

Modelización del mercado energético argentino mediante *cluster analysis*

César Bucci¹, Walter Legnani¹

1 Universidad Tecnológica Nacional, Facultad Regional Buenos Aires,
Secretaría de Ciencia, Tecnología e Innovación productiva,
Medrano 951 (C1179AAQ) Ciudad Autónoma de Buenos Aires, Argentina

ingbucci@yahoo.com.ar

Recibido el 13 de Abril de 2012, aprobado el 18 de Agosto de 2012

Resumen

El presente trabajo reseña un estudio pormenorizado y minucioso sobre la matriz energética argentina que tiene en cuenta todas las fuentes de energía (primarias y secundarias) siguiendo la metodología y la agrupación de valores según la serie histórica de datos del balance energético nacional entre los años 1970 y 2005 que han sido elaborados por la Secretaría de Energía de la Nación. Luego de ello, se realizó un completo análisis de los modelos energéticos y económicos del país, haciendo especial hincapié en las fuentes de energía que gravitan con mayor peso dentro de la matriz energética. Para tal fin, se implementó una herramienta computacional y estadística conocida como *cluster analysis*, utilizada frecuentemente para la identificación de patrones de comportamiento de un determinado sistema. Ello permitió caracterizar a las fuentes de energía y al hábito de consumo, lo que posibilitó arribar a las conclusiones que contribuirán a un mejor y más eficiente uso técnico- económico de los recursos energéticos.

PALABRAS CLAVE: MATRIZ ENERGÉTICA PRIMARIA Y SECUNDARIA - BALANCE ENERGÉTICO NACIONAL - KMEANS CLUSTER ANALYSIS

Abstract

The developed work provides a detailed and careful study about Argentina's energy matrix that takes into account all energy sources (primary and secondary) by means of the methodology and the grouping of values according to the time data series of the national energy balance between 1970 and 2005. These have been developed by the Energy Secretary. After that, we proceeded with a complete analysis of energy and economic country models, with a particular emphasis on energy sources that weigh more heavily in the energy matrix, and thus being the most widely used for developing economic and social activity. To reach this aim it was implemented a computational and statistical tool known as cluster analysis that is frequently used to identify patterns of behaviour of a given system. It allowed us to characterize energy sources and the habit of consumption and subsequently to arrive at conclusions that will contribute to a better and more efficient use of technical - economical energy resources.

KEYWORDS: PRIMARY AND SECONDARY ENERGY MATRIX - NATIONAL ENERGY BALANCE - KMEANS CLUSTER ANALYSIS

* El presente artículo forma parte del trabajo de tesis "Modelización del mercado energético argentino mediante cluster analysis" para optar al grado de Magister en Administración de Negocios, bajo la dirección del Dr. Walter Legnani.

Introducción

Generalidades del mercado energético

Desde hace prácticamente un siglo, el desarrollo de la actividad económica e industrial en el país estuvo fuertemente asociado al consumo energético de fuentes de energía no renovables¹ (Serie histórica de datos de Balance Energético Nacional (BEN), Secretaría de Energía de la Nación, 2006).

Si se toma por ejemplo el período que va desde los años '70 hasta el presente, se observa que las fuentes de energía de origen fósil han tenido una participación cercana al 90% de la matriz energética (MEN) primaria. En aquel momento la utilización de las mismas representaba 71% para el petróleo y 18% para el gas natural. En la actualidad la tendencia del 90% para combustibles fósiles se mantiene constante, mientras que la participación de ambas fuentes se invirtió, quedando 50% para el gas natural y 39% para el petróleo² (BEN, Secretaría de Energía de la Nación, 2006).

Por su parte, la MEN secundaria³, también ha experimentado grandes cambios en cuanto a las proporciones y a la utilización de los recursos. En los años '70 el fuel oil representaba el 30% de la matriz energética y el gas oil el 18%, mientras que en el presente, luego de un notorio cambio en el hábito de consumo, ha quedado el gas distribuido por redes con un 48% y el gas oil con un 15% (BEN, Secretaría de Energía de la Nación, 2006).

Lo expuesto hasta el momento hace suponer que la matriz energética argentina ha evolucionado hacia la "gasificación" del mercado energético.

Generalidades de los métodos de agrupación de conglomerados

Los modelos del presente trabajo se implementaron mediante la técnica de estadística matemática conocida como *kmeans cluster analysis* (kca) (Timm; 2002). Para realizar este procesamiento se ha seleccionado el *software ClustanGraphics* dado que proporciona un ac-

ceso rápido y eficiente a las herramientas más importantes para realizar los estudios necesarios y por otra parte, el equipo de investigación que desarrolló el presente trabajo cuenta con una licencia académica para su utilización.

Visto que una de las características del análisis de conglomerados, es la determinación de las distancias entre agrupamientos o *clusters*, en el caso presente se adoptó la medida euclídea (Timm; 2002). Por igual motivo, mediante la metodología conocida como kca (Jain and Dubes; 1988 y Pelleg and Moore; 2000) se trabajó en la obtención de un modelo que fuera representativo de la estructura del mercado energético actual y que pudiera emplearse para el diseño de estrategias económicas, comerciales y geopolíticas a nivel nacional.

Los algoritmos diseñados para realizar los conglomerados de datos se suelen dividir en dos grandes categorías denominadas jerárquicas y no jerárquicas, en este trabajo se empleó el método no jerárquico *k-means* y también las técnicas tradicionales utilizadas para determinar el número de *clusters* que se describirán más adelante (Timm; 2002). La justificación del número de *clusters* fue validada mediante la técnica conocida como *bootstrapping*. A continuación se realizaron gráficos con la evolución de los distintos tipos de energía incluyendo los *clusters* obtenidos. Se verificaron las relaciones entre los tipos de energía y dichos *clusters*, y se buscaron patrones que permitieron determinar las cercanías entre *clusters* y los tipos de energía para establecer una relación entre los más cercanos o similares entre sí.

Una vez validada la cantidad y tipo de *clusters*, se establecieron concordancias entre *clusters* y tipos de energía para las relaciones o similitudes halladas entre ambos, las cuales contribuyeron a la comprensión del comportamiento del mercado energético argentino.

A su vez, los resultados también fueron validados mediante la comparación con otros modelos y estudios realizados tanto por consultores privados como del gobierno nacional.

1 Balance Energético Nacional: es una metodología de cálculo que permite realizar comparaciones con modelos energéticos de otros países o de instituciones relacionadas.

2 Se refiere a las fuentes de energía en el estado que se extrae o captura de la naturaleza.

3 Se refiere a los productos energéticos producidos a partir de energías primarias o secundarias en los centros de transformación, con la finalidad de hacerlas más aptas a los requerimientos del consumo.

Alcances y limitaciones

El presente análisis, abarcó tanto el mercado energético local como el internacional e incluyó todos los tipos de energía definidas y especificadas en el BEN en el período comprendido entre 1970 y 2005, además de otros datos provenientes de organismos estatales y privados, nacionales e internacionales, correspondientes al sector energético hasta el año 2008. Posteriormente en base a los datos suministrados, se realizó una comparación con algunas fuentes de energía del mercado energético mundial proporcionados por la Energy Information Administration (EIA-Official Energy Statistics from the US Government) a efectos de:

- Modelar y caracterizar el mercado energético argentino
- Identificar y comprender el comportamiento de los mercados energéticos
- Obtener información y resultados que permitan resolver la problemática actual
- Obtener información y resultados que permitan establecer una política energética sustentable a largo plazo
- Reducir y/o sustituir la utilización de hidrocarburos en la producción de energía
- Desarrollar fuentes alternativas de energía, especialmente las de origen renovable
- Fomentar el uso eficiente de la energía.

Cabe destacar que la fuente oficial de la cual se han tomado gran parte de los datos y las estadísticas fue la Secretaría de Energía de la Nación, cuya característica es que la realización de los informes oficiales requiere 2 años de elaboración hasta su emisión preliminar, mientras que los informes oficiales definitivos toman 3 años hasta su emisión. Por lo tanto, para la presentación de los resultados en el resumen final detallado por tipo de energía, se han utilizado los datos oficiales definitivos del BEN (1970-2005) desarrollados mediante la técnica conocida como *Cluster Analysis*. Además de datos del BEN, se contemplaron también otros datos puntuales de reconocidos organismos, que llegan hasta 2008.

El período en el cual se realizó el presente trabajo se extendió desde mediados de 2007 hasta mediados de 2009, ello implica que parte del análisis tiene en cuenta la escasez de energía sufrida en el período, debido a la gran actividad económica experimentada hasta mediados de 2008, momento en el cual comienza a declinar la actividad industrial y por ende, se normaliza el suministro energético a valores más estables.

