

UNIVERSIDAD DE CUENCA

Facultad de Ingeniería

Carrera de Ingeniería Eléctrica



“Zonificación del mercado de consumo eléctrico de la empresa eléctrica Azogues”

Trabajo de titulación previo a la obtención del título de
Ingeniero Eléctrico

AUTOR:

Juan Carlos Carvajal Parra

C.I: 0104467543

DIRECTOR:

Ing. Juan Bautista Sanango Fernández. Mg

C.I: 0301522892

Cuenca – Ecuador

2018



RESUMEN

El sobre dimensionamiento y las pérdidas de energía presentes en las redes eléctricas han ocasionado que los índices de confiabilidad y servicio de muchas empresas de comercialización y distribución de energía eléctrica se vean perjudicados, acarreando fuertes sanciones económicas y perjudicando considerablemente el costo de las inversiones futuras. Por ello, la gestión de estas empresas ha ido evolucionado paulatinamente con un enfoque direccionado a la eficiencia, en donde se pone especial énfasis en predecir, con la mayor exactitud posible, los patrones de demanda actual y futura de los usuarios pues, este es el dato principal en el cual está basado el dimensionamiento, diseño y operación del sistema.

La gran ventaja que se tiene al estimar la magnitud y la ubicación del consumo de energía futuro es que se puede actuar de una manera óptima en los procesos, además de abaratar los costos de las inversiones a largo plazo, pues se suprime de manera substancial las readecuaciones en las redes y equipos encargados de proporcionar energía eléctrica a los usuarios. Para ello, el primer paso es la zonificación del área en donde se encuentran los usuarios, para de esta manera evaluar la situación de la demanda eléctrica actual en cada sector.

La metodología de zonificación planteada en este documento evalúa la demanda promedio mensual de los consumidores para posteriormente elaborar un mapa de estratos y un mapa de densidad de carga, con lo cual se puede inspeccionar de una manera clara el consumo promedio y la concentración de carga presente en cada zona del área de concesión de la Empresa Eléctrica Azogues. Para ello, se presenta una alternativa de estratificación la cual pretende definir el número ideal de grupos en los cuáles dividir a los usuarios de cada tipo de tarifa (residencial, comercial e industrial) bajo el concepto de aglomerados jerárquicos y del método k-means, y con la ayuda del método propuesto por Dalenius & Hodges (1959) definir los límites (rangos) de cada uno de estos grupos. Estos conceptos son usados también en la elaboración del mapa de densidades a la hora de definir la cantidad de grupos en los que se divide la densidad de carga presente en el área de concesión, y los límites de cada uno de estos grupos.

Se calcula además la demanda máxima unitaria para cada estrato de clientes residenciales, basándose en los conceptos propuestos por el método REA (Rural Electrification Administration), para posteriormente elaborar una tabla de demandas máximas diversificadas tanto para el caso de clientes con cocina de inducción como para clientes sin cocina de inducción.

Palabras clave: aglomerados jerárquicos, ArcGIS, carga, cluster, cuadrícula, Dalenious & Hodges, demanda, dendograma, densidad, energía, estrato, georreferenciación, Hubert & Vandervieren, k-means, potencia, REA, Ward, zonificación.



ABSTRACT

Over sizing and energy losses of electrical grids have affected the reliability and service rate of many electric power distribution and marketing companies, these troubles had led severe economic sanctions and consequently the cost of future investments have been affected. For this reason, the management of these companies has gradually evolved to efficiency focused administration, with the special target of predicting, with the greatest possible accuracy, the current and future demand pattern of each user type, since this will be the central fact to sizing, design and operate the system.

A great advantage of estimate the scale and location of future demand is the possibility of optimal variations in the processes, in addition to reduce the long term investments cost, due to substantial reduction of readjustments in electric network and electric equipment that are main responsible for provide energy to users. So, the first step is the area zoning where the users are located, in order to evaluate the current electricity demand situation in each sector.

Zoning procedure proposed in this document evaluate the average consumer monthly demand to elaborate a strata map and density map about user's demand, with that can be clearly noted the average consumption and the density of currently load for each zone of concession area of Azogues Electric Company. To do this, a stratification alternative is presented which propose determine the ideal number of groups in which divide users of each type of tariff (residential, commercial and industrial) under the concept of hierarchical agglomerates and k-means method and with the help of the method proposed by Dalenius & Hodges (1959) define the limits (ranges) of each of these groups. These concepts are also used in the elaboration of the density map to define the number of groups in which the load density present in the concession area will be divided, and limits of each of these groups.

Was calculated also the maximum unit demand for each stratum of residential customers, based on concepts proposed by REA method (Rural Electrification Administration), to later elaborate a table of maximum diversified demands for users with induction cooking and users without induction cooking

Keywords: ArcGIS, cluster, Dalenious & Hodges, demand, dendogram, density, energy, georeferencing, grid, hierarchical agglomerates, Hubert & Vandervieren, k-means, load, power, REA, stratum, Ward, zoning.



ÍNDICE

RESUMEN	2
ABSTRACT.....	3
ÍNDICE.....	4
LISTA DE FIGURAS.....	6
AGRADECIMIENTOS	10
DEDICATORIA	11
CAPÍTULO 1: GENERALIDADES	12
1.1. INTRODUCCIÓN.....	12
1.2. ANTECEDENTES	13
1.3. OBJETIVOS.....	14
1.3.1. <i>Objetivo general</i>	14
1.3.2. <i>Objetivos específicos</i>	15
1.4. ALCANCES DEL PROYECTO	15
1.5. JUSTIFICACIÓN	15
2. CAPÍTULO 2: MARCO TEÓRICO.....	16
2.1. CONCEPTOS GENERALES	16
2.1.1. <i>Definiciones eléctricas básicas</i>	16
2.1.2. <i>Conceptos relacionados a la demanda</i>	19
2.1.3. <i>Conceptos básicos estadísticos</i>	23
2.2. MÉTODOS DE ANÁLISIS CLUSTER.....	25
2.2.1. <i>Medidas de distancia para individuos</i>	26
2.2.2. <i>Algoritmos de agrupación</i>	26
2.2.3. <i>Algoritmo de partición (K-MEANS)</i>	30
2.2.4. <i>Dendograma</i>	34
2.3. MÉTODOS DE ESTRATIFICACIÓN	36
2.3.1. <i>Dalenius & Hodges (1959)</i>	36
2.3.2. <i>Lavallée & Hidiroglou (1988)</i>	38
2.3.3. <i>Método de Kozak (2004)</i>	39
3. CAPÍTULO 3: MAPA DE ESTRATOS.....	40
3.1. INFORMACIÓN DE LA EMPRESA	40
3.1.1. <i>Zona de concesión</i>	40
3.1.2. <i>Usuarios y consumos</i>	40
3.1.3. <i>Datos técnicos de la empresa</i>	41
3.2. CRITERIOS Y DATOS NECESARIOS PARA EL PROCESO DE ZONIFICACIÓN.....	42
3.3. DETECCIÓN DE OUTLIERS EN LA BASE DE DATOS.....	43
3.3.1. <i>Método para detectar datos atípicos (outliers)</i>	43
3.3.2. <i>Resultados obtenidos</i>	45
3.4. DEFINICIÓN DEL NÚMERO DE ESTRATOS	49
3.4.1. <i>Resultados obtenidos usando el método K-means</i>	49
3.4.2. <i>Resultados obtenidos usando el método de Ward</i>	53
3.5. LÍMITES DE CADA ESTRATO	55
3.6. MAPA DE ESTRATOS	56
3.6.1. <i>ArcGIS</i>	56
3.6.2. <i>Georreferenciación de los usuarios</i>	58



3.6.3.	<i>Cuadrícula del área de concesión</i>	60
3.6.4.	<i>Resultados obtenidos</i>	63
3.7.	DEMANDA MÁXIMA DIVERSIFICADA PARA USUARIOS RESIDENCIALES.....	68
3.7.1.	<i>Método REA (Rural Electrification Administration)</i>	69
3.7.2.	<i>Criterios adicionales para el cálculo de la demanda máxima unitaria</i>	2
3.7.3.	<i>Resultados obtenidos</i>	3
3.7.4.	<i>Verificación de los resultados</i>	4
4.	CAPÍTULO 4: MAPA DE DENSIDAD.....	6
4.1.	CARACTERIZACIÓN DE LAS CURVAS DE CONSUMO.....	6
4.1.1.	<i>Campañas de medición</i>	7
4.1.2.	<i>Técnicas de muestreo</i>	7
4.1.3.	<i>Características de las mediciones</i>	8
4.1.4.	<i>Generación y normalización de las curvas de demanda</i>	8
4.2.	DETERMINACIÓN DE LA POTENCIA DE CADA CLIENTE EN BASE A LA HORA DE MÁXIMA DEMANDA.....	11
4.2.1.	<i>Metodología para la estimación de la demanda de cada cliente</i>	11
4.2.2.	<i>Ejemplo para la determinación de la demanda de un cliente</i>	12
4.3.	GEORREFERENCIACIÓN DE LOS USUARIOS EN EL MAPA	12
4.3.1.	<i>Filtrado de datos atípicos de la muestra</i>	13
4.3.2.	<i>Definición del número de grupos</i>	14
4.3.3.	<i>Límites de cada grupo</i>	16
4.3.4.	<i>Resultados obtenidos</i>	16
4.4.	CUADRICULACIÓN DEL ÁREA DE CONCESIÓN.....	17
4.5.	CLASIFICACIÓN DE CUADRÍCULAS	19
4.5.1.	<i>Determinación de la demanda de cada cuadrícula</i>	19
4.5.2.	<i>Definición del número de grupos</i>	19
4.5.3.	<i>Definición de los intervalos de los grupos</i>	21
4.6.	MAPA DE DENSIDADES.....	21
5.	CAPÍTULO 5: CONCLUSIONES Y RECOMENDACIONES.....	22
5.1.	CONCLUSIONES.....	22
5.2.	RECOMENDACIONES.....	23
	REFERENCIAS.....	24
6.	ANEXOS	26

LISTA DE FIGURAS

-FIGURA 2.1. TRIÁNGULO DE POTENCIAS.....	18
FIGURA 2.2. REPRESENTACIÓN GRÁFICA DE LA DEMANDA PROMEDIO DE UNA CARGA INTERMITENTE.....	19
FIGURA 2.3. CURVA DE CARGA DIARIA PARA USUARIOS RESIDENCIALES DE LA EEA.....	20
FIGURA 2.4. DEMANDA MÁXIMA PRESENTE EN LA CURVA DE CARGA PARA USUARIOS RESIDENCIALES DE LA EEA.	20
FIGURA 2.5. EJEMPLO DE DENDOGRAMA EN EL QUE SE DIVISAN LOS ENLACES DE OCHO ELEMENTOS	34
FIGURA 2.6. EJEMPLO DE LA DETERMINACIÓN DEL NÚMERO DE CLUSTERS EN UN DENDOGRAMA.	35
FIGURA 2.7. EJEMPLO DE GRÁFICA QUE RELACIONA EL NÚMERO DE CLUSTERS CON LAS DISTANCIAS DE ENLACE.....	35
FIGURA 2.8. DETERMINACIÓN DEL NÚMERO DE CLUSTERS ÓPTIMOS.....	36
FIGURA 3.1. ZONA DE CONCESIÓN DE LA EEA. (FUENTE: DATOS PROPORCIONADOS POR LA EEA).....	40
FIGURA 3.2. UBICACIÓN DE LAS SUBESTACIONES AZOGUES 1 Y 2 PERTENECIENTES A LA EEA.	41
FIGURA 3.3. EJEMPLO MÉTODO BOXPLOT.	43
FIGURA 3.4. DISTRIBUCIÓN DE DEMANDAS PARA USUARIOS CON TARIFA RESIDENCIAL DE LA EEA.	44
FIGURA 3.5. DISTRIBUCIÓN DE DEMANDAS PARA USUARIOS CON TARIFA COMERCIAL DE LA EEA.	45
FIGURA 3.6. DISTRIBUCIÓN DE DEMANDAS PARA USUARIOS CON TARIFA INDUSTRIAL DE LA EEA.	45
FIGURA 3.7. DISTRIBUCIÓN DE LA DEMANDA DE USUARIOS RESIDENCIALES UNA VEZ FINALIZADO CON EL PROCESO DE DETECCIÓN DE DATOS ATÍPICOS.....	46
FIGURA 3.8. DISTRIBUCIÓN DE LA DEMANDA DE USUARIOS COMERCIALES UNA VEZ FINALIZADO CON EL PROCESO DE DETECCIÓN DE DATOS ATÍPICOS.	47
FIGURA 3.9. DISTRIBUCIÓN DE LA DEMANDA DE USUARIOS CON TARIFA INDUSTRIAL UNA VEZ FINALIZADO CON EL PROCESO DE DETECCIÓN DE DATOS ATÍPICOS.....	48
FIGURA 3.10. VARIACIÓN DE LA DISTANCIA INTER-CLUSTER CON RELACIÓN AL NÚMERO DE CLUSTERS EN LOS QUE SE AGRUPAN LOS USUARIOS CON TARIFA RESIDENCIAL DE LA EEA.....	50
FIGURA 3.11. VARIACIÓN DE LA DISTANCIA INTRA-CLUSTER CON RELACIÓN AL NÚMERO DE CLUSTERS EN LOS QUE SE AGRUPAN LOS USUARIOS CON TARIFA RESIDENCIAL DE LA EEA.....	50
FIGURA 3.12. VARIACIÓN DE LA DISTANCIA INTER-CLUSTER CON RELACIÓN AL NÚMERO DE CLUSTERS EN LOS QUE SE AGRUPAN LOS USUARIOS CON TARIFA COMERCIAL DE LA EEA.	51
FIGURA 3.13. VARIACIÓN DE LA DISTANCIA INTRA-CLUSTER CON RELACIÓN AL NÚMERO DE CLUSTERS EN LOS QUE SE AGRUPAN LOS USUARIOS CON TARIFA COMERCIAL DE LA EEA.	52
FIGURA 3.14. VARIACIÓN DE LA DISTANCIA INTER-CLUSTER CON RELACIÓN AL NÚMERO DE CLUSTERS EN LOS QUE SE AGRUPAN LOS USUARIOS CON TARIFA INDUSTRIAL DE LA EEA.....	52
FIGURA 3.15. VARIACIÓN DE LA DISTANCIA INTRA-CLUSTER CON RELACIÓN AL NÚMERO DE CLUSTERS EN LOS QUE SE AGRUPAN LOS USUARIOS CON TARIFA INDUSTRIAL DE LA EEA.	53
FIGURA 3.16. DENDOGRAMAS OBTENIDOS TRAS APLICAR EL MÉTODO DE WARD	54
FIGURA 3.17. APLICACIONES PRINCIPALES DEL SOFTWARE ARCGIS.....	57
FIGURA 3.18. GEORREFERENCIACIÓN DE LOS USUARIOS CON TARIFA RESIDENCIAL.....	58
FIGURA 3.19. GEORREFERENCIACIÓN DE USUARIOS CON TARIFA COMERCIAL.	59
FIGURA 3.20. GEORREFERENCIACIÓN PARA USUARIOS CON TARIFA INDUSTRIAL.	60
FIGURA 3.21. EJEMPLO DE LA CUADRICULACIÓN DE LA ZONA DE CONCESIÓN DE LA EEA.....	61
FIGURA 3.22. CARACTERIZACIÓN DEL ESTRATO DE UNA CUADRÍCULA PERTENECIENTE A LA ZONA DE CONCESIÓN DE LA EEA.	62
FIGURA 3.23. MAPA DE ESTRATOS PARA LOS USUARIOS CON TARIFA RESIDENCIAL.	64
FIGURA 3.24. VARIACIÓN DEL MAPA DE ESTRATOS PARA USUARIOS RESIDENCIALES.	64
FIGURA 3.25. MAPA DE ESTRATOS USANDO CUADRÍCULA DE 300x300M PARA CLIENTES CON TARIFA COMERCIAL.	66
FIGURA 3.26. MAPA DE ESTRATOS USANDO CUADRÍCULA DE 500x500M PARA CLIENTES CON TARIFA COMERCIAL.	66
FIGURA 3.27. CONTRASTE ENTRE LA DENSIDAD DE USUARIOS PRESENTES EN ZONAS URBANAS Y ZONAS RURALES PARA CLIENTES CON TARIFA INDUSTRIAL.....	67
FIGURA 3.28. MAPA DE ESTRATOS PARA USUARIOS CON TARIFA INDUSTRIAL.....	68
FIGURA 3.29. NOMOGRAMA CARACTERÍSTICO DEL MÉTODO REA.....	2
FIGURA 4.1. CURVA TÍPICA DE DEMANDA PARA DÍAS LABORABLES DE USUARIOS RESIDENCIALES.	9
FIGURA 4.2. CURVA TÍPICA DE DEMANDA PARA DÍAS SÁBADOS DE USUARIOS RESIDENCIALES.	9



FIGURA 4.3. CURVA TÍPICA DE DEMANDA PARA DOMINGOS Y FERIADOS DE USUARIOS RESIDENCIALES..... 10

FIGURA 4.4. GEORREFERENCIACIÓN DE LOS USUARIOS DE LA EEA..... 13

FIGURA 4.5. DISTRIBUCIÓN DE DATOS SIN LA DETERMINACIÓN DE OUTLIERS..... 13

FIGURA 4.6. DISTRIBUCIÓN DE DATOS UNA VEZ FILTRADOS LOS OUTLIERS. 14

FIGURA 4.7. VARIACIÓN DE LA DISTANCIA INTER-CLUSTER CON RELACIÓN AL NÚMERO DE CLUSTERS EN LOS QUE SE AGRUPAN LOS USUARIOS PARA LA ELABORACIÓN DEL MAPA DE DENSIDADES..... 15

FIGURA 4.8. VARIACIÓN DE LA DISTANCIA INTRA-CLUSTER CON RELACIÓN AL NÚMERO DE CLUSTERS EN LOS QUE SE AGRUPAN LOS USUARIOS PARA LA ELABORACIÓN DEL MAPA DE DENSIDADES..... 15

FIGURA 4.9. USUARIOS GEORREFERENCIADOS DIVIDIDOS EN GRUPOS DE CONSUMO. 16

FIGURA 4.10. DIVISIÓN DEL ÁREA DE CONCESIÓN DE LA EEA MEDIANTE CUADRÍCULAS DE 800x800M. 18

FIGURA 4.11. DIVISIÓN DEL ÁREA DE CONCESIÓN DE LA EEA MEDIANTE CUADRÍCULAS DE 600x600M. 18

FIGURA 4.12. DIVISIÓN DEL ÁREA DE CONCESIÓN DE LA EEA MEDIANTE CUADRÍCULAS DE 400x400M. 19

FIGURA 4.13. VARIACIÓN DE LA COHESIÓN ENTRE ELEMENTOS SEGÚN EL NÚMERO DE GRUPOS..... 20

FIGURA 4.14. VARIACIÓN DE LA SEPARACIÓN ENTRE CLUSTERS SEGÚN LA CANTIDAD DE GRUPOS QUE SE DISPONGA. 20

FIGURA 4.15. MAPA DE DENSIDADES. 21

FIGURA A.1. ICONO "CONNECT TO FOLDER". 27

FIGURE A.2. VENTANA "CONNECT TO FOLDER". 27

FIGURA A.3. ARCHIVO DE DONDE SE EXTRAE LA BASE DE DATOS..... 27

FIGURA A.4. MENÚ PARA LA CREACIÓN DE UNA TABLA CON LA INFORMACIÓN DE LA BASE DE DATOS..... 28

FIGURA A.5. VENTANA CREATE FEATURE CLASS FROM XY TABLE. 28

FIGURA A.6. VENTANA PARA LA ELECCIÓN DEL SISTEMA DE COORDENADAS EN LAS QUE ESTARÁ FUNDAMENTADA LA BASE DE DATOS. 29

FIGURA A.7. CAPA DE PUNTOS GENERADA A PARTIR DE LA BASE DE DATOS..... 29

FIGURA A.8 UBICACIÓN DE MEDIDORES PARA EL SECTOR INDUSTRIAL DE LA EEA 30

FIGURA A.9. PROPIEDADES DE CAPA. 30

FIGURA A.10. NÚMERO DE CLASES Y COLORES QUE SE TENDRÁ EN LA CLASIFICACIÓN. 31

FIGURA A.11. PARÁMETROS DE CLASIFICACIÓN..... 31

FIGURA A.12. HERRAMIENTA PARA LA CREACIÓN DE UNA CUADRICULA EN UNA ZONA DETERMINADA. 32

FIGURA A.13. VENTANA DE CREACIÓN DE CUADRICULA..... 33

FIGURA A.14. CUADRICULA GENERADA PARA LA ZONA URBANA DE LA EEA..... 33

FIGURA A.15. DIRECCIÓN DE LA HERRAMIENTA SPATIAL JOIN..... 34

FIGURA A.16. VENTANA DE PARÁMETROS DE LA HERRAMIENTA SPATIAL JOIN. 35

FIGURA A.17. ELECCIÓN DE ATRIBUTOS PARA EL SPATIAL JOIN. 36

FIGURA A.18. RESULTADO DEL SPATIAL JOIN. 36

FIGURA A.19. CONDICIÓN PARA ELIMINAR CELDAS INNECESARIAS 37

FIGURA A.20. RESULTADO DEL SPATIAL JOIN TOMANDO EN CUENTA EL CONDICIONAL IMPUESTO EN LA VENTANA QUERY BUILDER. 37

FIGURA A.21. EJEMPLO CONFIGURACIÓN DE VARIABLES EN SPSS..... 41

FIGURA A.22. PROCESO SEGUIDO PARA CLASIFICAR LOS DATOS. 41

FIGURA A.23. VENTANA DE CONFIGURACIÓN PARA ANÁLISIS CLUSTER. 42

FIGURA A.24. PARAMETROS DE LA SECCIÓN DE GRÁFICOS PARA LA CLASIFICACIÓN DE DATOS EN SPSS. 42

FIGURA A.25. PARÁMETROS DE LA SECCIÓN DE MÉTODOS PARA LA CLASIFICACIÓN DE DATOS EN SPSS..... 42

FIGURA A. A.26. RESULTADOS OBTENIDOS POR EL MÉTODO DE WARD. 43

Cláusula de Propiedad Intelectual

Yo, **Juan Carlos Carvajal Parra**, autor del trabajo de titulación “**Zonificación del mercado de consumo eléctrico de la empresa eléctrica Azogues**”, certifico que todas las ideas, opiniones y contenidos expuestos en el presente documento son de exclusiva responsabilidad de su autor.

