MERCADOS ENERGÉTICOS CONSULTORES

ESTUDIO INTEGRAL DE TARIFAS ELÉCTRICAS

TAREA 1.2.5 INFORME N° 14: SELECCIÓN DE SISTEMAS ELÉCTRICOS REPRESENTATIVOS DE DISTRIBUCIÓN

Preparado para:

CRE
COMISIÓN REGULADORA DE ENERGÍA

SELECCIÓN DE SISTEMAS ELECTRICOS REPRESENTATIVOS DE DISTRIBUCIÓN

CONTENIDO

ÍNDICE DE FIGURAS Y TABLAS

GLOSARIO

AT: Alta tensión

BT: Baja tensión

CFE: Comisión Federal de Electricidad

Cluster: agrupación de elementos con características similares o conglomerado

CRE: Comisión Reguladora de Energía

Curva monótona de carga: corresponde a la curva de demanda horaria anual de un sistema, ordenada en forma decreciente.

MT: Media tensión

PESED: Procedimiento para la determinación de pérdidas de energía en el sistema eléctrico de distribución

SCADA: Supervisory Control and Data Acquisition (en español, registro de datos y control de supervisión)

SEN: Sistema Eléctrico Nacional

SENER: Secretaria de Energía

SER: Sistemas Eléctricos Representativos

SICOM: Sistema Comercial de CFE

TDR: Términos de referencia

Transformador AT/MT: Transformador reductor de Alta tensión a Media tensión

Transformador MT/BT: Transformador reductor de Media tensión a Baja tensión

SELECCIÓN DE SISTEMAS ELECTRICOS REPRESENTATIVOS DE DISTRIBUCIÓN

RESUMEN EJECUTIVO

El presente documento tiene por objeto cumplir con lo indicado en los TDR, Tarea, 1.2.5 Integración de la información para el cálculo del Valor de Reposición a Nuevo de los activos de distribución, punto a). *El Consultor seleccionará una muestra de zonas de distribución que representen el espectro de dispersiones de carga de todo el país*.

La metodología que se utilizó para la determinación de las zonas de distribución, denominadas "Sistemas Eléctricos Representativos – SER", consistió en el análisis de conglomerados o *clusters*. Mediante este método se agrupan las distintas zonas de las divisiones de la Comisión Federal de Electricidad (CFE) en grupos homogéneos sobre la base de variables observadas.

Previamente a la aplicación de la metodología descripta, se analizó la información recibida, tanto su completitud respecto a lo solicitado como el contenido y la consistencia de los datos enviados por las divisiones de CFE.

Además del análisis de consistencia de la información recibida, se realizó un análisis estadístico y gráfico de las variables, definiendo rangos de razonabilidad de los datos en base a la experiencia del Consultor. Este análisis previo permitió detectar la presencia de valores anormales (atípicos).

En función de la información depurada se plantearon indicadores que permitieron medir la intensidad del consumo, estimar la densidad lineal de la demanda, calcular el factor de uso de las instalaciones; y obtener indicadores de estructura (por ej,: si la red es mayormente urbana o rural, aérea o subterránea). A los efectos de decidir cuales indicadores serán utilizados en el análisis de cluster, se realizaron análisis de correlación entre los mismos para confirmar su independencia.

Dado que la configuración de la red se refleja necesariamente en los costos -las redes urbanas pueden costar el doble que las redes rurales, y las redes subterráneas pueden ser hasta 10 veces más costosas que las redes áreas-, se realizó una primera clasificación de las zonas, de la cual surgieron tres grandes grupos, separados según el porcentaje de red subterránea en MT respecto al total de la red MT:

i. Grupo 1: conformado por las zonas que tienen menos de 0.5% de red subterránea.

ii. Grupo 2: conformado por las zonas con red subterránea mayor o igual a 0.5% y menor a 10%.

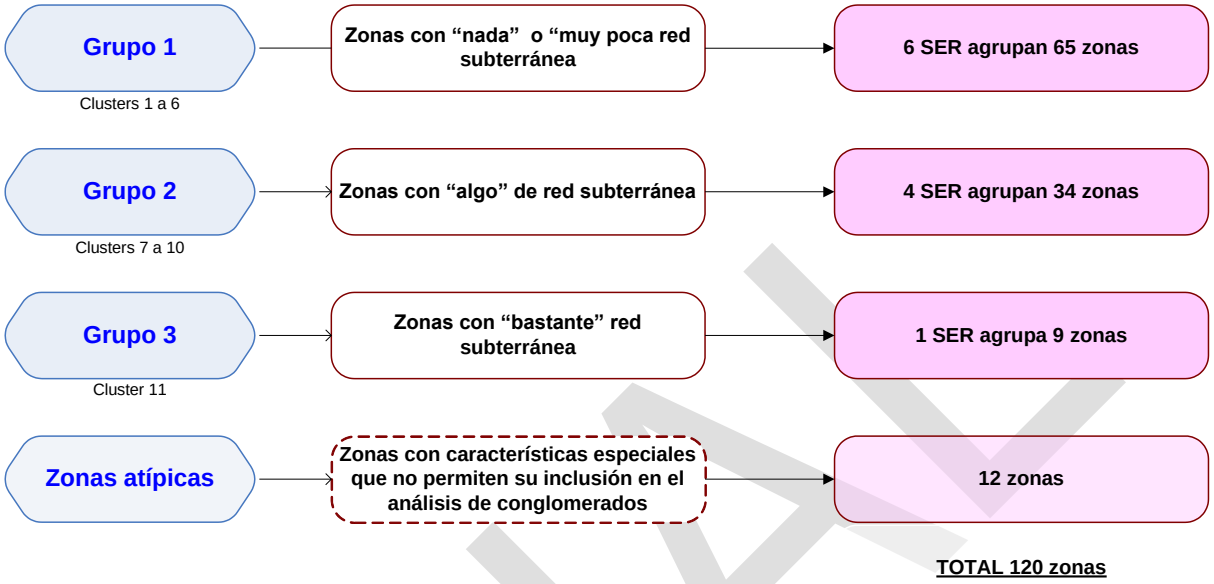
iii. Grupo 3: conformado por las zonas que tienen una red subterránea mayor o igual a 10%.

El Grupo 1 quedó finalmente conformado por 65 zonas y seis *clusters*, mientras el Grupo 2 por 34 zonas y cuatro *clusters*. El Grupo 3, por su parte, quedó conformado por 9 zonas, las cuales por sus características forman un *cluster* en si mismo. Finalmente, 12 zonas fueron tratadas como zonas atípicas, y no se incluyeron en ningún *cluster*.

Para los grupos 1 y 2 se realizó un análisis de *clusters*, utilizando el algoritmo de las *k-medias,* según la metodología descripta en el capítulo 2 del presente informe. El grupo 3 forma un conjunto homogéneo cada uno.

En la figura siguiente se muestra un esquema de los resultados:

ESQUEMA DE LOS RESULTADOS



En la siguiente tabla se presentan los resultados obtenidos en cuanto a la determinación de los *cluster*s y el correspondiente SER:

RESULTADOS SER

Cluster	Cantidad de zonas	SER	Características
1	12	Durango	Alta intensidad del consumo en MT; bajo factor de utilización en MT; una densidad de la demanda similar a la media del Grupo en MT, y por debajo de la media en BT; y redes urbanas en cantidades similares a la media del Grupo tanto en MT como en BT.
2	22	Poza Rica	Baja intensidad del consumo en MT; bajo factor de utilización en MT; baja densidad de la demanda tanto en MT como en BT; y redes menos urbanas que la media del Grupo tanto en MT como en BT.
3	5	Sabinas	Alta intensidad del consumo en MT; alto factor de utilización en MT; con una densidad de la demanda alta tanto en MT como en BT; y redes más urbanas que la media del Grupo, tanto en MT como en BT.
4	12	Tuxtla	Intensidad del consumo en MT baja; un factor de utilización en MT similar a la media del Grupo; con una densidad de la demanda baja en MT y media en BT; y redes urbanas más bajas que la media del Grupo en MT y en BT.
5	7	Navojoa	Intensidad del consumo en MT media; bajo factor de utilización en MT; con una densidad de la demanda alta en MT y en BT; y redes menos urbanas que la media del Grupo en MT, y similares a la media en BT.
6	7	Tepic	Intensidad del consumo en MT alta; alto factor de utilización en MT; con una densidad de la demanda media en MT y alta en

Cluster	Cantidad de zonas	SER	Características
			BT; y redes menos urbanas que la media del Grupo 1 en MT, y más urbanas que la media en BT.
7	3	Culiacán	Alta intensidad del consumo en MT; un factor de utilización en MT similar a la media del grupo; una densidad de la demanda similar a la media del grupo en MT, y alto en BT; redes urbanas similares a la media del Grupo en MT, y mayores en BT; y una relación redes subterráneas / redes urbanas mayor a la media del Grupo en MT y en BT.
8	7	Ensenada	Baja intensidad del consumo en MT; alto factor de utilización en MT; baja densidad de la demanda en MT, y similar a la media en BT; redes menos urbanas que la media del Grupo en MT y en BT; y una relación redes subterráneas / redes urbanas menor a la media del Grupo en MT, y mayor en BT.
9	13	Puebla Oriente	Alta intensidad del consumo en MT; alto factor de utilización en MT; alta densidad de la demanda tanto en MT como en BT; redes urbanas mayores a la media del Grupo en MT y en BT; y una relación redes subterráneas / redes urbanas menor a la media del Grupo en MT y similar a la media en BT.
10	11	Victoria	Baja intensidad del consumo en MT; bajo factor de utilización en MT; baja densidad de la demanda tanto en MT como en BT; redes menos urbanas que la media del Grupo tanto en MT como en BT; y una relación redes subterráneas / redes urbanas menor a la media del Grupo en MT y en BT.
11	8	Guadalajara	Porcentaje de red subterránea sobre el total mayor a 10%. Zonas con alta intensidad del consumo, y muy urbanas.

El procedimiento de *clusters* aplicado para la selección de los SER permite determinar las zonas que serán analizadas en detalle como representativas de cada uno de los *clusters* definidos, a partir del análisis de indicadores determinados en base a información característica de las instalaciones de las etapas de red MT, transformación MT/BT y red BT.

Para cada uno de los SER seleccionadas, se requerirá información detallada que será utilizada para adaptar las instalaciones a la demanda. Este proceso se desarrollará para las instalaciones de MT (redes y equipos) y para el conjunto centros de transformación MT/BT y red BT.

Como resultado de las adaptaciones realizadas, se obtendrán las cantidades de instalaciones optimizadas que serán el resultado del análisis de la demanda a abastecer y el mínimo costo de inversión de las unidades constructivas disponibles.

A partir de la información requerida se determinarán relaciones que permitirán expandir o extrapolar los resultados al resto de las zonas pertenecientes al *cluster* de la zona representativa o SER.