Ello implica que cierta parte de las tendencias previstas en su momento puedan verse afectadas por el cambio abrupto e inesperado en la economía mundial y por ende en los precios de

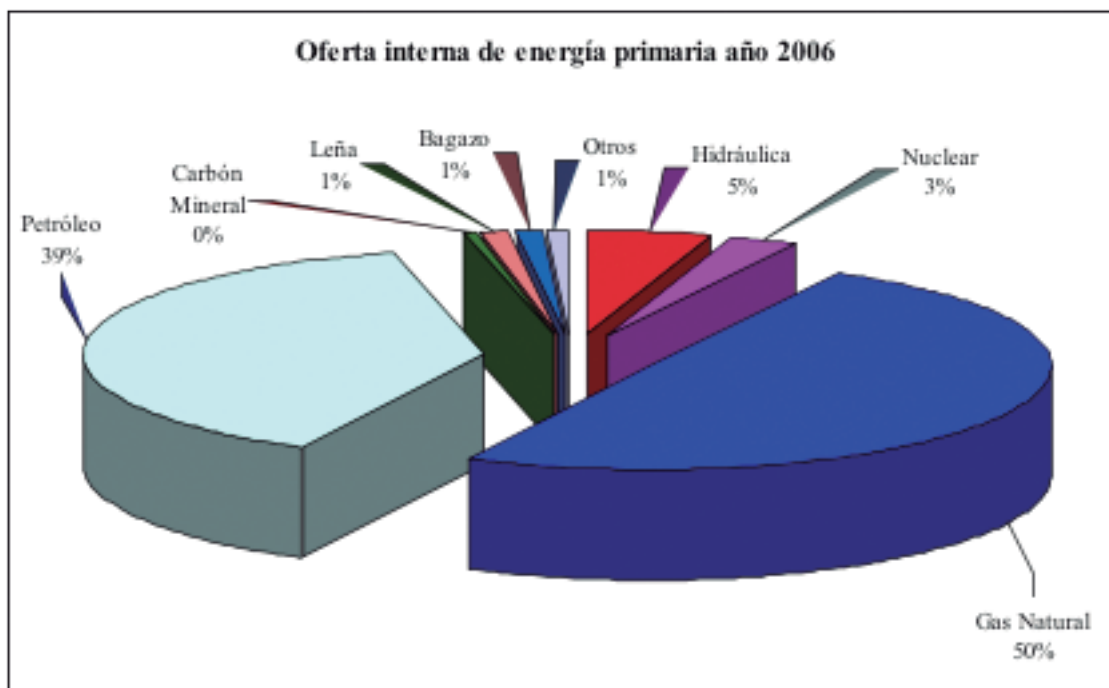


Fig. 1. Matriz energética primaria año 2006 (Fuente: Secretaría de Energía de la Nación)

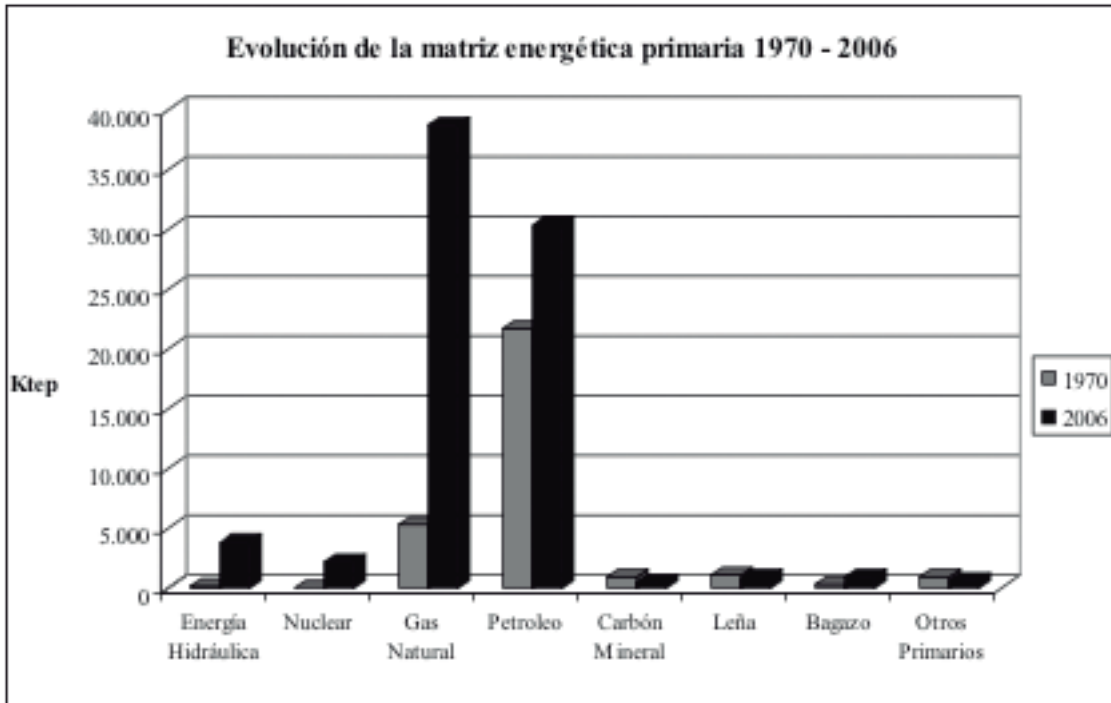


Fig. 2. Evolución de la matriz energética primaria 1970-2006, en Ktep⁴ (Fuente: Secretaría de Energía de la Nación)

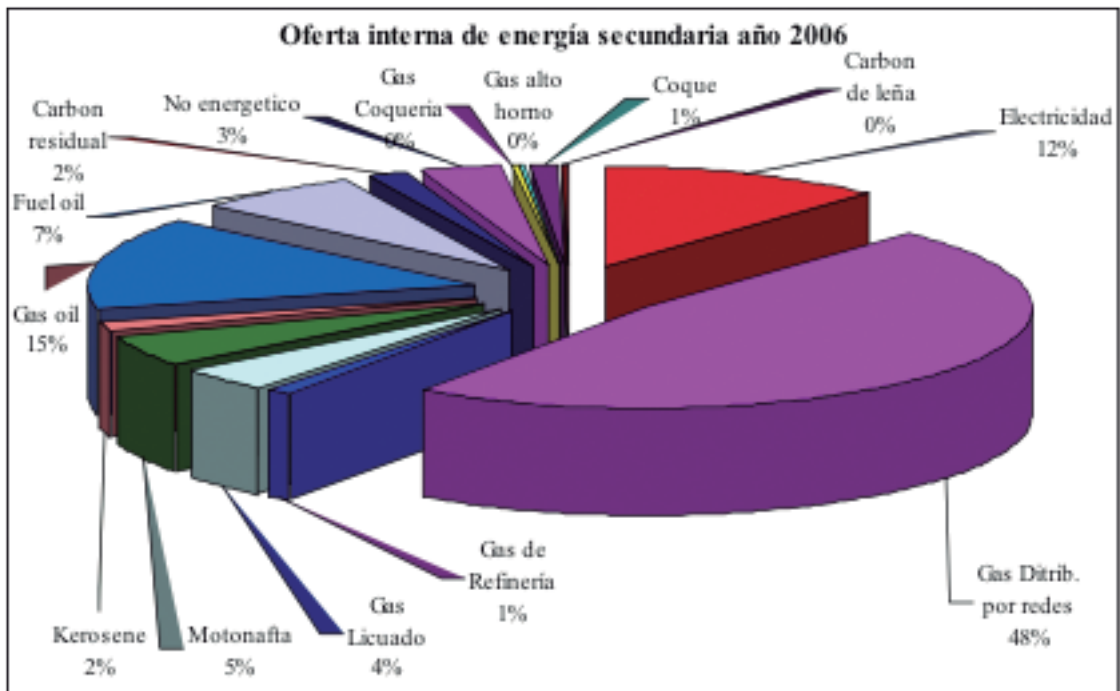


Fig. 3. Matriz energética secundaria año 2006 (Fuente: Secretaría de Energía de la Nación)

⁴ Kilo Toneladas Equivalentes de Petróleo

la energía en todas sus formas. Es de esperar que la desaceleración de la economía también haya traído aparejada una disminución del consumo energético, lo cual tiene repercusiones en los modelos matemáticos utilizados y en las ecuaciones económicas que definen las inversiones en materia energética hacia distintos tipos de fuentes de energía.

El mercado energético argentino

Descripción de la matriz energética

El concepto de matriz energética está asociado fundamentalmente a la agrupación por tipo de energía, las cuales a su vez son segregadas dependiendo que se trate de un recurso primario o secundario y comprende a las distintas fuentes y recursos renovables y no renovables que se utilizan para generar la energía necesaria para movilizar la industria, el transporte, la economía en general y dar bienestar a la sociedad. Estos grandes grupos llamados fuentes de energía primaria y secundaria a su vez están divididos en subgrupos por tipo o fuente de energía. La función de los mismos es discriminar el origen y el uso de cada fuente de energía a efectos de clasificarlas y posteriormente analizarlas con claridad (según la clasificación de la Secretaría de Energía de la Nación).

Fuentes de energía primaria

El grupo de la llamada Energía Primaria está subdividido en ocho tipos de energía y representa la energía extraída de la naturaleza en estado bruto, sin ningún tratamiento ni proceso que le genere valor agregado.

En la figura 1 se observa la matriz energética primaria argentina que corresponde al año 2006. Si bien los porcentajes pueden variar año a año, es importante resaltar que los tipos de energía se mantienen iguales en todo el período analizado.

En la figura 1 se denota claramente que existen dos actores preponderantes en la escena de la matriz energética primaria, ellos son el petróleo con un aporte del 39% y el gas natural con el 50%. Es decir que entre ambos dominan la escena energética argentina con casi el 90% del total, en cuanto a recursos primarios. Luego aparece un segundo grupo, conformado por seis tipos de energía con una participación cercana al 10%. Lo que implic-

que hay una tendencia que acentúa más el uso de energías de origen fósil.

En la figura 2 se puede apreciar un crecimiento abrupto en la utilización de gas natural como combustible (período 1970-2006), registrándose también un crecimiento en la utilización de petróleo, aunque en menor medida.

Existen cuatro tipos de energía que no han sufrido prácticamente variación: el carbón mineral, la leña, el bagazo y otros primarios.

Por último, el tipo de energía preponderante es el gas natural que ha aumentado su utilización en un 620% en los últimos 35 años. El mayor incremento se dio por la ampliación de los gasoductos y la utilización masiva de este tipo de energía para uso doméstico, industrial y, fundamentalmente, el vuelco del parque automotor y las centrales termoeléctricas al gas natural.

Fuentes de energía secundaria

El grupo de la llamada Energía Secundaria está subdividido en catorce tipos y representa la energía que ha sido extraída de la naturaleza en estado bruto y que ha sido transformada de energía primaria a energía secundaria mediante el empleo de procesos o tratamiento que la convierten en un bien utilizable por el consumidor.

En la figura 3 se observa la matriz energética secundaria de la República Argentina que corresponde al año 2006. Como se mencionó anteriormente, si bien los porcentajes pueden variar año a año, es importante resaltar que los tipos de energía continúan siendo los mismos durante el período analizado.