Cuenca, 31 de octubre de 2018

A handwritten signature in blue ink, appearing to read "Juan Carlos Carvajal Parra", written over a horizontal line.

Juan Carlos Carvajal Parra

C.I. 0104467543

Cláusula de licencia y autorización para publicación en el Repositorio Institucional

Yo, **Juan Carlos Carvajal Parra**, en calidad de autor y titular de los derechos morales y patrimoniales del trabajo de titulación "**Zonificación del mercado de consumo eléctrico de la empresa eléctrica Azogues**", de conformidad con el Art. 114 del CÓDIGO ORGÁNICO DE LA ECONOMÍA SOCIAL DE LOS CONOCIMIENTOS, CREATIVIDAD E INNOVACIÓN, reconozco a favor de la Universidad de Cuenca una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos.

Asimismo, autorizo a la Universidad de Cuenca para que realice la publicación de este trabajo de titulación en el repositorio institucional, de conformidad en lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Cuenca, 31 de octubre de 2018



Juan Carlos Carvajal Parra

C.I: 0104467543



AGRADECIMIENTOS

El presente trabajo de titulación lo dedico primeramente a Dios por darme la fuerza y la inspiración para concluir esta etapa de formación y concederme la posibilidad de cumplir uno de mis más grandes anhelos.

A mi querida universidad por haberme inculcado en sus aulas valores y enseñanzas necesarias que me acompañarán para toda la vida. Un agradecimiento especial para el Ing. Juan Sanango por su apoyo y tiempo brindado en la culminación de este trabajo.

A la empresa Eléctrica Azogues por darme la posibilidad de desarrollar este trabajo en su medio. Un especial agradecimiento al Ing. Galo Ríos y al Ing. Luis Quevedo por haber compartido su conocimiento y su tiempo conmigo con el fin de elaborar este trabajo.

Juan Carvajal P.



DEDICATORIA

Dedico el presente trabajo de titulación principalmente a Dios por darme salud, fortaleza y sabiduría para alcanzar una meta más en mi vida y por haberme dado a las personas que hoy comparten conmigo, en especial a mi familia.

A mi madre, que ha sido madre y padre a la vez, ya que sin sus consejos, paciencia y enseñanzas nada de esto habría sido posible. Es y siempre será un privilegio ser su hijo. A mis hermanos, por sus sonrisas y apoyo moral, en especial para mi hermana Jessy por haber estado a mi lado incondicionalmente.

Juan Carvajal P.



CAPÍTULO 1: GENERALIDADES

1.1. INTRODUCCIÓN

El servicio de energía eléctrica hoy en día se ha convertido en un aspecto fundamental para la vida de un gran porcentaje de personas, tanto así que actualmente el desarrollo económico de un país depende directamente de la disponibilidad de energía eléctrica con la que se cuente. Por esta razón, la presencia de una fuente continua, barata, eficiente y confiable de energía tiene una gran relevancia en casi todos los ámbitos, sobre todo en temas económicos y sociales.

Tomando en cuenta la importancia de la energía eléctrica, se vuelve necesario realizar un trabajo cuidadoso en las fases de planificación, diseño, construcción, operación, control y supervisión de todas las partes que conforman el sistema eléctrico, prestando especial atención al sistema de distribución dado el impacto directo que tiene en cuanto a la calidad y confiabilidad del servicio eléctrico. Todo este trabajo tiende a ser una actividad ardua y laboriosa, y que hasta hace algunos años se lo hacía con métodos heurísticos y formulaciones empíricas. Sin embargo, las actuales bondades tecnológicas permiten aprovechar de una mejor manera los datos, métodos y algoritmos, lo que irremediablemente conduce a una mejor explotación de la infraestructura eléctrica.

La gestión de las empresas de energía eléctrica paulatinamente se ha integrado con los avances tecnológicos y de esta manera ha ido evolucionado con un enfoque direccionado a la eficiencia. En el caso particular de las empresas de distribución se pone especial énfasis en predecir con la mayor exactitud posible los patrones de demanda de sus usuarios pues, debido a la finalidad que tiene el sistema eléctrico se debe prestar singular atención a la demanda de energía que se va a requerir, ya que este es el dato principal en el cual está basado el dimensionamiento, diseño y operación del sistema.

Uno de los principales problemas que se tiene al momento de establecer la demanda de energía es que esta obedece al consumo que tiene cada usuario del sistema, por consiguiente, es afectada por numerosas variables tales como clima, época del año, situación socioeconómica, costumbres, etc. Siendo los dos últimos factores muy relevantes en el análisis del consumo eléctrico ya que están fuertemente ligados a la demanda energética de cada sector.

En este marco, la *zonificación de la demanda eléctrica* para la zona de concesión de una empresa de distribución es crucial pues, de esta manera se obtiene una interfaz georreferenciada, creada a partir de criterios técnicos y datos de consumo, y que a su vez permite un análisis práctico, interactivo y de fácil interpretación visual de la demanda energética en cada zona, lo cual contribuye de manera significativa al entendimiento de los actuales y futuros patrones de demanda eléctrica presente en cada sector y a su vez conforma una base para la toma de decisiones coordinadas que aseguren la calidad y robustez del sistema.

Finalmente se debe recordar que un alto porcentaje de las interrupciones y problemas ocasionados a los clientes se deben a fallas en el sistema de distribución de energía eléctrica, por esta razón del total de la inversión realizada en los sistemas eléctricos una gran parte es destinada a la distribución eléctrica. Así pues, un buen control, mantenimiento y sobre todo una buena planificación conlleva a un mejor aprovechamiento de los recursos técnicos y económicos además del aumento en las tasas de eficiencia de las empresas de distribución, lo cual se resume en un mejoramiento del servicio eléctrico en el cual los consumidores son los más beneficiados.



1.2. ANTECEDENTES

La expansión de los sistemas de generación, transmisión y distribución dependen principalmente de la organización y de la proyección del consumo de energía eléctrica que se tenga en cada área, sin embargo, hasta inicios de la década de los 60's el desarrollo del sistema eléctrico ecuatoriano fue totalmente desorganizado. En esos años los municipios de cada ciudad eran los encargados del abastecimiento eléctrico pero el trabajo que realizaron fue calificado por muchos como deficiente y falta de planificación pues tan solo el 35% de la población tenía acceso al servicio eléctrico, esto sumado al hecho de que existieron 1100 centrales térmicas de generación, las mismas que producían tan solo 164 MW daban como resultado un escenario nada alentador (Flores Taipe, 2011).

Este panorama obligo al Estado a emitir el Decreto de Emergencia No.24 mediante el cual se creó el Instituto Nacional de Electrificación (INECEL). Este organismo se formó en el año de 1961 con el claro objetivo de ser la entidad encargada de crear un sistema eléctrico flexible y eficiente, capaz de adaptarse a las condiciones de crecimiento de la población, además tenía en sus manos la realización de los estudios necesarios para la expansión de la capacidad de generación de energía en el país.

El Instituto Nacional de Electrificación conjuntamente con técnicos ecuatorianos y extranjeros fueron los que dieron forma a lo que hoy se conoce como sector eléctrico ecuatoriano. Una de las primeras acciones del INECEL fue recopilar la información necesaria del sistema de energía, para lo cual llevo a cabo el primer censo energético del Ecuador. Una vez obtenidas las estadísticas del sector eléctrico, este organismo planificó y elaboró el primer Plan Maestro de electrificación (1967-1976). En este documento se contemplaba la creación de proyectos tales como el sistema nacional interconectado (SNI), el anillo de transmisión y por el lado de la distribución de energía, se planteaba la modificación de los sistemas regionales existentes para de esta manera obtener una empresa de distribución en cada provincia. Este Plan Maestro no pudo cumplir con sus objetivos pues para hacerlo el estado debía abonar alrededor de \$ 285.61 millones, pero la severa crisis política y económica que atravesaba el país en esos momentos hizo que las finanzas del estado sean insuficientes para cubrir tal monto (Arosemena, 2008).

Posteriormente, entre los años 1972 y 1983 se obtuvieron grandes ingresos gracias a los altos precios del petróleo, estos recursos dieron paso para que el INECEL arranque con la creación del segundo Plan Maestro de electrificación (1973-1980) el cual estaba enfocado en garantizar el suministro de energía a bajo costo (Moncada, 1974). Durante el tiempo que tuvo vigencia este segundo plan nacional se pudieron cumplir todos sus objetivos y la actividad eléctrica en el país tuvo un crecimiento de 16.7% anual, lo que dio paso a la construcción de obras de gran relevancia tales como Paute, Agoyán, Pisayambo y Termoesmeraldas, por otro lado, se concluyó el anillo de transmisión lo cual incremento a un 70% los indicadores de cobertura nacional.

En estos años, el sector eléctrico se convirtió en un actor notable en la economía, lo que permitió la modificación de los sistemas de distribución pasando de siete operadoras regionales a 20 operadoras a nivel nacional. Sin embargo, en 1983 el país cayó nuevamente en una crisis económica lo cual puso fin a los recursos monetarios del sector eléctrico y por ende significó el cese operativo de INECEL.

En 1988, INECEL presentó su último Plan Maestro de Electrificación (1989-2000) en el cual se habla de los inconvenientes que se suscitaron por la falta de recursos, y por otro lado, pone en evidencia uno de los mayores problemas que se gestó en esas décadas: las pérdidas de eficiencia, por esta razón ese documento pone especial cuidado en demostrar el crecimiento y los posibles inconvenientes que podría acarrear la no erradicación de este gran problema. Este Plan Maestro indica también que se deberá mantener la política energética renovable y que esto sería un punto clave para la expansión de la generación en el sector eléctrico y a su vez permitirá cubrir la demanda rural, el cuál era un tema que no fue muy discutido en esos años.



Con el pasar del tiempo la situación del país fue decayendo a tal punto que la demanda eléctrica se volvió insostenible, esta realidad fue evidenciada por la ciudadanía en las temporadas de apagones (1992, 1993 y 1996) que generaron malestar en las personas y le costó al país cerca de \$1 800 millones (Flores Taípe, 2011). A partir de ese punto, en las temporadas de escasa lluvias Ecuador se veía obligado a declarar estados de emergencia para poder buscar soluciones rápidas a este gran problema, esto condujo a que se implementen más centrales de generación termoeléctrica dejando a un lado los proyectos hidroeléctricos elaborados por el INECEL.

Tras haber superado la primera crisis energética en el año de 1992, el sistema eléctrico ecuatoriano volvió a sufrir un colapso en el año de 1993. Esta segunda crisis energética reforzó la idea de privatizar las empresas eléctricas con el fin de conseguir un mejor servicio y mejores utilidades para el país. Por esta razón, en el año de 1996 se promulgó la ley de Régimen del Sector Eléctrico (LRSE) la misma que impulsaba la privatización de dichas entidades, pero que no sería recibida como se esperaba por parte de los inversionistas.

Tras más de una década de soluciones momentáneas, como lo fue el aumento del parque generador termoeléctrico y de políticas enfocadas a la privatización de las empresas, el Ecuador se había mantenido en el subdesarrollo por su modelo económico. No obstante, tras la entrada del gobierno de turno en el año 2007, se trataría de modificar este panorama, para lo cual se elaboró una nueva constitución la misma que devolvió al estado la administración, gestión, regulación y control del sector energético, considerado uno de los sectores estratégicos para la economía del país.

Es así como el país volvió a un modelo semejante al implementado por INECEL en el cual el estado sería la única entidad encargada de la provisión de energía, con la diferencia que se reconocía la concesión a empresas privadas solo en casos excepcionales. El principal objetivo detrás de todo este cambio institucional fue la transformación de la matriz energética, aprovechando así los enormes recursos renovables del país, especialmente la capacidad hídrica con la cual se esperaba generar el 80% de la producción nacional. Todo este escenario dio paso a la creación de un nuevo Plan Maestro de Electrificación (2013-2022) el cual presta especial atención a la proyección de la demanda energética en el país con el objetivo de mejorar los índices de eficiencia de las empresas de distribución, con ello se esperaba también una reducción natural de las tarifas de energía eléctrica y de esta manera reducir el subsidio del estado, por otro lado se planeó que el aumento de la oferta eléctrica a bajos precios favorecería al crecimiento económico y a la inversión extranjera directa (Conelec, 2013).

Cabe destacar que alrededor del año 1998, el Consejo Nacional de Electricidad (CONELEC) aprobó y publicó el primer Plan de Electrificación (2000-2009) elaborado en base a la ley de Régimen del Sector Eléctrico (LRSE). En este plan se incorporó la Regulación No. CONELEC 003/99 de reducción anual de pérdidas no técnicas en las empresas de distribución, la misma que pretendía tener un 2% de pérdidas aceptables para el 2002 y de esa manera llegar a un total de pérdidas en distribución de un 10% en 2009.

Como era de esperar, tanto la regulación 003/99 emitida por CONELEC como el Plan Maestro de Electrificación (2013-2022) impulsaron a las empresas de distribución a optimizar su gestión reduciendo los niveles de pérdidas y mejorando los procesos de planificación, para de esta manera mejorar sus índices de eficiencia cumpliendo así con los estándares dictaminados por las leyes.

1.3. OBJETIVOS

1.3.1. Objetivo general

Tomando como base la información aportada por la empresa eléctrica Azogues además del Sistema de Información Geográfica (GIS) se segmentará el mercado de consumo eléctrico de la empresa eléctrica Azogues basándose en la demanda energética de cada zona, para de esta manera elaborar un mapa de consumo de energía eléctrica.



1.3.2. Objetivos específicos

- ♣ Estimar el tamaño de la cuadrícula de consumo y con ello realizar la cuadriculación de la zona de concesión general.
- ♣ Estimar la densidad de demanda por cuadrícula de consumo para toda la zona de concesión.
- ♣ Elaborar un mapa de estratos según las demandas promedio de cada zona.
- ♣ Validar las zonas determinadas mediante técnicas que permitan corroborar los resultados obtenidos.

1.4. ALCANCES DEL PROYECTO

Partiendo del análisis del consumo mensual de energía de cada usuario para el período comprendido entre 2016 y 2017 y de las coordenadas de cada medidor de energía, proporcionados por la empresa eléctrica Azogues, además de la información existente en el Sistema de Información Geográfica (GIS) y del sistema SICO, se agrupará a los usuarios de cada tipo de tarifa (residencial, comercial e industrial) en estratos para lo cual se utilizará el método de Dalenius-Hodges, además de los métodos jerárquicos conglomerados y el método *k-means* (método de agrupación) los cuáles ayudarán en la determinación del número óptimo de estratos en los que se deberán dividir los usuarios para cada tipo de tarifa, además de los límites que deberán tener cada uno de estos estratos.

Esto a su vez servirá de punto de partida para georreferenciar a cada usuario mediante el programa ArcGis, y posteriormente definir el tamaño de cuadrícula óptimo para toda la zona de concesión. Esto con el objetivo de elaborar un mapa de estratos en el cual se pueda analizar de una manera fácil el consumo medio de energía en cada zona geográfica.

Finalmente, se elaborará un mapa de densidades de carga con lo cual se evaluará la concentración de usuarios en las distintas áreas de la zona de concesión. Además, se calculará la demanda máxima unitaria para cada estrato de usuarios residenciales y con ello se elaborará un cuadro de demandas máximas diversificadas.

1.5. JUSTIFICACIÓN

El sobredimensionamiento que se ha dado al aplicar las guías de diseño actuales de la Empresa Eléctrica Azogues, sobre todo en el caso de los transformadores para clientes residenciales, es un tema de gran importancia económica y técnica para esta empresa. Este panorama sumado a los lineamientos planteados por el Plan Maestro de Electrificación (2013-2022) hacen necesaria la revisión de las metodologías usadas y sobre todo generan la necesidad de evaluar detenidamente el comportamiento de la carga.

Por tales necesidades se vuelve indispensable, como primer paso, la zonificación del mercado eléctrico de esta empresa a fin de analizar el consumo, comportamiento y densidad de usuarios en cada zona del área de concesión. A partir de estos datos se podrán implementar cambios en las guías de diseño, lo cual repercute directamente en la reducción de costos para las inversiones futuras, además de la obtención de un mejoramiento de los índices de confiabilidad y de servicio proporcionados por parte de la Empresa Eléctrica de Azogues.

CAPÍTULO 2: MARCO TEÓRICO

La planificación para los sistemas de distribución es un tema primordial cuyo eje fundamental es la estimación de la demanda que tienen los usuarios en cada una de las zonas del área de concesión. Por tal motivo, es importante conocer a profundidad las características y conceptos de esta magnitud para posteriormente examinar con mayor facilidad los algoritmos y etapas de análisis que son usados para la zonificación del mercado de consumo de la empresa eléctrica Azogues.

2.1. CONCEPTOS GENERALES

2.1.1. Definiciones eléctricas básicas

Demanda: Es definida como la cantidad de energía promedio que el usuario recibe en un periodo de tiempo específico. Esta magnitud puede ser dada en kilowatts (kW), kilovars (kVAR), kiloampers (kA) o kilovoltampers (kVA). Matemáticamente está dada por la ecuación 2.1 (Gönen, 2014).

$$D = \frac{E}{t_2 - t_1} \quad (2.1)$$

En donde:

D = Demanda.

E = Energía.

t_2 = Tiempo final del periodo de medición.

t_1 = Tiempo inicial del periodo de medición.

Si se toma en cuenta que la energía demandada por una carga puede ser expresada como la sumatoria de la potencia consumida en cada instante de tiempo, entonces puede ser calculada mediante una integral en el tiempo, tal como se muestra en la ecuación 2.2:

$$E = \int_{t_1}^{t_2} p(t) dt \quad (2.2)$$

En donde:

E = Energía.

$p(t)$ = Potencia instantánea.

t_2 = Tiempo final del periodo de medición.

t_1 = Tiempo inicial del periodo de medición.

Si se reemplaza la ecuación 2.2 en el numerador de la ecuación 2.1, la demanda puede ser expresada como lo indica la ecuación 2.3.

$$D = \frac{E}{t_2 - t_1} = \frac{\int_{t_1}^{t_2} p(t) dt}{t_2 - t_1} \quad (2.3)$$

Potencia: Es una de las magnitudes fundamentales de la física y está definida como la rapidez con la que se consume energía o se realiza un trabajo. En el contexto eléctrico, esta magnitud es dividida en potencia instantánea, promedio, activa, reactiva y aparente (Westinghouse Electric Corporation, 1959).



Potencia instantánea: Está definida como la potencia suministrada a una instalación o a un sistema en un instante de tiempo determinado. En un contexto matemático se la expresa como la relación entre el voltaje y la corriente instantánea, tal como se muestra en la ecuación 2.4 (Westinghouse Electric Corporation, 1959):

$$p(t) = v(t) * i(t) \quad (2.4)$$

En donde:

$p(t)$: Potencia instantánea

$v(t)$: Voltaje instantáneo

$i(t)$: Corriente instantánea

Potencia activa: Denominada también como potencia real; es aquella potencia “útil” que es aprovechada por los sistemas para generar trabajo. Esta es la verdadera potencia contratada y la cual se registra en los medidores. Matemáticamente se la expresa como lo indica la ecuación 2.5 (Westinghouse Electric Corporation, 1959):

$$P = VI \cos\varphi \quad (2.5)$$

En donde:

P: Potencia activa.

V: Voltaje.

I: Corriente.

φ : Ángulo de desfase entre voltaje y corriente.

Potencia reactiva: Esta potencia es consumida por cualquier elemento que contenga bobinas (motores, transformadores, etc.) para generar su campo magnético. Este tipo de potencia fluctúa por la red y no genera trabajo como tal. Su unidad de medida es el VAR (voltamperio reactivo) y se puede calcular matemáticamente mediante la ecuación 2.6 (Westinghouse Electric Corporation, 1959):

$$Q = VI \sin\varphi \quad (2.6)$$

En donde:

Q: Potencia reactiva

V: Voltaje

I: Corriente

φ : Ángulo de desfase entre voltaje y corriente

Potencia aparente: Está definida como la potencia total que llega a la carga y, matemáticamente, se la obtiene del producto fasorial del voltaje y la corriente, tal como se presenta en la ecuación 2.7 (Westinghouse Electric Corporation, 1959):

$$\tilde{S} = \dot{V} I^* \quad (2.7)$$

Si se toman en cuenta las ecuaciones 2.8 y 2.9 que describen los fasores de voltaje y corriente eléctrica respectivamente

$$\dot{V} = V e^{j\varphi_v} \quad (2.8)$$

$$I^* = I e^{j\varphi_i} \quad (2.9)$$

Entonces la ecuación 2.7 queda expresada mediante las ecuaciones 2.10 o 2.11.

$$\tilde{S} = V e^{j\varphi_v} * I e^{j\varphi_i} \quad (2.10)$$

$$\tilde{S} = VI e^{j(\varphi_v - \varphi_i)}$$

$$\tilde{S} = VI e^{j\varphi}$$

$$\tilde{S} = VI (\cos\varphi + j\text{sen}\varphi) \quad (2.11)$$

$$\tilde{S} = VI \cos\varphi + j VI \text{sen}\varphi$$

En donde:

\tilde{S} = Potencia aparente.

