

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/275465919>

# Aplicación de Técnicas de Clustering a la Serie Temporal de los Precios de la Energía en el Mercado Eléctrico

Conference Paper · September 2007

CITATIONS

0

READS

273

4 authors:



**Francisco Martínez-Álvarez**

Universidad Pablo de Olavide

153 PUBLICATIONS 2,771 CITATIONS

[SEE PROFILE](#)



**Alicia Troncoso**

Universidad Pablo de Olavide

137 PUBLICATIONS 2,393 CITATIONS

[SEE PROFILE](#)



**Jesús M. Riquelme Santos**

Universidad de Sevilla

90 PUBLICATIONS 2,511 CITATIONS

[SEE PROFILE](#)



**José C. Riquelme**

Universidad de Sevilla

278 PUBLICATIONS 3,852 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Special Issue: Big data and natural disasters. Computers & Geosciences (IF 2.47) [View project](#)



Transfer Learning [View project](#)

# Aplicación de Técnicas de Clustering a la Serie Temporal de los Precios de la Energía en el Mercado Eléctrico

F. Martínez-Álvarez	A. Troncoso	J. C. Riquelme	J. M. Riquelme
Dpto. Lenguajes y Sistemas Informáticos. Univ. de Sevilla. fmartinez@lsi.us.es	Área de Lenguajes y Sistemas Informáticos. Univ. Pablo de Olavide. ali@upo.es	Dpto. Lenguajes y Sistemas Informáticos. Univ. de Sevilla. riquelme@lsi.us.es	Dpto. Ingeniería Eléctrica. Univ. de Sevilla. jsantos@us.es

## Resumen

La principal tarea de las técnicas de *clustering* es formar grupos de elementos que presenten una conducta similar a partir de una base de datos. Se trata de generar un modelo capaz de simular el comportamiento del sistema de la manera más precisa posible. Así, en primer lugar, se procede a extraer el conocimiento necesario del sistema para, en segunda instancia, buscar los patrones capaces de conservar la información original y conseguir de este modo un modelo válido de descripción del sistema. En este trabajo se presentan dos técnicas de clustering particionales ampliamente conocidas, *K-means* y *Expectation Maximization*, para el análisis de la serie temporal de los precios de la energía eléctrica. Se demuestra que aplicar dichas técnicas resulta efectivo a la hora de dividir un año completo en diferentes grupos de días. Concretamente, se distinguen dos tipos de clusters fundamentales: los que identifican a los días laborables y aquéllos que representan a los días festivos y los fines de semana. Igualmente resaltable es la similitud mostrada en el comportamiento de los días pertenecientes a una misma estación.

## 1. Introducción

Debido al proceso liberalizador que está sufriendo el mercado eléctrico español, las estrategias dirigidas a obtener ofertas óptimas están adquiriendo mucho protagonismo entre las compañías generadoras de energía eléctrica

[14]. Este trabajo se centra en la extracción de información útil de la serie temporal de los precios de la energía en el mercado eléctrico, haciendo uso de técnicas de *clustering* [16] ya existentes. Dichas técnicas son la base de muchos algoritmos de clasificación y de modelado de sistemas. Para ello generan subgrupos de datos homogéneos con la intención de obtener una representación del comportamiento del sistema de la manera más precisa posible. En concreto, los algoritmos *K-means* [11] y *Expectation Maximization* (EM) [10] se utilizan aquí para encontrar aquellos días que tienen una conducta similar.

Diversas técnicas de predicción han sido ya utilizadas en series temporales de precios recientemente [5, 12, 13]. De este modo, A. J. Conejo et al. [2] utilizaron modelos ARIMA y transformadas wavelet mientras que R. C. García et al. [6] presentaron una técnica de predicción basada en un modelo GARCH. Igualmente, una combinación de redes neuronales artificiales y lógica difusa fue propuesta en [1]. En [18] se presenta una propuesta basada en una regresión no paramétrica adaptativa aplicada al mercado eléctrico de Ontario en Canadá. Un modelo basado en los vecinos más cercanos se propone en [15]. Finalmente, en [7] se propone un método que combina distintos modelos ARIMA para realizar predicciones de los precios de la energía en el mercado eléctrico. Pese a la diversidad de métodos encontrados en la literatura, ninguno de ellos han utilizado técnicas de clustering para mejo-

rar la predicción de la serie temporal formada por los precios de la energía.