SELECCIÓN DE SISTEMAS ELECTRICOS REPRESENTATIVOS DE DISTRIBUCIÓN

1. INTRODUCCIÓN

El presente documento tiene por objeto cumplir con lo indicado en los TDR, Tarea, 1.2.5 Integración de la información para el cálculo del Valor de Reposición a Nuevo de los activos de distribución, punto a). *El Consultor seleccionará una muestra de zonas de distribución que representen el espectro de dispersiones de carga de todo el país*.

En este marco, este informe presenta la metodología utilizada y los resultados alcanzados para la determinación de zonas de distribución, cada una de las cuales se denomina "Sistemas Eléctricos Representativos - SER" haciendo referencia a sistemas eléctricos que tienen características que pueden considerarse homogéneas.

Estos sistemas pueden caracterizarse con base en una variable o un conjunto de variables o indicadores que pueden tener en cuenta una o más de las siguientes características: densidad de clientes, intensidad de consumo eléctrico, orografía, tecnología usada en las redes, entre otras.

El presente informe contiene 5 incisos además de esta introducción. El inciso 2 presenta el enfoque metodológico utilizado para la selección de los SER. El Inciso 3 contiene los resultados de la determinación de los SER. El Inciso 4 incluye una propuesta de expansión de las instalaciones. En el inciso 5 se presentan las conclusiones.

Por último, se incluye dos anexos: el Anexo 1 contiene los resultados del análisis de la información recibida y utilizada en el presente estudio. El Anexo 2 se presenta el desarrollo teórico para el análisis de los conglomerados o *clusters*.

2. ENFOQUE METODOLÓGICO PARA LA SELECCIÓN DE LOS SER

La selección de la muestra de zonas de distribución se basó en el análisis de conglomerados o *clusters*, cuyo objeto es agrupar elementos en grupos lo más homogéneos posible en función de las similitudes entre ellos y sobre la base de las variables observadas. En este caso, los elementos son las 120 zonas pertenecientes a las 13 divisiones de la Comisión Federal de Electricidad (CFE).

Antes de iniciar un análisis de *cluster* deben tomarse tres decisiones: (i) selección de las variables relevantes para identificar a los grupos, (ii) elección de la medida de proximidad entre los grupos, y (iii) elección del criterio para agrupar individuos en conglomerados.

La selección de variables es decisiva para identificar adecuadamente a los grupos, de acuerdo con el objetivo del estudio. En el presente estudio se definieron una serie de indicadores que representan las características que se desean evaluar. Así, la elaboración de estos indicadores se basó en las características que definen la valorización de las instalaciones de la empresa. Asimismo, antes de definir los indicadores que se utilizaron como variables para la determinación de los *clusters*, se analizó la correlación entre los mismos a efectos de seleccionar aquellos que resulten independientes entre sí.

También se analizó la consistencia de la información recibida, realizándose un análisis estadístico y gráfico de las variables, así como una definición de rangos de razonabilidad de

los datos en base a la experiencia del Consultor. Este análisis previo permitió detectar la presencia de valores anormales (atípicos). Las zonas identificadas como atípicas no han sido consideradas en el análisis de cluster.

En el ANEXO 1 – ANALISIS DE LA INFORMACIÓN~~ANEXO 1 – ANALISIS DE LA INFORMACIÓN~~ se presenta un resumen del análisis de la información.

El algoritmo utilizado para el análisis de cluster fue el denominado *k-medias*, el cual permite dividir la muestra en *k clusters*, los cuales incluyen zonas relativamente homogéneas ente sí, basándose en las características seleccionadas. Este algoritmo es el más importante desde los puntos de vista conceptual y práctico, y resulta el más apropiado cuando la cantidad de elementos (zonas) es grande.

Así, el procedimiento de *k-medias* se puede describir con las siguientes etapas:

- Seleccionar *k* "centros" alrededor de los cuales se formarán los *clusters* (grupos) iniciales de los "puntos" más próximos a cada uno;
- Calcular las distancias euclídeas de cada punto con respecto a los k centros y asignar cada elemento al *cluster* cuyo centro esté más próximo;
- Definir un criterio de homogeneidad u optimalidad y comprobar si reasignando alguno de los elementos a otro *cluster* mejora el criterio y en caso afirmativo redefinir así los *clusters*;
- Cuando no es posible encontrar ningún punto que pueda ser reasignado para mejorar el criterio de optimalidad, el proceso ha finalizado.

El criterio de optimalidad que se utiliza en el algoritmo de *k-medias* es minimizar la suma de cuadrados dentro de los *clusters* para todas las variables, o lo que es equivalente, la suma ponderada de las varianzas de las variables en los *clusters*. Estas últimas, son una medida de la heterogeneidad de la clasificación y al minimizarlas obtendremos *clusters* más homogéneos. La interpretación geométrica es minimizar las distancias al cuadrado entre los puntos y el centro del *cluster*, o sea, la distancia euclídea. Ambos criterios son equivalentes.

Cabe aclarar que para la aplicación de este algoritmo es conveniente estandarizar los indicadores calculados, previo al análisis, para evitar que el resultado dependa de cambios irrelevantes en la escala de medida. La estandarización de las variables consiste en la transformación de cada uno de sus valores restándole la media y dividiendo por la desviación estándar (de manera tal que la media de la variable estandarizada tome valor 0 y su varianza 1).

En la aplicación del algoritmo de las *k-medias* hay que fijar *a priori* el número de *clusters* k, pero puede llegar a determinarse mediante pruebas sucesivas el número de grupos más recomendable para el análisis, de acuerdo a cierto criterio numérico de optimalidad.

Para seleccionar el número óptimo de grupos, en este estudio se utilizó el criterio propuesto por Calinski y Harabasz, que consiste en seleccionar el valor de *k* que maximiza la razón entre la variabilidad entre los *clusters* y la variabilidad dentro de los *clusters* (es decir, lo que busca es que los elementos clasificados dentro de un *cluster* sean lo más similares entre sí dentro de lo posible, sobre la base de las variables observadas, y asimismo que los elementos de uno a otro *cluster* sean lo más diferentes posibles).

Es recomendable observar un equilibrio práctico entre la cantidad de *clusters* y la exactitud en los cálculos para la valorización de las redes. En efecto, a mayor número de sistemas eléctricos representativos analizados, posiblemente sean más exactos los valores calculados, dado que es posible reflejar más detalladamente las variaciones en los indicadores; pero en el extremo, esto es igual a hacer un cálculo detallado por todas y cada

una de las zonas, perdiendo una de las grandes ventajas comparativas de este método que reside en lograr una economía de recursos y tiempos de análisis sin perder significativamente precisión.

En el ANEXO 2 – ANALISIS DE CONGLOMERADOS se presenta una descripción teórica más detallada del algoritmo de las *k-medias.*

3. DETERMINACIÓN DE LAS ZONAS REPRESENTATIVAS

3.1. Cálculo de indicadores y análisis de la muestra

Para el análisis de *clusters* realizado con información de 120 zonas se calcularon indicadores a partir de las siguientes variables:

- N-MT: es el número de clientes en MT -conformado por todos los clientes MT más los centros de transformación MT/BT considerando que cada uno representa un punto de demanda puntual en la red de MT-.
- N-BT: es el número de clientes en BT
- E-MT: es la energía eléctrica total vendida (MT+BT), en MWh.
- E-BT: es la energía vendida en BT, en MWh.
- P: es la potencia máxima anual registrada en los alimentadores de la red MT, en MW.
- Pn: es la potencia nominal instalada en los transformadores de clientes MT y en los centros de transformación MT/BT, en MVA.
- L-MT: es la longitud de la red MT, en km.
- L-BT: es la longitud de la red BT, en km.
- Lu-MT: es la longitud de la red MT urbana, en km.
- Lu_BT: es la longitud de la red BT urbana, en km.
- Ls-MT: es la longitud de la red MT subterránea, en km.
- Ls-BT: es la longitud de la red BT subterránea, en km.

Con esta información, se estimaron indicadores en MT y BT, con el objeto de medir la intensidad del consumo, estimar la densidad lineal de la demanda, calcular el factor de uso de las instalaciones; y obtener indicadores de estructura (por ej., si la red es mayormente urbana o rural, aérea o subterránea).

Así, se calcularon los siguientes indicadores:

(i) Intensidad del Consumo:

- E-MT/N-MT, para medir la intensidad del consumo por punto de entrega en MT.
- E-BT/N-BT, para medir la intensidad del consumo por punto de entrega en BT.

(ii) Densidad Lineal de la Demanda:

- Pn/L-MT, para medir la densidad lineal de la demanda de la red MT.

- Pn/L-BT, para medir la densidad lineal de la demanda de la red BT.
- P/L-MT, para medir la densidad lineal de la demanda de la red MT.
- E-MT/L-MT, para medir la densidad lineal de la demanda de la red MT.
- E-BT/L-BT, para medir la densidad lineal de la demanda de la red BT.
- N-BT/L-BT, para medir la densidad lineal de la demanda de la red BT.
- P/L-MT, para medir la densidad lineal de la demanda de la red MT.

(iii) Factor de Uso:

- E-MT/P, como indicador del factor de uso de las instalaciones MT.
- Pn/N-BT, es la potencia nominal por cliente en BT

(iv) Estructura:

- Lu-MT/L-MT, es el porcentaje de red urbana sobre el total de red MT.
- Lu-BT/L-BT, es el porcentaje de red urbana sobre el total de red BT.
- Ls-MT/L-MT, es el porcentaje de red subterránea sobre el total de red MT.
- Ls-BT/L-BT, es el porcentaje de red subterránea sobre el total de red BT.
- Ls-MT/Lu-MT, es la relación entre la red subterránea MT y la red urbana MT, y permite determinar si la red existente en el área urbana es más o menos costosa.
- Ls-BT/Lu-BT, es la relación entre la red subterránea BT y la red urbana BT.

Posteriormente, a los efectos del análisis de *clusters* y sobre la base de la matriz de correlación entre las distintas variables, se seleccionaron algunos de los indicadores arriba incluidos.

Específicamente, las variables utilizadas para el análisis de *cluster* fueron:

- Variables en MT: E-MT/N-MT, E-MT/P, P/L-MT, Lu-MT/L-MT, y Ls-MT/Lu-MT.
- Variables en BT: E-BT/L-BT, Lu-BT/L-BT, Ls-BT/Lu-BT y N-BT/L-BT.

Dado que la configuración de la red se refleja necesariamente en los costos -las redes urbanas pueden costar el doble que las redes rurales, y las redes subterráneas pueden ser hasta 10 veces más costosas que las redes áreas urbanas-, se realizó una primera clasificación de las zonas considerando esta primera diferenciación por costos, de la cual surgieron tres grandes grupos, separados según el porcentaje de red subterránea en MT respecto al total de la red MT (Ls-MT/L-MT):

iv. Grupo 1: 0.0% ≤ Ls-MT/L-MT < 0.5%

v. Grupo 2: 0.5% ≤ Ls-MT/L-MT < 10.0%

vi. Grupo 3: Ls-MT/L-MT ≥ 10.0%

Así, el Grupo 1 queda conformado por 69 zonas, mientras el Grupo 2 por 42 zonas.