V = Voltaje.

I = Corriente eléctrica.

φ = Desfase entre voltaje y corriente.

Descrita de otra manera, la potencia aparente puede ser obtenida mediante la suma vectorial de la potencia real y la potencia reactiva, tal como se expresa en la ecuación 2.12.

$$\tilde{S} = (\text{Potencia activa}) + j (\text{Potencia reactiva}) \quad (2.12)$$

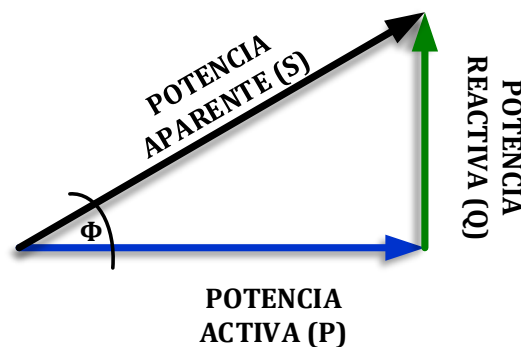


Figura 2.1. Triángulo de potencias.
(Fuente: Propia)

La figura 2.1 muestra el vector de potencia reactiva perpendicular al vector de potencia activa. La suma de estos dos vectores da como resultante al vector de potencia aparente, el mismo que tiene un ángulo φ con respecto al vector de potencia activa, dicho ángulo es muy usado para el cálculo del factor de potencia.

Potencia media (o Potencia promedio): Está definida como el promedio de la potencia activa en un periodo de tiempo definido. Matemáticamente se expresa con la ecuación 2.13 (Pérez Miñaca & Salazar Yépez, 2014):

$$P = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} p(t) dt \quad (2.13)$$

En donde:

P : Potencia media.

t_1 : Tiempo inicial del intervalo de medición.

t_2 : Tiempo final del intervalo de medición.

$p(t)$: Potencia instantánea.

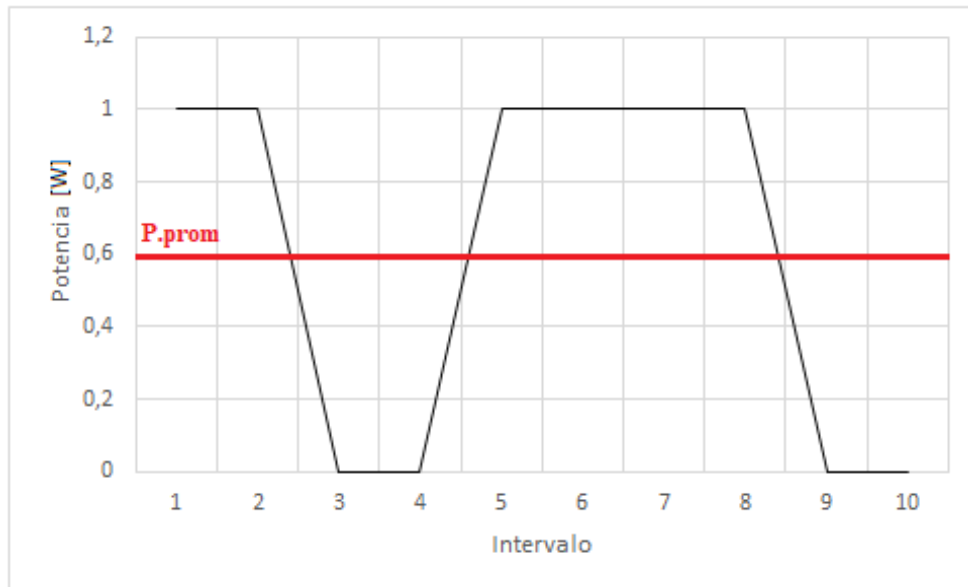


Figura 2.2. Representación gráfica de la demanda promedio de una carga intermitente (Fuente: Propia)

La figura 2.2 muestra el consumo de una carga intermitente cuya demanda de potencia varía de 0 a 1W en ciertos intervalos de tiempo. Si se considera la ecuación 2.13, la potencia promedio de esta carga en el intervalo de tiempo mostrado es de 0.6W.

2.1.2. Conceptos relacionados a la demanda

Carga: En lo referente a sistemas de distribución, se entiende por carga al elemento o sistema que consume energía (Pérez Miñaca & Salazar Yépez, 2014).

Intervalo de demanda: Es el periodo de tiempo sobre el cuál la demanda es promediada. Los intervalos de tiempo más usados son de 15 minutos, 30 minutos y 1 hora. Cuando se habla de demanda debe estar implícito el intervalo de demanda (Gönen, 2014). La figura 2.2 muestra un ejemplo claro de la importancia de este concepto.

Curva de carga: La curva de carga es la representación gráfica de la variación de demanda que se tiene en el sistema para un periodo de tiempo específico. Este tipo de gráficas puede aportar mucha información acerca del elemento estudiado tales como los picos de demanda y su duración. Las más comunes son las curvas de carga diarias, mensuales y anuales (Gönen, 2014).

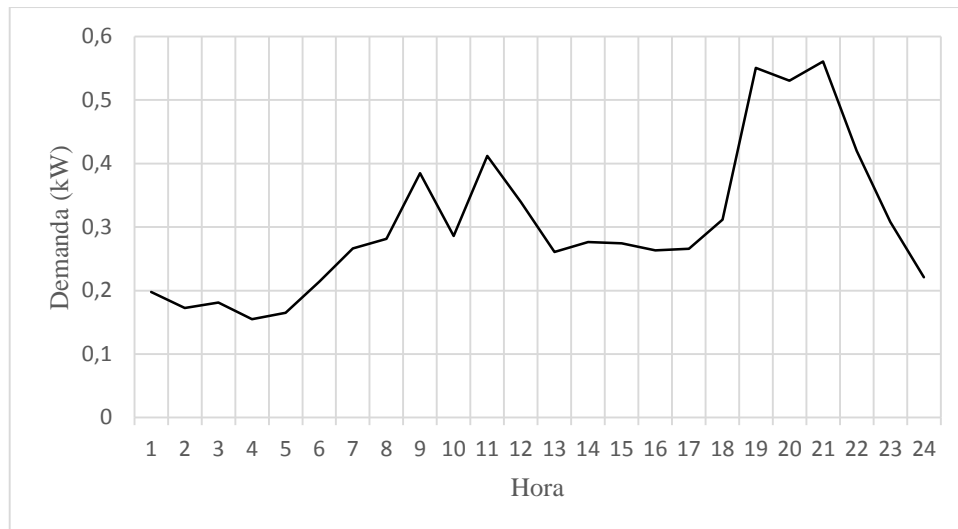


Figura 2.3. Curva de carga diaria para usuarios residenciales de la EEA.
(Fuente: Propia)

La figura 2.3 muestra la curva de carga típica para clientes con tarifa residencial de la Empresa Eléctrica Azogues en un día laborable común. Esta es una curva de carga diaria con intervalos de 1 hora la misma que muestra de una manera sencilla la variación de demanda según la hora del día, con lo cual se puede estimar una serie de características tales como:

- La demanda máxima del usuario
- La potencia promedio consumida.

Demanda máxima: Se refiere a la magnitud más grande de demanda que se ha presentado en un periodo de tiempo específico (Gönen, 2014). En la figura 2.4 se muestra la demanda máxima presente en la curva de carga típica para usuarios residenciales de la EEA, la misma que alcanza los 0.56kW a las 21h00.

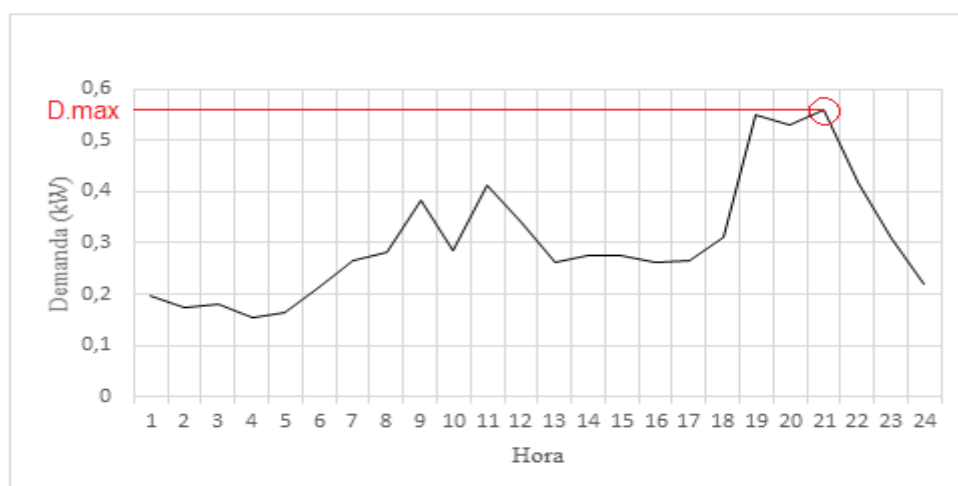


Figura 2.4. Demanda máxima presente en la curva de carga para usuarios residenciales de la EEA.
(Fuente: Propia)

Demanda máxima unitaria: Es la demanda máxima de un solo elemento del sistema o circuito, en lo referente a sistemas de distribución representa la demanda máxima de un cliente en particular (Pérez Miñaca & Salazar Yépez, 2014).



Demanda diversificada (o demanda coincidente): Es la demanda total de un grupo de cargas tomadas como un todo, sobre un periodo de tiempo. El valor de la máxima demanda diversificada juega un papel muy importante dentro de la ingeniería de distribución pues, este es uno de los datos principales con los que se lleva a cabo el dimensionamiento del sistema (Westinghouse Electric Corporation, 1959).

Carga instalada: Se trata de la suma de los valores nominales de potencia de los equipos que se encuentran conectados a un circuito determinado. Generalmente se encuentra expresado en kVA y MVA (Pérez Miñaca & Salazar Yépez, 2014).

Factor de utilización: Este factor cuantifica cuan bien se está aprovechando un dispositivo o un sistema eléctrico. Matemáticamente puede ser expresado como el cociente entre la demanda máxima y la capacidad nominal del sistema o dispositivo, tal como lo indica la ecuación 2.14 (Gönen, 2014).

$$F_u = \frac{D_{max}}{C_n} \quad (2.14)$$

En donde:

F_u = Factor de utilización.

D_{max} = Demanda máxima del sistema.

C_n = Capacidad nominal de sistema.

Factor de demanda: El factor de demanda representa la cantidad de elementos que se encuentran consumiendo energía cuando se da la demanda pico en el sistema (Westinghouse Electric Corporation, 1959). En términos matemáticos, se refiere a la relación entre la máxima demanda y la carga total conectada (ecuación 2.15).

$$F_{dem} = \frac{D_{max}}{C_{inst}} \quad (2.15)$$

En donde:

F_{dem} = Factor de demanda.

D_{max} = Demanda máxima del sistema

C_{inst} = Carga instalada o potencia nominal instalada en el sistema.

Factor de carga: Es el cociente entre la demanda promedio sobre un periodo de tiempo y el pico de demanda ocurrido en ese periodo (ecuación 2.16). Esta magnitud representa cuán bien se han aprovechado las instalaciones. El valor óptimo de este factor es uno (Gönen, 2014).

$$F_{carga} = \frac{D_{prom}}{D_{max}} \quad (2.16)$$

En donde:

F_{carga} = Factor de carga.

D_{prom} = Demanda promedio en un intervalo de tiempo.

D_{max} = Demanda máxima del sistema.

Se debe tener en cuenta que mientras más extenso es el periodo de medición el resultado es menor, esto es debido a que para un mismo valor pico de demanda, el consumo de energía cubre un periodo de tiempo más largo.

Factor de diversidad: Es definido como el cociente entre la suma de las demandas máximas individuales y la máxima demanda del sistema completo (Gönen, 2014). Matemáticamente se lo expresa mediante la ecuación 2.17:

$$F_{div} = \frac{D_{1\ max} + D_{2\ max} + D_{3\ max} + \dots + D_{n\ max}}{D_{max.\ total}} = \frac{\sum_1^n D_i\ max}{D_{max.\ total}} \quad (2.17)$$

En donde:

F_{div} = Factor de diversidad.

$D_i\ max$ = Demanda máxima de cada carga.

$D_{max.\ total}$ = Demanda máxima del sistema.

Si se considera la ecuación 2.15, entonces la demanda máxima puede ser expresada a través de la ecuación 2.18:

$$D_{max} = C_{inst} * F_{dem} \quad (2.18)$$

Por lo tanto, el factor de diversidad puede ser obtenido también con la ecuación 2.19:

$$F_{div} = \frac{\sum_1^n C_{inst} * F_{dem}}{D_{max.\ total}} \quad (2.19)$$

Es preciso resaltar que este factor siempre es igual o mayor que uno.

Factor de coincidencia (simultaneidad): Es el cociente entre la demanda coincidente máxima de un grupo de consumidores y la suma de las demandas máximas de energía de cada uno de los consumidores en un periodo de tiempo específico (Gönen, 2014). Esta definición es expresada matemáticamente con la ecuación 2.20.

$$F_c = \frac{D_{max.\ total}}{\sum_1^n D_i\ max} \quad (2.20)$$

En donde:

F_c = Factor de coincidencia.

$D_i\ max$ = Demanda máxima de cada carga.

$D_{max.\ total}$ = Demanda máxima del sistema.

Tal como lo indica la ecuación 2.20, el factor de coincidencia es el recíproco del factor de diversidad, por consiguiente, puede también ser expresado con la ecuación 2.21:

$$F_c = \frac{1}{F_{div}} \quad (2.21)$$

Densidad de carga: Se refiere a la concentración de carga que existe en un área, en otras palabras, es la demanda por unidad de área. Generalmente se lo expresa en $\frac{kW}{km^2}$, $\frac{kVA}{km^2}$ o en ocasiones $\frac{kVA}{m}$, si se considera una construcción lineal (Pérez Miñaca & Salazar Yépez, 2014).



2.1.3. Conceptos básicos estadísticos

Media aritmética: Conocida simplemente como media, es definida como el promedio de los valores de una muestra y se la expresa matemáticamente utilizando la ecuación 2.22 (Walpole, Myers, & Myers, 2012):

$$\bar{x} = \sum_{i=1}^n \frac{x_i}{n} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (2.22)$$

En donde:

\bar{x} = Media de la muestra.

x_i = Valor de cada elemento de la muestra.

n = Número de elementos de la muestra.

Cuando la media toma en cuenta a toda la población se la denota como μ y cuando solo toma en cuenta a una muestra de la población se utiliza \bar{x} . Para el caso de datos agrupados (tabla de frecuencias) la media se calcula con la ayuda de la ecuación 2.23:

$$\bar{x} = \frac{\sum_{i=1}^k m_i f_i}{n} \quad (2.23)$$

En donde:

m_i = Valor medio de la clase i .

f_i = Frecuencia de la clase.

k = número de intervalo de clases.

Media Geométrica: Es utilizada generalmente en porcentajes o tasas de crecimiento y se calcula por medio de la ecuación 2.24 (Walpole, Myers, & Myers, 2012):

$$MG = \sqrt[n]{x_1 x_2 \dots x_n} \quad (2.24)$$

En donde:

MG = Media Geométrica.

x_i = Valor de la i -ésima muestra.

n = Tamaño de la muestra.

Cabe señalar que la media geométrica es siempre igual o menor a la media aritmética.

Media Armónica: La media armónica se suele utilizar para promediar variaciones con respecto al tiempo, siendo una de sus ventajas el ser poco influenciada por valores irregularmente grandes en comparación al resto de la muestra, pero si por valores pequeños. Está expresada matemáticamente por la ecuación 2.25 (Walpole, Myers, & Myers, 2012).

$$MA = \frac{1}{\frac{1}{n} \left(\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n} \right)} \quad (2.25)$$

En donde:

MA = Media armónica.

x_i = Valor de la i -ésima muestra.

n = Tamaño de la muestra.

La media armónica es siempre menor o igual a la media geométrica y, por tanto, siempre es menor o igual a la media aritmética.

Mediana: Representa el valor medio de una muestra ordenada, en otras palabras cumple con que la mitad de datos están por encima y la otra mitad por abajo de este valor. Una de sus principales características es que no se ve afectada por los valores extremos y matemáticamente se expresa mediante la ecuación 2.26 (Walpole, Myers, & Myers, 2012).

$$\tilde{x} = \begin{cases} x_{(n+1)/2} & \text{si } n \text{ es impar} \\ \frac{1}{2} \left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right) & \text{si } n \text{ es par} \end{cases} \quad (2.26)$$

En donde:

\tilde{x} = Mediana.

x_i = Valor del elemento i -ésimo de la muestra.

n = Longitud de la muestra

Moda: Se refiere al valor de la muestra que se repite con mayor frecuencia. Así pues, se habla de una distribución bimodal cuando existen dos modas o de una distribución trimodal cuando hay tres datos que se repiten con una frecuencia absoluta similar (Cowan, 1998).

Cuantiles: Se define a los cuantiles de orden k como los valores de una variable ordenada, que la dividen en k partes iguales. Así pues, el cuantil 20 de orden 100 deja por debajo de sí el 20% del total de valores (Walpole, Myers, & Myers, 2012).

Los cuantiles más utilizados son:

1. **Percentiles:** Son los 99 puntos que dividen a la muestra en 100 partes iguales, tales que dentro de cada percentil se alberga el 1% de total de la muestra.
2. **Cuartiles:** Son los 3 valores que dividen a la muestra en 4 partes iguales, es decir, se divide a la muestra en 4 intervalos dentro de los cuáles se alberga el 25% del total de la muestra. Así pues, el cuartil 1 (Q_1) es igual al percentil 25 (P_{25}) y el cuartil 2 (Q_2) es igual al percentil 50 (P_{50}) el cual también es conocido como mediana.
3. **Deciles:** Son los 9 puntos que dividen a la muestra en 10 partes iguales, de tal manera que, dentro de cada decil se albergue el 10% de la muestra total.

Varianza: Es una medida de dispersión definida como la esperanza del cuadrado de la desviación de dicha variable respecto a su media. Se expresa matemáticamente a través de la ecuación 2.27 y su unidad de medida es el cuadrado de la unidad de medición de la variable (Cowan, 1998).

$$s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n - 1} \quad (2.27)$$

En donde:

x_i = i -ésimo elemento de la variable estudiada.

\bar{x} = Media de la muestra

n = Número de elementos de la muestra.

Desviación estándar: Corresponde a la raíz cuadrada positiva de la varianza e indica que tan dispersos se encuentran los datos con respecto a la media. Se expresa matemáticamente mediante la ecuación 2.28 (Cowan, 1998).

$$\sigma = \sqrt{s^2} = \sqrt{\sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n - 1}} \quad (2.28)$$

2.2. MÉTODOS DE ANÁLISIS CLUSTER

El *análisis cluster* es una técnica multivariante cuya idea básica es formar grupos de elementos (o variables) llamados clusters o conglomerados, con el objetivo de conseguir la máxima homogeneidad en cada grupo y al mismo tiempo la mayor heterogeneidad entre grupos, en otras palabras los elementos al interior de cada grupo deben ser muy semejantes pero los grupos deben ser distintos entre sí (Aggarwal & Reddy, 2014).

Básicamente, el proceso que se sigue para realizar un análisis cluster es el siguiente:

- i. Partir de un conjunto de n individuos con m variables
- ii. Establecer un criterio de similitud que permita relacionar la semejanza de los individuos entre sí. Existen una gran cantidad de índices sin embargo los más usados son aquellos basados en la distancia, los cuáles toman en cuenta los elementos de la variable estudiada como componentes de un vector, así pues una distancia pequeña entre elementos denota una alta similitud y viceversa.
- iii. Escoger un algoritmo de agrupación. Estos algoritmos se pueden clasificar en dos grandes grupos:
 - a. **Algoritmo de partición:** Divide los elementos en k grupos (clusters), en donde k es definido previamente.
 - b. **Algoritmos jerárquicos:**
 - **Jerárquicos aglomerativos:** En este caso, se parte con todos los elementos de la variable analizada y en cada iteración se une un elemento a un conglomerado, la operación se repite hasta que todos los elementos se engloben en un solo cluster .
 - **Jerárquicos disociativos:** Sigue un sentido inverso al algoritmo aglomerativo pues se parte de un solo cluster que engloba todos los elementos y en cada iteración se dividen los clusters hasta al final obtener tantos agrupamientos como elementos tenga la variable analizada.
- iv. Especificar una estructura mediante diagramas arbóreos, dendogramas, u otros métodos gráficos.