La principal y más novedosa aportación de este trabajo radica, precisamente, en la utilización de técnicas de clustering para hacer una agrupación inicial del comportamiento de los elementos pertenecientes a la serie temporal de los precios de la energía para, posteriormente, ser capaces de predecir cómo evolucionará la curva de los precios en días posteriores. Como variable de entrada se utiliza el precio de la energía horaria, que está disponible en [4].

El resto del trabajo se organiza como sigue. En la sección 2 se presentan los algoritmos utilizados, K-means y EM, y se discute el número de clusters que se debe seleccionar para hacer el análisis. La sección 3 muestra los resultados obtenidos por cada método, proporcionando una medida de calidad de ellos. Finalmente, la sección 4 expone las conclusiones derivadas del estudio de los resultados y esboza las líneas de futuros trabajos.

## 2. Técnicas de clustering particionales

El proceso de agrupar un conjunto de objetos abstractos o físicos en clases similares recibe el nombre de *clustering*. Un cluster es una colección de datos parecidos entre ellos y diferentes a los datos pertenecientes a otros clusters.

Las técnicas de clustering son técnicas de clasificación no supervisada de patrones en clusters. El problema del clustering ha sido abordado por gran cantidad de disciplinas y es aplicable en multitud de contextos, hecho que refleja su utilidad como uno de los pasos en el análisis experimental de datos. Se aborda aquí una categoría de clustering, el particional, que será aplicada al análisis de los precios de la energía en el mercado eléctrico.

Dada una base de datos con  $n$  objetos, un método particional construye  $M$  grupos de los datos, donde cada partición representa a un cluster y  $M \leq n$ . Esto es, clasifica a los datos en  $M$  grupos que satisfacen los siguientes requisitos:

- Cada grupo debe contener, al menos, un elemento.

- Cada elemento debe pertenecer únicamente a un grupo. Nótese que este segundo requerimiento se relaja en ciertas técnicas particionales difusas.

En esta sección se presentan los dos métodos seleccionados para realizar clustering: K-means y EM. El número de clusters en los que se quiere dividir (particionar) los datos de entrada es el parámetro más crítico en tanto en cuanto un número demasiado elevado puede hacer que los resultados no sean suficientemente claros y un número demasiado bajo puede conducir a resultados poco relevantes o triviales. Ya que este tipo de algoritmos requieren que este parámetro sea introducido como parámetro de entrada, se discutirá en esta sección también la forma de elegir su valor óptimo.

### 2.1. K-means

El algoritmo K-means fue propuesto por MacQueen en el año 1968 [11]. Este algoritmo coge el parámetro de entrada,  $K$ , y particiona el conjunto de  $n$  datos en  $K$  clusters de tal manera que la similitud entre los elementos que pertenecen a un mismo cluster es elevada y la similitud entre los elementos de distintos clusters es baja. Dicha similitud se mide en relación al valor medio de los objetos en el cluster, es decir, con respecto a su centro de gravedad.

El algoritmo procede como sigue. En primer lugar, escoge aleatoriamente  $K$  objetos haciendo que éstos representen el centro del cluster. Cada uno de los objetos restantes se va asignando al cluster que sea más similar basándose en la distancia del objeto a la media del cluster. Entonces computa la nueva media de cada cluster y el proceso sigue iterando hasta que se consigue la convergencia. Es decir:

1. *Paso 1.* En cada iteración se evalúan todos los puntos y se reasignan a su cluster más cercano.
2. *Paso 2.* La reasignación de los puntos se lleva a cabo sólo si la suma de las distancias se reduce. A continuación, se recalculan los centros de los clusters.

El proceso itera hasta que se encuentra un mínimo de la función objetivo considerada que suele ser el error cuadrático medio, definido por:

$$E = \sum_{i=1}^K \sum_{x \in C_i} |x - m_i|^2 \quad (1)$$

donde  $x$  es el punto en el espacio que representa al objeto dado y  $m_i$  es la media del cluster  $C_i$ . Este criterio busca que los  $K$  clusters resultantes sean lo más compacto posible y lo más distanciado posible entre ellos.

El método es relativamente escalable y eficiente para el procesamiento de conjuntos de datos grandes ya que la complejidad computacional del algoritmo es  $O(nKt)$ , donde  $n$  es el número de objetos,  $K$  el número de clusters y  $t$  el número de iteraciones. Normalmente  $K$  y  $t$  suelen tener órdenes de magnitud sensiblemente inferiores a  $n$  y  $N$ , respectivamente.