El Grupo 3, por su parte, queda conformado por 9 zonas (Tijuana, La Paz, Metropolitana Norte, Veracruz, Coatzacoalcos, Cancún, Rivera Maya, Guadalajara, Vallarta), y, como se explicará con mayor detalle más adelante, forma un *cluster* en si mismo.

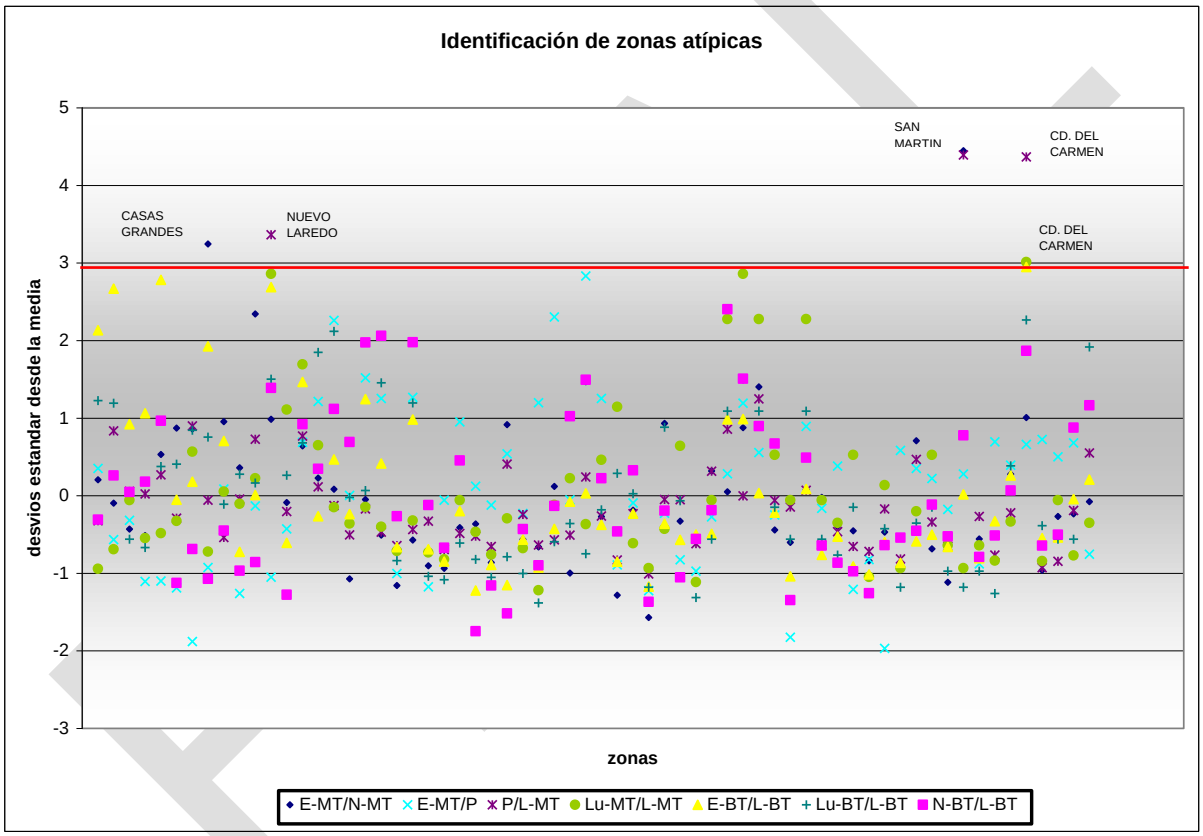
Para los grupos 1 y 2 se realizó un análisis de *clusters* según la metodología descripta en el capítulo 2.

Para ambos grupos, antes de realizar el análisis de *cluster* se realizó una verificación previa a los efectos de detectar zonas con valores atípicos, las cuales deben ser tratadas como zonas atípicas (casos particulares), eliminándolas del análisis de *cluster* y estudiándolas individualmente.

Así, se estimó para cada grupo el valor medio y la desviación estándar, y se consideró el criterio de identificar como zonas atípicas aquellas con una desviación estándar dentro de la muestra mayor a 3.

En la figura siguiente se grafican las variables seleccionadas para el análisis de *clusters* de las 69 zonas que conforman el Grupo 1:

FIGURA 1 IDENTIFICACIÓN DE ZONAS ATÍPICAS DEL GRUPO 1



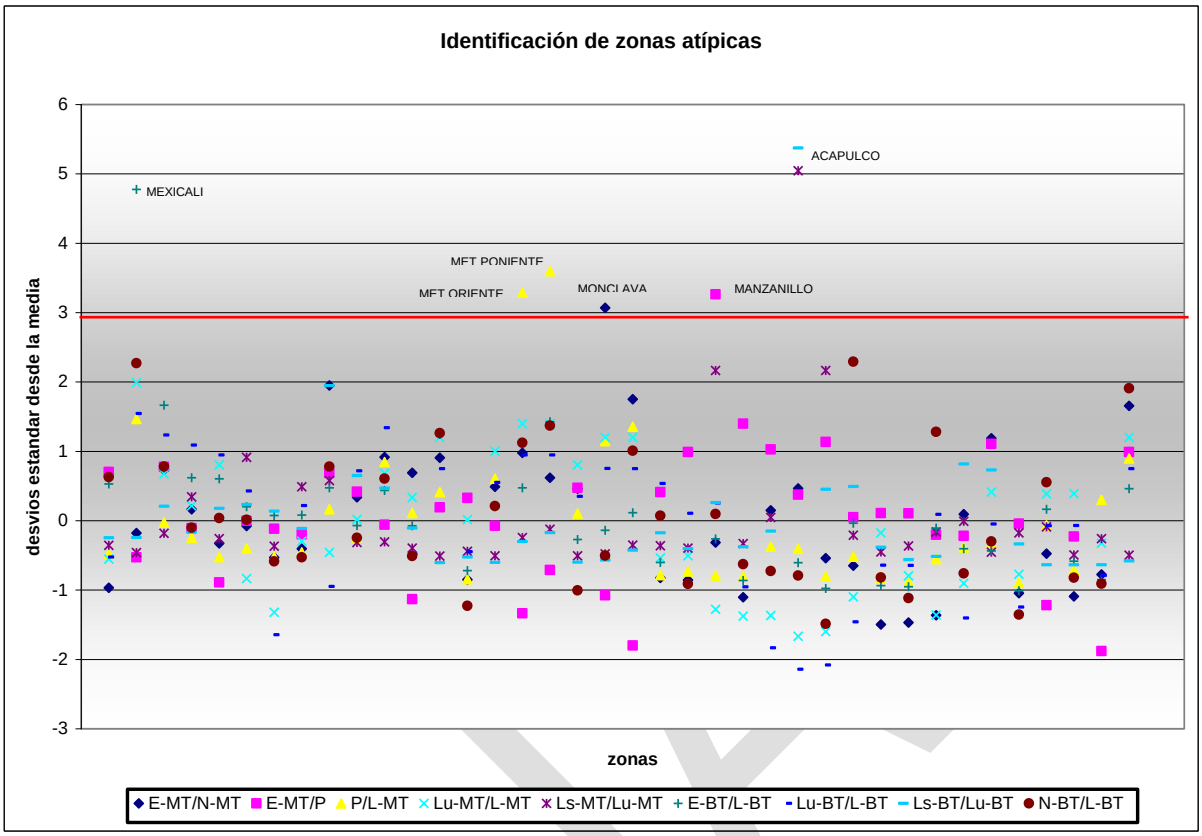
Fuente: elaboración propia sobre la base de datos de CFE

Dado el criterio expuesto, fueron eliminadas de la muestra y serán tratadas como zonas atípicas las siguientes 4: Casas Grandes, Nuevo Laredo, San Martín y CD del Carmen.

Considerando la exclusión de las cuatro zonas consideradas atípicas, el Grupo 1 queda formado por 65 zonas.

En la figura siguiente se grafican las variables seleccionadas para el análisis de *cluster* de las 42 zonas que conforman el Grupo 2:

FIGURA 2 IDENTIFICACIÓN DE ZONAS ATÍPICAS DEL GRUPO 2



Fuente: elaboración propia sobre la base de datos de CFE

Dado el criterio expuesto, fueron eliminadas de la muestra y serán tratadas como zonas atípicas las siguientes 6: Mexicali, Metropolitana Oriente, Metropolitana Poniente, Monclova, Manzanillo y Acapulco. Adicionalmente, a las zonas de Nogales y Zihuatanejo se les dio tratamiento de zonas atípicas. En el caso de Zihuatanejo, se observa que la relación entre la red subterránea y urbana en MT, es superior al 50%, lo que claramente la diferencia del resto (excepto Acapulco y Manzanillo, que ya fueron identificadas como atípicas). En el caso de Nogales, la variable que la diferencia sustancialmente del resto (excepto Acapulco) es la relación entre la red subterránea y urbana en BT.

De esta forma, el Grupo 2 queda conformado por 34 zonas.

3.2. Análisis de *clusters* y resumen de resultados

3.2.1. GRUPO 1

Como ya se mencionó, para el análisis de *cluster*s del denominado Grupo 1 se consideraron los siguientes indicadores: E-MT/N-MT, E-MT/P, P/L-MT, Lu-MT/L-MT, E-BT/L-BT, Lu-BT/L-BT, N-BT/L-BT; y 65 zonas (recuérdese que 4 de las 69 zonas fueron consideradas zonas atípicas).

Se hicieron sucesivas aplicaciones del algoritmo *k-medias* considerando de 4 a 10 *clusters* o conglomerados. A partir de los resultados obtenidos en cada caso, y utilizando el criterio propuesto por Calinski y Harabasz, se seleccionó el número *k* finalmente adoptado. En este caso, se obtuvieron originalmente 6 *clusters.*

En la tabla siguiente se muestra el indicador de Calinsky y Harabasz resultante de la aplicación del algoritmo de *k-medias* considerando k de 4 a 10. Allí es posible identificar que el número óptimo de *k* es *k* =6, aquel que tiene el indicador más alto:

TABLA 1 GRUPO 1 NÚMERO DE CLUSTERS ÓPTIMO

Número de *clusters [k]*	Calinsky y Harabasz
4	19.67
5	19.34
6	**19.80**
7	14.99
8	16.85
9	14.82
10	15.39

En la siguiente tabla se muestran los principales indicadores de cada cluster (*k* =6):