En la misma figura se ve claramente que existe un actor preponderante en la escena de la matriz energética secundaria, el gas distribuido por redes con un 48% de participación. Es decir que, este último sumado al gas oil y a la electricidad son claros dominadores de la escena energética argentina con el 75% en cuanto a recursos secundarios se refiere. Luego aparece un segundo grupo, conformado por once tipos de energía con una participación del 25%. Lo que implica que hay una tendencia más que acentuada al uso de energías de origen fósil y, especialmente, orientado al uso del gas distribuido por redes.

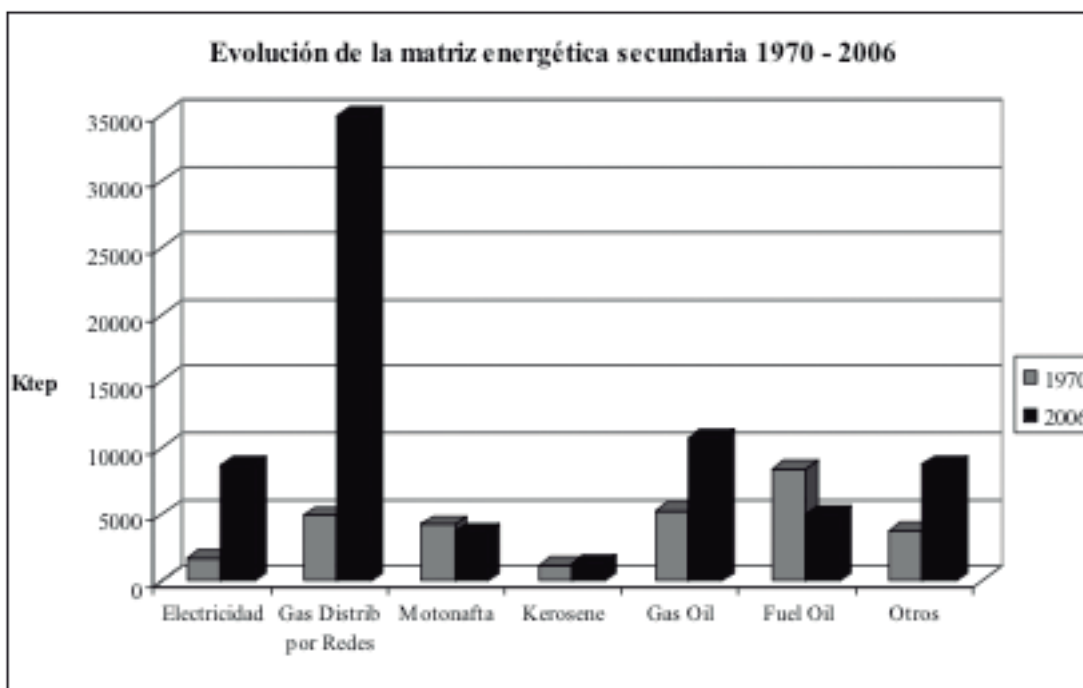


Fig. 4. Evolución de la matriz energética secundaria 1970-2006
(Fuente: Secretaría de Energía de la Nación)

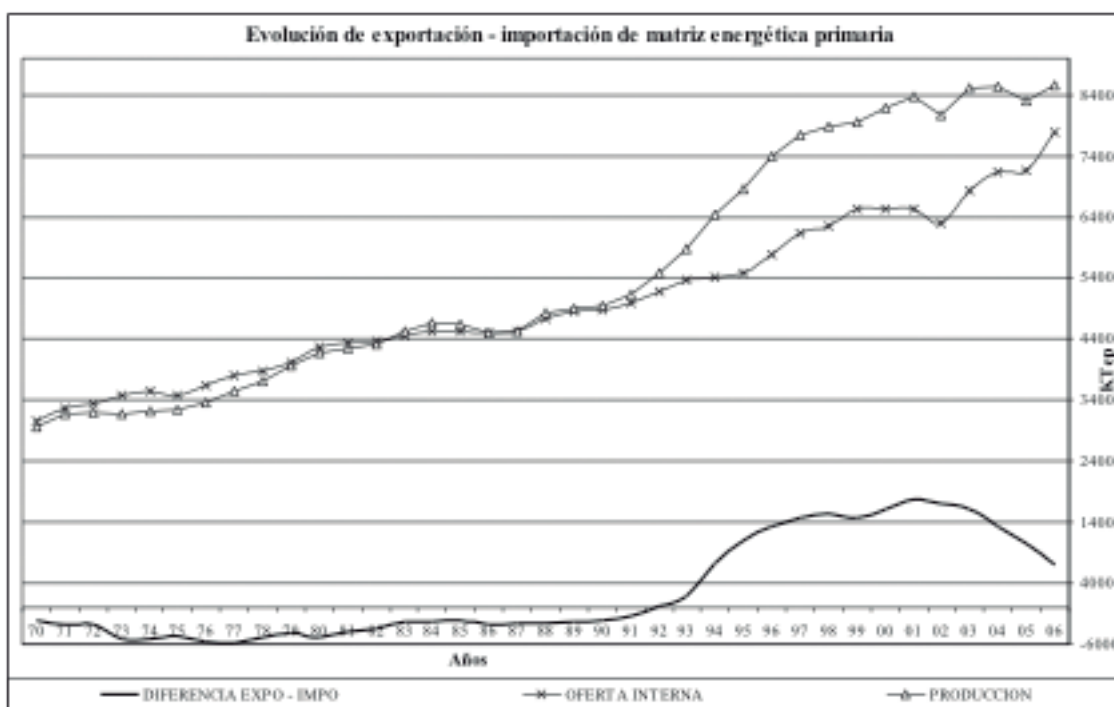


Fig. 5. Evolución de la exportación - importación de energía primaria
(Fuente: Secretaría de Energía de la Nación)

En la figura 4 de la página siguiente se observa un crecimiento abrupto en el uso de gas distribuido por redes como combustible, y por otra parte, la menor utilización de combustibles líquidos. Lo que implica tanto en la matriz energética primaria como en la secundaria, una "gasificación" del mercado energético.

De aquí se desprende que existe un grupo compuesto por el gas distribuido por redes y la electricidad que denota el mayor crecimiento, y un segundo grupo con una variación más moderada, compuesto por el resto de los tipos de energía.

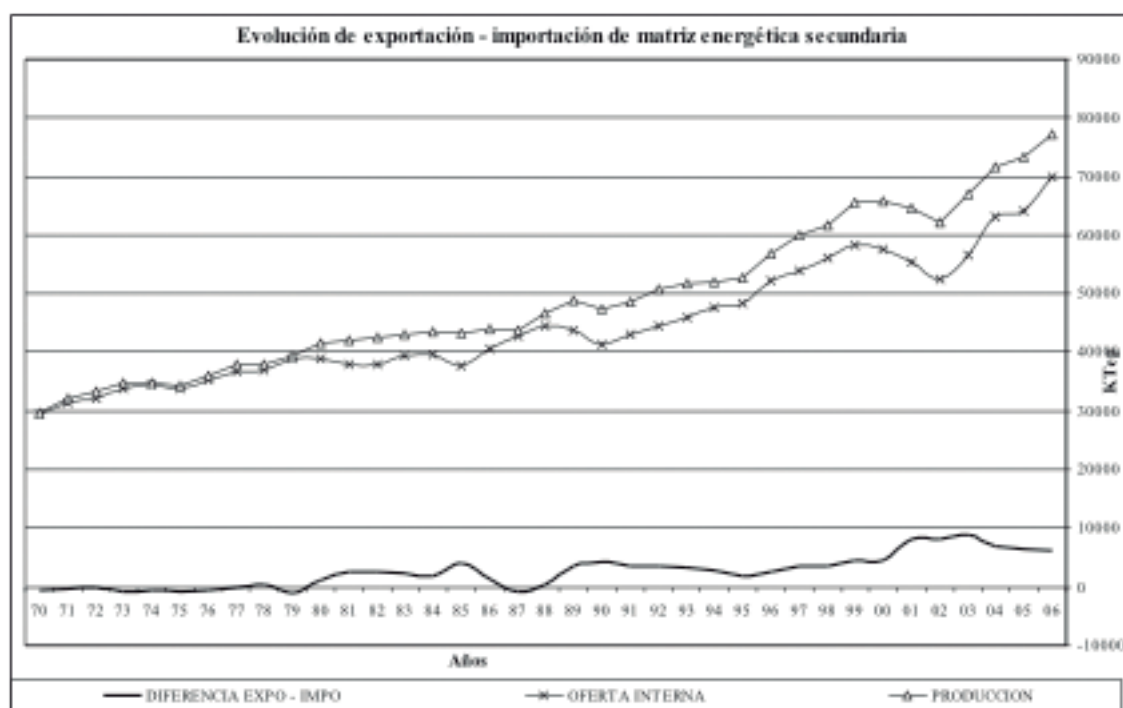


Fig. 6. Evolución de la exportación – importación de energía secundaria (Fuente: Secretaría de Energía de la Nación)

Mercado energético de energía primaria

De acuerdo con lo planteado los principales actores del mercado energético argentino son el petróleo y el gas en el primer grupo, y el gas oil y el gas distribuido por redes, en el segundo. Este mercado está integrado por los productores de energía que se encuentran en el lugar de producción o extracción del elemento energético.

Como se observa en la figura 5, en Argentina se ha crecido sostenidamente en los últimos treinta y siete años a un ritmo cercano al 3% y al 2% anual promedio en cuanto a producción y oferta interna de energía primaria respectivamente. Existe una estrecha relación entre la evolución de la energía primaria y el PBI, por eso, mientras que el país experimenta un crecimiento en materia de desarrollo económico e industrial, las fuentes de energía primaria acompañan ese crecimiento.