El algoritmo K-means se puede aplicar sólo cuando la media de un cluster puede ser definida, esto es, no es de aplicación en los casos en que los atributos sean categóricos. Otro inconveniente es su sensibilidad al ruido y a los *outliers*. Además, la necesidad de dar el valor de  $K$  a priori resulta uno de sus mayores puntos débiles, tal y como se detalla en la siguiente sección.

### 2.1.1. Número de clusters en K-means

La función *silhouette* [8] proporciona una medida de calidad de los clusters obtenidos usando el algoritmo K-means. En concreto, calcula la separación existente entre los clusters. El valor de esta función está acotada entre  $-1$  y  $+1$ , donde  $+1$  denota una clara separación de los clusters y  $-1$  marca aquellos puntos cuya asignación puede ser cuestionable. Para concluir que un clustering ha sido realizado con éxito, esta función debe presentar un valor medio mayor que  $0,5$  para todos sus clusters. Sin embargo, para las series temporales reales es casi imposible alcanzar este valor y el hecho de no tener valores negativos en la gráfica suele ser suficiente para decidir cuántos clusters escoger.

En la Figura 1 se muestran los valores de la

función *silhouette* para 4 clusters aplicados a los precios de la energía en el mercado eléctrico en el año 2005. La métrica usada fue la euclídea y el número de clusters escogido fue 4, ya que sólo un cluster presentaba valores negativos significativos (cluster 3) y el valor medio de la función *silhouette* era el más elevado, como se muestra en la Figura 2.

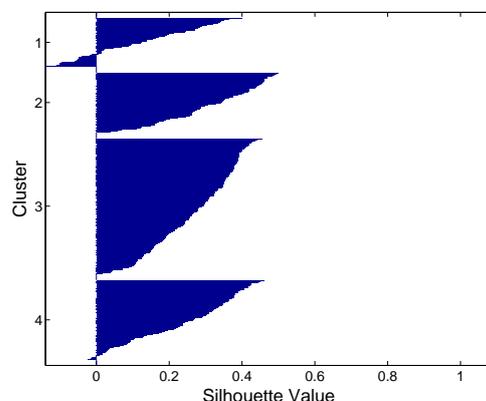


Figura 1: Representación de la función *silhouette* obtenida al aplicar K-means a los precios de la energía eléctrica en el año 2005 con 4 clusters.

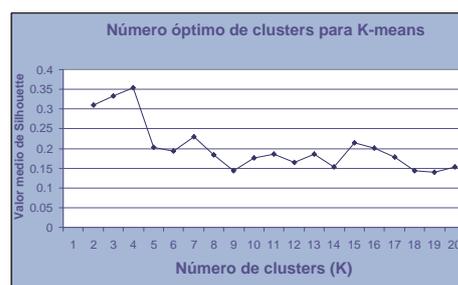


Figura 2: Función *silhouette* para la obtención del número de clusters con el algoritmo K-means.

## 2.2. Expectation Maximization

Este algoritmo es una variante del K-means y fue propuesto por Lauritzen en 1995 [10]. Se trata de obtener la función de densidad de pro-

babilidad (FDP) [17] desconocida que representa la distribución del conjunto completo de datos. Esta FDP se puede aproximar mediante una combinación lineal de  $NC$  componentes que dependen de una serie de parámetros  $\Theta_j$  que hay que estimar. Es decir:

$$P(x) = \sum_{j=1}^{NC} \pi_j p(x; \Theta_j) \quad (2)$$

con

$$\sum_{j=1}^{NC} \pi_j = 1 \quad (3)$$

donde  $\pi_j$  son las probabilidades a priori de pertenencia a cada cluster,  $P(x)$  denota la FDP arbitraria y  $p(x; \Theta_j)$  la función de densidad que representa la distribución de los datos que pertenecen al cluster  $j$ . Se pueden elegir distintas funciones de densidad que representen a los clusters, como FDP normales n-dimensionales, t-Student, Bernoulli, Poisson, y log-normales. En este trabajo cada uno de estos clusters vendrá representado por una distribución normal. La estimación de los parámetros del modelo requiere una medida de su bondad, es decir, cómo de bien se ajustan los datos a la distribución que los representa. Este valor de bondad se conoce como función de verosimilitud de los datos. Se trata entonces de estimar los parámetros  $\Theta_j$  maximizando esta función (criterio conocido como Máxima Verosimilitud, *ML-Maximum Likelihood*) [3]. Normalmente, se calcula el logaritmo de la función de verosimilitud ya que éste transforma los productos en sumas y los cocientes en restas. La solución obtenida es la misma gracias a la propiedad de monotonicidad del logaritmo. La forma de esta función es:

$$L(\Theta, \pi) = \log \prod_{n=1}^{NI} P(x_n) \quad (4)$$

donde  $NI$  es el número de instancias, que suponemos independientes entre sí. El algoritmo EM, procede en dos pasos que se repiten de forma iterativa:

1. **Expectation.** Utiliza los valores de los parámetros iniciales o proporcionados por

el paso Maximization de la iteración anterior, para calcular la probabilidad de que cada objeto pertenezca a un cluster.