2.2.1. Medidas de distancia para individuos

Sean, $i = (x_{i1}, x_{i2}, \dots, x_{ip})$ y $j = (x_{j1}, x_{j2}, \dots, x_{jp})$ vectores de dimensión p , en donde x_{in} y x_{jn} con $n=1,2,3,\dots,p$ denotan los elementos de cada vector. La distancia entre dichos elementos puede ser dada por (Pérez Miñaca & Salazar Yépez, 2014):

- **Distancia euclidiana:** Es la distancia más popular y puede ser calculada mediante la ecuación 2.29:

$$d(i, j) = \sqrt{|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{ip} - x_{jp}|^2} \quad (2.29)$$

- **Distancia Manhattan:** Está expresada mediante la ecuación 2.30:

$$d(i, j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{ip} - x_{jp}| \quad (2.30)$$

- **Distancia Minkowski:** Es una generalización de la distancia Euclidiana y Manhattan y está definida por la ecuación 2.31:

$$d(i, j) = (|x_{i1} - x_{j1}|^q + |x_{i2} - x_{j2}|^q + \dots + |x_{ip} - x_{jp}|^q)^{1/q} \quad (2.31)$$

Así pues, si $q=1$ se obtiene la distancia Manhattan y si $q=2$ se obtiene la distancia Euclidiana. El principal inconveniente con este método es que la distancia de los datos de mayor magnitud tiende a dominar el resultado. Para contrarrestar esta desventaja es común que se normalicen los datos dentro de un rango predefinido.

- **Distancia Mahalanobis:** Surge como una solución al inconveniente de la distancia Minkowski pues aquí se le asigna diferentes pesos a cada característica basándose en la varianza y en la correlación lineal de los pares. Así pues, se tiene que la distancia se encuentra definida matemáticamente por la ecuación 2.32:

$$d(i, j) = \sqrt{(x_i - x_j)s^{-1}(x_i - x_j)} \quad (2.32)$$

En donde x_i y x_j son vectores fila y s^{-1} se entiende como la matriz de covarianza de los patrones. Por otro lado, si a cada variable se le asigna un peso de acuerdo con su importancia se puede calcular la nueva distancia ponderada con la ecuación 2.33:

$$d(i, j) = \sqrt{w_1|x_{i1} - x_{j1}|^2 + w_2|x_{i2} - x_{j2}|^2 + \dots + w_p|x_{ip} - x_{jp}|^2} \quad (2.33)$$

2.2.2. Algoritmos de agrupación

2.2.2.1. Métodos jerárquicos aglomerativos

Los métodos aglomerativos, o conocidos también como ascendentes, comienzan el análisis con tantos grupos como individuos se tenga. A partir de estos individuos se elaboraron grupos (clusters) de forma ascendente, hasta que al final se cumpla uno de los siguientes enunciados (Hair, Black, Babin, & Anderson, 2010):

- Formar un solo grupo conglomerado
- Alcanzar un número de grupos prefijado
- Detectar con base estadística que los grupos más similares no son lo suficientemente homogéneos como para unirse en una misma agrupación

Para lograr tal objetivo se utiliza cualquiera de los siguientes métodos o estrategias:

- **Estrategia de la distancia mínima o similitud máxima:** También conocida como el método de amalgamamiento simple (single linkage). En este método se tiene que la distancia o similitud entre clusters viene dada por la mínima distancia (máxima similitud) entre componentes.

Se tiene que la distancia entre el cluster C_i y el cluster C_j está dado por la ecuación 2.34.

$$d(C_i, C_j) = \min_{\substack{x_l \in C_i \\ x_m \in C_j}} \{d(x_l, x_m)\} \quad l = 1, \dots, n_i ; m = 1, \dots, n_j \quad (2.34)$$

En donde:

x_l = Elementos del cluster C_i .

x_m = Elementos del cluster C_j .

n_i = Número de elementos del cluster C_i .

n_j = Número de elementos del cluster C_j .

- **Estrategia de la distancia máxima o similitud mínima:** En este procedimiento, también conocido como proceso de amalgamamiento completo (*complete linkage*), se considera que la distancia (similitud) entre dos clusters viene dada por sus elementos más dispares, en otras palabras la distancia entre clusters está dada por la máxima distancia entre sus componentes tal como se expresa en la ecuación 2.35.

$$d(C_i, C_j) = \max_{\substack{x_l \in C_i \\ x_m \in C_j}} \{d(x_l, x_m)\} \quad l = 1, \dots, n_i ; m = 1, \dots, n_j \quad (2.35)$$

- **Estrategia de la distancia (similitud) promedio no ponderada:** Aquí la distancia o similitud del cluster C_i con el cluster C_j se obtiene como la media aritmética entre la distancia de los componentes de cada cluster. Así pues, el cluster C_i está compuesto a su vez por dos clusters C_{i1} y C_{i2} (con n_i y n_j elementos respectivamente) y el cluster C_j posee n_j elementos, de tal manera que la distancia es calculada utilizando la ecuación 2.36:

$$d(C_i, C_j) = \frac{d(C_{i1}, C_j) + d(C_{i2}, C_j)}{2} \quad (2.36)$$

- **Estrategia de la distancia (similitud) promedio ponderada:** Esta estrategia es, hasta cierto punto, similar a la anterior teniendo como diferencia que la distancia o similitud entre clusters está dada por el promedio ponderado de las componentes de un cluster respecto al otro. De esta manera y al igual que en el caso anterior se tiene un cluster C_i el cuál se divide en dos clusters C_{i1} y C_{i2} (con n_i y n_j elementos respectivamente) y el cluster C_j que posee n_j elementos, de tal manera que la distancia es determinada con la ecuación 2.37:

$$d(C_i, C_j) = \frac{1}{(n_{i1} + n_{i2})n_j} \sum_{i=1}^{n_{i1}+n_{i2}} \sum_{j=1}^{n_j} d(x_i, x_j) = \frac{n_{i1}d(C_{i1}, C_j) + n_{i2}d(C_{i2}, C_j)}{n_{i1} + n_{i2}} \quad (2.37)$$

- **Métodos basados en el centroide:** En este tipo de método jerárquico la similitud entre dos clusters se encuentra dada por la semejanza de sus centroides. Los centroides están definidos como los vectores de medias, en otras palabras el promedio de los elementos medidos en cada cluster. Existe básicamente dos variantes:

i. Método del centroide ponderado

En este caso los tamaños de los clusters son tomados en cuenta a la hora de efectuar los cálculos. Así pues, si se tiene dos clusters C_i y C_j en donde el cluster C_i se subdivide en C_{i1} y C_{i2} (con n_i y n_j elementos respectivamente). Además, sea m^{i1} , m^{i2} y m^j vectores de dimensión n que representan los centroides de los clusters citados anteriormente.

Entonces, el centroide del cluster C_i se encuentra dado por la ecuación 2.38.

$$m^i = \frac{n_{i1}m^{i1} + n_{i2}m^{i2}}{n_{i1} + n_{i2}} \quad (2.38)$$

En donde:

m^i = Centroide del cluster C_i

n_{ix} = i -ésimo elemento del cluster C_{i1}

n_{jx} = i -ésimo elemento del cluster C_{i2}

m^{ix} = Centroide del cluster C_{ix}

Para determinar las componentes de los centroides se utiliza la ecuación 2.39.

$$m_l^i = \frac{n_{i1}m_l^{i1} + n_{i2}m_l^{i2}}{n_{i1} + n_{i2}} \quad l = 1, \dots, n \quad (2.39)$$

De esta manera se tiene que la distancia euclidiana al cuadrado entre los clusters se encuentra expresada por la ecuación 2.40.

$$d^2(C_i, C_j) = \sum_{l=1}^n (m_l^j - m_l^i)^2 = \sum_{l=1}^n \left(m_l^j - \frac{n_{i1}m_l^{i1} + n_{i2}m_l^{i2}}{n_{i1} + n_{i2}} \right)^2 \quad (2.40)$$

El resultado de la ecuación 2.40 puede ser expresado tal como lo indica la ecuación 2.41.

$$d^2(C_i, C_j) = \frac{n_{i1} d^2(C_{i1}, C_j)}{n_{i1} + n_{i2}} + \frac{n_{i2} d^2(C_{i2}, C_j)}{n_{i1} + n_{i2}} - \frac{n_{i1}n_{i2} d^2(C_{i1}, C_{i2})}{(n_{i1} + n_{i2})^2} \quad (2.41)$$

El inconveniente que se tiene al trabajar con este procedimiento es que si los tamaños (n_{i1} y n_{i2}) de los clusters C_{i1} y C_{i2} son muy distintos entre sí, la componente de mayor tamaño tiende a influenciar de sobre manera el centroide (m^i) del cluster.

ii. Método del centroide no ponderado

Este método es conocido también como método de la mediana y a diferencia del método mostrado previamente aquí no se toma en cuenta el tamaño de los clusters. La estrategia de este método es dividir al cluster C_i en dos elementos de igual dimensión ($n_{i1} = n_{i2}$) provocando que el centroide del cluster C_i se encuentre entre los clusters C_{i1} y C_{i2} lo cual provoca a su vez que el centroide del cluster (C_i, C_j) se encuentre en el punto central o mediana del triángulo formado por los clusters C_{i1}, C_{i2} y C_j . Esta es la única diferencia entre este método y el citado anteriormente, por lo tanto el resto de sus características son las mismas y de esta manera se obtiene la ecuación 2.42.

$$d(C_i, C_j) = \frac{d(C_{i1}, C_j) + d(C_{i2}, C_j)}{2} - \frac{d(C_{i1}, C_{i2})}{4} \quad (2.42)$$

Se debe notar que una de las características más importantes de los métodos basados en el centroide es que el valor de similitud o de distancia asociada con los clusters enlazados puede variar de una etapa a otra.

- **Método de Ward:** Para cada etapa de este algoritmo se unen dos clusters para los cuáles se tenga el menor error en el valor total de la suma de los cuadrados de las diferencias dentro de cada cluster de cada individuo, al centroide del cluster. Así pues, primeramente, se define que:

x_{ij}^k = El valor de la j-ésima variable sobre el i-ésimo individuo del k-ésimo cluster.

m^k = Centroide del cluster k con componentes m_j^k .

E_k = Es la distancia euclidiana al cuadrado entre cada individuo del cluster k a su centroide, en otras palabras, es la suma de cuadrados de los errores del cluster k, la misma que se expresa matemáticamente en la ecuación 2.43.

$$E_k = \sum_{i=1}^{nk} \sum_{j=1}^n (x_{ij}^k - m_j^k)^2 = \sum_{i=1}^{nk} \sum_{j=1}^n (x_{ij}^k)^2 - \sum_{j=1}^n (m_j^k)^2 \quad (2.43)$$

E = Suma de cuadrados de los errores para todos los clusters y se determina mediante la ecuación 2.44

$$E = \sum_{k=1}^h E_k \quad (2.44)$$

El proceso empieza con m clusters, cada uno de los cuáles está compuesto por un solo individuo ($E_k = 0, E = 0$). El objetivo de este método es encontrar en cada etapa aquellos dos clusters cuya unión proporcione el menor incremento en la suma total de errores (E). Una vez que los clusters C_p y C_q se han unido resultando un cluster C_t se tiene que el incremento en E es determinado por la ecuación 2.45.

$$\Delta E_{pq} = E_t - E_p - E_q = n_p \sum_{j=1}^n (m_j^p)^2 + n_q \sum_{j=1}^n (m_j^q)^2 - n_t \sum_{j=1}^n (m_j^t)^2 \quad (2.45)$$

En donde:

E_t = Error total.

E_p = Error del cluster p.

E_q = Error del cluster q.

n_p = Número de elementos del cluster p.

n_q = Número de elementos del cluster q.

n_t = Número de elementos de la unión de los clusters p y q.

La ecuación 2.45 puede ser simplificada hasta obtener la ecuación 2.46.

$$\Delta E_{pq} = \frac{n_p n_q}{n_t} \sum_{j=1}^n (m_j^p - m_j^q)^2 \quad (2.46)$$

Se concluye de la ecuación 2.46 que el menor incremento de los errores cuadráticos es proporcional a la distancia euclidiana al cuadrado, de los centroides de los clusters unidos. Por otro lado E_{pq} no es decreciente y por consiguiente este método no presenta los problemas de los métodos basados en centroides.

2.2.2.2. Métodos jerárquicos disociativos

Los métodos disociativos, también llamados descendentes, constituyen el proceso inverso al de los métodos asociativos. Comienzan con un conglomerado que engloba a todos los elementos y a partir de este grupo inicial se forman grupos cada vez más pequeños a través de divisiones sucesivas. Al final de este proceso se tiene tantas agrupaciones como elementos. Sin embargo, dado que en este documento no se hace uso de este tipo de algoritmos no se profundiza en su explicación.

2.2.3. Algoritmo de partición (K-MEANS)

Uno de los métodos más populares, y generalmente usados para la explicación de los algoritmos de partición en general es el método *k-means*, el cuál es muy utilizado en el análisis exploratorio y minería de datos; pero puede ser empleado en cualquier campo de investigación gracias a la facilidad de implementación, eficiencia computacional y bajo consumo de memoria.

Este método pertenece a las técnicas de agrupamiento basadas en particiones, las cuáles se fundamentan en la relocalización iterativa de datos entre clusters. Su objetivo principal es producir grupos de casos o variables con un alto grado de similitud intergrupual y un bajo nivel de similitud intragrupal. Para lograr tal objetivo el presente algoritmo resuelve un problema de optimización en el que se minimiza la suma de distancias cuadráticas de cada dato, al centroide de su grupo. Matemáticamente el problema puede ser expresado como lo indica la ecuación 2.47 (Hans-Hermann, 2008).

$$\min E(\mu_i) = \min \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \mu_i\|^2 \quad (2.47)$$

En donde, se define un conglomerado S el cual contiene los elementos x_j , los mismos que son divididos en k grupos con su correspondiente centroide μ_i

Para determinar la similitud, denominada en algunos casos como “proximidad”, existen muchos métodos de cálculo entre los cuales destaca la distancia Euclidiana, la distancia Euclidiana cuadrada y la distancia Manhattan, aunque en ciertos casos es usada también la distancia Mahalanobis.

El método k-means presenta una gran cantidad de variantes, sin embargo entre los más utilizados se encuentran el algoritmo de Forgy/Lloyd, el algoritmo de MacQueen y el algoritmo Hartigan & Wong (Morissette & Chartier, 2013).

- **Algoritmo Forgy/Lloyd**

El algoritmo Lloyd (Lloyd, 1982) y el algoritmo Forgy (Forgy, 1965) son ambos modelos de centroides de grupo. Un centroide es definido como el centro geométrico de un objeto y puede ser tomado como la generalización de la media. Los algoritmos de grupo son aquellos en los que un paso transformativo es aplicado a todos los casos al mismo tiempo. Esto es muy adecuado para analizar grandes grupos de datos.



La única diferencia entre el algoritmo Forgy y el algoritmo Lloyd es que este último considera una distribución de datos discreta mientras que el algoritmo Forgy considera una distribución continua de elementos (ecuaciones 2.48 y 2.49 respectivamente). Así pues, para un conjunto de elementos $[x_1, x_2, \dots, x_n] \in R^d$, en donde R^d es el espacio de datos de dimensión d , el algoritmo intenta encontrar un grupo de k centros de clusters $C = [c_1, c_2, \dots, c_k] \in R^d$ que representan una solución a la minimización del problema:

$$E = \sum_{i=1}^k \sum_{j=1}^n d(c_i, x_{ij}), \quad \text{distribución discreta} \quad (2.48)$$

$$E = \sum_{i=1}^k \int \rho(x) d(c_i, x_{ij}) dx, \quad \text{distribución continua} \quad (2.49)$$

En donde $\rho(x)$ es la función densidad de probabilidad y d es la función de distancia.

Los pasos básicos del algoritmo son los siguientes:

- i. Escoger los k centroides iniciales
- ii. Asignar cada dato a un cluster basado en la distancia a los centroides
- iii. Actualizar el valor de los centroides usando la media de los valores de cada cluster (confirmados en el paso ii)

Los pasos ii y iii se repiten hasta que los centroides dejen de variar o hasta que se cumpla un rango de tolerancia definido previamente.

- **Algoritmo MacQueen**

El algoritmo de MacQueen es un algoritmo iterativo y su principal diferencia con el método propuesto por Forgy/Lloyd es que los centroides son recalculados cada vez que el subespacio es cambiado y también después de que se pasa por todos los casos. De esta manera, si el centroide del subespacio al que pertenece actualmente es el más cercano, no se realiza ningún cambio pero si otro centroide es el más cercano, el caso se reasigna al otro centroide y los centroides tanto para el subespacio viejo como para el nuevo se recalculan (MacQueen, 1967).

Cabe señalar que este algoritmo es más eficiente pues actualiza los centroides más frecuentemente y por lo general necesita realizar una iteración completa a través de los casos para poder converger en una solución.

- **Algoritmo Hartigan & Wong**

Este algoritmo busca la partición del espacio de datos con óptimo local basándose en la suma de cuadrados de los errores (SSE) dentro del cluster. Esto significa que puede asignar un caso a otro subespacio aun si este pertenece a un subespacio de un centroide más cercano, de esta manera minimiza la suma de cuadrados dentro del cluster (Hatigan & Wong, 1979). Los pasos básicos del algoritmo son los siguientes:

- i. Escoger los k centroides iniciales.
- ii. Asignar a cada caso al centroide más cercano.
- iii. Calcular la media de cada cluster formado en el paso anterior.

Si los centroides han sido actualizados en el último paso.

- iv. Calcular SSE dentro del cluster.
- v. Si SSE del cluster k es menor al SSE del cluster j (expresión 2.50), el dato cambia de cluster.

$$SSE2 = \frac{N_i \sum_j \|x_{ij} - c_i\|^2}{N_i - 1} < SSE1 = \frac{N_1 \sum_j \|x_{1j} - c_1\|^2}{N_1 - 1} \quad (2.50)$$

Pese al gran número de variantes que se tiene del método K-means, no existe un algoritmo absoluto para la resolución de problemas, pues la elección del algoritmo más óptimo para cada caso depende de las características del grupo de datos, por esta razón se recomiendan tratar cada problema con diferentes algoritmos de agrupamiento (Jain, Diun, & Mao, 2000).

2.2.3.1. Limitaciones del método en general

- El algoritmo *k-means* siempre converge, pero es probable que se encuentre una solución mínima local en lugar de una solución global, por lo cual puede que no se encuentre una partición óptima.
- La familia de métodos al que pertenece el algoritmo *k-means* son buscadores locales heurísticos, lo que conlleva a que sean sensibles a la elección inicial de centroides. Para contrarrestar esta limitación es recomendable repetir los procesos con diferentes centroides iniciales.
- En el caso particular del método Forgy/Floyd, es posible crear clusters vacíos si todos los elementos son movidos al mismo tiempo de un subespacio a otro.
- Tanto el método de MacQueen como el método de Hartigan son sensibles al orden en que los puntos son ubicados, pudiendo dar diferentes resultados según el orden de los datos.
- El algoritmo *k-means* en general tiene la tendencia de crear clusters de igual magnitud, aun si no se genera la mejor distribución de grupos en los datos.
- Este algoritmo *k-means* tiene problemas con los *outliers* ya que está basado en la *media*, la cuál es una estadística descriptiva que se ve muy afectada por los valores extremos de una muestra.

2.2.3.2. Verificación de los resultados obtenidos con el método k-means

En general, existen dos maneras de evaluar los resultados obtenidos por el método k-means. El primero es un criterio interno el cual está basado únicamente en las características propias de los datos y es normalmente usado para escoger el algoritmo de agrupamiento, así como el número óptimo de clusters que son usados. El segundo método de evaluación es un criterio externo basado en una comparación entre la solución encontrada e información externa conocida de antemano y generalmente es usado para la elección de un método de agrupamiento óptimo sobre un conjunto de datos específico.

- **Métricas de validación interna**

El objetivo principal de los métodos de agrupación en general (*clustering*) es unir los elementos iguales en un mismo cluster y ubicar los elementos diferentes en distintos clusters. Tomando lo anterior, las métricas de validación interna están basadas en los siguientes criterios:

- **Cohesión:** El miembro de cada cluster debe ser lo más cercano posible al resto de elementos contenidos en ese cluster. Para la cuantificación de este criterio, se utiliza el método de la suma del cuadrado de la distancia interna (*Sum of Squared Within*) la misma que matemáticamente es expresada como lo indica la ecuación 2.51

$$SSW = \sum_{i=1}^k \sum_{x \in C_i} dist^2(m_i, x) \quad (2.51)$$



En donde:

- k = Número de clusters
- x = Elemento del cluster C_i
- m_i = Centroide del cluster C_i .

- **Separación:** Los clusters deben estar lo más separados posible entre sí. Uno de los métodos más comúnmente utilizados para cuantificar este criterio es la suma de los cuadrados de la distancia entre clusters (*Sum of Squared Between*), lo cual se expresa con la ecuación 2.52:

$$SSB = \sum_{j=1}^k n_j \text{dist}^2(c_j - \bar{x}) \quad (2.52)$$

En donde:

- k = Número de clusters.
- n_j = Número de elementos en el cluster j
- c_j = Centroide del cluster j
- \bar{x} = Media del conjunto de datos.

2.2.3.3. Índices de evaluación para métricas internas

Existe una gran variedad de índices para la evaluación de métricas internas de los resultados de un proceso de agrupamiento (clustering), algunos de los más usados se encuentran resumidos en la tabla 2.1.

