante para el buen funcionamiento del mismo.

III. Todos los programas informáticos empleados deberán contar con una licencia validada por el fabricante y se deberán cumplir los términos y acuerdos que en ella figuran. En caso de emplear *software* con licencia GNU[1], también existe la obligación de cumplir con sus condiciones y clausulas.

IV. El equipo informático debe de satisfacer las exigencias de rendimiento recomendadas por el fabricante de los programas que son necesarios para su ejecución.

V. El disco duro del sistema informático debe de disponer de una capacidad de al menos 35 GB.

[1]La Licencia Pública General de GNU es una licencia creada por la *Free Software Foundation* y está orientada principalmente a proteger la libre distribución, modificación y uso de software.

2.2. Condiciones *Software*

I. El sistema operativo utilizado deberá de ser Windows XP Professional o superior.

II. Todos las versiones utilizadas de los distintos programas informáticos deberán de ser compatibles con el sistema operativo utilizado.

III. La versión de Matlab instalada deberá ser la versión de Matlab 7.5 (R2007b) o superior.

IV. La versión de la MCR de Matlab instalada deberá ser la versión 7.7.

V. La versión de Microsoft Access instalada deberá ser la versión 2003 o superior.

2.3. Normas de seguridad e higiene

La realización del proyecto cumplirá en todo momento con la Ley 31/1995 de Prevención de Riesgos Laborales.

DOCUMENTO III

PRESUPUESTO

Índice

1. Mediciones **3**

1.1. *Hardware* 3

1.2. *Software* 3

1.3. Mano de obra 4

2. Precios unitarios **5**

2.1. *Hardware* 5

2.2. *Software* 5

2.3. Mano de obra 6

3. Sumas parciales **7**

3.1. *Hardware* 7

3.2. *Software* 8

3.3. Mano de obra 8

4. Presupuesto general **9**

Capítulo 1

Mediciones

A lo largo de este capítulo se recogen de forma desglosada los equipos y sistemas utilizados, así como las horas de utilización de cada uno a lo largo del proyecto.

1.1. *Hardware*

Elemento	Cantidad	Tiempo(h)
Ordenador personal	1	640

Tabla 1. Mediciones *Hardware*.

1.2. *Software*

Programa	Cantidad	Tiempo(h)
Matlab 7.5 (R2007b)	1	400
Databababase Toolbox	1	400
Microsoft Access 2003	1	300
MikTeX 2.7 (Distribución LaTeX	1	150
TeXnicCenter v1.0	1	150
Adobe Acrobat 9 Pro Extended	1	150

Tabla 2. Mediciones *Software*.

1.3. Mano de obra

Actividad	**Tiempo(h)**
Documentación del proyecto	50
Diseño Bases de Datos	150
Construcción de modelos	100
Programación	400
Redacción Memoria	150
Total Horas	850

Tabla 3. Mediciones Mano de Obra.

Capítulo 2

Precios unitarios

En este capítulo se van a detallar los costes referidos a los elementos necesarios para llevar a cabo el proyecto, citados anteriormente.

2.1. *Hardware*

Elemento	**Cantidad**	**Precio(€)**
Ordenador personal	1	699

Tabla 4. Costes *Hardware*.

2.2. *Software*

Programa	**Cantidad**	**Precio(€)**
Matlab 7.5 (R2007b)	1	2000
Database Toolbox	1	1000
Microsoft Office	1	639
MikTeX 2.7 (Distribución LaTeX	1	*Software libre*
TeXnicCenter v1.0	1	*Software libre*
Adobe Acrobat 9 Pro Extended	1	659.62

Tabla 5. Costes *Software*.

2.3. Mano de obra

Actividad	**Tiempo(h)**
Documentación del proyecto	30
Diseño Bases de Datos	20
Construcción de modelos	20
Programación	20
Redacción Memoria	30

Tabla 6. Costes Mano de Obra.

Capítulo 3

Sumas parciales

Utilizando los cálculos anteriores, se va a imputar el coste de cada uno de los elementos parciales que han sido necesarios para realizar el proyecto

3.1. *Hardware*

Se aconseja amortizar los equipos en cuatro años, amortizando cada año la mitad del importe pendiente, es decir el primer año un 50 %, el segundo un 25 % y los dos últimos un 12,5 %.

Elemento	**Cantidad**	**Precio(€)**	**Amortización**	**Coste(€)**
Ordenador personal	1	699	50 %	349.5
			TOTAL	349.5

Tabla 7. Total *Hardware*.

3.2. *Software*

En el Software el tiempo de amortización es de 8 años, algo más amplio que en el caso anterior.

Programa	**Cantidad**	**Precio(€)**	**Amortización**	**Coste(€)**
Matlab 7.5 (R2007b)	1	2000	12.5 %	250
Database Toolbox	1	1000	12.5 %	125
Microsoft Office	1	639	12.5 %	79.88
Adobe Acrobat 9 Pro Extended	1	659.62	12.5 %	82.45
			TOTAL	537.33

Tabla 8. Total *Software*.

3.3. Mano de obra

Actividad	**Tiempo(h)**	**Precio/horas**	**Coste(€)**
Documentación del proyecto	50	30	1500
Diseño Bases de Datos	150	20	3000
Construcción de modelos	100	20	2000
Programación	400	20	8000
Redacción Memoria	100	30	4500
		TOTAL	19000

Tabla 9. Total Mano de Obra.

Capítulo 4

Presupuesto general

Sumando todos los costes anteriormente calculados, se concluye que el coste total del proyecto requiere un desembolso de[1]:

Concepto	**Coste(€)**
Hardware	349.5
Software	537.33
Mano de obra	19000
TOTAL	19886.83

Tabla 10. Presupuesto general

[1] Se supone que todos los datos utilizados (consumos de energía, temperaturas, laboralidad, etc) se encuentran en disponibles en otras aplicaciones o son de dominio público