2. **Maximization.** Obtiene nuevos valores de los parámetros a partir de las probabilidades obtenidas en el paso anterior.

Después de una serie de iteraciones, el algoritmo EM tiende a un máximo local de la función  $L$ , obteniéndose un conjunto de clusters que agrupan el conjunto de datos original.

### 2.2.1. Número de clusters en EM

Para este método el número óptimo de clusters se ha calculado utilizando validación cruzada [9]. La validación cruzada consiste en emplear  $k$  subconjuntos del conjunto de datos de los cuales  $k/2$  se emplean para entrenamiento y  $k/2$  para la validación del esquema de aprendizaje. Aquí se ha usado una variante llamada *n-fold cross validation*, o validación cruzada de  $n$  particiones. En ella se dividen los datos en tantas particiones como indique el parámetro  $n$  y se entrena  $n$  veces promediando el error de cada prueba. Presenta un inconveniente y es que no se consigue una representación equitativa de todas las clases. Para el presente análisis se escogió  $n = 10$ , es decir, se dividieron los datos de entrada en diez grupos dejando uno como validación y los nueve restantes como entrenamiento.

La Figura 3 muestra la evolución del logaritmo de la función de verosimilitud (*log-ML*). Así, el número de clusters escogido es 11, ya que el *log-ML* presenta el valor máximo para este número.

## 3. Resultados

Los algoritmos K-means y EM descritos en la sección precedente se han aplicado a diferentes experimentos con el fin de obtener una predicción de la serie temporal de los precios de la energía en el mercado eléctrico español correspondiente al año 2005 [4].

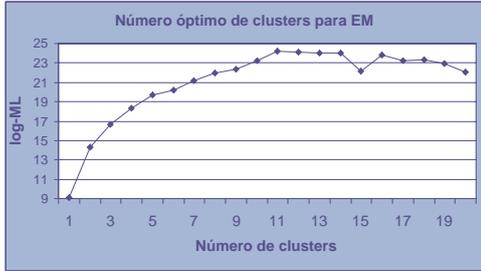


Figura 3: Validación cruzada para determinar el número de clusters en el algoritmo EM.

### 3.1. Resultados con K-means

La Tabla 1 muestra el porcentaje de pertenencia de cada día de la semana a los distintos clusters. Por el contrario, la Figura 4 muestra

Tabla 1: Distribución de los días de la semana en los diferentes clusters generados.

Día	Clus. 1	Clus. 2	Clus. 3	Clus. 4
Lunes	36.54 %	51.92 %	3.85 %	7.69 %
Martes	31.48 %	57.41 %	3.70 %	7.41 %
Miércoles	30.77 %	63.46 %	3.85 %	1.92 %
Jueves	32.69 %	59.62 %	5.77 %	1.92 %
Viernes	28.85 %	59.62 %	3.85 %	7.69 %
Sábado	11.32 %	0.00 %	39.62 %	49.06 %
Domingo	0.00 %	0.00 %	44.23 %	55.77 %

el año 2005 clasificado en 4 clusters mediante al algoritmo K-means. De un simple vistazo, se pueden diferenciar claramente dos tipos de clusters:

- Clusters 1 y 2 agrupan todos los días laborables.
- Clusters 3 y 4 agrupan los fines de semana y los días festivos.

Sin embargo, existen ciertos días que presentan un comportamiento aparentemente discordante. Existen 22 días laborables que han sido clasificados en los clusters 3 ó 4. Pero un meticuloso análisis revela que la mayor parte de estos días fueron vacaciones. La Tabla 2 muestra cada uno de estos días en detalle.