Tabla 2.1. Índices para verificar la eficiencia de un proceso de agrupamiento

(Fuente: Propia)

ÍNDICES DE VALIDACIÓN INTERNA	
Método	Expresión matemática
Ball y Hall (1965)	$\frac{SSW}{k}$
Calinski y Harabasz (1974)	$\frac{SSB/(k-1)}{SSW/(n-k)}$
Hartigan (1975)	$\log\left(\frac{SSB}{SSW}\right)$
Xu (1997)	$d * \log\left(\sqrt{\frac{SSW}{dN^2}}\right) + \log(k)$

Uno de los más usados al momento de evaluar los resultados provenientes del método k-means es el indicador Dunn (Dunn, 1979) el cual es aproximadamente igual a la relación entre la similitud inter-cluster y la similitud intra-cluster (ecuación 2.53)

$$DI = \min_{i=1 \dots m} \left\{ \min_{j=1 \dots m, i \neq j} \left\{ \frac{d(c_i, c_j)}{\max_{k=1 \dots m} \Delta_k} \right\} \right\} \quad (2.53)$$

En donde:

$d(c_i, c_j)$ = Distancia entre los centroides de cada cluster.

Δk = Medida de variación de inercia de cluster.

Es necesario resaltar que la solución con el valor más alto de indicador Dunn es considerado el mejor.

2.2.3.4. Métricas de validación externa

Como evaluador externo se tiene el indicador Jaccard (Jaccard, 1901) el cuál evalúa el grado de similitud entre dos conjuntos y es casi siempre usado cuando se tiene una clasificación previa confiable de los datos. Así pues, se calcula la similitud entre la solución encontrada y un índice de referencia como un porcentaje de clasificación correcta (ecuación 2.54).

$$J = \frac{|A \cap B|}{|A \cup B|} \quad (2.54)$$

En donde:

A = Solución encontrada

B = Índice de referencia

2.2.4. Dendograma

Un dendograma es básicamente un resumen del proceso de agrupación de un análisis cluster, el mismo que se presenta en forma de árbol de una manera gráfica. Así pues, los objetos similares se conectan mediante enlaces, cuya posición en el diagrama está de acuerdo al nivel de similitud entre objetos. Un ejemplo de dendograma se muestra en la figura 2.5, en la cual se tiene ocho elementos y la distancia del enlace en que se une cada uno de ellos, hasta que al final se forma un solo conglomerado (Hair, Black, Babin, & Anderson, 2010).

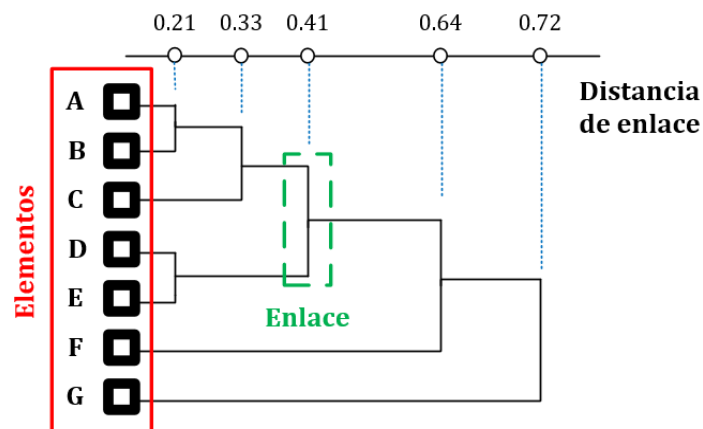


Figura 2.5. Ejemplo de Dendograma en el que se divisan los enlaces de ocho elementos (Fuente: Propia)

Dado que la decisión acerca del número óptimo de clusters en los que se debe agrupar los datos es muy subjetiva, ya que si se seleccionan muy pocos clusters la resultante tienden a ser muy artificial y heterogénea, mientras que si se seleccionan muchos clusters la interpretación se vuelve complicada.

El uso de dendogramas puede ofrecer una solución heurística al problema del número óptimo de clusters. Para ello, se corta el dendograma con una línea la misma que determina el número de clusters (grupos) que se tienen a este nivel.

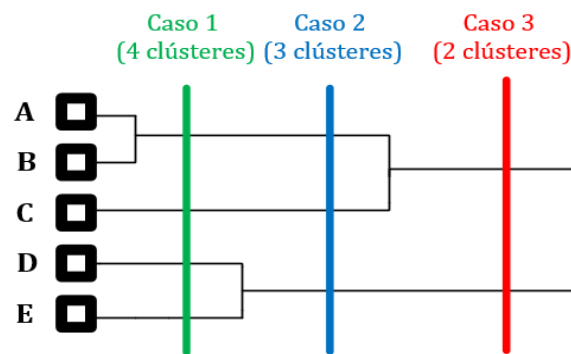


Figura 2.6. Ejemplo de la determinación del número de clusters en un dendograma. (Fuente: Propia)

Tal como se distingue en la figura 2.6, para el *Caso 1* la línea corta al dendograma en 4 puntos lo que significa que existen cuatro clusters en ese nivel de agrupación, en el *Caso 2* se tiene dos clusters y en el *Caso 3* se obtienen dos clusters.

Posteriormente se debe representar en una gráfica la relación entre el número de clusters observados en cada nivel y la distancias a la que dichos clusters se unen en cada etapa, tal como se muestra en la figura 2.7.

Es usual que en los primeros pasos la variación en las distancias sea pequeña, denotando que la unión de los clusters no aporta mayor información, mientras que en los últimos pasos la variación es mayor. Así pues, el punto de corte óptimo es el punto en donde comienzan a producirse variaciones bruscas, este a su vez es considerado como el número de clusters óptimos (Pérez Miñaca & Salazar Yépez, 2014). La figura 2.8 muestra la variación en la distancia para cuatro enlaces, en la cual se puede notar que a partir del tercer enlace (recuadro rojo) la distancia varía de forma abrupta por lo que se coloca una línea cerca de este punto y se determina que el punto 3 es el punto óptimo.

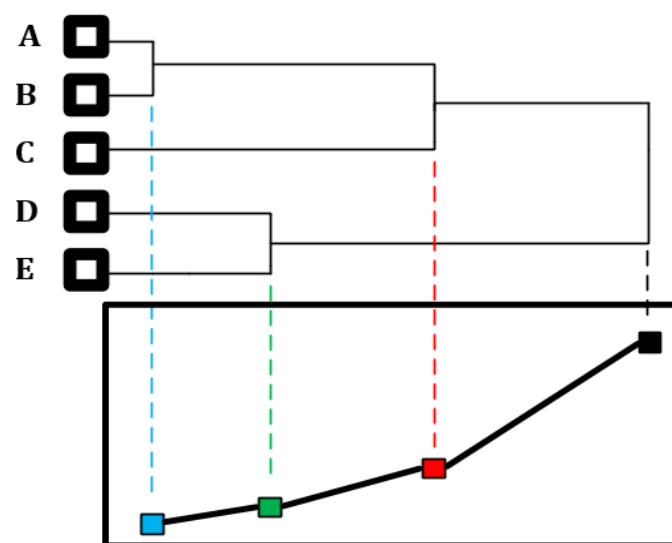


Figura 2.7. Ejemplo de grafica que relaciona el número de clusters con las distancias de enlace. (Fuente: Propia)

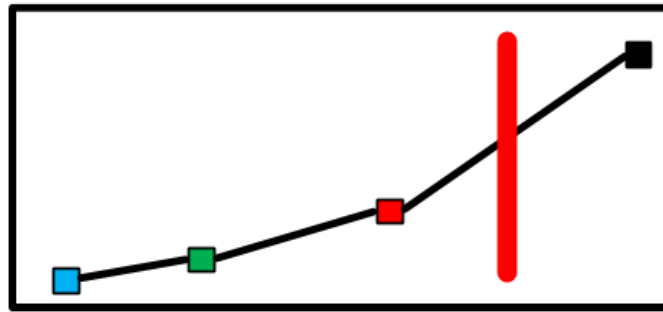


Figura 2.8. Determinación del número de clusters óptimos.
(Fuente: Propia)

2.3. MÈTODOS DE ESTRATIFICACIÒN

2.3.1. Dalenius & Hodges (1959)

El método de Dalenius & Hodges está basado en la construcción de intervalos iguales tomando la raíz cuadrada acumulada de las frecuencias de la variable de estratificación para de esta manera obtener puntos muy cercanos a los óptimos (Dalenius & Hodges, 1959).

Si se considera una función de densidad $f(t)$ con una media $\mu = \int tf(t) dt$. El rango $x_0 - x_L$ de la variable estudiada x es dividida en L partes en los puntos $x_1 < \dots < x_h < \dots < x_{L-1}$ en el que cada parte representa un estrato. Así pues, para el estrato h la ecuación 2.55 expresa la ponderación de este estrato (W_h).

$$W_h = \int_{x_{h-1}}^{x_h} f(t) dt \quad (2.55)$$

Si a la ecuación 2.55 se la relaciona con la media de elementos del estrato h (μ_h), la expresión resultante está dada por la ecuación 2.56.

$$W_h \mu_h = \int_{x_{h-1}}^{x_h} tf(t) dt \quad (2.56)$$

Con la expresión descrita por la ecuación 2.56 se puede calcular la varianza en el estrato h (σ_h^2) mediante la ecuación 2.57.

$$\sigma_h^2 = \frac{\int_{x_{h-1}}^{x_h} t^2 f(t) dt}{W_h} - \mu_h^2 \quad (2.57)$$

Si se toma una muestra de $f(x)$, la media de los elementos de la muestra puede ser calculados a través de la ecuación 2.58. Además, la varianza de la muestra se estima mediante la ecuación 2.59.

$$\bar{x} = \sum_h W_h \bar{x}_h \quad (2.58)$$

$$\sigma_s^2(\bar{x}) = \sum_h W_h^2 \frac{\sigma_h^2}{n_h} \quad (2.59)$$

Con la ayuda del método de Neyman se puede encontrar la varianza mínima de los elementos, la misma que está expresada mediante la ecuación 2.60.

$$\sigma_{min}^2(\bar{x}) = \frac{1}{n} \left(\sum_h W_h \sigma_h \right)^2 \quad (2.60)$$

Sin embargo, en la práctica no se suele conocer la distribución de la variable de estudio. Partiendo de este punto, Dalenius & Hodges realizaron algunas aproximaciones e introdujeron la transformación presentada por la ecuación 2.61.

$$Y(u) = \int_{-\infty}^u \sqrt{f(t)} dt \quad (2.61)$$

Si $u \rightarrow \infty$, $Y(u)$ se aproxima a un límite superior que puede ser expresado por la ecuación 2.62.

$$Y(u) = \frac{h}{L} H \quad h = 1, \dots, L - 1 \quad (2.62)$$

Además, si L tiene una magnitud considerable, los estratos van a ser estrechos y por lo tanto la distribución en cada uno de los estratos va a ser aproximadamente uniforme, así la varianza en cada estrato está definida por la ecuación 2.63.

$$\sigma_h \approx \frac{x_h - x_{h-1}}{\sqrt{12}} \quad (2.63)$$

Además, se puede simplificar la ponderación del estrato h de tal manera que quede expresada mediante la ecuación 2.64

$$W_h \approx f_h(b_h - b_{h-1}) \quad (2.64)$$

Por el teorema del valor medio se sabe que hay un valor f_h en el estrato h , tal que se cumple la ecuación 2.65.

$$\sqrt{12} \sum W_h \sigma_h \approx \sum [\sqrt{f_h}(x_h - x_{h-1})]^2 \approx \sum [Y_h - Y_{h-1}]^2 \quad (2.65)$$

Ya que $Y_L - Y_0$ es un intervalo fijo es fácil de verificar que la suma de la ecuación anterior es minimizada para hacer $Y_h - Y_{h-1}$ constante, lo cual es equivalente a hacer $W_h \sigma_h$ constante.

El siguiente algoritmo es generalmente aplicado para construir los h estratos de una variable X , se debe resaltar que la cantidad de estratos (h) debe ser previamente definida, de esta manera los pasos a seguir son los siguientes:

- i. Ordenar de forma ascendente los datos de la variable X que debe ser analizada.
- ii. Agrupar X en J clases, en donde $J = \min\{h * 10, n\}$
- iii. Calcular los límites de cada clase utilizando las ecuaciones 2.66 y 2.67

$$\lim \inf C_k = \min\{x_i\} + (k - 1) * \frac{\max\{x_i\} - \min\{x_i\}}{j} \quad (2.66)$$

$$\lim \sup C_k = \min\{x_i\} + (k) * \frac{\max\{x_i\} - \min\{x_i\}}{j} \quad (2.67)$$

Cabe señalar que los intervalos son tomados abiertos por la izquierda y cerrados por la derecha, con la excepción del primer intervalo cuyos límites son considerados cerrados en ambos lados.

- iv. Obtener la frecuencia (f_i) de observación en cada una de las clases.
- v. Calcular la raíz cuadrada de frecuencias en cada clase.
- vi. Calcular la raíz cuadrada acumulada de frecuencias de cada clase mediante la ecuación 2.68.

$$\sum_{i=1}^J \sqrt{f_i} \quad (2.68)$$

- vii. Dividir la raíz cuadrada acumulada de cada frecuencia por el número de estratos como lo indica la ecuación 2.69.

$$Q = \frac{1}{h} \sum_{i=1}^J \sqrt{f_i} \quad (2.69)$$

- viii. Finalmente, los puntos de corte de cada estrato son tomados sobre la raíz cuadrada acumulada de las frecuencias en cada clase, según lo indica la expresión 2.70.

$$Q, 2Q, \dots, (h - 1)Q \quad (2.70)$$

Si el valor de Q llega a quedar entre dos clases, se considera como punto de corte aquella clase que tenga la menor distancia a Q .

Este es un método simple, sin embargo, uno de sus principales problemas es la elección del número óptimo de clases en los que se debe dividir la variable. Además, ya que en este método una buena aproximación depende de la distribución uniforme de los estratos, (Cochran, 1980) se aconseja tener una cantidad sustancial de clases en la frecuencia original de lo contrario la verdadera estratificación óptima puede fallar y el cálculo de los límites dentro del estrato se ve afectado por los errores de agrupación.

2.3.2. Lavallèe & Hidiroglou (1988)

Lavallèe & Hidiroglou desarrollaron un algoritmo iterativo para poblaciones sesgadas (asimétricas), tal que el tamaño de la muestra se minimiza para un nivel dado de precisión expresado en términos del coeficiente de variación, el mismo que por lo general se encuentra entre el 1% y 10% (Hidiroglou & Lavalleé, 1988). Este método es numéricamente comparable en términos de tamaño de muestra y valor de límites al método de Dalenius & Hodges. Este algoritmo recursivo es de fácil programación y tiende a converger rápidamente a los puntos del límite óptimo.

De esta manera, el método de Lavallèe & Hidiroglou trata de encontrar los valores de la expresión 2.71 que minimicen n considerando un estrato superior que contenga todo, tomando en cuenta el nivel de precisión (CV) y el esquema de asignación representado por a_h . De esta manera se obtiene la ecuación 2.72.

$$b_1 < b_2 < \dots < b_{L-1} \quad (2.71)$$

$$n = N_L + \left(\sum_{h=1}^{L-1} \frac{N_h^2 \sigma_h^2}{a_h} \right) \left(N^2 \mu^2 CV^2 + \sum_{h=1}^{L-1} N_h \sigma_h^2 \right)^{-1} \quad (2.72)$$

Lavallèe & Hidiroglou expresan el esquema de asignación (a_h) tal como se muestra en la ecuación 2.73

$$a_h = \frac{(W_h \mu_h)^p}{\sum_{h=1}^{L-1} (W_h \mu_h)^p} \quad (2.73)$$

En donde $0 < p < \infty$. Con lo cual se obtiene la ecuación 2.74.

$$n = NW_L + \frac{N \left[\sum_{h=1}^{L-1} (W_h \sigma_h)^2 \right] \left[\sum_{h=1}^{L-1} (W_h \mu_h)^p \right]}{N CV^2 \mu^2 + \sum_{h=1}^{L-1} W_h \sigma_h^2} \quad (2.74)$$



Ya que la ecuación 2.74 es minimizada para encontrar los límites, las primeras derivaciones son tomadas con respecto a $b_1 < b_2 < \dots < b_{L-1}$ y después son igualadas a cero. Usando el método de Sethi (Sethi, 1963) esas ecuaciones son resueltas usando un método iterativo inicial con límites arbitrarios hasta que dos conjuntos consecutivos difieran por magnitudes muy pequeñas. Bajo suposiciones simples, los coeficientes de variación en cada estrato tienden a ser igualados sin un incremento significativo en la varianza de la media de la muestra.

Sin embargo este método tiene algunos problemas. Ya que los resultados dependen de los límites iniciales, de esta forma se puede encontrar un mínimo local que no necesariamente es el global. También se ha encontrado que el algoritmo converge rápidamente solo cuando se tiene un número bajo de estratos.

2.3.3. Método de Kozak (2004)

El método de Kozak, conocido también como *Kozak's Random Search Method*, es un algoritmo en que se trata de encontrar los valores de la expresión 2.75 (Zimmer, Kim, & Nusser, 2012).

$$b_1 < b_2 < \dots < b_{L-1} \quad (2.75)$$

Tales valores deben minimizar la función objetivo expresada por la ecuación 2.76.

$$n = N_L + \left(\sum_{h=1}^{L-1} W_h \sigma_h \right)^2 \left(\mu^2 CV^2 + \frac{1}{N} \sum_{h=1}^{L-1} W_h \sigma_h^2 \right)^{-1} \quad (2.76)$$

Bajo las restricciones $N_h \geq 2$ para $h = 1, 2, \dots, L$ y $2 \leq n_h \leq N_h$ para $h = 1, 2, \dots, L - 1$, en donde n es el tamaño de la muestra minimizada requerida para alcanzar la precisión dada (CV) de la media estimada de la estratificación, en donde los límites de estrato son $b = (b_1, b_2, \dots, b_{L-1})^T$.

Kozak's Random Search Method escoge un conjunto inicial de límites de estrato y calcula los valores de la función n . Así pues, para un cierto número de iteraciones se repiten los siguientes pasos:

- i. Generar el conjunto de límites de estrato b , de donde se escoger un límite de estrato b_h , para posteriormente variarlo como lo indica la ecuación 2.77

$$b'_h = b_h + j \quad (2.77)$$

En donde, j es un entero aleatorio $j \in (-p, -1) \cup (p, 1)$ y p es un entero dado acorde con las dimensiones de la población.

- ii. Calcular el valor de la función n' .
- iii. Si las restricciones son satisfechas y $n' < n$, entonces el nuevo conjunto de límites b'_h es aceptado.

El algoritmo termina cuando los criterios se cumplen. Ya que el algoritmo no garantiza que el mínimo global es alcanzado, el rendimiento del algoritmo ha sido comparado con otros métodos disponibles. Kozak & Verma han comparado la eficiencia de este método con las del método geométrico de estratificación encontrando que en todos los casos el presente método fue más eficiente que el método geométrico.

CAPÍTULO 3: MAPA DE ESTRATOS

3.1. INFORMACIÓN DE LA EMPRESA

3.1.1. Zona de concesión

La empresa eléctrica de Azogues conformada el 24 de marzo de 1972, actualmente cuenta con un área de concesión de alrededor de 1187 km², la misma que tiene su centro de operaciones en la ciudad de Azogues y cubre las parroquias: Cojitambo, Guapán, Borrero, Javier Loyola, Bayas, Luis Cordero, Taday, Pindilig, Rivera, San Miguel de Porotos y el cantón Déleg con su parroquia Solano, así como también la parroquia Sageo del cantón Biblián. La zona total de concesión se puede examinar en la figura 3.1 en la cual se exponen los centros urbanos más importantes.

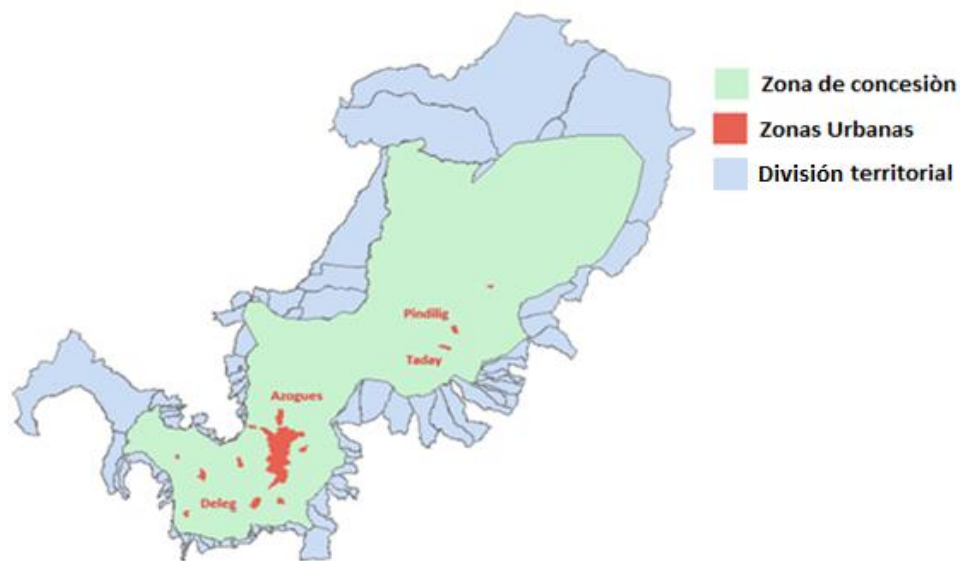


Figura 3.1. Zona de concesión de la EEA.
(Fuente: Datos proporcionados por la EEA)

3.1.2. Usuarios y consumos

La empresa eléctrica Azogues tiene un total de 37 029 usuarios registrados hasta el año 2017. Tal como lo muestra la tabla 3.1, el mercado de esta empresa es predominantemente residencial (90.5%), seguido de un consumo comercial (6.13%), industrial (1.2%) y otros (2.07%):

Tabla 3.1. Distribución de usuarios según su tarifa
(Fuente: Datos proporcionados por la EEA)

NÚMERO DE USUARIOS A DICIEMBRE DE 2017	
RESIDENCIAL	33512
COMERCIAL	2273
INDUSTRIAL	477
OTROS	767
TOTAL	37029

Tabla 3.2. Consumo por tipo de usuario de la EEA para el año 2017
(Fuente: Datos proporcionados por la EEA)

TIPO DE USUARIO	ENERGÍA CONSUMIDA [kWh]
Residencial	2 645 931.3
Comercial	854 574.2
Industrial	4 509 436.6
Otros	397 562.2
Total	8 407 819.9

El promedio de consumo mensual de los usuarios de cada tipo de tarifa en el año 2017 se detalla en la tabla 3.2. En dicha tabla se puede notar que los mayores consumos se dan en el sector industrial y residencial.