Por otra parte, se observa que nuestro país ha sido un importador de energía primaria hasta el año 1991. A partir del año 1992 se ha revertido esa tendencia pasando a ser un exportador hasta el año 2001, con una tasa de crecimiento promedio anual en dicho período de 60%. A partir del año 2002 y hasta el 2006, debido prin-

cipalmente a la declinación de las reservas petrolíferas, el decrecimiento en las exportaciones ha registrado una caída promedio anual de 17%. Según datos oficiales preliminares del 2007, las exportaciones han continuado con la tendencia decreciente llegando a valores de exportación de tan solo 400 Ktep⁵, lo cual representa una caída cercana al 90% respecto al año anterior.

Todo país que tenga en sus planes el desarrollo de su economía, deberá contemplar indefectiblemente un plan energético que guarde relación con la planificación de las expectativas de crecimiento a futuro.

Mercado energético de energía secundaria

Este mercado está integrado por las empresas encargadas de transformar la energía primaria en secundaria, para luego ser fraccionada y comercializada al mercado consumidor. Donde la más importante es el gas distribuido por redes y seguida por el gas oil y luego por la electricidad.

En la figura 6 se observa que el consumo de energía secundaria también presenta una tendencia prácticamente constante de crecimiento cercana al 5% anual en promedio, haciendo la salvedad del período entre los años 2001 a



Fig. 7. Esquema de dendograma y agrupación jerárquica

2003 en el que por efectos de la disminución de la actividad económica, es evidente una caída en el consumo de energía secundaria del orden del 15%, lo cual trajo aparejado indudablemente que los volúmenes sobrantes de combustibles fundamentalmente líquidos se hayan destinado hacia la exportación.

En consecuencia si se compara este gráfico con el de exportación de gas oil, se puede inferir que la componente de la matriz energética constituida por la energía secundaria guarda mayor relación con los combustibles líquidos, hecho que es de suponer, debido al peso relativo que tienen los mismos sobre el total de los valores de las fuentes que participan en dicha componente.

Como contrapartida de lo expresado en el balance de la electricidad, se puede observar en la figura 6 que el país, salvo hechos puntuales, ha sido constantemente un exportador de energía secundaria, manteniendo desde hace seis años niveles prácticamente constantes de exportación.

Elaboración de los modelos en base a *cluster analysis* y al balance energético nacional

Métodos de agrupación de conglomerados o *cluster analysis*

El modelado se implementó mediante la técnica de estadística matemática conocida como

kmeans cluster analysis (kca) que en español recibe el nombre de agrupación de conglomerados y que también se emplea en ciencias de la computación en lo que se denomina *data mining* (Timm; 2002).

El *cluster analysis* está relacionado con la identificación de grupos en conjuntos de datos. El principal objetivo de estas técnicas es la división o partición de un grupo de observaciones en un número de subgrupos o agrupamientos que en inglés se denomina *cluster* de tal manera que todas las observaciones dentro de un subgrupo sean similares entre sí y se diferencien en la mayor medida posible de las observaciones pertenecientes a otros *clusters*, mientras que las observaciones de diferentes agrupamientos no tienen que ser similares entre sí (Timm; 2002).

De alguna manera el objetivo de esta metodología es detectar estructuras en los conjuntos de datos que se hallan ocultas.

Los métodos de aglomeración jerárquica emplean elementos de una matriz de proximidad para generar un diagrama de árbol o dendograma como se muestra en la figura 7.

El proceso comienza con 5 *clusters* los cuales se dividen en tres ramas. Combinando el ítem del nivel 1 con el nivel 2 se reduce el número de *clusters* en uno, es decir se pasa de 5 a 4. Uniendo luego los ítems entre los *clusters* 3 y 4 se reduce aún más llegando a 3. El siguiente

paso es unir el ítem 5 con el *cluster* (3,4), resultando de esta manera 2 *clusters*. Finalmente, todos los ítems son combinados para formar un *cluster* único, que se constituirá en la raíz del árbol o tronco principal del mismo.

Como los diferentes ítems se han movido de izquierda a derecha los grupos de ítems se combinaron de tal manera que nunca más serán separados. La figura 7 representa el proceso de agrupación mediante un proceso de tipo aglomerativo.

El título en la parte superior del diagrama representa el número de *clusters* en cada paso del proceso. El mismo diagrama de la figura 7 puede interpretarse como un método de agrupación de conglomerados o *clusters* jerárquicos de tipo divisivo, si uno se mueve de derecha a izquierda para crear cada una de las ramas del árbol.

De una forma más general, dada una matriz de proximidad $D = [d_{rs}]$, los pasos para crear un algoritmo de aglomeración jerárquico se pueden esquematizar como sigue:

1. Comenzar con n *clusters*, cada uno conteniendo un objeto simple.
2. Buscar en la matriz de disimilaridad el par de objetos más similares. Este par estará asociado con el elemento d_{rs} de tal manera que el objeto r y el objeto s son seleccionados.
3. Combinar los objetos r y s dentro de un nuevo *cluster* (rs) empleando algún criterio que permita reducir el número de *clusters* en 1 eliminando la fila y la columna de los objetos r y s . Calcular las disimilaridades entre el *cluster* (rs) y todos los restantes *clusters* utilizando para ello el criterio de distancia seleccionado y agregar una fila y una nueva columna a la nueva matriz de disimilaridad.
4. Repetir los pasos 2 y 3, $(n-1)$ veces hasta que todos los objetos formen un *cluster* simple. En cada paso se deben identificar los *clusters* que van apareciendo y los valores asociados en la matriz de disimilaridad actualizada.

Si se cambia el criterio de distancia en el paso 3, se obtienen diversos métodos de aglomeración jerárquica, entre los que se pueden mencionar: vinculación simple o del vecino más cercano, vínculo completo o *fasthest neighbor*, método del vínculo promedio, método del centroide, método de Ward (o suma incremental de cua-

drados) para más detalles ver (Timm; 2002).

Si se está interesado en métodos de *clusterización* no jerárquicos, el número de *clusters* o conglomerados es desconocido, el proceso comienza con una matriz de proximidad o disimilaridad y a un objeto se lo asigna a un *cluster* y no se lo vuelve a relocalizar. Cuando se emplean métodos de vinculación se pueden *clusterizar* ítems o variables, en cambio cuando se emplean métodos de agrupación de conglomerados no jerárquicos solo se pueden agrupar ítems (Timm; 2002). El proceso comienza utilizando la matriz de datos originales Y , en lugar de hacerlo con una matriz de disimilaridad D . En esta forma de agrupamiento se debe conocer de antemano el número de *clusters* k en los cuales se ubicarán las semillas o centroides de los *clusters*, y las observaciones se reagrupan reasignándolas a alguno de los *clusters* en base a algún criterio de reubicación, continuando con este proceso hasta que se llega a algún fin dado en base a determinada regla predefinida.

Los procedimientos de *clusterización* no jerárquica siguen usualmente los siguientes pasos (Everitt, 1993 y Decker y Lenz; 2007).

1. Seleccionar los k (n -dimensionales) centroides de los *clusters* o semillas.
2. Asignar cada observación al centroide más cercano usando alguna norma y alguna, medida de distancia, usualmente la distancia euclídea.
3. Reasignar cada observación a uno de los k *clusters* basándose en algún criterio.
4. Finalizar el proceso si no se puede realizar ninguna relocalización de las observaciones en ningún *cluster* o si la reasignación satisface algún criterio de convergencia, en otro caso retornar al paso 2.

La implementación de agrupamientos no jerárquicos puede realizarse mediante una variada cantidad de métodos, que surgen de diferentes formas de seleccionar la determinación de los k *clusters*, la elección de diferentes normas de medida, y los diversos criterios que se pueden adoptar para la reubicación de las observaciones (Witten and Frank; 2005 y Gan et al.; 2007).

Tal como se describió previamente el método de agrupamiento no jerárquico comienza seleccionando k centroides o semillas. Estas k semillas pueden ser las primeras k observaciones, que pueden estar tomadas con cierto nivel de

separación, o en forma aleatoria, los algoritmos comienzan con semillas aleatorias y luego mediante algún algoritmo muy bien definido, las reubican. Una vez que las semillas son seleccionadas, cada una de las observaciones es evaluada para asignarla a uno de los *clusters* definidos por las mismas.

Los pasos se pueden resumir de la siguiente manera:

1. Seleccionar k semillas.
2. Asignar cada una de las $(n \times k)$ observaciones a la semilla más próxima y con esta incorporación recalcular el centroide del *cluster*.
3. Repetir el paso 2 hasta que todas las observaciones son asignadas a un *cluster* o los cambios en los centroides son pequeños (o dicho de otra forma no se produce una reasignación de *cluster* a las observaciones).

En el paso 2 la semilla puede o no ser actualizada. De esta forma se pueden emplear dos pruebas para determinar si se debe realizar el reemplazo. Una observación puede reemplazar a una o un par de semillas si la distancia entre las semillas es menor que la distancia entre una observación y la siguiente semilla más próxima. La semilla desplazada se transforma en una observación para el recálculo de los centroides. Si una observación falla esta prueba se puede plantear una segunda instancia de evaluación (Timm; 2002).

La observación reemplaza la semilla más próxima si la distancia euclídea de la misma a todas las semillas próximas es mayor que la menor distancia de la semilla más cercana al resto de las semillas.

En una primera instancia todas las observaciones son asociadas con k *clusters*. Luego este proceso se repite hasta que los cambios en las semillas de los *clusters* son pequeños basados en algún criterio de convergencia.

Como primer paso se puede emplear un esquema de tipo jerárquico para identificar tanto las semillas iniciales como el número de *clusters*, estas luego se pueden emplear como dato de entrada dentro de algún procedimiento de aglomeración no jerárquico para refinar la solución. Esta es justamente la metodología empleada en el presente trabajo (para más detalles se puede consultar a Jain and Dubes; 1988 y el trabajo de Pelleg and Moore; 2000).