TABLA 2 CLUSTERS # 1 A 6 (GRUPO 1) PRINCIPALES MEDIDAS DE LOS *CLUSTERS*

Variables	E-MT/N-MT	E-MT/P	P/L-MT	Lu-MT/L-MT	E-BT/L-BT	Lu-BT/L-BT	N-BT/L-BT
				Cluster 1			
N	12	12	12	12	12	12	12
media	104.742	3896.757	0.0569167	0.2236917	154.2457	0.403075	64.40283
mínimo	63.1589	2504.73	0.0282	0.1367	37.19	0.1452	33.4224
máximo	175.2195	4860.711	0.089	0.4	316.7463	0.9088	92.1018
rango	112.0606	2355.981	0.0608	0.2633	279.5563	0.7636	58.6794
cv	0.2753682	0.1777512	0.3174915	0.3590871	0.4938457	0.4988728	0.2907162
				Cluster 2			
N	22	22	22	22	22	22	22
media	46.40986	3923.799	0.0265091	0.1152773	96.05607	0.1248727	66.5432
mínimo	11.1901	2418.484	0.0084	0.0011	26.5574	0.0011	23.3464
máximo	77.3677	5504.801	0.0448	0.406	161.1927	0.406	95.022
rango	66.1776	3086.317	0.0364	0.4049	134.6353	0.4049	71.6756
cv	0.3465676	0.2097483	0.3827924	0.8599983	0.3792978	0.8191905	0.2654471
				Cluster 3			
N	5	5	5	5	5	5	5
media	102.5637	5041.117	0.07584	0.6	317.6219	0.60024	155.1183
mínimo	79.1078	4612.341	0.0509	0.5	215.446	0.5012	121.9332
máximo	135.8367	5498.857	0.1037	0.7	431.2506	0.7	206.1978
rango	56.7289	886.5156	0.0528	0.2	215.8046	0.1988	84.2646
cv	0.2329868	0.0664078	0.2997578	0.1178511	0.2970656	0.1170984	0.2109123
				Cluster 4			
N	12	12	12	12	12	12	12
media	60.17503	4838.033	0.0336417	0.1870333	173.9397	0.2608417	114.2988
mínimo	32.0084	4093.16	0.0153	0.0342	129.3791	0.1079	78.1483
máximo	88.6259	6581.646	0.0485	0.3	249.6499	0.4293	145.4705
rango	56.6175	2488.485	0.0332	0.2658	120.2708	0.3214	67.3222
cv	0.2807276	0.1446202	0.292163	0.4694676	0.191091	0.3367801	0.1740705
				Cluster 5			
N	7	7	7	7	7	7	7
media	74.05081	3626.772	0.0626429	0.1456429	420.8368	0.4628	120.8901
mínimo	55.3792	2776.706	0.0373	0.0486	212.138	0.174	86.7589
máximo	99.2717	4679.986	0.0863	0.2864	628.9453	0.8006	151.638
rango	43.8925	1903.281	0.049	0.2378	416.8073	0.6266	64.8791
cv	0.2027727	0.1701106	0.2645321	0.5342384	0.4063085	0.5068445	0.2036973
				Cluster 6			
N	7	7	7	7	7	7	7
media	81.00307	5874.286	0.0454	0.2055571	280.2575	0.5818	165.8909
mínimo	53.0047	5020.227	0.0312	0.1414	170.588	0.1545	115.6098
máximo	138.5635	7097.247	0.0612	0.3211	397.7268	0.8498	191.0625
rango	85.55881	2077.02	0.03	0.1797	227.1388	0.6953	75.4527
cv	0.3497199	0.1201763	0.2435789	0.3673187	0.2778467	0.4224936	0.1624074

Fuente: elaboración propia sobre la base de datos de CFE
Nota: N es el número de elementos (zonas) de cada grupo, media es el valor medio o promedio de los N elementos del grupo, mínimo es el valor mínimo, máximo es el valor máximo, rango es la diferencia entre el valor máximo y el mínimo, y cv es el coeficiente de variación (igual al desvío estándar / media).

Respecto a la conformación interna del Grupo 1 se puede observar que, en términos relativos dentro de Grupo:

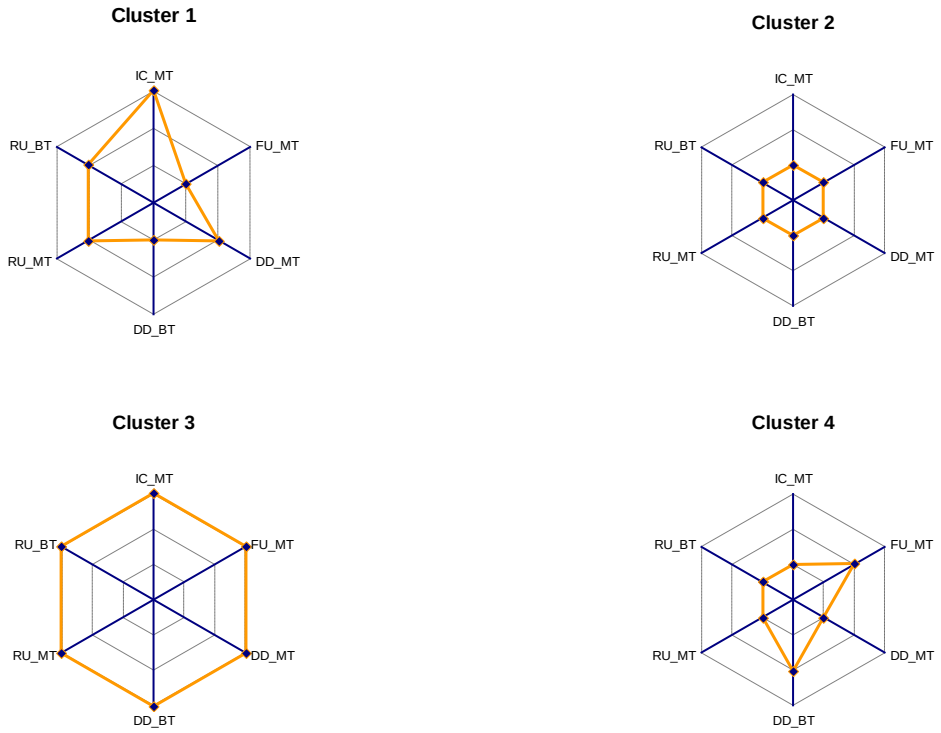
- El *cluster* 1 está formado por 12 zonas, con alta intensidad del consumo en MT; bajo factor de utilización en MT; una densidad de la demanda similar a la media del Grupo

en MT, y por debajo de la media en BT; y redes urbanas en cantidades similares a la media del Grupo tanto en MT como en BT.

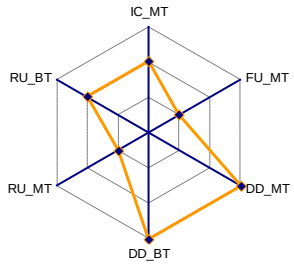
- El *cluster* 2 está también formado por 22 zonas, con baja intensidad del consumo en MT; bajo factor de utilización en MT; baja densidad de la demanda tanto en MT como en BT; y redes menos urbanas que la media del Grupo tanto en MT como en BT.
- El *cluster* 3 está formado por 5 zonas, con alta intensidad del consumo en MT; alto factor de utilización en MT; con una densidad de la demanda alta tanto en MT como en BT; y redes más urbanas que la media del Grupo, tanto en MT como en BT.
- El *cluster* 4 está formado por 12 zonas, con una intensidad del consumo en MT baja; un factor de utilización en MT similar a la media del Grupo; con una densidad de la demanda baja en MT y media en BT; y redes urbanas más bajas que la media del Grupo en MT y en BT.
- El *cluster* 5 está formado por 7 zonas, con una intensidad del consumo en MT media; bajo factor de utilización en MT; con una densidad de la demanda alta en MT y en BT; y redes menos urbanas que la media del Grupo en MT, y similares a la media en BT.
- El *cluster* 6 también está formado por 7 zonas, con una intensidad del consumo en MT alta; alto factor de utilización en MT; con una densidad de la demanda media en MT y alta en BT; y redes menos urbanas que la media del Grupo 1 en MT, y más urbanas que la media en BT.

En la figura siguiente se pueden apreciar gráficamente las diferencias entre los distintos *clusters* del Grupo 1. La línea más cercana al centro representa “baja”, la del medio “media” y la última “alta”:

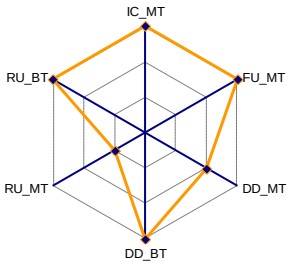
FIGURA 3 C*LUSTERS* # 1 A 6 (GRUPO 1) DIFERENCIAS ENTRE *CLUSTERS*



Cluster 5



Cluster 6



Nota: IC_MT es la intensidad del consumo en MT; FU_MT es el factor de uso en MT, DD_MT es la densidad de la demanda en MT, RU_MT y RU_BT son redes urbanas en MT y BT respectivamente.

3.2.2. GRUPO 2

Para el análisis de *clusters* del denominado Grupo 2 se consideraron las mismas variables consideradas para el caso del Grupo 1: E-MT/N-MT, E-MT/P, P/L-MT, Lu-MT/L-MT, E-BT/L-BT, Lu-BT/L-BT, N-BT/L-BT; más las variables Ls-MT/Lu-MT y Ls-BT/Lu-BT; y 34 zonas.

Las variables Ls-MT/Lu-Mt y Ls-BT/Lu-BT se incluyeron en este grupo, el cual posee redes subterráneas significativas, a los efectos de considerar el sobrecosto que esto genera, y las diferencias en costos existentes en cada zona.

Al igual que en el caso del Grupo 1 se hicieron sucesivos procesos *k-medias* considerando de 4 a 10 *cluster*s o conglomerados, utilizándose luego el criterio propuesto por Calinski y Harabasz para seleccionar el número apropiado del *clusters*.

En este caso se obtuvieron 4 *clusters*. En la tabla siguiente se muestra el indicador de Calinsky y Harabasz resultante de la aplicación del algoritmo de *k-medias* considerando k de 4 a 10. Allí es posible identificar que el número óptimo de *k* es *k*=4, aquel que tiene el indicador más alto:

TABLA 3 GRUPO 2 NÚMERO DE CLUSTERS ÓPTIMO

Número de *clusters [k]*	Calinsky y Harabasz
4	**9.37**
5	9.18
6	7.26
7	7.10
8	8.02
9	8.19
10	7.35

En la siguiente tabla se muestran los principales indicadores de cada cluster (*k* =5):