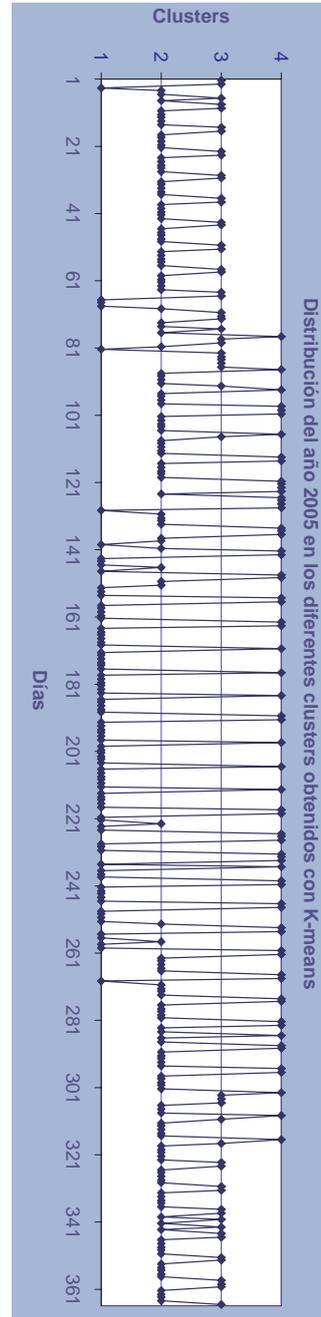


Figura 4: Distribución de los días en los diferentes clusters obtenidos por medio del método K-means.

Tabla 2: Días laborables mal clasificados con K-means.

Nº de día	Fecha	Festividad
6	06/01	Epifanía
70	11/03	Ninguna
75	16/03	Ninguna
77	18/03	Viernes antes de Sem. Sta.
82	23/03	Semana Santa
83	24/03	Semana Santa
84	25/03	Semana Santa
87	28/03	Lunes después de Sem. Sta.
98	08/04	Ninguna
122	02/05	Día del trabajo
123	03/05	Día Comunidad de Madrid
125	05/05	Puente del 1 de mayo
126	06/05	Puente del 1 de mayo
227	15/08	Día de la Asunción
231	19/08	Ninguno
235	23/08	Ninguno
285	12/10	Día de la Hispanidad
304	31/10	Puente de Todos los Santos
305	01/11	Día de Todos los Santos
340	06/12	Día de la Constitución
342	08/12	Día de la Inmaculada
360	26/12	Lunes después de Navidad

Antes de continuar con el estudio de los resultados, debe hacerse un comentario sobre la primera semana de mayo. Los días festivos reales son el 1 de mayo (Día del Trabajo) y el 2 de mayo (Día de la Comunidad de Madrid). Sin embargo, el 1 de mayo fue domingo en 2005 y ambas festividades fueron pospuestas un día. Con referencia a los fines de semana, existen 6 sábados que se clasificaron como si fueran días laborables, concretamente en el cluster 1. Este hecho queda reflejado en la Tabla 3. Nótese

Tabla 3: Fines de semana mal clasificados con K-means.

Nº de día	Fecha
169	18 Junio
176	25 Junio
183	2 Julio
197	16 Julio
204	23 Julio
211	30 Julio

que la mayoría de los sábados son consecutivos y pertenecen al verano, excepto el 9 de julio que se clasificó en el cluster 4. El año completo se divide en 261 días laborables y 104 días

pertenecientes a los fines de semana o festividades. De la Tabla 2 se deduce que 5 fueron los días mal clasificados (11 de marzo, 16 de marzo, 8 de abril, 19 de agosto y 23 de agosto). Por consiguiente el error medio cometido en los días laborables es de 1.92% (5 días de 261). En lo que respecta a los fines de semana y festividades, 6 fueron los sábados que se clasificaron erróneamente (18 de junio, 25 de junio, 2 de julio, 16 de julio, 23 de julio y 30 de julio), por lo que el error medio cometido fue de 5.77% (6 días de 104). Así, se concluye que el error total promedio es de 3.01% (11 días de 365).

La siguiente tarea consiste en explicar cuándo un día laborable pertenece al cluster 1 o cuándo pertenece al cluster 2 y, equivalentemente, cuándo un día festivo pertenece al cluster 3 ó 4. En la Figura 4 se pueden diferenciar tres zonas tanto para días festivos como laborables. Desde principio de año hasta el 18 de mayo (día número 144) la mayor parte de los días laborables pertenecen al cluster 2. Desde este día hasta el 20 de septiembre (día número 263) pertenecen al cluster 1. Finalmente, desde el 21 de septiembre hasta final de año, los días laborables vuelven a pertenecer al cluster 2. En lo relativo a las festividades y fines de semana se da una situación similar. Desde principios de año hasta el 27 de marzo (día número 86) la mayor parte de estos días pertenecen al cluster 3. Desde ese fin de semana hasta el fin de semana del 30 de octubre (día número 303), pertenecen al cluster 4. Finalmente, la última parte del año vuelven a ser fines de semana asociados al cluster 3. Por consiguiente, se puede apreciar un comportamiento estacional asociada a la serie temporal de los precios de la energía eléctrica.