La confirmación del número de los *clusters* que se tomaron como modelos representativos del mercado energético fue validada estadísticamente mediante la técnica conocida como *bootstrapping* (Efron; 1979, 1982, Efron y Tibshirani 1998). El *bootstrap* es un procedimiento estadístico que muestrea modelos de una población mediante el proceso de remuestreo de la muestra, y fue introducido formalmente por Efron en 1982.

Si una determinada cantidad estadística se expresa como función de una distribución desconocida, entonces el estimador *bootstrap* tiene la misma funcionalidad que la función de distribución empírica (Timm; 2002 y Efron; 1982).

Este método está basado en el procedimiento estadístico de muestreo sobre el conjunto de muestras, ya sea medidas en el laboratorio o resultado de un experimento, que bien puede ser con o sin reemplazo. En el caso del presente trabajo cuando se toma una muestra del conjunto de datos para formar un grupo se realiza sin reemplazo. Es decir, la misma muestra, una vez seleccionada, no puede ser seleccionada otra vez. Es como elegir equipos de fútbol: no se puede elegir la misma persona dos veces (Witten and Frank; 2005).

Como es claro en los datos del presente trabajo, no se empleó la reposición de las muestras. La mayoría de los métodos de aprendizaje, entre los que se halla el *bootstrapping*, pueden utilizar la misma muestra dos veces, pero esta produce una diferencia en el resultado final (Witten and Frank; 2005).

La idea de *bootstrap* es muestrear el conjunto de datos para formar un conjunto de entrenamiento. Para ello, un grupo de datos de n casos es muestreado n veces, sin sustitución.

Cada elemento de la muestra tiene una probabilidad de $1/n$ de ser elegido para el grupo de entrenamiento y , por tanto, una probabilidad de $(1 - 1/n)$ de no ser elegido. Multiplicar estas probabilidades conjuntas según el número de oportunidades, que es n , da como resultado la selección:

$$\left(1 - \frac{1}{n}\right)^n \approx e^{-1} = 0.368$$

donde e es la base de los logaritmos naturales y n la tasa de error.

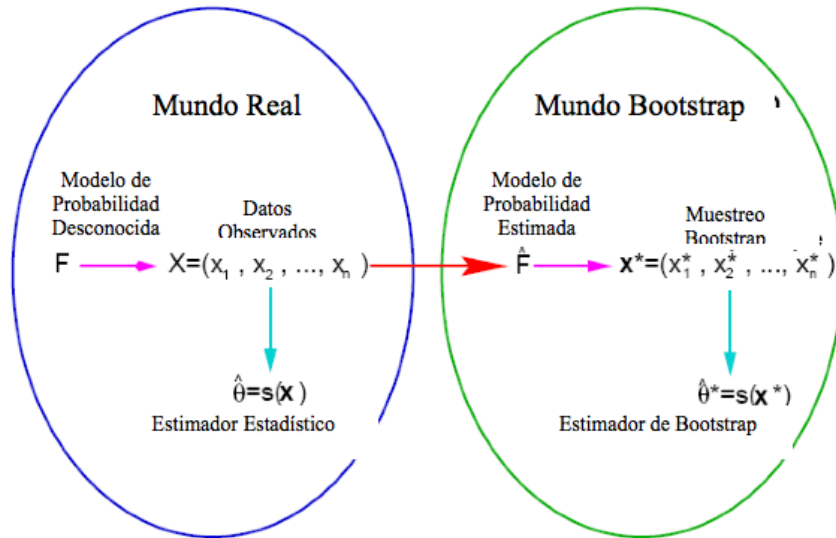


Fig. 8. Procedimiento de *bootstrap*

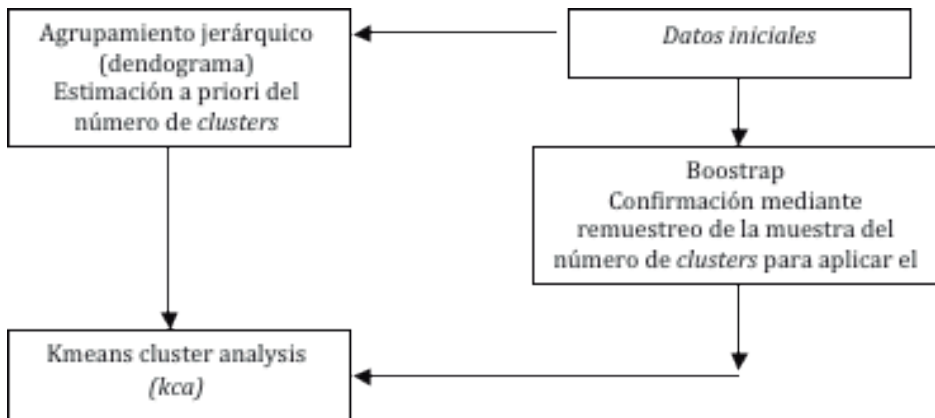


Fig. 8a. Diagrama lógico de *bootstrap*

De este modo, para una base de datos de tamaño razonable, la prueba contendrá aproximadamente el 36,8% de los casos y el conjunto de entrenamiento contendrá aproximadamente el 63,2% de ellos (este es el motivo por el cual se denomina *bootstrap* 0.632) (Witten and Frank; 2005).

En el gráfico de la figura 8 se puede apreciar el concepto que subyace en la técnica del *bootstrap*.

La cifra obtenida mediante la formación de un sistema de aprendizaje sobre un grupo de entrenamiento y calculando este error fuera del grupo de prueba será una estimación pesimista de la verdadera estimación de la tasa de error, ya que el conjunto de formación, tiene tamaño n , sin embargo, sólo contiene el 63% de los casos, que no es mucho en comparación

con otras técnicas (Witten and Frank; 2005).

Para un análisis pormenorizado de las tasas de error en los procedimientos de *bootstrap* se puede ver: Davison and Hinkley; 2006, Witten and Frank; 2005 y Last et al.; 2004.

Resumiendo, la técnica de estadística de *bootstrap* que se presentó sucintamente en el párrafo anterior, se ha usado en el presente trabajo para validar la cantidad de conglomerados o *clusters* que se emplearon en la etapa de agrupación *kca*. A su vez como resultado de la primera instancia de *clusterización* este dato también se comparó a partir del diagrama de dendograma, como el que se mostró previamente.

El diagrama lógico se muestra en las figuras 8a y 8b.