TABLA 4 CLUSTERS # 7 A 10 (GRUPO 2) PRINCIPALES MEDIDAS DE LOS *CLUSTERS*

Variables	E-MT/N-MT	E-MT/P	P/L-MT	Lu-MT/L-MT	Ls-MT/Lu-MT	E-BT/L-BT	Lu-BT/L-BT	Ls-BT/Lu-BT	N-BT/L-BT
Cluster 7									
N	3	3	3	3	3	3	3	3	3
media	102.7253	4041.586	0.103	0.4185	0.2209	457.7415	0.7574667	0.1096667	126.2138
mínimo	86.7281	3958.807	0.0897	0.2847	0.1743	396.6772	0.6669	0.0833	108.1287
máximo	117.2254	4127.564	0.1207	0.55	0.2859	546.6497	0.886	0.1534	138.4591
rango	30.4973	168.7576	0.031	0.2653	0.1116	149.9725	0.2191	0.0701	30.33041
cv	0.1489766	0.0208889	0.1549662	0.3170011	0.2627011	0.1720797	0.1510006	0.3477864	0.1266603
Cluster 8									
N	7	7	7	7	7	7	7	7	7
media	74.85861	4457.931	0.0806429	0.2075143	0.0676857	312.4807	0.3350429	0.1164429	149.8229
mínimo	35.2394	3940.188	0.039	0.1478	0.0344	132.3221	0.152	0.0212	95.0026
máximo	116.3844	5343.745	0.1022	0.356	0.116	521.8781	0.6348	0.2571	266.7317
rango	81.145	1403.557	0.0632	0.2082	0.0816	389.556	0.4828	0.2359	171.7291
cv	0.4151461	0.1283144	0.2603394	0.3828114	0.4875403	0.4321996	0.5137932	0.738403	0.4550463
Cluster 9									
N	13	13	13	13	13	13	13	13	13
media	143.8506	4154.854	0.1951538	0.7104385	0.0274	443.0031	0.7847231	0.0891846	155.4497
mínimo	90.814	2570.771	0.0799	0.5004	0.0063	251.8613	0.4911	0.0053	81.1713
máximo	202.2047	5090.272	0.3553	0.8695	0.0705	839.3676	0.9486	0.2416	245.1644
rango	111.3907	2519.501	0.2754	0.3691	0.0642	587.5063	0.4575	0.2363	163.9931
cv	0.2376204	0.1881572	0.4066192	0.1498356	0.7832798	0.4013667	0.1667947	1.018443	0.3120882
Cluster 10									
N	11	11	11	11	11	11	11	11	11
media	59.84199	3777.901	0.0729636	0.4248727	0.0394	176.9356	0.5627545	0.0305182	97.84987
mínimo	28.2072	2500.008	0.0258	0.2948	0.0093	90.6444	0.3	0	61.6018
máximo	88.2028	4989.503	0.2007	0.5945	0.0892	419.3283	0.8	0.0813	168.8811
rango	59.9956	2489.496	0.1749	0.2997	0.0799	328.6839	0.5	0.0813	107.2793
cv	0.3172407	0.222981	0.859624	0.2467405	0.6343909	0.5166732	0.273818	0.9981721	0.3556516

Fuente: elaboración propia sobre la base de datos de CFE

Respecto a la conformación interna del Grupo 2, los distintos *clusters* se pueden describir en términos cualitativos:

- El *cluster* 7 está formado por 3 zonas, con alta intensidad del consumo en MT; un factor de utilización en MT similar a la media del grupo; una densidad de la demanda similar a la media del grupo en MT, y alto en BT; redes urbanas similares a la media del Grupo en MT, y mayores en BT; y una relación redes subterráneas / redes urbanas mayor a la media del Grupo en MT y en BT.

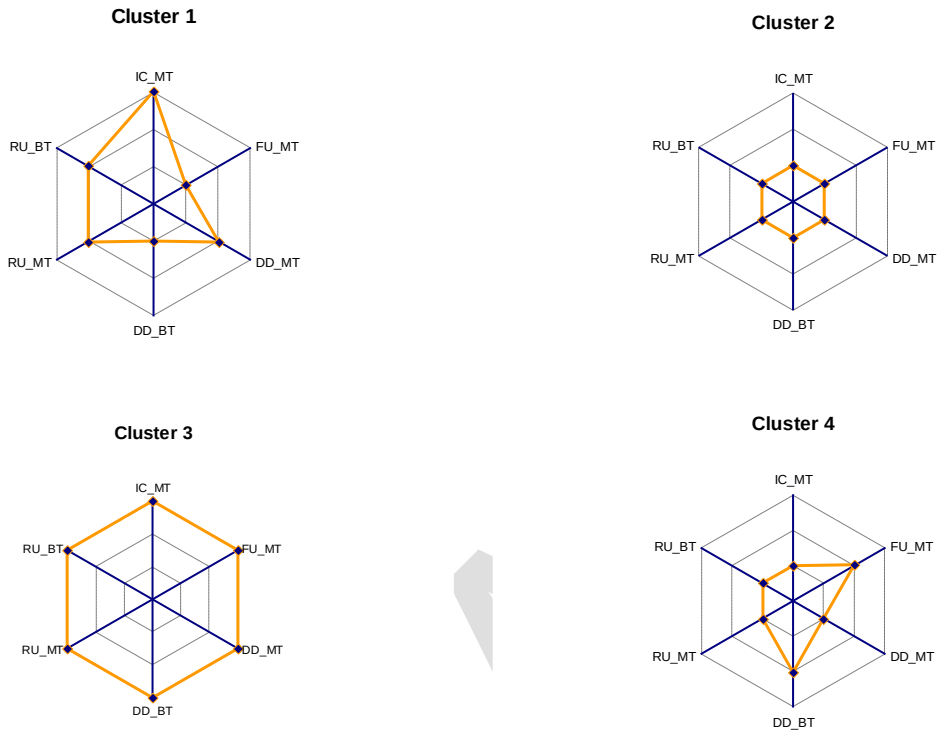
- El *cluster* 8 está formado por 7 zonas, con baja intensidad del consumo en MT; alto factor de utilización en MT; baja densidad de la demanda en MT, y similar a la media en BT; redes menos urbanas que la media del Grupo en MT y en BT; y una relación redes subterráneas / redes urbanas menor a la media del Grupo en MT, y mayor en BT.

- El *cluster* 9 está formado por 13 zonas, con alta intensidad del consumo en MT; alto factor de utilización en MT; alta densidad de la demanda tanto en MT como en BT; redes urbanas mayores a la media del Grupo en MT y en BT; y una relación redes subterráneas / redes urbanas menor a la media del Grupo en MT y similar a la media en BT.

- El *cluster* 10 está formado por 11 zonas, con baja intensidad del consumo en MT; bajo factor de utilización en MT; baja densidad de la demanda tanto en MT como en BT; redes menos urbanas que la media del Grupo tanto en MT como en BT; y una relación redes subterráneas / redes urbanas menor a la media del Grupo en MT y en BT.

En la figura siguiente se pueden apreciar gráficamente las diferencias entre los distintos *clusters* del Grupo 2:

FIGURA 4 CLUSTERS # 7 A 10 (GRUPO 2) DIFERENCIAS ENTRE *CLUSTERS*



Nota: IC_MT es la intensidad del consumo en MT; FU_MT es el factor de uso en MT, DD_MT es la densidad de la demanda en MT, RU_MT y RU_BT son redes urbanas en MT y BT respectivamente, y SU_MT y SU_BT es la relación red subterránea / red urbana en MT y BT respectivamente.

3.3. Propuesta de zonas representativas

3.3.1. GRUPOS 1 Y 2

En las dos tablas siguientes se presenta la identificación de las zonas que integran los seis *clusters* del Grupo 1 y los cuatro *clusters* del Grupo 2, respectivamente; así como la caracterización de cada zona en términos de:

- distancia del centro del cluster,
- cantidad de clientes en MT y BT
- energía vendida en MT y BT
- SER: zona representativa propuesta. La misma debe ser una zona que cumpla con los siguientes criterios: (i) que su proximidad (distancia) del centro del cluster sea mínima; y (ii) que sea una zona para la cual se posea información completa.

  En este marco, cabe destacar que la elección de la zona representativa se realizó en conjunto con personal de CFE, que posee un conocimiento profundo de las distintas zonas. Así, en primer lugar, MEC entregó a CFE los resultados del análisis, incluyendo la información sobre las distancias de cada zona al centro del *cluster* en el que la misma quedó agrupada. Sobre la base de dicha información, CFE procedió a hacer su propuesta de zona representativa (SER). MEC analizó dicha propuesta y conformó sobre la base de la misma los SER de cada cluster, considerando que la propuesta de CFE respondía a los criterios expuestos en el párrafo precedente.

TABLA 5 CLUSTERS # 1 A 2 (GRUPO 1) RESULTADOS DEL ANÁLISIS DEL *CLUSTERS* Y CARACTERÍSTICAS DE LOS SER

Cluster #	Zona	División	Distancias	Cant. de clientes [dic.2007]		Energía anual vendida [MWh]		SER
				MT	BT	MT	BT	
	DURANGO	**NORTE**	**1.18**	**2,201**	**275,658**	**394,357**	**486,360**	
	TEHUANTEPEC	SURESTE	1.47	411	180,885	79,445	307,472	
	SAN JUAN DEL RIO	BAJIO	1.60	5,117	163,905	418,235	243,154	
	TLAXCALA	CENTRO ORIENTE	1.80	1,678	325,386	585,429	494,139	
	CUAUHTÉMOC	NORTE	1.83	1,778	127,552	229,146	509,067	
1	CHONTALPA	SURESTE	2.77	3,715	227,314	186,548	526,529	**DURANGO**
	CERRALVO	GOLFO NORTE	3.20	487	48,093	92,667	129,844	
	VALLE DE BRAVO	CENTRO SUR	4.11	1,835	165,866	76,358	184,563	
	PARRAL	NORTE	4.15	905	94,269	188,716	371,010	
	DELICIAS	NORTE	5.45	1,417	100,735	240,122	341,901	
	MINAS	JALISCO	5.92	835	103,058	143,930	156,214	
	GÓMEZ PALACIO	NORTE	6.98	2,916	156,135	865,298	528,924	
	ALTAMIRANO	CENTRO SUR	0.42	1,642	122,459	76,118	189,273	
	CHILPANCINGO	CENTRO SUR	0.43	1,661	222,106	96,305	260,159	
	MATAMOROS DE IZUCARD	CENTRO ORIENTE	0.46	1,793	170,335	91,824	244,952	
	TECAMACHALCO	CENTRO ORIENTE	0.67	2,358	213,127	301,512	290,085	
	Poza Rica	**ORIENTE**	**0.74**	**2,067**	**358,158**	**261,285**	**552,318**	
	PATZCUARO	CENTRO OCCIDENTE	1.00	754	106,365	41,495	133,132	
	MANTE	GOLFO CENTRO	1.04	1,026	81,286	106,420	171,302	
	RIOVERDE	GOLFO CENTRO	1.08	1,214	123,674	101,860	159,506	
	LOS RIOS	SURESTE	1.12	393	151,465	76,341	270,390	
2	ZACAPU	CENTRO OCCIDENTE	1.38	728	99,298	51,302	111,434	**Poza Rica**
	FRESNILLO	BAJIO	1.45	5,707	202,209	465,240	268,956	
	HUEJUTLA	GOLFO CENTRO	1.64	532	171,139	84,487	180,459	
	TOLUCA	CENTRO SUR	1.87	2,469	240,225	409,012	273,266	
	TICUL	PENINSULAR	1.98	228	68,914	35,159	143,044	
	SANTIAGO	JALISCO	2.12	574	96,945	55,998	189,954	
	TIZIMIN	PENINSULAR	2.32	1,365	78,469	53,179	139,937	
	MATEHUALA	GOLFO CENTRO	2.93	1,286	66,729	151,366	85,592	
	Teziutlán	ORIENTE	3.51	1,897	276,097	175,036	324,313	
	HUAJUAPAN	SURESTE	3.58	926	206,936	24,700	177,695	
	IXMIQUILPAN	BAJIO	3.58	795	97,135	102,975	128,754	
	VALLES	GOLFO CENTRO	4.01	1,001	158,319	156,740	264,076	
	SAN CRISTOBAL	SURESTE	6.37	487	351,789	74,346	361,988	