Las curvas características de cada cluster se muestran en la Figura 5. Especialmente destacable es que las curvas asociadas a los fines de semana y festividades (clusters 3 y 4) poseen unos precios al comienzo y término del día mayores que los de los días laborables. Igualmente, presentan sus valores más altos a últimas horas de la tarde, hecho que es debido a que la población consume más electricidad por la noche durante los fines de semana. Por otro lado,

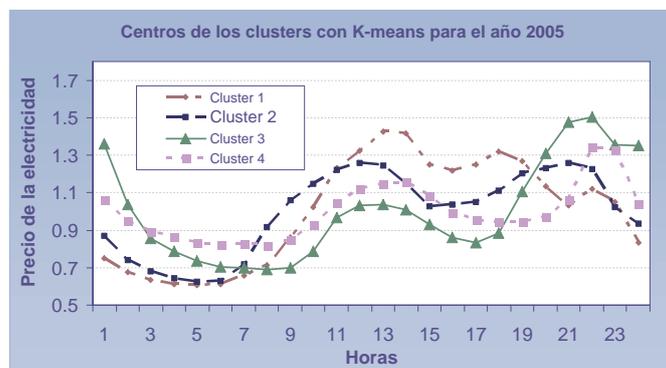


Figura 5: Curvas características de los clusters obtenidos con el método K-means para el año 2005.

las curvas de los días laborables tienen sus picos a mediodía, cuando las industrias, comercios y empresas están a pleno rendimiento.

### 3.2. Resultados con EM

La Figura 6 presenta los 11 patrones encontrados por el algoritmo EM para los precios de la energía eléctrica durante el año 2005. La Tabla 4 muestra la siguiente información:

- Clusters 1, 2, 3, 5 y 7 agrupan los días laborables
- Clusters 4, 6, 8, 9, 10 y 11 agrupan los fines de semana y los días festivos. Además, se puede destacar que:
  - Clusters 4, 10 y 11 son fundamentalmente domingos.
  - Clusters 8 y 9 son fundamentalmente sábados.
  - Cluster 6 comprende determinados sábados y la primera semana de agosto.

La asociación de los días a los clusters con el algoritmo EM no resulta tan sencilla como lo es con sólo 4 clusters. Así, la dispersión a través de los clusters es mucho más elevada, hecho que se manifiesta por medio de un error más elevado ya que un sábado y 16 días laborables fueron clasificados erróneamente. Por tanto, el error cometido es de 4.38 %.

En contraste con lo que sucedía con K-means, estos 16 días no se corresponden con días festivos. Por el contrario, este fenómeno aparece aleatoriamente y sin causas aparentes. No obstante, el sábado mal clasificado (se clasificó en el cluster 5) es, una vez más, el 2 de julio: el comienzo de las vacaciones para muchas personas en España.

Las curvas características de los 11 clusters se muestran en la Figura 7.

## 4. Conclusiones

Se ha demostrado que utilizar técnicas de clustering particionales resulta útil para encontrar patrones en las curvas de precios de la energía en el mercado eléctrico. El análisis se ha llevado a cabo por medio de dos técnicas, K-means y Expectation Maximization, que han proporcionado información relevante ya que han encontrado patrones interesantes en las curvas de los precios.

El error medio cometido en las clasificaciones fue de 3.01 % (11 días) para K-means y de 4.38 % (16 días) para EM, lo que significa un elevado grado de fiabilidad. Son varios los factores que afectan a la clasificación, incrementando así la tasa de errores. A continuación, se destacan algunos de los principales factores:

1. *Hora*. El precio de la energía en el mercado eléctrico puede fluctuar dependiendo

Tabla 4: Grado de pertenencia de los días a los diferentes clusters generados con EM.