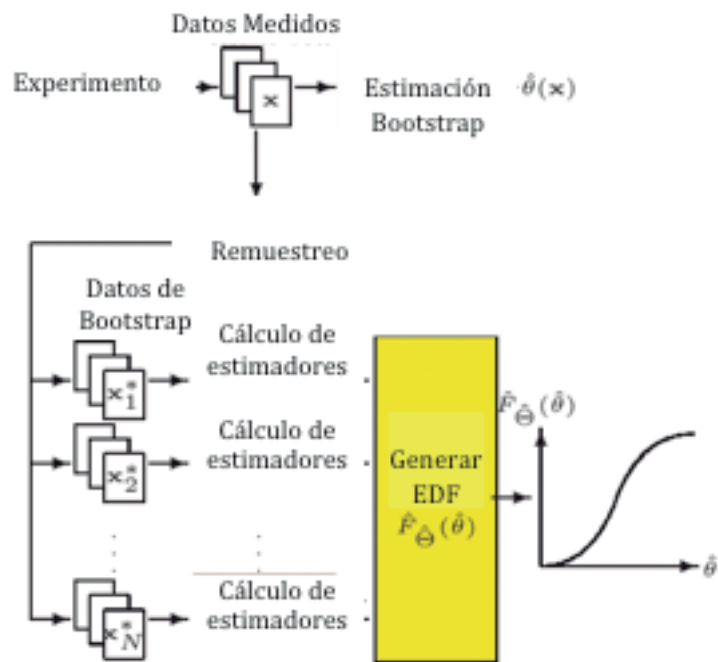


Fig. 8b. Diagrama lógico de *bootstrap*

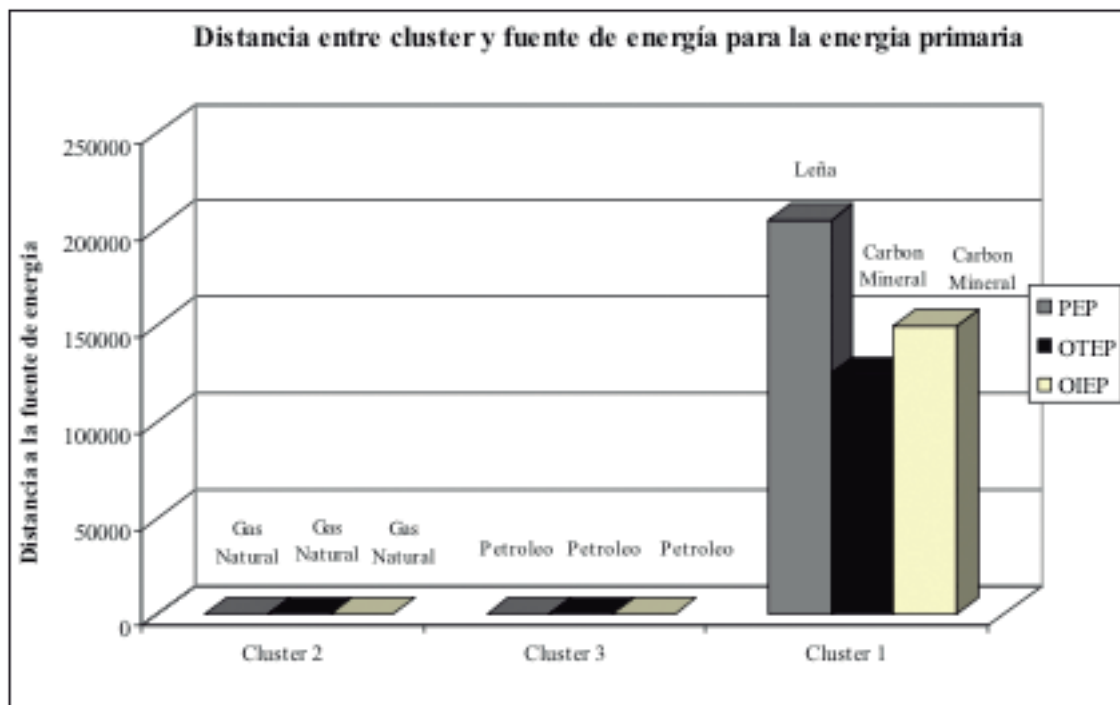


Fig. 9. Distancia de aproximación para la energía primaria

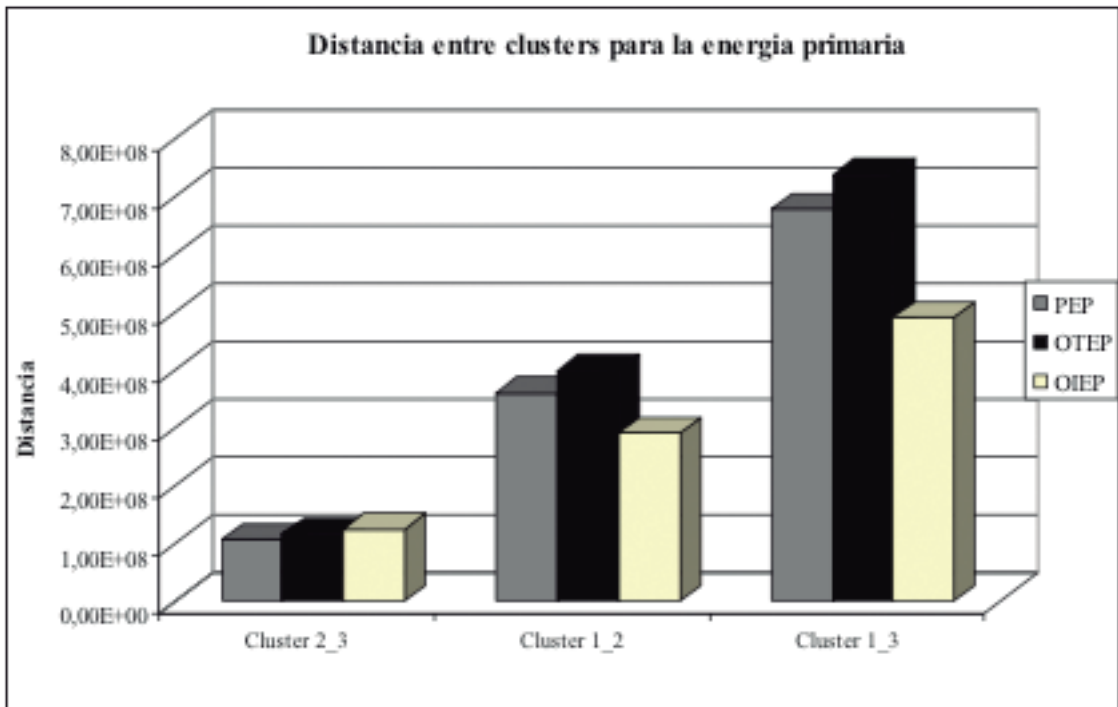


Fig. 10. Distancia entre *clusters* para la energía primaria

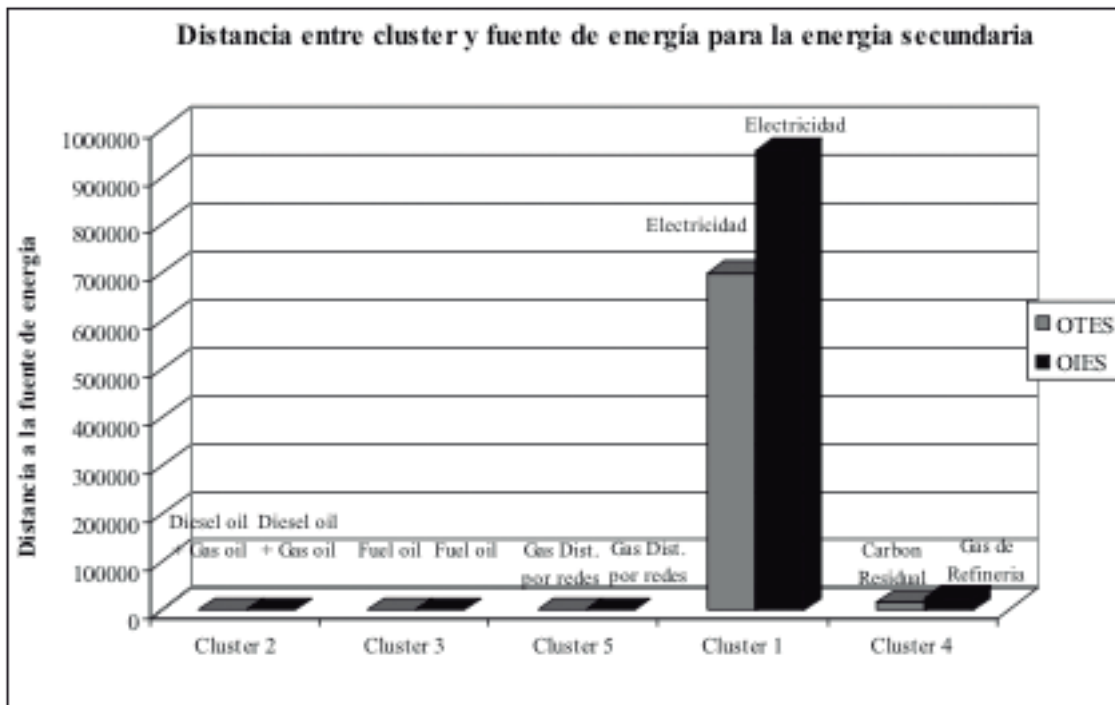


Fig. 11. Distancia de aproximación para la energía secundaria

Descripción de los datos utilizados

De acuerdo con lo expresado en el apartado anterior los datos utilizados para el presente análisis fueron tomados de los documentos elaborados por la Secretaría de Energía de la Nación en base a la información que diversos actores del sector aportan al mencionado organismo.

Para un mejor entendimiento y análisis estos datos han sido ordenados en forma matricial dentro de una planilla que resume la actividad energética y los valores que la caracterizan en forma de una tabla de doble entrada, según Tabla 1.

La energía primaria y la energía secundaria están representadas en forma de fila, y cada una de ellas a su vez está compuesta por subgrupos de energía según lo expresado anteriormente. En el caso de la energía primaria los subgrupos serán: la energía hidráulica, la nuclear, el gas natural, entre otras, y en el caso de la energía secundaria los subgrupos serán: la electricidad, el gas distribuido por redes, el gas de refinería, entre otras.

Por su parte en forma de columna están expresadas las formas en las cuales se encuentra la energía desde el inicio de la cadena de producción hasta su comercialización.

Sumario Final

El objetivo del presente apartado es hacer hincapié en la situación actual del modelo energético del país, sin perder de vista el comportamiento y la evolución del mercado energético de la región y del mundo. En particular, exponer la detección y modelado de patrones de comportamiento de la MEN mediante la implementación de técnicas matemáticas y computacionales de origen estadístico cuya característica fundamental es la determinación de esquemas cuantitativos de comportamiento, para estudio y análisis de la MEN.

Los resultados obtenidos como resumen general de la MEN primaria se muestran en la figura 9.

Luego de analizados tres casos de la MEN

primaria en base los datos provistos por el BEN, es decir la Producción de Energía Primaria (PEP), la Oferta Total de Energía Primaria (OTEP) y la Oferta Interna de Energía Primaria (OIEP), se han obtenido dos clusters que se ajustan a dos fuentes de energía primaria, que en este caso resultan ser el *cluster 2* en representación del gas natural y por su parte el *cluster 3* al petróleo (puesto que como se ha visto, la distancia de aproximación entre *cluster* y fuente de energía es cero). Por lo tanto, una conclusión importante a tener en cuenta es que gracias a este método de análisis el gas natural y el petróleo son elementos característicos e imprescindibles a tener en cuenta cuando se requiera estudiar el comportamiento de la MEN primaria.

Un elemento interesante que también valida los resultados de la figura anterior respecto de los *clusters* obtenidos se muestra en la figura 10 en donde la distancia entre los *clusters 2* y *3* (en el gráfico rotulado como "*Cluster 2_3*") guardan estrecha relación para los modelos de la PEP, la OTEP y la OIEP, ya que los tres se encuentran en valores cercanos a $1,17 \cdot 10^8$ con un margen de $\pm 10\%$.

Otro elemento que refuerza el resultado de los *clusters* obtenidos puede observarse en la figura 1 en donde se había determinado que el gas natural representaba un 50% de la MEN primaria y el petróleo por su parte un 39%, siendo ellos los de mayor participación en la misma. Esta conclusión denota la fuerte polarización hacia la utilización de los combustibles fósiles en el orden del 90% del total de la MEN primaria desde por lo menos los años '70 aunque con proporciones distintas y con una participación en total del 95%.