3.1.3. Datos técnicos de la empresa

La empresa eléctrica Azogues cuenta con dos subestaciones de transformación: La subestación *Azogues 1* que se encuentra ubicada en la comunidad de Zhizhiquin y tiene una capacidad de 32.5 MVA, y, por otro lado, la subestación *Azogues 2* que se encuentra ubicada en la parroquia Borrero de la provincia del Cañar y cuenta con una capacidad de 26 MVA. La figura 3.2 muestra la ubicación de las subestaciones pertenecientes a la EEA.



Figura 3.2. Ubicación de las subestaciones Azogues 1 y 2 pertenecientes a la EEA.
(Fuente: Google Maps)

Desde estas subestaciones se conectan cerca de 26.87 km de líneas de subtransmisión y 789.16 km de redes de medio voltaje, que conjuntamente con los 2 001 transformadores de distribución instalados (hasta 2017), sirven de medio para servir de energía a los 37 029 usuarios registrados. Esta empresa dispone anualmente de 110 591 MWh y cubre una demanda máxima de 19 051 MW con un porcentaje de pérdidas de energía del 4.56%. Estos valores son resumidos en la tabla 3.3.

Tabla 3.3. Resumen técnico de la empresa eléctrica de Azogues.
(Fuente: Rendición de cuentas de la EEA 2017)

Resumen técnico EEA 2017	
Subestaciones de distribución	2
Capacidad de subestaciones	26 – 32.5 MVA
Líneas de subtransmisión	26.87 km
Redes de medio voltaje	789.16 km
Nivel de voltaje primario	22 kV
Transformadores de distribución	2 001
Número de luminarias	14 567
Capacidad instalada de transformadores	52 MVA
Demanda máxima del sistema	19 051 MW
Energía anual disponible	110 591.988 MWh
Pérdidas de energía	4.56%
Clientes	37 029
Número de trabajadores y servidores	149

3.2. CRITERIOS Y DATOS NECESARIOS PARA EL PROCESO DE ZONIFICACIÓN.

Para dar inicio al proceso de zonificación, es necesario obtener los siguientes datos:

- Código de identificación del usuario.
- Consumo de energía por cliente de los años 2016 y 2017.
- Categoría tarifaria de cada usuario.
- Localidad a la que pertenece cada consumidor.
- Coordenadas “X” y “Y” (latitud y longitud) de cada medidor.
- Nivel de tensión a la que se encuentra conectado cada usuario.
- Identificación del transformador al que se encuentra conectado cada consumidor.
- Curvas de carga características por categoría tarifaria.
- Datos de la zona de concesión de la empresa.
- Clasificación de los sectores rurales y urbanos.

A continuación, se detallan los criterios en los que se basa la zonificación:

- La información base para la zonificación corresponde al promedio de energía facturada de cada usuario en el año 2017.
- Los usuarios cuyas coordenadas (X, Y) no estén registradas no son considerados en la elaboración del mapa de estratos.
- Los usuarios cuya demanda promedio anual sea menor o igual a cero son excluidos de todo el análisis.
- Solo se consideran aquellos usuarios cuyo voltaje de servicio es clasificado como de bajo voltaje.
- Se considera un total tres grupos para la zonificación correspondientes a los grupos tarifarios: residenciales, comerciales e industriales.
- Se excluye del análisis aquellos usuarios denominados como “otros” dado a la gran diversidad de tipos de cargas (iglesias, edificios estatales, etc.) y pequeño número de consumidores dentro de esta tarifa.

3.3. DETECCIÓN DE OUTLIERS EN LA BASE DE DATOS

Los datos atípicos también conocidos como *outliers*, son en términos generales; datos significativamente distintos al resto de la muestra. Hawkins define formalmente el concepto de *outlier* como:

“ Un *outlier* es una observación que se desvía tanto de otras observaciones que despierta sospechas de que fue generado por otros mecanismos.” (Hawkins 1980)

En otras palabras, cuando se habla de datos atípicos u outliers se hace referencia a anomalías, discordancias o desviaciones. Por esta razón, se vuelve imprescindible en ciertas aplicaciones la detección de este tipo de datos pues pueden llegar a ocasionar errores serios en los cálculos. Por otro lado, se debe tomar en cuenta que no siempre es necesario suprimir este tipo de datos de la muestra tomada pues muy a menudo contienen información útil acerca de las características anormales del sistema y de sus elementos.

3.3.1. Método para detectar datos atípicos (outliers)

Uno de los métodos más frecuentemente utilizados para examinar conjuntos de datos univariados es el método de caja y bigotes, también conocido como *boxplot*, propuesto por Turkey (1977). Este método indica que si se tiene un conjunto de datos univariados, tal como se muestra en la ecuación 3.1.

$$X_n = \{x_1, x_2, \dots, x_n\} \quad (3.1)$$

Se debe crear el diagrama de caja y bigotes (*boxplot*), el cual se construye tomando en cuenta los siguientes pasos:

1. Poner una línea a la altura de la mediana Q_2 (segundo cuartil)
2. Dibujar un recuadro desde el primer cuartil Q_1 hasta el tercer cuartil Q_3 . La longitud de este recuadro se conoce como *rango intercuartil* (ecuación 3.2).

$$IQR = Q_3 - Q_1 \quad (3.2)$$

3. Dibujar los límites (bigotes) en los extremos del intervalo presentado por la expresión 3.3.

$$[Q_1 - 1.5IQR ; Q_3 + 1.5IQR] \quad (3.3)$$

Así pues, se define como posible dato atípico (outlier) a cualquier valor que esté fuera del intervalo mostrado en la expresión 3.3.

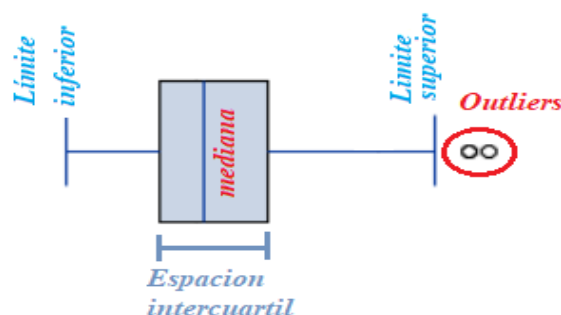


Figura 3.3. Ejemplo método boxplot.
(Fuente: Propia)

Este es un método sencillo que muestra información acerca de la localización, asimetría, así también como de la cola que tiene el conjunto de datos examinado. En la figura 3.3 se muestra un ejemplo del método de caja y bigotes, en el que se puede notar que existen dos puntos fuera de los límites (bigotes) de la figura, los cuales son considerados como datos atípicos

El panorama cambia cuando se tiene un conjunto asimétrico de datos, ya que muchos datos exceden los límites (bigotes) y una gran parte de estos datos son denominados erróneamente como *outliers*.

Para contrarrestar este problema Hubert & Vandervieren ajustan el método de caja y bigotes para de esta manera tomar en cuenta el grado de asimetría de un conjunto de datos, concluyendo que los límites con los que se debe trabajar están dados por las expresiones 3.4 y 3.5 (Hubert & Vandervieren, 2007).

$$[Q_1 - 1.5e^{-4MC}IQR ; Q_3 + 1.5e^{3MC}IQR] \quad \text{para } MC \geq 0 \quad (3.4)$$

$$[Q_1 - 1.5e^{-3MC}IQR ; Q_3 + 1.5e^{4MC}IQR] \quad \text{para } MC < 0 \quad (3.5)$$

En donde MC (*medcouple*), término introducido por Brys (Brys, Hubert, & Struyf, 2003), cuantifica el grado de asimetría de una muestra univariable $\{x_1, x_2, \dots, x_n\}$. Este se encuentra definido por la ecuación 3.6.

$$MC = \text{med}_{x_i \leq Q_2 \leq x_j} h(x_i, x_j) \quad (3.6)$$

En donde Q_2 es la media de la muestra y para toda $x_i \neq x_j$ la función h está definida por la ecuación 3.7.

$$h(x_i, x_j) = \frac{(x_j - Q_2) - (Q_2 - x_i)}{x_j - x_i} \quad (3.7)$$

Cabe mencionar que dada la naturaleza asimétrica de la base de datos empleada para realizar la zonificación en el área de concesión de la EEA (figura 3.4, 3.5 y 3.6), se utiliza el método expuesto por Hubert & Vandervieren para la detección de datos atípicos.

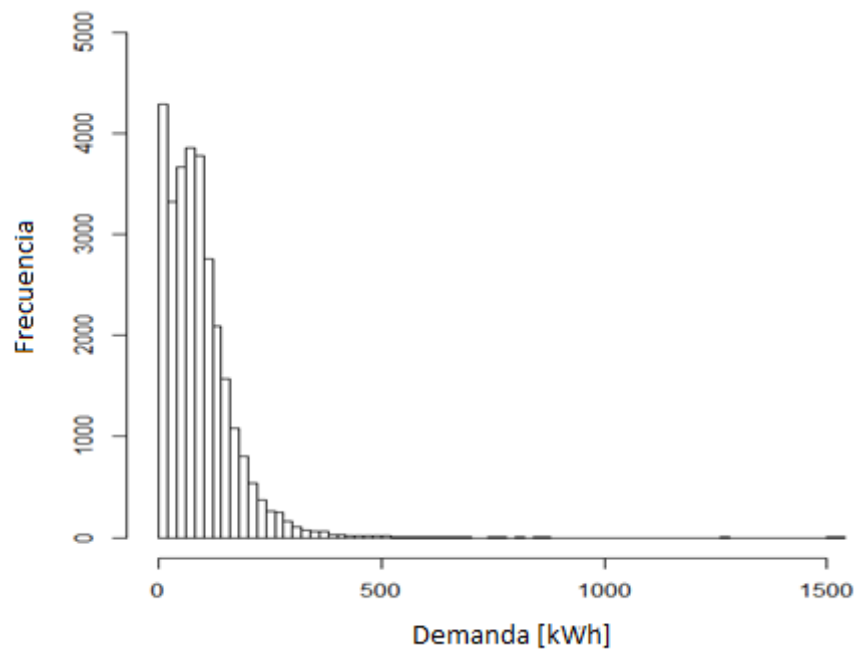


Figura 3.4. Distribución de demandas para usuarios con tarifa residencial de la EEA. (Fuente: Propia)

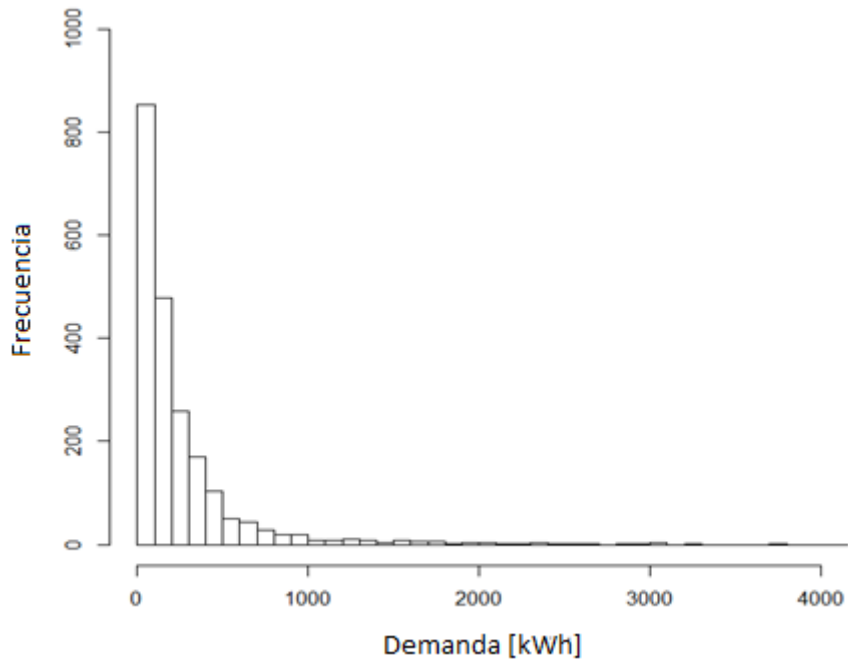


Figura 3.5. Distribución de demandas para usuarios con tarifa comercial de la EEA.
(Fuente: Propia)

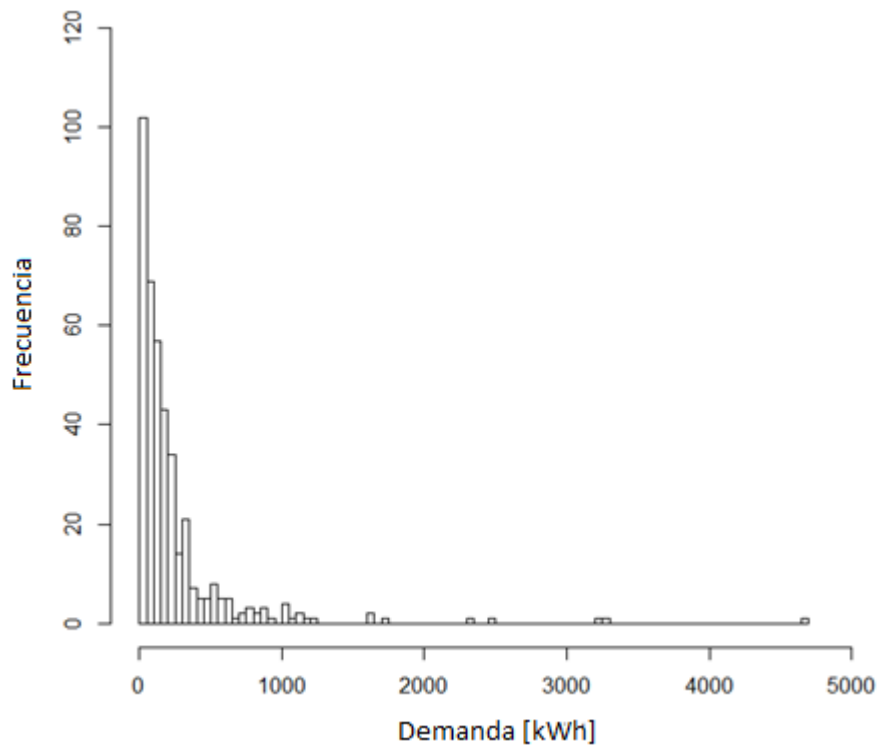


Figura 3.6. Distribución de demandas para usuarios con tarifa industrial de la EEA.
(Fuente: Propia)

3.3.2. Resultados obtenidos

Una vez ejecutado el método de detección de datos atípicos expuesto en el punto anterior los resultados para cada grupo de usuarios son los siguientes:

- **Usuarios Residenciales**

El método propuesto para la detección de datos atípicos, en este caso da como resultado dos límites $[-44.9, 309]$, esto quiere decir que todo valor menor a -44.9 y mayor a 309 es considerado como dato atípico. Así pues, de los $30\ 226$ usuarios registrados inicialmente, 447 (1.47%) usuarios superan una demanda promedio mensual de 309 kWh, quedando así un total de $29\ 779$ (98.53%) usuarios que son considerados para los posteriores análisis. En la figura 3.7 se muestra la distribución de demandas para los usuarios con tarifa residencial luego de filtrar los datos atípicos. Cabe señalar que en la base de datos no existe valores menores o iguales a cero dado los criterios que se han tomado en cuenta a la hora de realizar la base de datos, por lo tanto el límite inferior (-44.9) no tiene mucha relevancia. En la tabla 3.4 se muestra un resumen de los datos obtenidos en el proceso de detección de datos atípicos.

Tabla 3.4. Resumen del proceso de detección de datos atípicos para usuarios con tarifa residencial de la EEA.
(Fuente: Propia)

Resumen Residencial	
Muestra inicial	30 226
Muestra Final	29 779
Datos atípicos	447
Valor extremo 1	309
Valor extremo 2	-44.9
Medcouple	0.124
Cuartil 1	34.1
Cuartil 3	121

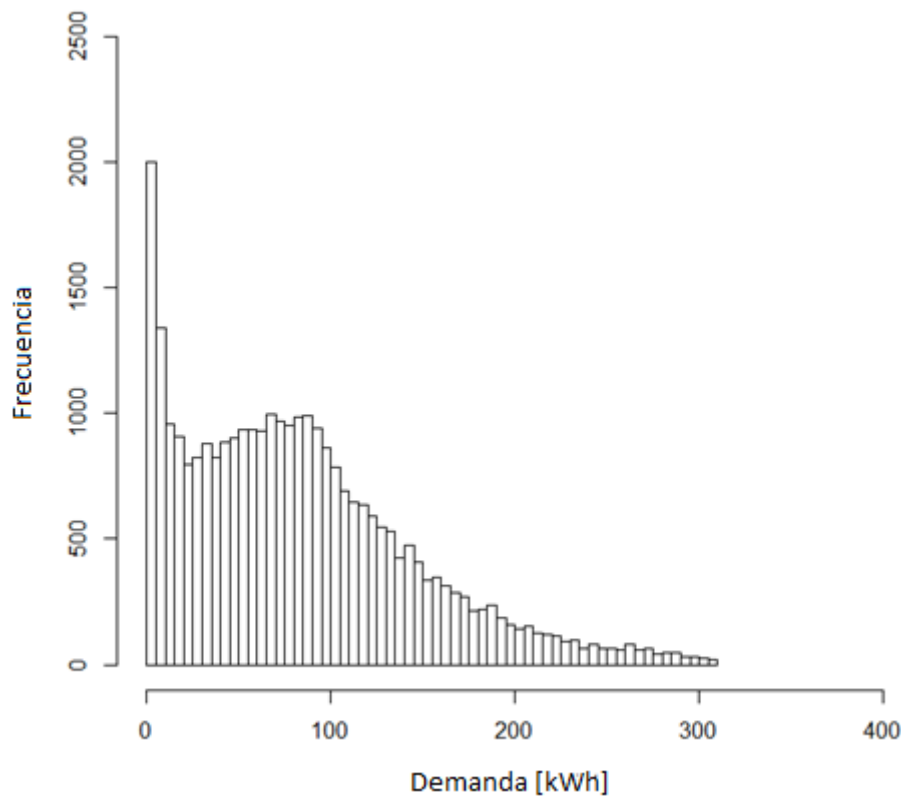


Figura 3.7. Distribución de la demanda de usuarios residenciales una vez finalizado con el proceso de detección de datos atípicos.
(Fuente: Propia)

- **Usuarios Comerciales**

Para el caso de los usuarios con tarifa comercial los límites usados para localizar datos atípicos son $[-11.7, 1661]$. Así pues, del total de usuarios con esta tarifa se han encontrado 42 (1.98%) usuarios cuya demanda promedio mensual sobrepasa los 1661 kWh, quedando así un total de 2079 (98.2%) usuarios que son considerados para los análisis posteriores. En la tabla 3.5 se muestra un resumen de los datos más importantes en el proceso de detección de outliers.