Fuente: elaboración propia sobre la base de datos de CFE

TABLA 5 CONTINUACIÓN *CLUSTERS* # 3 A 6 (GRUPO 1) RESULTADOS DEL ANÁLISIS DEL *CLUSTERS* Y CARACTERÍSTICAS DE LOS SER

Cluster #	Zona	División	Distancias	Cant. de clientes [dic.2007]		Energía anual vendida [MWh]		SER
				MT	BT	MT	BT	
3	**SABINAS**	**GOLFO NORTE**	**1.64**	**681**	**59,028**	**270,006**	**180,694**	SABINAS
	LEON	BAJIO	2.54	8,928	402,681	1,513,873	867,051	
	IRAPUATO	BAJIO	2.58	14,092	346,242	1,149,166	602,544	
	AGUASCALIENTES	BAJIO	2.76	15,356	388,092	1,354,340	711,553	
	CELAYA	BAJIO	3.68	10,575	358,084	1,118,325	551,669	
4	IGUALA	CENTRO SUR	0.34	729	118,448	95,997	177,692	TUXTLA
	TUXTLA	**SURESTE**	**1.00**	**1,267**	**386,849**	**407,708**	**591,833**	
	TEHUACAN	CENTRO ORIENTE	1.14	868	160,845	181,284	227,954	
	Córdoba	ORIENTE	1.16	1,324	167,237	232,497	233,672	
	ALTOS	JALISCO	1.26	2,722	151,608	390,372	222,543	
	ZITACUARO	CENTRO OCCIDENTE	1.44	517	109,532	36,078	146,177	
	MOTUL	PENINSULAR	1.69	284	72,953	58,350	120,778	
	Los Tuxtlas	ORIENTE	1.76	460	98,685	40,543	135,005	
	SALVATIERRA	BAJIO	1.79	5,574	224,290	311,524	307,122	
	CAMPECHE	PENINSULAR	2.01	969	158,299	215,960	382,619	
	COSTA	JALISCO	2.01	449	78,035	49,027	114,719	
	Papaloapan	ORIENTE	4.08	2,257	248,456	165,244	384,146	
5	**NAVOJOA**	**NOROESTE**	**2.51**	**2,027**	**89,275**	**330,359**	**304,293**	NAVOJOA
	GUASAVE	NOROESTE	2.60	2,050	125,975	174,428	430,804	
	CABORCA	NOROESTE	3.69	1,537	34,625	327,529	152,408	
	SAN LUIS	BAJA CALIFORNIA	4.11	2,610	133,665	479,179	731,685	
	ZAPOTLÁN	JALISCO	4.84	1,353	110,830	191,422	165,107	
	CIÉNEGA	JALISCO	4.85	1,471	88,966	168,436	141,846	
	CONSTITUCION	BAJA CALIFORNIA	5.03	1,238	44,105	67,268	270,159	
6	**TEPIC**	**JALISCO**	**1.25**	**2,472**	**190,917**	**199,641**	**309,912**	TEPIC
	COLIMA	CENTRO OCCIDENTE	1.96	2,204	183,999	321,393	345,647	
	LA PIEDAD	CENTRO OCCIDENTE	1.98	2,964	151,333	267,251	216,166	
	APATZINGAN	CENTRO OCCIDENTE	2.06	1,637	114,865	142,650	219,479	
	LAZARO CARDENAS	CENTRO OCCIDENTE	2.12	495	67,733	77,232	143,786	
	MORELIA	CENTRO OCCIDENTE	3.85	2,599	338,968	476,687	500,164	
	Orizaba	ORIENTE	9.41	817	141,713	211,085	183,501	

Fuente: elaboración propia sobre la base de datos de CFE

TABLA 6 CLUSTERS # 7 A 10 (GRUPO 2) RESULTADOS DEL ANÁLISIS DEL *CLUSTER* Y CARACTERÍSTICAS DE LOS SER

Cluster #	Zona	División	Distancia	Cant. de clientes [dic.2007]		Energía anual vendida [MWh]		
				MT	BT	MT	BT	
7	**CULIACAN**	**NOROESTE**	**0.70**	**5,710**	**343,009**	**843,950**	**1,258,351**	**CULIACAN**
	GUAYMAS	NOROESTE	1.80	1,083	63,905	300,954	264,541	
	MAZATLAN	NOROESTE	1.97	1,408	211,485	498,672	656,634	
8	LOS MOCHIS	NOROESTE	2.07	2,114	176,195	369,549	664,167	**ENSENADA**
	ENSENADA	**BAJA CALIFORNIA**	**3.40**	**3,160**	**149,214**	**322,961**	**449,744**	
	MORELOS	CENTRO SUR	4.60	2,706	426,985	560,027	753,615	
	JIQUILPAN	CENTRO OCCIDENTE	4.91	720	88,170	57,149	113,851	
	TAPACHULA	SURESTE	5.91	4,630	274,380	199,180	449,155	
	VILLAHERMOSA	SURESTE	6.31	3,561	295,040	663,671	807,946	
	Xalapa	ORIENTE	6.51	1,549	270,630	257,201	369,075	
9	MATAMOROS	GOLFO NORTE	1.79	3,421	201,440	911,165	578,569	**PUEBLA ORIENTE**
	PUEBLA ORIENTE	**CENTRO ORIENTE**	**3.36**	**1,264**	**353,585**	**670,245**	**553,304**	
	JUÁREZ	NORTE	3.91	5,557	412,351	2,377,720	1,190,392	
	PIEDRAS NEGRAS	GOLFO NORTE	4.46	1,676	118,916	569,731	436,388	
	TORREÓN	NORTE	4.97	4,519	276,273	1,120,824	894,984	
	REYNOSA	GOLFO NORTE	5.12	3,431	245,523	1,456,641	848,919	
	OBREGÓN	NOROESTE	5.57	2,546	152,432	510,494	591,018	
	CHIHUAHUA	NORTE	5.97	4,290	300,873	1,180,009	858,689	
	MERIDA	PENINSULAR	6.28	4,660	371,691	1,044,523	970,398	
	HERMOSILLO	NOROESTE	7.69	6,972	266,651	1,339,090	1,232,192	
	PUEBLA PONIENTE	CENTRO ORIENTE	8.18	2,204	302,778	1,062,719	621,215	
	QUERETARO	BAJIO	9.47	5,885	320,976	1,877,139	668,436	
	SALTILLO	GOLFO NORTE	11.04	3,133	240,692	953,830	501,657	
10	**VICTORIA**	**GOLFO CENTRO**	**1.23**	**2,938**	**159,146**	**306,700**	**365,645**	**VICTORIA**
	MONTEMORELOS-LINARES	GOLFO NORTE	1.59	1,606	118,593	198,636	297,018	
	OAXACA	SURESTE	1.61	6,424	409,243	184,073	498,372	
	CHETUMAL	PENINSULAR	1.96	838	94,311	125,556	226,838	
	HUATULCO	SURESTE	2.07	2,903	142,260	77,327	203,884	
	URUAPAN	CENTRO OCCIDENTE	2.97	1,091	139,074	132,915	201,459	
	ZAMORA	CENTRO OCCIDENTE	3.09	1,396	177,278	188,559	240,538	
	ZACATECAS	BAJIO	3.31	5,240	248,059	408,920	365,008	
	SAN LUIS POTOSÍ	GOLFO CENTRO	6.07	11,545	373,959	1,023,114	585,502	
	CHAPALA	JALISCO	6.42	3,077	160,407	384,789	249,213	
	TAMPICO	GOLFO CENTRO	7.03	3,287	325,453	686,650	808,093	

Fuente: elaboración propia sobre la base de datos de CFE

3.3.2. GRUPO 3

Las 9 empresas identificadas como Grupo 3, las cuales resultan bastantes similares en sus características, constituyen el *cluster* # 12. Los indicadores o variables características de dichas zonas se muestran en la siguiente tabla:

TABLA 7 CLUSTER # 12 (GRUPO 3) PRINCIPALES MEDIDAS

Zona	E-MT/N-MT	E-MT/P	P/L-MT	Lu-MT/L-MT	Ls-MT/Lu-MT	E-BT/L-BT	Lu-BT/L-BT	Ls-BT/Lu-BT	N-BT/L-BT
CASAS GRANDES	213.0227	3,434.4037	0.0486	0.0862	0.0001	500.1033	0.5188	0.0185	53.1422
NUEVO LAREDO	118.2315	3,311.2288	0.1930	0.7000	0.0071	614.8686	0.7002	0.0061	161.5027
SAN MARTIN	263.3942	4,608.3089	0.2365	0.0500	-	213.0887	0.0500	0.0265	134.6268
CD. DEL CARMEN	119.1830	4,981.4306	0.2354	0.7257	0.0054	654.2118	0.8849	0.0337	182.5588
MEXICALI	98.7766	3,673.8695	0.3716	1.0000	0.0161	1,710.0167	1.0000	0.0692	265.4346
METROPOLITANA ORIENTE	160.7642	2,972.5393	0.6380	0.8500	0.0588	505.9141	0.8501	0.0588	200.9470
METROPOLITANA PONIENTE	141.5205	3,515.4861	0.6833	0.8500	0.0824	772.9088	0.8502	0.0820	214.8337
MONCLOVA	272.7082	3,198.8387	0.3249	0.8000	0.0125	335.0558	0.8008	0.0120	109.3924
MANZANILLO	91.5301	6,959.6300	0.0410	0.1729	0.5321	300.0923	0.6758	0.1584	143.2086
ACAPULCO	133.2809	4,454.9805	0.0977	0.0738	1.0978	203.8688	0.0748	1.0620	93.3223
NOGALES	212.7416	4,724.8882	0.1810	0.3804	0.2206	506.3095	0.3743	0.4567	181.6980
ZIHUATANEJO	79.3883	5,113.6646	0.0389	0.0907	0.5323	100.1887	0.0904	0.1925	54.0997

Fuente: elaboración propia sobre la base de datos de CFE

Se aprecia que está formado en promedio por zonas con alta intensidad del consumo en MT, y zonas muy urbanas.

En la tabla siguiente se presenta la caracterización de cada zona en términos de cantidad de clientes y energía vendida, y se identifica la zona representativa elegida (centro), que es la que tiene mayor número de clientes y ventas de energía.