El gas natural fue el de mayor crecimiento en los últimos 35 años según la figura 2 con un valor superior al 600%. Pero paradójicamente es la fuente de energía que más ha declinado sus reservas llegando a 9,36 años de horizonte en el caso más favorable, esto indica que se ha consolidado una tendencia negativa. En ese sentido, el gran aumento en su utilización pudo haberse sustentado por un lado por el buen precio interno que favoreció el reemplazo de otros fluidos por el gas natural y, por otra parte, a los buenos precios de venta para

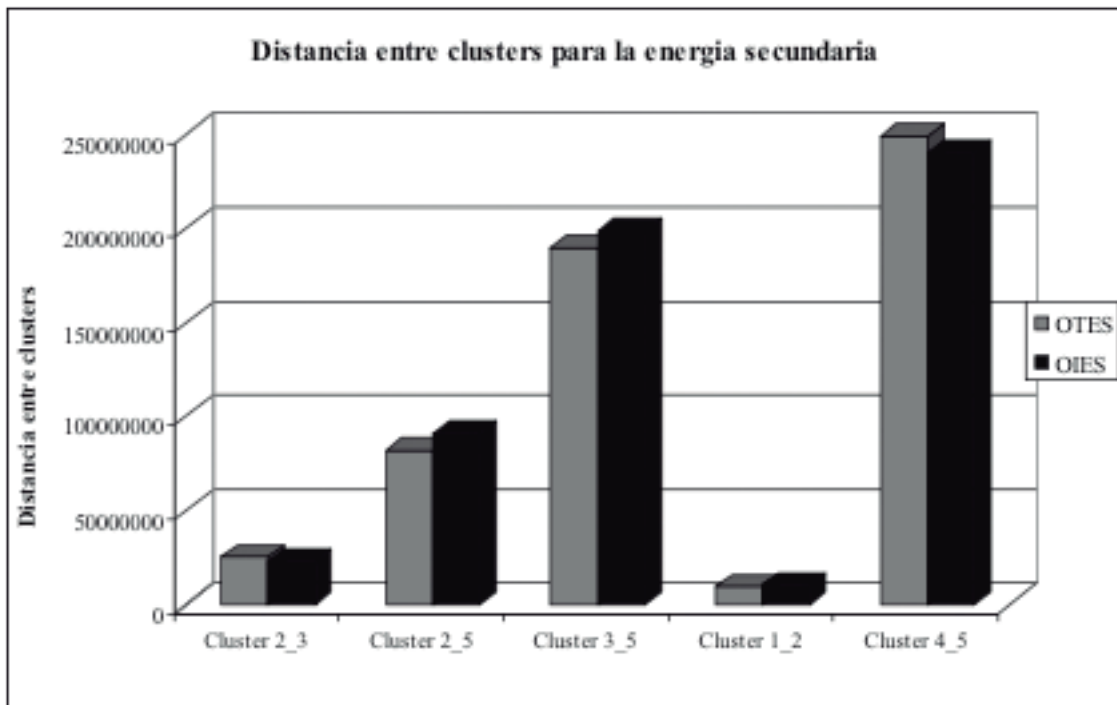


Fig. 12. Distancia entre *clusters* para la energía secundaria

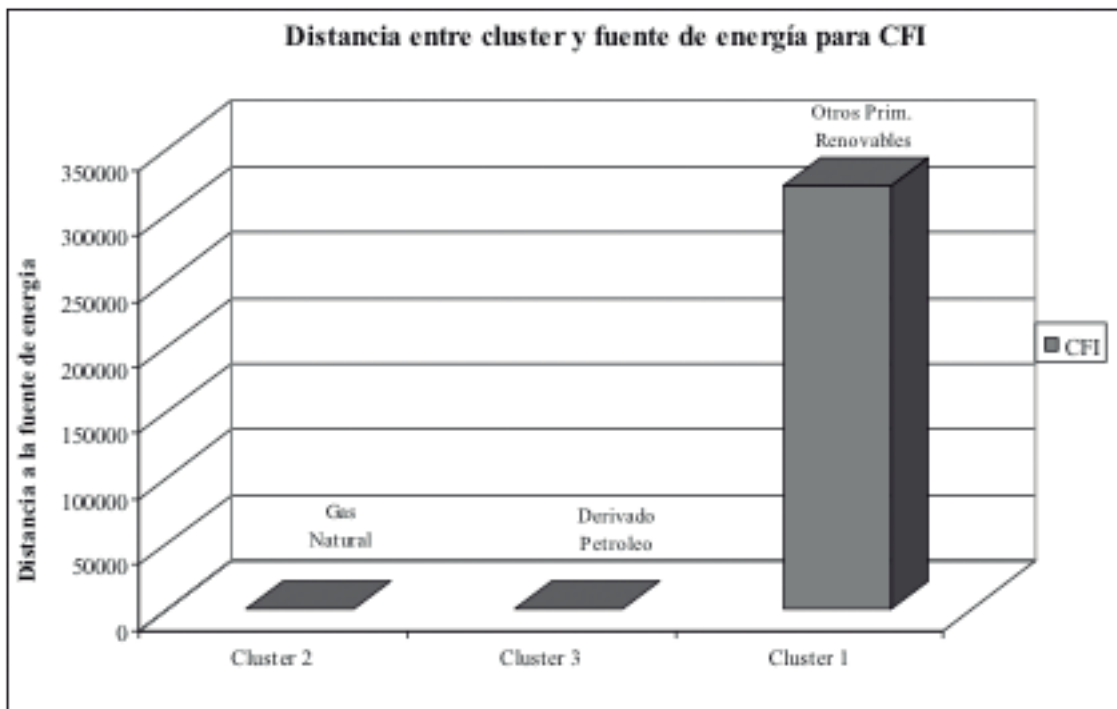


Fig. 13. Distancia de aproximación para el CFI

el mercado externo que derivaron en su exportación.

Este razonamiento, en coincidencia con la opinión de varios analistas del mercado energético lleva a la inevitable conclusión que la

Argentina se halla en vías de convertirse en un claro importador de energía primaria de mantenerse la tendencia graficada en la figura 5, hecho que no ocurría desde hace aproximadamente 20 años.

El resultado del análisis realizado para hallar

Cluster / Fuente	Energía Primaria		Energía Secundaria			Consumo a usuario final	
	Gas Natural	Petroleo	Diesel Oil + Gas Oil	Fuel Oil	Gas Dist. por Redes	Gas Natural	Derivado Petroleo
Cluster 2	3		2			1	
Cluster 3		3		2			1
Cluster 5					2		

Tabla 2. Resumen de *clusters* hallados en la MEN

los *clusters* de la MEN secundaria se muestra en la figura 11.

En el caso de la MEN secundaria se han analizado dos agrupamientos energéticos, es decir la OTES y la OIES. Los tres *clusters* hallados son iguales para los dos casos, puesto que el diesel oil con gas oil están representados por el *cluster 2*, el fuel oil por el *cluster 3* y por su parte el gas distribuido por redes por el *cluster 5*. En el caso del *cluster 1* y *4* al ser su distancia de aproximación a la fuente de energía distinta de cero, no se consideran como *clusters* válidos.

Lo obtenido en la figura anterior para los tres *clusters* puede contrastarse con los resultados de la figura 12 de la página siguiente en donde se observa que la distancia entre *clusters* para el caso de la OTES y la OIES tiene diferencia entre el *cluster 2* y *3* menor al 10% al igual que entre el *cluster 2* y *5*, mientras que para el *cluster 3* y *5* es menor a 5%. Esto refleja la paridad que existe entre los *clusters* en cuestión.

En consecuencia, se observa que el gas distribuido por redes representa el 48% y el diesel oil con el gas oil el 15% siendo estas las dos fuentes de energía más importantes, seguidas por último por el fuel oil 7% en cuarto lugar. En la misma figura es sencillo entonces determinar que la suma de las fuentes de energía de origen fósil representan un valor cercano al 88% a diferencia de lo que ocurría en los años '70 cuando los mismos eran del 93%.

Desde ese momento hasta la fecha, los combustibles líquidos de origen fósil han disminuido su participación en la MEN, pasando el conjunto diesel oil - gas oil desde el 18% hasta una participación del 15% en la actualidad, y en el caso del fuel oil la reducción ha sido aún más drástica pasando de 30% al actual 7%.

Un combustible de gran expansión en la MEN secundaria ha sido sin dudas, el gas distribuido por redes, que ha crecido su participación alrededor de 170 % en el mismo período. En particular, este último combustible ha cambiado su patrón de consumo ya que evidencia una nueva distribución de utilización por sector en los últimos 15 años; los casos más relevantes son el consumo industrial que ha mermado 4 %, y las centrales eléctricas donde ha avanzado 6 %. En ese sentido, las centrales eléctricas se han convertido en importantes demandantes de gas natural debido a que se logra una mejor combustión, un menor costo de mantenimiento y en consecuencia mayor rendimiento en el ciclo completo, que se suma a un menor costo de combustible y mayor flexibilidad para el transporte almacenamiento y acondicionamiento.

Por lo visto hasta el momento en el desarrollo del presente capítulo, se han identificado una determinada cantidad de *clusters*, los que ya sean para energía primaria o secundaria, siempre se encuentran representando a combustibles fósiles que pueden estar en estado líquido o gaseoso, y siguiendo con ese lineamiento, no resultaría nada llamativo que el consumo del producto proveniente de la oferta interna continuará en ese sentido.

La figura 13 valida entonces la suposición, puesto que aquí se ha hallado que el *cluster 2* se ha ajustado al consumo de gas natural y el *cluster 3* al derivado de petróleo.

Todo lo expresado hasta el momento confluye hacia una conclusión clara y contundente: todos los *clusters* hallados para cualquier tipo de agrupamiento ya sea energía primaria o secundaria provienen de fuentes de energía de origen fósil. Para sustentar esta afirmación se ha realizado un cuadro de análisis que resume la

totalidad de los *clusters* hallados en la Tabla 2.

En donde se observa que:

La MEN primaria compuesta por tres agrupamientos (PEP, OTEP y OIEP) ha quedado modelada por un total de seis *clusters* según se muestra en la columna denominada Energía Primaria, y tendrá como componente característica la serie de datos correspondiente al gas natural y al petróleo dependiendo de cada agrupamiento.

La segunda columna correspondiente a la MEN secundaria compuesta por dos agrupamientos Oferta Total de Energía Secundaria (OTES) y Oferta Interna de Energía Secundaria (OIES) ha sido modelada por seis *clusters* según se muestra en la columna denominada Energía Secundaria, y cuyo factor común será la serie de datos del diesel oil con gas oil, el fuel oil y el gas distribuido por redes discriminado por su agrupamiento.

Es decir que los *clusters* hallados no hacen otra cosa más que sacar a la luz que la matriz energética del país esta estrechamente relacionada con la producción, la transformación y el consumo de productos energéticos de origen fósil. Esto queda evidenciado puesto que no han aparecido *clusters* representativos de fuentes de energía renovables como por ejemplo la energía hidráulica, la leña y el bagazo entre otros.