Tabla 3.5. Resumen del proceso de detección de datos atípicos para usuarios con tarifa comercial de la EEA.
(Fuente: Propia)

Resumen Comercial	
Muestra inicial	2 121
Muestra Final	2 079
Datos atípicos	42
Valor extremo 1	1 661
Valor extremo 2	-11.7
Medcouple	0.413
Cuartil 1	56.5
Cuartil 3	299

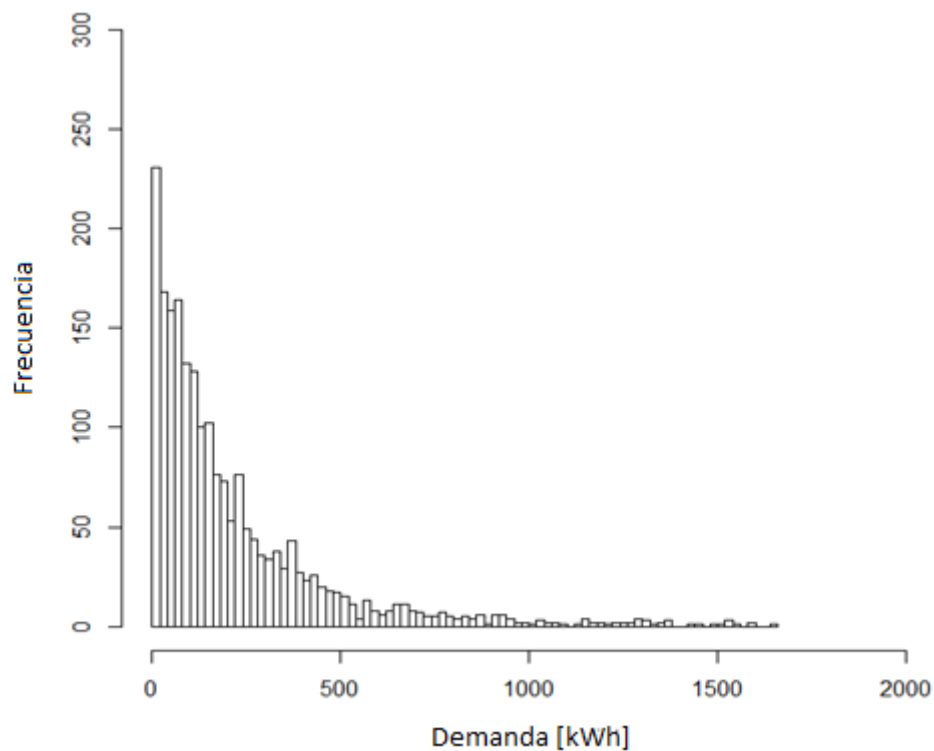


Figura 3.8. Distribución de la demanda de usuarios comerciales una vez finalizado con el proceso de detección de datos atípicos.
(Fuente: Propia)

La figura 3.8, muestra la distribución de demandas para usuarios registrados con tarifa comercial en la EEA, una vez que se han extraído los datos considerados como atípicos de la muestra.

- **Usuarios Industriales**

Para el caso de los usuarios con tarifa industrial los límites calculados son $[-20.1, 1130]$. Así pues, se tiene una muestra total de 404 usuarios de los cuales 11 (2.72%) tienen una demanda promedio mensual superior a los 1130 kWh, obteniéndose al final del proceso de detección de datos atípicos un total de 393 (97.27%) usuarios que conforman la base de datos definitiva de usuarios con tarifa industrial. La distribución de los datos resultantes de todo el proceso para este tipo de usuarios es mostrado en la figura 3.9, además se presenta un resumen de los datos obtenidos en este proceso en la tabla 3.6.

Tabla 3.6. Resumen del proceso de detección de datos atípicos para usuarios con tarifa industrial de la EEA.
(Fuente: Propia)

Resumen industrial	
Muestra inicial	404
Muestra Final	393
Datos atípicos	11
Valor extremo 1	1130
Valor extremo 2	-20.1
Medcouple	0.3625
Cuartil 1	49.7
Cuartil 3	248

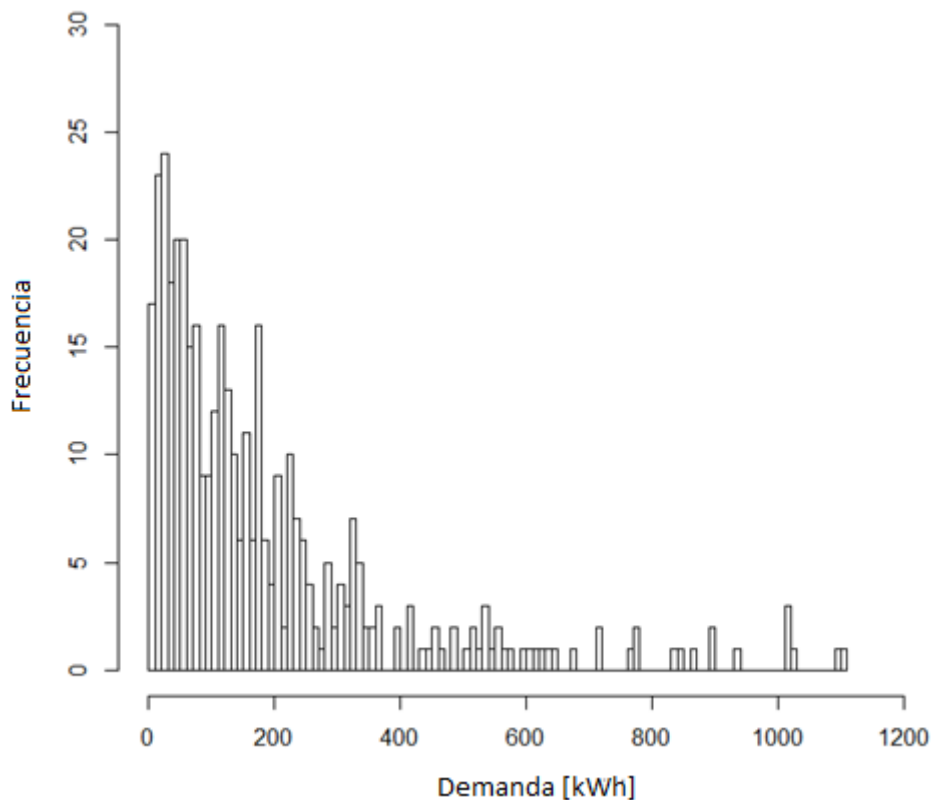


Figura 3.9. Distribución de la demanda de usuarios con tarifa industrial una vez finalizado con el proceso de detección de datos atípicos.
(Fuente: Propia)



3.4. DEFINICIÓN DEL NÚMERO DE ESTRATOS

Uno de los principales problemas a la hora de agrupar una muestra es definir la cantidad idónea de clusters, que en este caso significa la cantidad óptima de estratos en los que se debe dividir a los usuarios. Para la resolución de este problema se recurre al método *k-means* con el cuál se puede agrupar los elementos de la base de datos variando el número de clusters en los que se clasifican los elementos, en otras palabras, se empieza por agrupar los datos en dos clusters, luego en tres clusters y así sucesivamente. En cada caso se calcula las métricas internas de los clusters resultantes para evaluar tanto la dispersión intra-cluster como la dispersión inter-cluster. Esto da como resultado dos imágenes, en las cuáles se examina la variación de las métricas internas según la variación del número de agrupaciones, con lo cual se obtiene una base gráfica para poder tomar la decisión del número óptimo de estratos.

Para poder corroborar los resultados obtenidos con el método *k-means*, se procede al uso de los métodos jerárquicos conglomerados, más específicamente del método de Ward, el cual proporciona un dendograma en el cual se puede verificar los enlaces de elementos realizados en cada paso, así también como de la distancia a la que ocurre cada unión y de esta manera verificar las respuestas obtenidas con el método *k-means*.

Es necesario que antes de ejecutar los algoritmos mencionados previamente se depure completamente la base de datos con el objetivo de remover los datos atípicos pues los métodos presentados son muy susceptibles a outliers.

3.4.1. Resultados obtenidos usando el método *K-means*

Es necesario recordar que un análisis cluster busca agrupar elementos teniendo en cuenta que los grupos entre sí debe ser tan distintos como sea posible pero los elementos que conforman cada grupo deben ser muy semejantes entre sí, en otras palabras, la distancia inter-cluster debe ser lo más grande posible pero la distancia intra-cluster debe ser muy pequeña. Los resultados obtenidos se presentan a continuación:

- **Usuarios con tarifa residencial**

La figura 3.10 muestra la variación de la distancia inter-grupos con relación al número de clusters en los que se han agrupado los elementos. Así pues, se puede apreciar que la distancia inter-grupos tiene un comportamiento logarítmico según como se varía el número de clusters, obteniéndose así un cambio pequeño a partir del cuarto punto (4 clusters). Esto indica que los grupos se aproximan de forma significativa a su máxima separación cuando se agrupe a los elementos en cuatro o más clusters.

En la figura 3.11, se puede notar que el espacio intra-cluster tiene un comportamiento exponencial decreciente, con variaciones pequeñas a partir del tercer punto (3 clusters). Esto indica que los datos que forman cada uno de los grupos son muy semejantes (tienen una distancia pequeña) entre sí cuando son agrupados en tres o más clusters.

Se concluye entonces que cuatro estratos es el número óptimo de grupos en los que se debe dividir a los usuarios con tarifa residencial, para de esta manera obtener índices de cohesión y separación aceptables, con el uso del menor número de estratos posible.

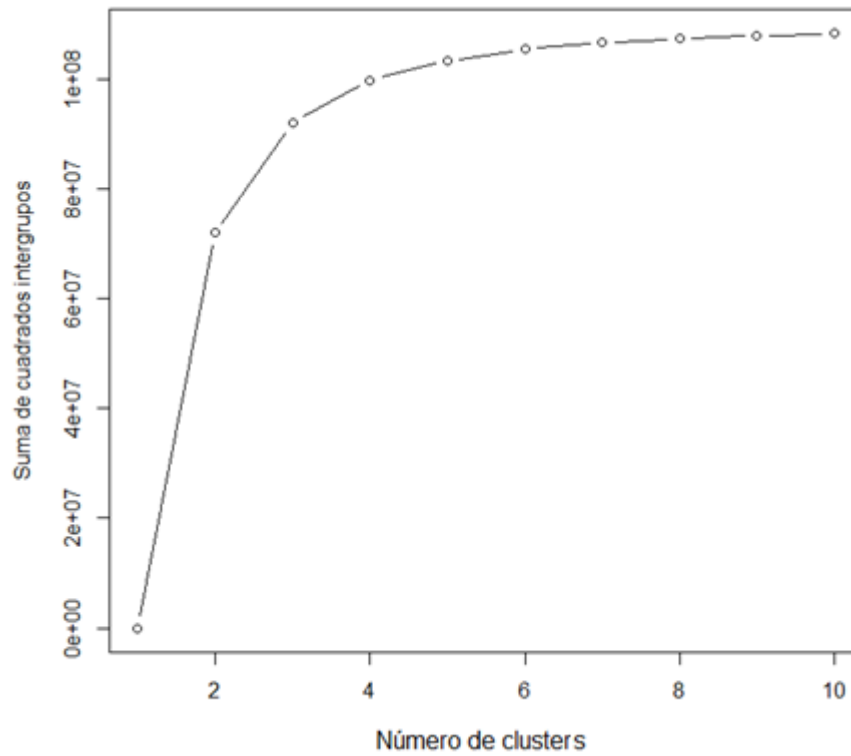


Figura 3.10. Variación de la distancia inter-cluster con relación al número de clusters en los que se agrupan los usuarios con tarifa residencial de la EEA.
(Fuente: Propia)

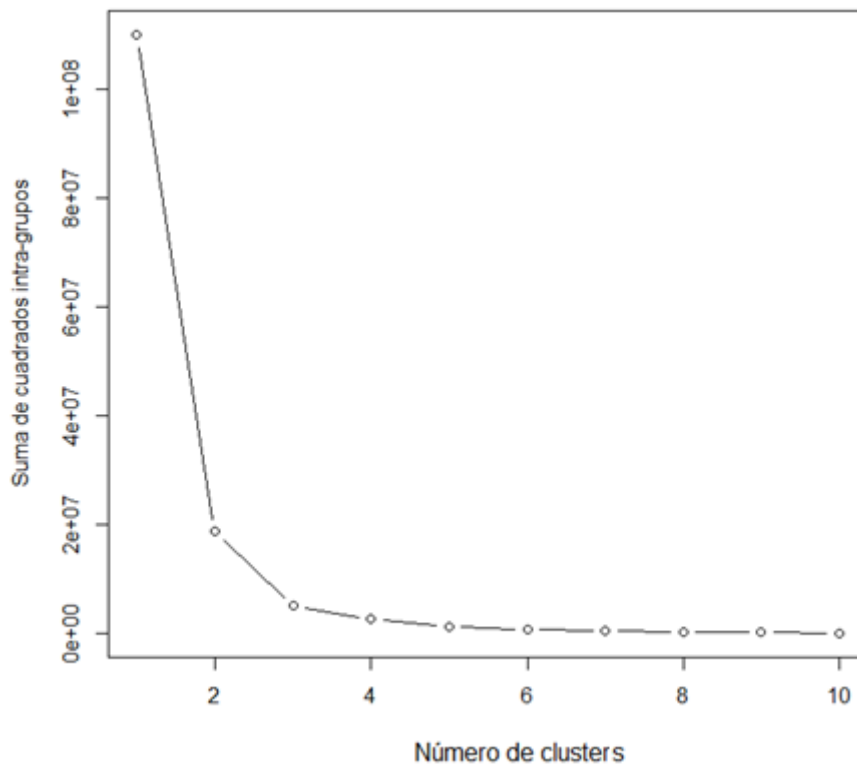


Figura 3.11. Variación de la distancia intra-cluster con relación al número de clusters en los que se agrupan los usuarios con tarifa residencial de la EEA.
(Fuente: Propia)

- **Usuarios con tarifa comercial**

Al igual que para el caso de los usuarios con tarifa residencial, la distancia inter-cluster varía de una forma logarítmica alcanzado una buena aproximación a su valor máximo a partir del cuarto punto (figura 3.12), en otras palabras, se tiene un buen índice de separación entre clusters si los elementos son agrupados en cuatro o más clusters.

Para el caso de la distancia intra-cluster, esta tiene una forma exponencial decreciente según se varía la cantidad de clusters en los que se agrupan los datos (figura 3.13), obteniéndose a partir del cuarto punto una buena aproximación al valor mínimo de distancias entre elementos.

Así pues, tomando los mismos criterios que se impuso para los usuarios residenciales, se puede concluir que cuatro estratos es el número óptimo de grupos en los que se debe dividir a los usuarios con tarifa comercial de la empresa eléctrica Azogues.

- **Usuarios con tarifa industrial**

Al igual que en los casos anteriores, las figuras 3.14 y 3.15 muestran la variación de la distancia intra-cluster e inter-cluster respectivamente. Para el caso de la distancia intra-cluster se puede apreciar que a partir de cuatro grupos se alcanza un valor muy cercano al valor máximo de separación entre clusters. Por otro lado, para el caso de la distancia intra-cluster se tiene que a partir de tres clusters se alcanza un índice de cohesión muy cercano al mínimo.

Con esto se puede concluir que los usuarios con tarifa industrial deben ser divididos en 4 estratos, pues, al igual que en los casos anteriores, esta cantidad de agrupamientos garantiza índices de separación y cohesión aceptables con un mínimo número de agrupaciones.

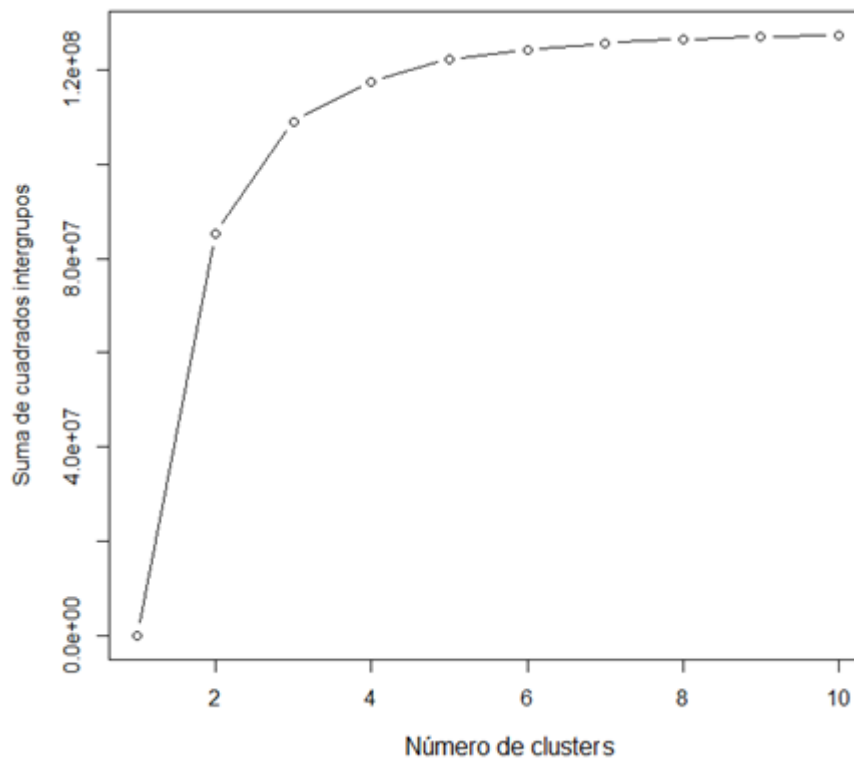


Figura 3.12. Variación de la distancia inter-cluster con relación al número de clusters en los que se agrupan los usuarios con tarifa comercial de la EEA.

(Fuente: Propia)

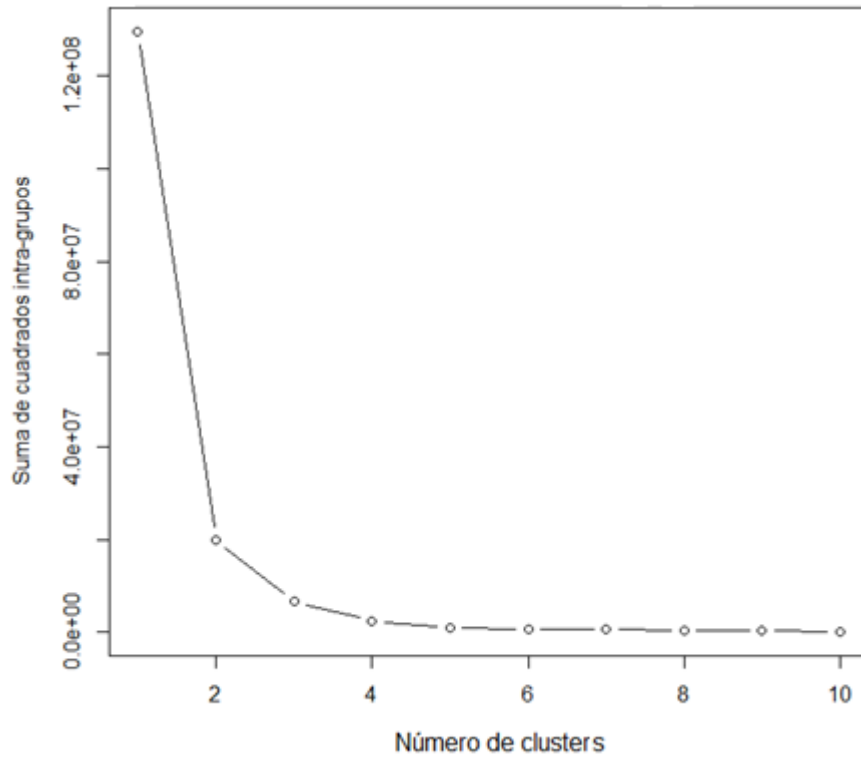


Figura 3.13. Variación de la distancia intra-cluster con relación al número de clusters en los que se agrupan los usuarios con tarifa comercial de la EEA.
(Fuente: Propia)

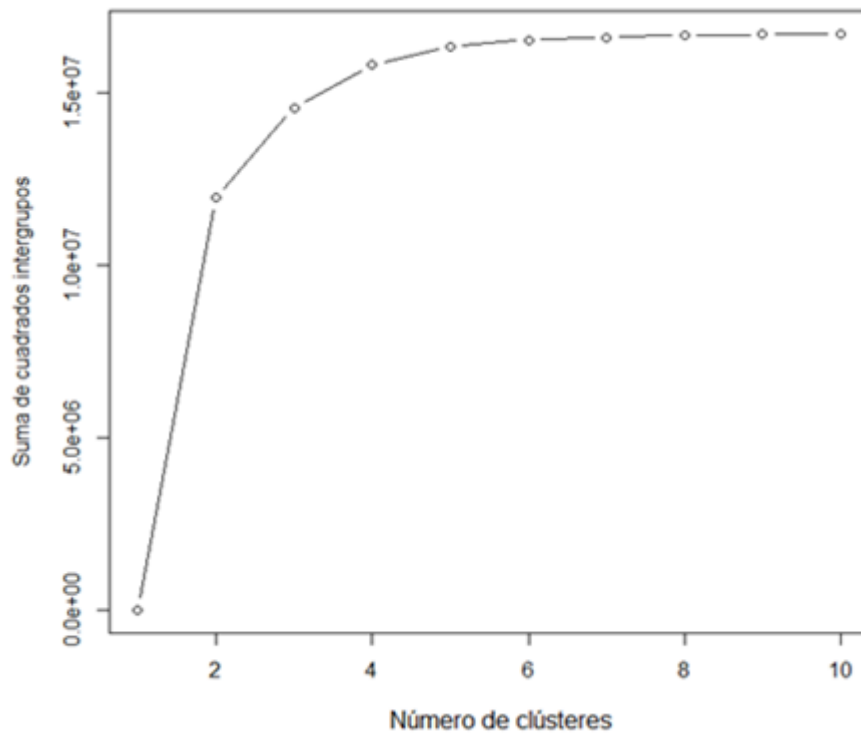


Figura 3.14. Variación de la distancia inter-cluster con relación al número de clusters en los que se agrupan los usuarios con tarifa industrial de la EEA.
(Fuente: Propia)

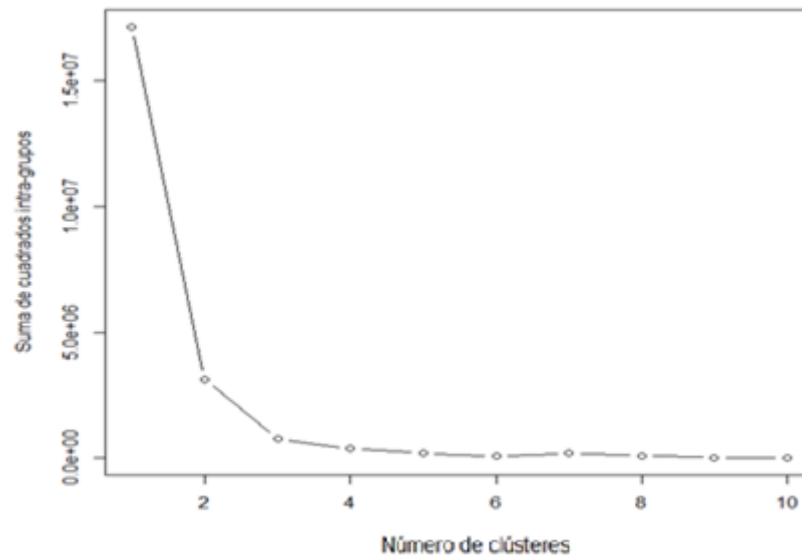


Figura 3.15. Variación de la distancia intra-cluster con relación al número de clusters en los que se agrupan los usuarios con tarifa industrial de la EEA. (Fuente: Propia)

3.4.2. Resultados obtenidos usando el método de Ward

A fin de corroborar los resultados obtenidos en el punto anterior se utiliza el método de Ward, el cuál es un algoritmo perteneciente al grupo de conglomerados jerárquicos que agrupa en forma ascendente los elementos considerando la distancia euclidiana cuadrada entre cada uno de ellos. Junto a este método son usados también los dendogramas, en los cuáles se expone de forma gráfica las uniones que se han suscitado en todo el proceso al igual que la distancia a la que han ocurrido dichas uniones. Esto sirve para encontrar el punto en donde las distancias de unión de los clusters tienden a variar considerablemente, en otras palabras, este método ayuda a encontrar el punto en donde se unen clusters de similitud baja (distancia grande), tomando el método expuesto en el capítulo dos de este documento.