TABLA 8 CLUSTER # 12 (GRUPO 3) CARACTERÍSTICAS DEL SER

Zona	División	Cant. de clientes [dic.2007]		Energía anual vendida [MWh]		Centro
		MT	BT	MT	BT	
GUADALAJARA	**JALISCO**	**9,066**	**1,163,911**	**2,807,061**	**2,234,789**	
TIJUANA	BAJA CALIFORNIA	5,311	512,945	1,869,431	1,150,010	
Coatzacoalcos	ORIENTE	2,224	336,727	524,525	676,621	
Veracruz	ORIENTE	4,209	343,875	851,892	887,104	
CANCUN	PENINSULAR	2,799	201,684	1,141,445	630,900	**GUADALAJARA**
LA PAZ	BAJA CALIFORNIA	2,076	158,473	538,324	685,732	
RIVIERA MAYA	PENINSULAR	1,356	84,524	581,525	400,882	
VALLARTA	JALISCO	4,064	164,082	521,066	442,993	
METROPOLITANA NORTE	GOLFO NORTE	9,240	390,619	2,452,476	1,270,989	

Fuente: elaboración propia sobre la base de datos de CFE

3.3.3. TRATAMIENTO DE *ZONAS* ATÍPICAS

En la tabla siguiente se muestran las zonas consideradas zonas atípicas y que han quedado sin clasificación.

TABLA 9 ZONAS ATÍPICAS

Zona	División	Cant. de clientes [dic.2007]		Energía anual vendida [MWh]	
		MT	BT	MT	BT
CASAS GRANDES	NORTE	1,064	43,531	145,268	409,656
NUEVO LAREDO	GOLFO NORTE	2,505	152,459	496,884	580,436
SAN MARTIN	CENTRO ORIENTE	727	120,114	271,309	190,118
CD. DEL CARMEN	PENINSULAR	693	58,689	153,073	210,316
MEXICALI	BAJA CALIFORNIA	4,380	287,463	1,618,487	1,851,931
METROPOLITANA ORIENTE	GOLFO NORTE	5,753	402,296	2,028,420	1,012,840
METROPOLITANA PONIENTE	GOLFO NORTE	9,735	369,944	2,031,732	1,330,949
MONCLOVA	GOLFO NORTE	1,530	113,659	395,274	348,123
MANZANILLO	CENTRO OCCIDENTE	1,404	102,794	204,262	215,404
ACAPULCO	CENTRO SUR	2,795	243,385	767,167	531,691
NOGALES	NOROESTE	2,283	147,388	597,796	410,703
ZIHUATANEJO	CENTRO SUR	1,537	113,644	173,064	210,461

Fuente: elaboración propia sobre la base de datos de CFE

Para estas zonas, consideradas atípicas desde el punto de vista de las variables e indicadores elegidos para el análisis, se pedirá información adicional, sobre la base de la cual se decidirá si se incluyen en alguno de los *clusters*.

4. EXTRAPOLACIÓN O EXPANSIÓN DE LAS INSTALACIONES

El procedimiento de *clusters* aplicado para la selección de los SER, permite determinar las zonas que serán analizadas en detalle como representativas cada uno de los grupos definidos, a partir del análisis de indicadores determinados en base a información característica de las instalaciones de las etapas de red MT, transformación MT/BT y red BT.

Para cada una de los SER, se requerirá información detallada que será utilizada para adaptar las instalaciones a la demanda. Este proceso se desarrollará para las instalaciones de MT (redes y equipos) y para el conjunto transformador MT/BT y red BT.

Como resultado de las adaptaciones realizadas, se obtendrán las cantidades de instalaciones optimizadas que serán el resultado del análisis de la demanda a abastecer y el mínimo costo de inversión de las unidades constructivas disponibles, entendiéndose como costo de inversión el valor presente del costo de las instalaciones más el costo de las pérdidas de energía más los gastos de operación y mantenimiento.

Para el caso de MT, la optimización se realizará sobre el tipo de unidad constructiva utilizada y el equipamiento requerido, manteniendo invariable la longitud real de los circuitos. Como resultado se obtendrán la longitud de la red y el equipo asociado, que responde a las distintas unidades constructivas consideradas. Si llamamos $RMTi$ a la longitud de cada tipo de unidad constructiva y $EQMTi$ el equipo requerido para cumplir con los estándares de calidad y operación de la red de MT asociada, para cada SER se cumple lo siguiente:

$$RMT_i = \sum_j RMT_{i,j}$$

$$EQMT_i = \sum_j EQMT_{i,j}$$

Donde:

$RMTi$: longitud total de la red MT del SER i

$RMTi,j$: longitud de la red MT asociada a la unidad constructiva j del SER i

$EQMTi$: equipos optimizados de la red MT del SER i

$EQMTi,j$: equipos optimizados tipo j de la red MT del SER i

A partir de esta información se determinarán relaciones que permitirán expandir al resto de las zonas pertenecientes al mismo cluster de la zona representativa o SER. Así, para una zona cualquiera perteneciente al mismo cluster que el SER i, las instalaciones resultantes serán:

$$RMT_{k,j} = \frac{RMT_{i,j}}{RMT_i} \times RMT_k$$

$$EQMT_{k,j} = \frac{EQMT_{i,j}}{RMT_i} \times RMT_k$$

Donde:

$RMTk,j$: longitud de la red de MT tipo j de la zona k (perteneciente al mismo cluster que el SER i)

$RMTk$: longitud total de la red de MT de la zona k

$EQMTk,$j: cantidad de equipos tipo j de la zona k (perteneciente al mismo cluster que el SER i)

$EQMTi,$j: cantidad de equipos tipo j del SER i

$RMTi$: longitud total de la red de MT del SER i

Para el caso del nivel de BT, la optimización se realiza para el conjunto transformador MT/BT y red de BT asociada para cada uno de los SER seleccionados. En tal sentido, a partir del listado de los transformadores y de la demanda a abastecer, se determina el requerimiento de potencia en kVA, considerando un grado de reserva razonable (generalmente se adopta un 80% de la carga máxima). Para cada módulo de transformación y tipo constructivo se determina la red óptima necesaria para abastecer la demanda de BT.

De esta manera para cada SER se determina el inventario de transformadores de MT/BT por tipo constructivo y capacidad [kVA], el cual tendrá asociada una red de BT adaptada a la demanda y que considera las particularidades de la zona abastecida (zonas urbanas, rurales o mixtas).

La expansión de los transformadores MT/BT y su red asociada, se realizará considerando la energía abastecida en BT. De esta manera se cumple lo siguiente:

$$CTMTBT_{k,j} = \frac{CTMTBT_{i,j}}{EBT_i} \times EBT_k$$

Donde:

$CTMTBTi,j$: cantidad de transformadores MT/BT del tipo constructivo y capacidad j del SER i

$CTMTBTk,j$: cantidad de transformadores MT/BT del tipo constructivo y capacidad j de la zona k (perteneciente al mismo cluster que el SER i)

$EBTi$: energía suministrada a usuarios de BT del SER i

$EBTk$: energía suministrada a usuarios de BT de la zona k (perteneciente al mismo cluster que el SER i)

Por otro lado cada centro de transformación MT/BT del SER i tendrá asociada una red de BT adaptada a la demanda de una determinada longitud por tipo constructivo y sección ($RBTi,j$) de la siguiente manera:

$$CTMTBT_j \Rightarrow RBT_{i,j}$$

Por lo tanto, para cada zona dentro del grupo de cada SER se considera que la red total de BT, resulta de multiplicar la cantidad de transformadores de un determinado tipo, resultante de la expansión, por la red de BT asociada a ese transformador MT/BT. La red de BT resultante para cada zona será:

$$RBT_{k,j} = CTMTBT_{k,j} \times RBT_{i,j}$$

Donde:

$RBTk,j$: longitud de la red de BT para un tipo constructivo y sección j de la zona k (perteneciente al mismo cluster que el SER i)

$CTMTBTk,j$: cantidad de transformadores MT/BT del tipo constructivo y capacidad j de la zona k (perteneciente al mismo cluster que el SER i)

$RBTi,j$: longitud de la red de BT para un tipo constructivo y sección j del SER i

5. CONCLUSIONES

Según los resultados preliminares obtenidos mediante el análisis de *clusters* descrito, se propone el agrupamiento de las zonas en 14 *clusters*, cada uno con su correspondiente SER, según el siguiente detalle:

TABLA 10 RESULTADOS SER

Cluster	Cantidad de zonas	SER	Características
1	12	Durango	Alta intensidad del consumo en MT; bajo factor de utilización en MT; una densidad de la demanda similar a la media del Grupo en MT, y por debajo de la media en BT; y redes urbanas en cantidades similares a la media del Grupo tanto en MT como en BT.
2	22	Poza Rica	Baja intensidad del consumo en MT; bajo factor de utilización en MT; baja densidad de la demanda tanto en MT como en BT; y redes menos urbanas que la media del Grupo tanto en MT como en BT.
3	5	Sabinas	Alta intensidad del consumo en MT; alto factor de utilización en MT; con una densidad de la demanda alta tanto en MT como en BT; y redes más urbanas que la media del Grupo,

Cluster	Cantidad de zonas	SER	Características
			tanto en MT como en BT.
4	12	Tuxtla	Intensidad del consumo en MT baja; un factor de utilización en MT similar a la media del Grupo; con una densidad de la demanda baja en MT y media en BT; y redes urbanas más bajas que la media del Grupo en MT y en BT.
5	7	Navojoa	Intensidad del consumo en MT media; bajo factor de utilización en MT; con una densidad de la demanda alta en MT y en BT; y redes menos urbanas que la media del Grupo en MT, y similares a la media en BT.
6	7	Tepic	Intensidad del consumo en MT alta; alto factor de utilización en MT; con una densidad de la demanda media en MT y alta en BT; y redes menos urbanas que la media del Grupo 1 en MT, y más urbanas que la media en BT.
7	3	Culiacán	Alta intensidad del consumo en MT; un factor de utilización en MT similar a la media del grupo; una densidad de la demanda similar a la media del grupo en MT, y alto en BT; redes urbanas similares a la media del Grupo en MT, y mayores en BT; y una relación redes subterráneas / redes urbanas mayor a la media del Grupo en MT y en BT.
8	7	Ensenada	Baja intensidad del consumo en MT; alto factor de utilización en MT; baja densidad de la demanda en MT, y similar a la media en BT; redes menos urbanas que la media del Grupo en MT y en BT; y una relación redes subterráneas / redes urbanas menor a la media del Grupo en MT, y mayor en BT.
9	13	Puebla Oriente	Alta intensidad del consumo en MT; alto factor de utilización en MT; alta densidad de la demanda tanto en MT como en BT; redes urbanas mayores a la media del Grupo en MT y en BT; y una relación redes subterráneas / redes urbanas menor a la media del Grupo en MT y similar a la media en BT.
10	11	Victoria	Baja intensidad del consumo en MT; bajo factor de utilización en MT; baja densidad de la demanda tanto en MT como en BT; redes menos urbanas que la media del Grupo tanto en MT como en BT; y una relación redes subterráneas / redes urbanas menor a la media del Grupo en MT y en BT.
11	8	Guadalajara	Porcentaje de red subterránea sobre el total mayor a 10%. Zonas con alta intensidad del consumo, y muy urbanas.