Una característica que es común a todos los *clusters* hallados y que no debe pasarse por alto es que siempre está presente el gas natural y el petróleo en cualquiera de sus formas ya sea crudo o refinado.

Desde una óptica pesimista se puede inferir entonces que, ante una mínima variación en las reservas o en el precio de los combustibles fósiles, esto traerá aparejado indefectiblemente un impacto en la economía local debido a la alta dependencia que existe en este tipo de combustibles, ya que, como se mencionó, cerca del 90% de la matriz energética depende de recursos no renovables, siendo el gas natural prácticamente el 50% de la misma.

El estado actual de las reservas petroleras y gasíferas plantean una delicada situación que abre una serie de interrogantes sobre la seguridad y factibilidad del abastecimiento ener-

gético a mediano plazo en una economía con un crecimiento moderado cercano al 2%. En el caso de una expectativa de crecimiento económico del orden del 6% la situación se ubica en un escenario más complejo, en donde deberían aparecer problemas coyunturales de abastecimiento lo que luego desembocaría en la necesidad de buscar el preciado fluido en el mercado internacional a un precio que seguramente dependerá del humor de la economía mundial.

Ante la falta de incentivos que precipiten una fuerte y decidida inversión del sector privado en la exploración y producción de hidrocarburos, se plantea la alternativa de recurrir entonces a las fuentes de energía alternativas o más bien llamadas fuentes de energía renovables, que como efecto rebote traerá mejoras sustanciales en la matriz energética y a la economía en general.

En ese sentido, los organismos oficiales en los últimos años han encarado una serie de medidas que tienen como finalidad favorecer la inversión directa sobre proyectos que generen energía a través de fuentes renovables como por ejemplo, la energía eólica, la energía solar, la energía hidráulica, la utilización de biocombustibles y programas destinados al ahorro energético entre otros.

Como consecuencia de ello se espera lograr una diversificación en la matriz energética con los siguientes beneficios:

- Obtener mayores volúmenes de inversión en proyectos energéticos lo cual motoriza la economía nacional y la creación de fuentes de empleo.
- Incorporar n fuentes de energía en lugares en donde actualmente no llegan las redes con lo cual se contribuye y motoriza el crecimiento de la economía de la región.
- Obtener mayor confiabilidad de abastecimiento energético lo cual atrae la inversión extranjera que se traduce en la instalación de nuevas plantas industriales de todo tipo.
- Mejorar la imagen y la proyección del país hacia el mundo, puesto que las principales economías han abierto grandes líneas de crédito para ser destinadas a proyectos de energía limpia "Green Projects".
- Obtener una disminución en la emisión de contaminantes hacia la atmósfera.
- Lograr una cierta inmunización ante abruptas

crecidas de precios de una fuente de energía puntual y por ende los efectos sobre la economía son menores.

- Resultar una vía alternativa ante la escasez de fuentes de energía tradicionales.
- Elegir entre varias fuentes de energía dependiendo de la oportunidad y del precio.

En otros países de la región como es el caso de Brasil, Venezuela, Paraguay, Colombia, Chile y Perú la participación de la energía hidroeléctrica en la generación de electricidad alcanza valores cercanos al 60%, mientras que en Argentina es de sólo un 30%, siendo el resto de la energía generada a partir de combustibles fósiles de un 65% y el 5% restante por energía nuclear. A efectos de tener menor dependencia de la energía hidroeléctrica (y por ende de las lluvias) Brasil ha decidido modificar su matriz energética incorporando a su parque de generación una cantidad importante de centrales termoeléctricas.

El resultado de lo obtenido en el presente trabajo deja planteado para el futuro la necesidad de encontrar una herramienta que pueda predecir cómo debería orientarse una matriz energética en función a sus recursos energéticos y a los precios, a partir de los *clusters* que caracterizan a un modelo energético actual.

Conclusiones

Con la elaboración del presente estudio se ha puesto en evidencia que la metodología de análisis de conglomerados por *kca* es una propuesta viable que se aplica en un campo de estudio de forma novedosa y ha resultado ser muy consistente con referencia a los resultados obtenidos. Además, mostró ser robusta dado que la determinación del número de *clusters* o agrupamientos a emplear se llevó a cabo mediante dos procesos independientes que arrojaron en todos los casos resultados absolutamente similares. Destacándose como cualidad, el haber presentado una metodología de determinación objetiva de los restantes agrupamientos que constituyen los diferentes modelos.

Tradicionalmente el *data mining* que se emplea en las ciencias de la computación o las técnicas de análisis estadístico o de series de tiempo, no se cruza con el estudio de mercados como los que se han mostrado en el

presente trabajo, es más, en la literatura son escasas las aplicaciones halladas a casos por lejos similares.

La caracterización de la MEN mediante la técnica de agrupación de conglomerados ha mostrado ser eficiente para ahondar en el estudio de las relaciones estructurales de los componentes de la misma.

Los resultados a los cuales se ha arribado luego de la caracterización, han permitido identificar el comportamiento de la MEN. Este interesante hallazgo es de vital importancia puesto que se convierte en el punto de partida para tomar decisiones estratégicas en materia energética.

El hecho de conocer la caracterización de la MEN también permite anticiparse a eventuales inconvenientes de desabastecimiento y de precios. Como consecuencia de ello se podrán implementar a tiempo todas las medidas que se consideren necesarias para corregirlo, avanzando al mismo tiempo con políticas energéticas de mediano y largo plazo en una dirección clara y definida.

En virtud de lo expresado, resultaría lógico suponer que se contaría con mayor previsibilidad para la planificación de un horizonte energético, que sin dudas, se traducirá en una mayor seguridad y factibilidad en el abastecimiento energético, convirtiéndose en un factor limitante e ineludible del crecimiento económico.

Referencias

- ANDERBERG, M.R. (1973) "Cluster analysis for applications", New York: Academic Press.
- Boyle, G., (1996). Editor, "Renewable Energy, Oxford, England: Oxford University Press, 1996.
- DAVISON, A. C.; HINKLEY, D. (2006)"Bootstrap Methods and their Application" (8th ed.). Cambridge: Cambridge Series in Statistical and Probabilistic Mathematics.
- DUKE, R. D., AND KAMMEN, D. M. (1999) "The economics of energy market transformation initiatives". The Energy Journal, 20 (4), 15 – 64.
- DECKER, R., AND LENZ, H. (2007) "Advances in Data Analysis", Springer Verlag, Hidelberg, Berlin.
- EFRON, B. (1982) "The Jackknife, the Bootstrap and Other Resampling Plans", SIAM, Bristol.
- EFRON, B. (1979) "Computers and the Theory of Statistics", SIAM Rev. 21, 460.
- EFRON, B. and Tibshirani, R. J. (1998) "An Introduction to Bootstrap", Chapman & Hall.
- EVERITT, B.S. (1993) "Cluster analysis", Third Edition. London: Edward
- GAN, G., MA, CH. AND WU, J., (2007) Data Clustering Theory, Algorithms, and Applications. SIAM.
- HALL, R. E. December (2001)"The Stock Market and Capital Accumulation" American Economic Review, 91(5), pp. 1185–202..
- HARTIGAN, J.A. (1975) "Clustering algorithms". New York: John Wiley.
- JAIN, A. K. AND DUBES, R. C.(1988) "Algorithms for Clustering Data" Prentice Hall, Englewood Cliffs, NJ.
- KAMMEN, D. M., SHLYAKHTER, A. I., BROIDO, C. AND WILSON, R. (1994) "Quantifying credibility of energy projections from trends in past data: the U. S. energy sector". Energy Policy, 22, 119 – 131.
- LAST, M., KANDEL, A. AND BURKE, H. (2004) Data Mining in Time Series Databases. World Scientific.
- PELLEG, D. AND MOORE, A.(2000) x-means: Extending k-means with efficient estimation of the number of clusters. In Proceedings of the Seventeenth International Conference on Machine Learning, Palo Alto, CA.
- SHAW, S.H. (2000)"Modeling Energy Markets with the Intermediate Future Forecasting System" Forthcoming in a Handbook of Engineering-Economic Modeling. EPA, USA.
- TIMM, NEIL H. (2002) "Applied Multivariate Analysis", Springer Berlag.
- WITTEN I. H.& FRANK E. (2005) "Data Mining Practical Machine Learning Tools and Techniques", Morgan Kaufman – 2nd Edition. USA.

Informes y Reportes Técnicos

1. International Energy Outlook 2009 (IEO2009), Energy Information Administration (EIA - Official Energy Statistics from the US Government)
2. Informe Anual 2007, Consejo Federal de Energía Eléctrica (CFEE).
3. Energías Renovables 2008 -Pequeños Aprovechamientos Hidroeléctricos, Subsecretaría de Energía Eléctrica
4. BEN, Secretaría de Energía de la Nación, 2006
5. Una estrategia energética sustentable para el siglo XXI, Fundación Creer y Crecer, 2005
6. Régimen Nacional de Energía Eólica y Solar, Ley 25.019, Honorable Congreso de la Nación, 2008
7. Plan energético nacional, Ministerio de Planificación Federal, Inversión pública y Servicios, 2008
8. Transporte y distribución de gas natural en Argentina, TGN, TGS, 2003
9. Información del sector, website de Natural Energy, www.naturalenergy.com.ar
10. Informe Anual 2005, Secretaría de Energía de la Nación
11. Anexo III - Licenciatarias del servicio de gas, Informe ENARGAS, 2007
12. Repote técnico de Hidrocarburos, Secretaría de Energía de la Nación
13. Mapa geográfico de empresas transportistas en alta tensión, ATEERA, 2009
14. Tablas dinámicas de *downstream* - Precio de venta por localidad, Secretaría de Energía de la Nación
15. Ministerio de justicia de la Nación, web site Infoleg