Cluster	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
Cluster 1	7.69 %	9.62 %	15.38 %	15.38 %	26.92 %	0.00 %	0.00 %
Cluster 2	17.31 %	25.00 %	23.08 %	17.31 %	11.54 %	0.00 %	0.00 %
Cluster 3	25.00 %	28.85 %	30.77 %	34.62 %	25.00 %	0.00 %	0.00 %
Cluster 4	0.00 %	1.92 %	0.00 %	0.00 %	1.92 %	3.77 %	19.23 %
Cluster 5	30.77 %	17.31 %	21.15 %	17.31 %	17.31 %	1.89 %	0.00 %
Cluster 6	5.77 %	11.54 %	3.85 %	7.69 %	9.62 %	11.32 %	0.00 %
Cluster 7	1.92 %	3.85 %	0.00 %	1.92 %	1.92 %	0.00 %	0.00 %
Cluster 8	5.77 %	1.92 %	3.85 %	3.85 %	1.92 %	39.62 %	9.62 %
Cluster 9	1.92 %	0.00 %	1.92 %	0.00 %	3.85 %	39.62 %	7.69 %
Cluster 10	3.85 %	0.00 %	0.00 %	0.00 %	0.00 %	1.89 %	44.23 %
Cluster 11	0.00 %	0.00 %	0.00 %	1.92 %	0.00 %	1.89 %	19.23 %

de la hora del día que se trate, lo que provoca una gran incertidumbre a la hora de conocer el precio futuro. Del estudio realizado, se concluye que durante los días laborables el precio de la energía alcanza los mayores valores durante el mediodía. Sin embargo, los picos de precios alcanzados en los fines de semana y días festivos se encuentran cercanos a las 12 de la noche.

2. *Día de la semana.* Como se ha ido esbozando a lo largo del estudio, la demanda eléctrica varía dependiendo del día de la semana. Una consecuencia inmediata es la variación producida en la serie temporal de los precios, tal y como se ha demostrado.
3. *Mes del año.* Se detectó, igualmente, un comportamiento asociado a la estación del año en el que se examine la serie temporal. Durante el invierno el día es más corto y se tiende a pasar más tiempo en casa. Todo lo contrario que sucede en verano. Así, el uso del aire acondicionado y de los calefactores tienen una influencia directa en la demanda y, consecuentemente, en los precios de la energía eléctrica.
4. *Otros factores no predecibles.* Otros factores aleatorios, como pueden ser huelgas o eventos deportivos, influyen en el precio

de la energía eléctrica. Además hay que añadir que el mercado eléctrico español no es un mercado de competencia perfecta.

Trabajos futuros irán encaminados a predecir los precios de la energía eléctrica, una vez conocida la clasificación hecha por medio de técnicas de clustering. Esto es, se obtendrán N clusters y se aplicarán diversos modelos a cada uno de ellos para mejorar la calidad de la predicción.

### Agradecimientos

Los autores quieren agradecer la financiación recibida por parte del Ministerio de Ciencia y Tecnología, proyectos TIN2004-00159 y ENE-2004-03342, así como por parte de la Junta de Andalucía, proyecto P05-TIC-00531.

### Referencias

- [1] N. Amjady. Day-ahead price forecasting of electricity markets by a new fuzzy neural network. *IEEE Transactions on Power Systems*, 21(2):887–896, 2006.
- [2] A. J. Conejo, M. A. Plazas, R. Espínola, and B. Molina. Day-ahead electricity price forecasting using the wavelet transform and arima models. *IEEE Transactions on Power Systems*, 20(2):1035–1042, 2005.

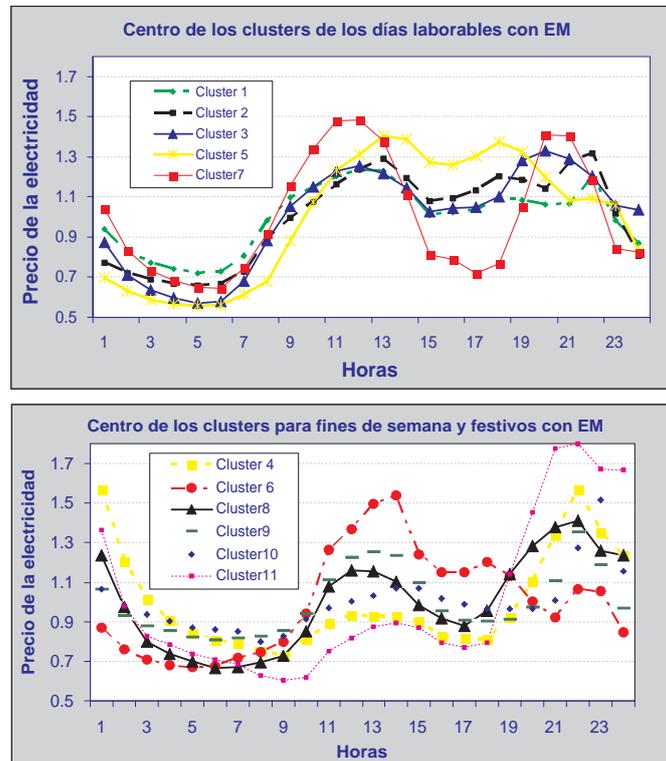


Figura 7: Curvas características de los clusters obtenidos con el método EM para el año 2005.