Fuente: elaboración propia sobre la base de datos de CFE

ANEXO 1 – ANALISIS DE LA INFORMACIÓN DE BASE

1. DESCRIPCIÓN DE LA INFORMACIÓN RECIBIDA

La información solicitada por zona de cada división consiste en:

- Área electrificada (km2 y % de electrificación);
- Número de clientes (MT y BT);
- Energía anual vendida en el año 2007 en MWh (Total, MT y BT);
- Pérdidas de energía totales;
- Demanda máxima anual registrada en el 2007 en MW;
- Red AT (Longitud red 115kV -incluyendo 161 y 138 kV- y Longitud red 69kV en km);
- Subestaciones AT/MT (Cantidad 115kV -incluyendo 161 y 138 kV-/ MT, Cantidad 69kV/MT, y su capacidad instalada en MVA);
- Alimentadores de MT (cantidad, longitud total, aéreo, subterráneo, urbano y rural en km);
- Transformadores MT/BT (cantidad total, aéreo, subterráneo, urbano y rural, y capacidad en MVA total, aéreo, subterráneo, urbano y rural);
- Red BT (longitud total, aéreo, subterráneo, urbano y rural en km);
- Indicación de la presencia de factores especiales que inciden en los costos (nivel isoceráunico alto, contaminación, vegetación frondosa, topografía accidentada y factores sociales);
- Otras observaciones que a criterio del responsable de la zona sea un elemento que signifique mayores costos en la prestación del servicio.

Se recibió información completa de todas las zonas de las Divisiones.

2. EVALUACIÓN DE LA CONSISTENCIA DE DATOS

Al analizar la información recibida, se encontraron algunas inconsistencias en los datos, algunas de las cuales fueron corregidas en base a información complementaria suministrada por CFE.

En principio, se presentan casos donde la información total de ciertos datos presenta diferencias significativas con la suma de los elementos que los componen, como ser:

- Energía anual vendida en MT y BT y la total por zonas: Tijuana, La Paz (División Baja California), Hermosillo, Obregón, Navojoa, Los Mochis, Guasave, Culiacán, Caborca, Nogales (División Noroeste), Nuevo Laredo, Reynosa, Cerralvo, Montemorelos-Linares, Matamoros, Metropolitana Norte, Metropolitana Oriente, Metropolitana Poniente, Piedras Negras, Monclova, Saltillo (División Golfo Norte), Chilpancingo, Morelos, Toluca, Altamirano, Valle de Bravo, Acapulco (Centro Sur) Poza Rica, Teziutlán, Veracruz, Papaloapan, Coatzacoalcos, Orizaba, Córdoba (División Oriente), Tehuantepec, Villa Hermosa, Chontalpa, Los Ríos (División Sureste), San Juan del Río, Irapuato, León, Celaya, Querétaro, Salvatierra,

Ixmiquilpán, Aguas Calientes, Fresnillo, Zacatecas (División Bajio), Tlaxcala, Tehuacán, San Martín, Tecamachalco, Puebla Poniente, Puebla Oriente (División Centro Oriente).

Pueden existir otros casos a los que no se hace mención cuyas diferencias no se consideran significativas.

Por otro lado, se observan inconsistencias en las unidades de medida como la capacidad de los transformadores. En el caso de Constitución de la división Baja California División Golfo Centro, se deduce que estos datos se presentaron en kVa, por lo cual se transformaron a MVA.

A su vez, se presenta información que resulta inconsistente al calcular indicadores, tal como en el caso del factor de carga en Coatzacoalcos cuyo valor resulta ser mayor a la unidad y los de Constitución y Ticul que parecen ser excesivamente bajos (0.08 y 0.04 respectivamente). En los dos últimos casos, se incorpora la información de energía por nivel de tensión de bases de datos suministradas por CFE, resultando sus factores de carga más cercanos a la media de las zonas analizadas (0.53 y 0.57 respectivamente).

Por último, en el caso de las zonas de la división Golfo Norte, se verificó que la cantidad de clientes informados representaba la suma de los clientes de enero a diciembre de 2007, y no la cantidad de cliente a diciembre de dicho año. Por lo tanto, se consideró como válida la información de la base de datos comercial suministrada por CFE como total de clientes en MT y BT a diciembre de 2007, manteniendo la relación entre los dos niveles de tensión que presenta la división en la base presentada.

ANEXO 2 – ANALISIS DE CONGLOMERADOS[1]

El objeto del análisis de conglomerados es agrupar elementos en grupos homogéneos en función de las similitudes entre ellos y se utiliza para estudiar tres tipos de problemas:

1. Partición de datos: se disponen de datos heterogéneos y se desea dividirlos en un número de grupos prefijado, de manera que:
    a. Cada elemento pertenezca a uno, y solo uno, de los grupos;
    b. Todo elemento quede clasificado;
    c. Cada grupo sea internamente homogéneo.
2. Construcción de jerarquías: se desea estructurar los elementos de un conjunto de forma jerárquica por su similitud. Estos métodos no definen grupos, sino la estructura de asociación en cadena que pueda existir entre los elementos. Sin embargo, la jerarquía construida permite obtener también una partición de los datos en grupos.
3. Clasificación de variables: en problemas con muchas variables es interesante hacer un estudio exploratorio inicial para dividir las variables en grupos y orientarnos para plantear los modelos formales para reducir la dimensión.

En este estudio el análisis de conglomerados se utiliza para dividir las zonas de las divisiones de CFE en grupos heterogéneos. Para ello se aplica el algoritmo de *k*-medias como método de partición.

1. FUNDAMENTOS DEL ALGORITMO DE K-MEDIAS

Supongamos una muestra de *n* elementos y *p* variables. El objetivo es dividir esta muestra en un número de grupos prefijado, *k*. El algoritmo de *k*-medias requiere cuatro etapas:

1. Seleccionar los *k* puntos como centros de los grupos iniciales, lo que puede hacerse:
    a. Asignando aleatoriamente los objetos a los grupos y tomando los centros de los grupos así formados;
    b. Tomando como centros los *k* puntos más alejados entre sí;
    c. Construyendo unos grupos iniciales con información *a priori* y calculando sus centros, o bien seleccionando los centros *a priori*.
2. Calcular las distancias euclídeas de cada elemento a los centros de los *k* grupos, y asignar cada elemento al grupo de cuyo centros esté más próximo. La asignación se realiza secuencialmente y al introducir un nuevo elemento en un grupo se recalculan las coordenadas de nuevo centro del grupo.
3. Definir un criterio de optimalidad y comprobar si reasignando alguno de los elementos mejora el criterio.
4. Si no es posible mejorar el criterio de optimalidad, terminar el proceso.

[1] Bibliografía consultada: Peña, Daniel (2002), Análisis de Datos Multivariantes, Madrid, España.

2. IMPLEMENTACIÓN DEL ALGORITMO

El criterio de homogeneidad o de optimalidad que se utiliza en el algoritmo de k-medias es minimizar *la suma de los cuadrados dentro de los grupos (SCDG)* para todas las variables:

$$SCDG = \sum_{k=1}^{K} \sum_{j=1}^{p} \sum_{i=1}^{n_g} (x_{ijg} - \bar{x}_{jg})^2$$

Donde x_{ijg} es el valor de la variable *j* en el elemento *i* del grupo *g* y $\bar{x}_{jg}$ la media de esta variable en el grupo. Este criterio es equivalente a la suma ponderada de las varianzas de las variables en los grupos:

$$\min SCDG = \min \sum_{k=1}^{K} \sum_{j=1}^{p} n_g s_{jg}^2$$

Donde n_g es el número de elementos del grupo *g* y s_{jg}^2 es la varianza de la variable *j* en dicho grupo.

Las varianzas de las variables en los grupos son claramente una medida de la heterogeneidad de la clasificación y al minimizarlas obtendremos grupos más homogéneos. Un criterio alternativo de homogeneidad sería minimizar las distancias al cuadrado entre los puntos y sus centros de grupo. Si medimos las distancias con al norma euclídea, este criterio se escribe así:

$$\min \sum_{k=1}^{K} \sum_{i=1}^{n_g} (x_{ig} - \bar{x}_g)'(x_{ig} - \bar{x}_g) = \min \sum_{k=1}^{K} \sum_{i=1}^{n_g} d^2(i,g)$$

Donde $d^2(i,g)$ es el cuadrado de la distancia euclídea entre el elemento *i* del grupo *g* y su media de grupo. Ambos criterios son idénticos. Como un escalar es igual a su traza, el último criterio se puede escribir como:

$$\min \sum_{k=1}^{K} \sum_{i=1}^{n_g} tr[d^2(i,g)] = \min tr[\sum_{k=1}^{K} \sum_{i=1}^{n_g} (x_{ig} - \bar{x}_g)(x_{ig} - \bar{x}_g)']$$

y llamando **W** a la matriz de suma de cuadrados dentro de los grupos

$$W = \sum_{k=1}^{K} \sum_{i=1}^{n_g} (x_{ig} - \bar{x}_g)(x_{ig} - \bar{x}_g)'$$

Tenemos que

$$\min tr(W) = \min SCDG$$

Y ambos criterios coinciden. Este criterio se denomina *criterio de la traza*.

El algoritmo de k-medias busca la partición óptima con la restricción de que en cada iteración sólo se permite mover un elemento de un grupo a otro. El algoritmo funciona como sigue:

1. Partir de la asignación inicial;
2. Comprobar si moviendo algún elemento se reduce el SCDG;
3. Si es posible reducir SCDG moviendo un elemento hacerlo, recalcular las medias de los dos grupos afectados por el cambio y volver al punto 2. Si no es posible reducir SCDG, terminar.

Cuando las variables se encuentran expresadas en distintas unidades de medida, es conveniente estandarizar, para evitar que el resultado del algoritmo de k-medias dependa de cambios irrelevantes en la escala de medida. Para ello se aplica la siguiente transformación lineal:

$$y_j = \frac{(x_j - \overline{x}_j)}{sd_j}$$

Donde $\overline{x}_j$ es la media y sd_j el desvío estándar y que transforma la variable x_j en la nueva y_j con media cero y varianza unitaria.

3. DETERMINACIÓN DEL NÚMERO DE GRUPOS

Generalmente, el número de grupos *K* es desconocido y se estima con los datos aplicando el algoritmo para distintos valores de *K* y seleccionando el mejor resultado. Compara las soluciones obtenidas no es simple, porque cualquiera de los criterios disminuirá si aumentamos el número de grupos. La variabilidad total puede descomponerse como:

$$T = W + B$$

Intuitivamente, el objetivo de la división en grupos es conseguir que B, la variabilidad entre los grupos, sea lo mayor posible, mientra que W, la variabilidad dentro de los grupos, sea lo menor posible. Dada una división cualquiera en grupos, si elegimos uno cualquiera de ellos podemos aplicarle de nuevo esta descomposición, con lo que reduciremos de nuevo la variabilidad descomponiendo más este grupo. Por lo tanto, no podemos utilizar ningún criterio basado en el tamaño de W para comparar soluciones con grupos distintos, ya que siempre podemos disminuir W haciendo más grupos.

Partiendo de la descomposición de la variabilidad total, Calinsky y Harabasz (1974) proponen seleccionar el valor de *K* maximizando:

$$CH = \max\left\{\frac{tr(B)/(G-1)}{tr(W)/(n-G)}\right\}$$

De esta manera, se pretende maximizar la variabilidad entre los grupos y minimizar la variabilidad dentro de los grupos.