- [3] H. Cramér. *Mathematical methods of statistics*. Princeton Univ. Press, 1946.
- [4] Operador del Mercado Eléctrico Español. On-line. <http://www.omel.es>.
- [5] A. Guillén et al. Improving clustering technique for functional approximation problem using fuzzy logic: Icf algorithm. *Lecture Notes in Computer Sciences*, 3512:272–279, 2005.
- [6] R. C. García, J. Contreras, M. van Akkeren, and J. B. García. A garch forecasting model to predict day-ahead electricity prices. *IEEE Transactions on Power Systems*, 20(2):867–874, 2005.
- [7] C. García-Martos, J. Rodríguez, and M. J. Sánchez. Mixed models for short-run forecasting of electricity prices: Application for the spanish market. *IEEE Transactions on Power Systems*, 22(2):544–552, 2007.
- [8] L. Kaufman and P. J. Rousseeuw. *Finding groups in Data: an Introduction to Cluster Analysis*. Wiley, 1990.
- [9] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of International Joint Conference on Artificial Intelligence*, pages 1137–1143, 1995.
- [10] S. L. Lauritzen. The em algorithm for graphical association models with missing data. *Computational Statistics and Data Analysis*, 19(2):191–201, 1995.

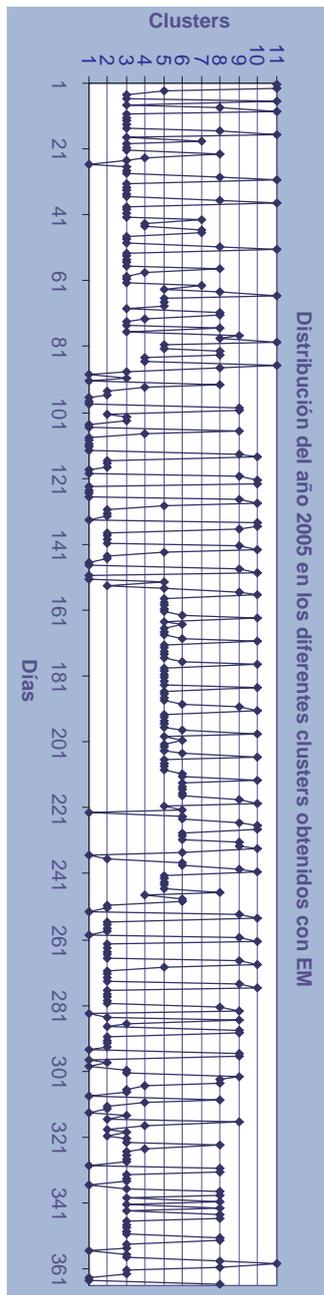


Figura 6: Distribución de los días en los diferentes clusters obtenidos por medio del método EM.

- [11] J. MacQueen. Some methods for classification and analysis of multivariate observations. *Proc. of the 5th Berkeley Symp. Math. Statist, Prob*, 1:281–297, 1968.
- [12] F. Martínez-Álvarez, A. Troncoso, J. C. Riquelme, and J. M. Riquelme. Discovering patterns in electricity prices using clustering techniques. In *Proceedings of International Conference on Renewable Energies and Power Quality*, 2007.
- [13] F. Martínez-Álvarez, A. Troncoso, J. M. Riquelme, and J. C. Riquelme. Partitioning-clustering techniques applied to the electricity price time series. *Lecture Notes in Computer Science*, 481:990–999, 2007.
- [14] M. A. Plazas, A. J. Conejo, and F. J. Prieto. Multimarket optimal bidding for a power producer. *IEEE Transactions on Power Systems*, 20(4):2041–2050, 2005.
- [15] A. Troncoso, J. C. Riquelme, J. M. Riquelme, J. L. Martínez, and A. Gómez. Electricity market price forecasting based on weighted nearest neighbours techniques. *IEEE Transactions on Power Systems*, in press, 2007.
- [16] R. Xu and D. C. Wunsch II. Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3):645–678, 2005.
- [17] S. Zacks. *The theory of statistical inference*. Wiley, 1946.
- [18] H. Zareipour, K. Bhattacharya, and C. A. Cañizares. Forecasting the hourly ontario energy price by multivariate adaptive regression splines. *IEEE Transactions on Power Systems*, 20(2):1035–1042, 2006.