La figura 3.16 (a) muestra el dendograma resultante para los usuarios con tarifa residencial de la EEA, en la que se puede distinguir los enlaces finales que se han efectuado hasta llegar a obtener un solo cluster. De aquí se puede concluir que el número ideal de estratos en los que se debería dividir a este tipo de usuarios sería cuatro o cinco estratos, pues después de ello las uniones se efectúan a distancias muy grandes.

Para el caso de los usuarios comerciales, el dendograma mostrado en la figura 3.16 (b) deja ver que entre cinco y seis clusters la distancia es pequeña, pero empieza a variar de una manera notable a partir de los cuatro clusters, denotando así, que cuatro estratos es el número óptimo de grupos en los que se debe dividir a este conglomerado de usuarios.

Finalmente, la figura 3.16 (c) muestra el dendograma resultante para los usuarios con tarifa industrial registrados en la EEA. Aquí se puede notar claramente que cuatro estratos es el número óptimo de grupos en los que se deberá dividir a este tipo de usuarios pues la siguiente unión sucede a una distancia mucho más grande y las uniones anteriores son tan pequeñas que el software con el que se realiza este proceso no las representa.

De esta manera, se puede ver que los dendogramas resultantes al usar el método de Ward exponen resultados muy similares a los obtenidos por el método k-means, por tal motivo se concluye que el número de estratos obtenidos con el método de Ward ratifica los resultados presentados en el anterior punto con el uso del método k-means.

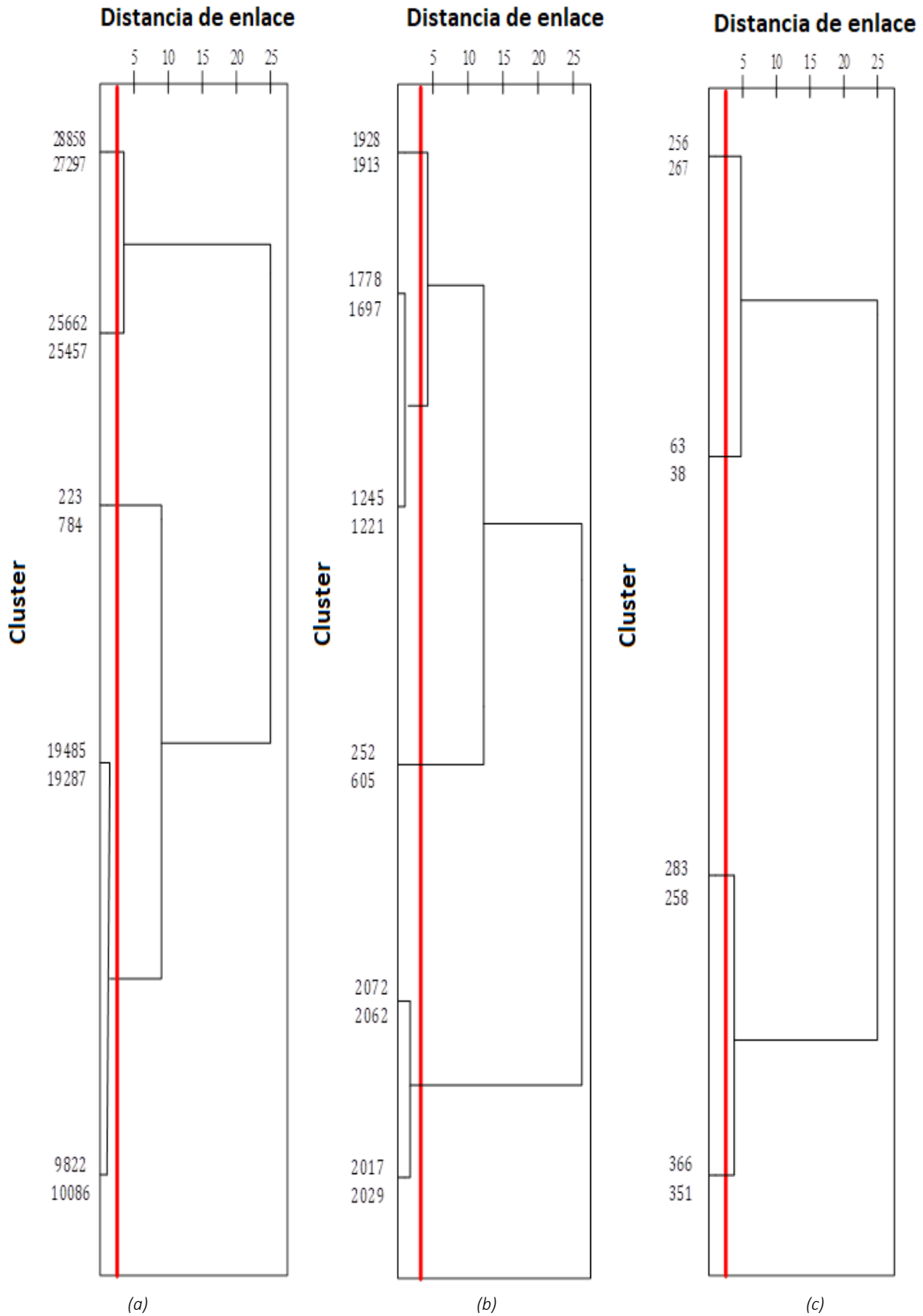


Figura 3.16. Dendrogramas obtenidos tras aplicar el método de Ward para el caso de: (a) Usuarios residenciales (b) Usuarios Comerciales (c) Usuarios industriales.

(Fuente: Propia)



3.5. LÍMITES DE CADA ESTRATO

Hasta el momento se ha realizado un análisis de la base de datos de los usuarios de la EEA, para de esta manera depurarla a fin de eliminar datos atípicos (outliers). Luego se realizó un proceso para determinar el número de estratos óptimo en los que se debe dividir a los usuarios de cada tipo de tarifa. Finalmente, en esta parte se analiza un método para definir los límites en los que están comprendidos los estratos, para ello se utiliza el método de estratificación idealizado por Dalenius & Hodges (Dalenius & Hodges, 1959).

Se debe recordar que el algoritmo descrito por Dalenius & Hodges está basado en la construcción de intervalos tomando la raíz cuadrada acumulada de las frecuencias de la variable de estratificación, de tal manera que la varianza obtenida al interior de cada estrato sea mínima y la varianza entre estratos sea máxima. Para ello, se debe tener bien definido el número de estratos en los que se dividirá la muestra, pues es un dato de suma importancia a la hora de ejecutar este algoritmo. Los detalles del presente método ya fueron descritos en el capítulo dos de este documento.

Los datos considerados como outliers, al inicio del análisis en este punto forman un estrato especial denominado como “estrato A” que es considerado de alto consumo dentro de cada grupo de usuarios.

- **Usuarios con tarifa residencial**

La tabla 3.7 muestra un resumen de los resultados obtenidos mediante el método de Dalenius & Hodges en los que se visualiza principalmente los rangos de cada estrato para el sector residencial, así como el número de clientes presentes en cada estrato.

Tabla 3.7. Límites y usuarios contenidos en cada estrato para clientes con tarifa residencial. (Fuente: Propia)

Rango de estratos para usuarios residencial			
Estrato	Límites [kWh]	Número de usuarios dentro del estrato	Porcentaje [%]
A	> 310	407	1.4
B	(160.35, 310]	3117	10.8
C	(93.56, 160.35]	7045	24.5
D	(47.32, 93.56]	9580	33.3
E	(0, 47.32]	8623	29.9

- **Usuarios con tarifa comercial**

En la tabla 3.8 se muestra los rangos obtenidos para cada estrato del sector comercial, así como el número de clientes dentro de cada uno de ellos

Tabla 3.8. Límites y usuarios contenidos en cada estrato para clientes con tarifa comercial. (Fuente: Propia)

Rango de estratos para usuarios comercial			
Estrato	Límites [kWh]	Número de usuarios dentro del estrato	Porcentaje [%]
A	> 1554.42	46	2.2
B	(621.97, 1554.42]	153	7.2
C	(285.61, 621.97]	361	16.9
D	(104.49, 285.61]	675	31.7
E	(0, 104.49]	891	41.9



- **Usuarios con tarifa industrial**

La tabla 3.9 muestra los rangos de cada estrato presentes para los usuarios con tarifa industrial, además presenta el número de clientes dentro de cada uno de los estratos definidos.

*Tabla 3.9. Límites y usuarios contenidos en cada estrato para clientes con tarifa industrial.
(Fuente: Propia)*

Rango de estratos para usuarios industrial			
Estrato	Límites [kWh]	Número de usuarios dentro del estrato	Porcentaje [%]
A	>1108.33	11	2.7
B	(461.97, 1108.33]	39	9.65
C	(222.27, 461.97]	69	17.1
D	(93.19, 222.27]	121	29.9
E	(0, 93.19]	164	40.6

3.6. MAPA DE ESTRATOS

El mapa de estratos es un medio técnico por el cual se representa de una manera fácil los distintos grupos de consumo conocidos como estratos, presentes en la zona de concesión. Dichos estratos están diferenciados generalmente por colores que representan un parámetro que los caracteriza, en este caso dicho parámetro es el consumo promedio mensual de los usuarios.

Los pasos básicos que se siguen para la elaboración del mapa de estratos son los siguientes:

- Georreferenciar los usuarios en el mapa
- Cuadricular el área de concesión
- Designar un estrato a cada cuadrícula

Ya que el desarrollo de esta parte del capítulo está basado en el uso del programa ArcGIS, se inicia por exponer este software incluyendo algunas de sus características y prestaciones más importantes para posteriormente presentar a mayor detalle los procedimientos, criterios y resultados referentes al mapa de estratos.

3.6.1. ArcGIS

El programa ArcGIS fue diseñado por la empresa Environmental System Research Institute (ESRI) entre los años de 1982 y 1999. Actualmente es una de las más importantes plataformas para el estudio de sistemas de información geográfica (SIG) ya que este software comprende un completo sistema que posibilita la recopilación, organización, administración, análisis y la distribución de información geográfica de una manera fácil e interactiva (Esri, n.d.).

Los primeros pasos de este software fueron dados en el año de 1982 cuando ESRI lanza su primer software comercial llamado ArcInfo, cuya interfaz era manejada a través de líneas de comando. Posteriormente, el año de 1990, la empresa ESRI evoluciona a un software con una interfaz gráfica denominada ArcView GIS la cuál facilitó la interacción del usuario con el software. Finalmente, en el año de 1999, ESRI transforma ArcInfo en una plataforma modular, de escritorio con la capacidad de escalarse a una plataforma empresarial, llamada ArcGis 8.x.

Actualmente, bajo el nombre ArcGis Desktop, se comercializan tres licencias: ArcInfo, ArcEditor, ArcView (Esri, n.d.). Estas comparten un mismo núcleo y están compuesta por las siguientes aplicaciones:

- **ArcMap:** Esta es la aplicación principal de ArcGis en la cual se crea, visualiza, edita, y explora los datos SIG del área de estudio, además es el lugar en donde se crea los diseños y se asignan los símbolos de los mapas en los que se trabaja.
- **ArcCatalog:** se trata de una ventana en la cual se organiza y administra información como por ejemplo geodatabases, archivos ráster, documentos de mapas, metadatos, etc. ArcCatalog organiza toda la información en una vista de árbol con lo cual se puede administrar de una manera sencilla los datos de proyectos y sus propiedades, además de tener acceso a un abanico de herramientas que permiten trabajar con esta clase de datos.
- **ArcToolbox:** Básicamente se trata de una ventana en donde se encuentra un compendio de todas las herramientas implementadas por ArcGis para el geoprocesamiento de los datos.

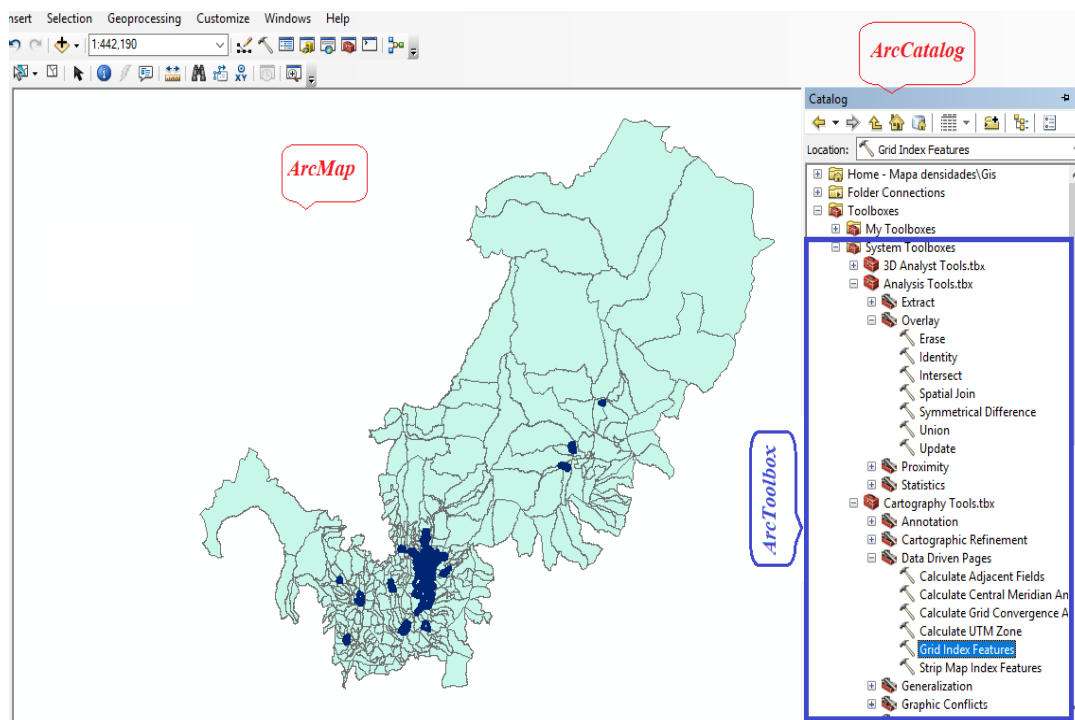


Figura 3.17. Aplicaciones principales del software ArcGis.
(Fuente: Propia)

El éxito de esta plataforma se debe a la gran cantidad de prestaciones que ofrece. Entre las cuáles se tiene:

- **Crear, utilizar y compartir mapas:** Un mapa de ArcGis muestra, integra y sintetiza capas de información geográfica y descriptiva de diversas fuentes. Estos mapas, a más de mostrar información, son usados para la consulta, análisis, planificación y administración de información. En otras palabras, un mapa en ArcGis es una ventana interactiva en donde los usuarios pueden visualizar, explorar, analizar y actualizar información geográfica.
- **Compilar información geográfica:** Este software permite sintetizar la información de muchas fuentes en una misma vista geográfica unificada. Esta capacidad de integrar múltiples datos en una misma base es de gran utilidad en muchos casos.

- **Crear y administrar bases de datos geográficas:** Las bases de datos son el pilar fundamental en la que se basa el trabajo SIG. En este aspecto, ArcGIS hace posible que la información geográfica se almacene en un formato estructurado para que de esta manera se facilite la administración, actualización, reutilización y la distribución de datos.
- **Resolver problemas de análisis espacial:** Este es uno de los aspectos más destacables e interesantes, pues ArcGIS permite extraer nueva información de los datos existentes aplicando operaciones geográficas, estadísticas y matemáticas. Así pues, ArcGIS pone a disposición una variedad de herramientas analíticas que permiten resolver una gran variedad de problemas.
- **Creación de aplicaciones basadas en mapas:** Una de las características más destacadas de ArcGIS es que no se necesita ser un desarrollador para crear aplicaciones personalizadas. ArcGIS proporciona los medios necesarios tales como plantillas de aplicaciones, interfaz de programación de aplicaciones (API), kits de desarrollo de software (SDK), además de bibliotecas de muestras de código, con el fin de que el usuario pueda crear aplicaciones ajustadas a su perfil de uso.

3.6.2. Georreferenciación de los usuarios

El primer paso para la creación del mapa de estratos es exportar la base de datos, en donde necesariamente deben estar registrada las coordenadas (longitud y latitud) del medidor de cada usuario, además del consumo promedio mensual o estrato al cual pertenece cada cliente. Así pues, en las coordenadas de cada usuario se coloca un punto de referencia, el cuál es coloreado posteriormente según el estrato al que pertenezca dicha carga. El proceso para la ejecución de este paso se explica en los Anexos uno y dos. Los resultados obtenidos para la tarifa residencial, comercial e industrial se muestran en las figuras 3.18, 3.19 y 3.20 respectivamente.

- **Usuarios con tarifa residencial**

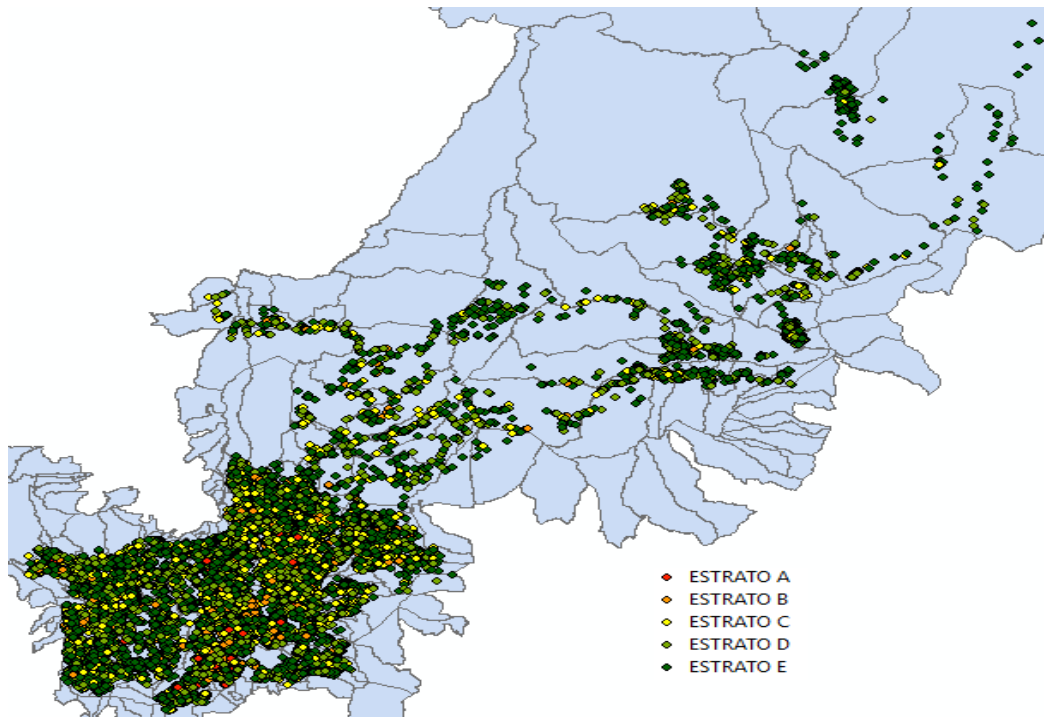


Figura 3.18. Georreferenciación de los usuarios con tarifa residencial
(Fuente: Propia)

Una vez que se exporta y georreferencia a todos los usuarios con tarifa residencial, se procede a colorear los puntos obtenidos según el estrato al cual pertenece cada usuario (tabla 3.8) tal como se muestra en la figura 3.18. En dicha imagen se puede analizar que las cargas tienen una conglomeración muy notable en la ciudad de Azogues y las parroquias próximas a esta, tales como Déleg, Javier Loyola y Guapán. La imagen muestra además que un gran porcentaje de usuarios se encuentra dentro de los estratos C, D y E, especialmente en el área rural.

- **Usuarios con tarifa comercial**

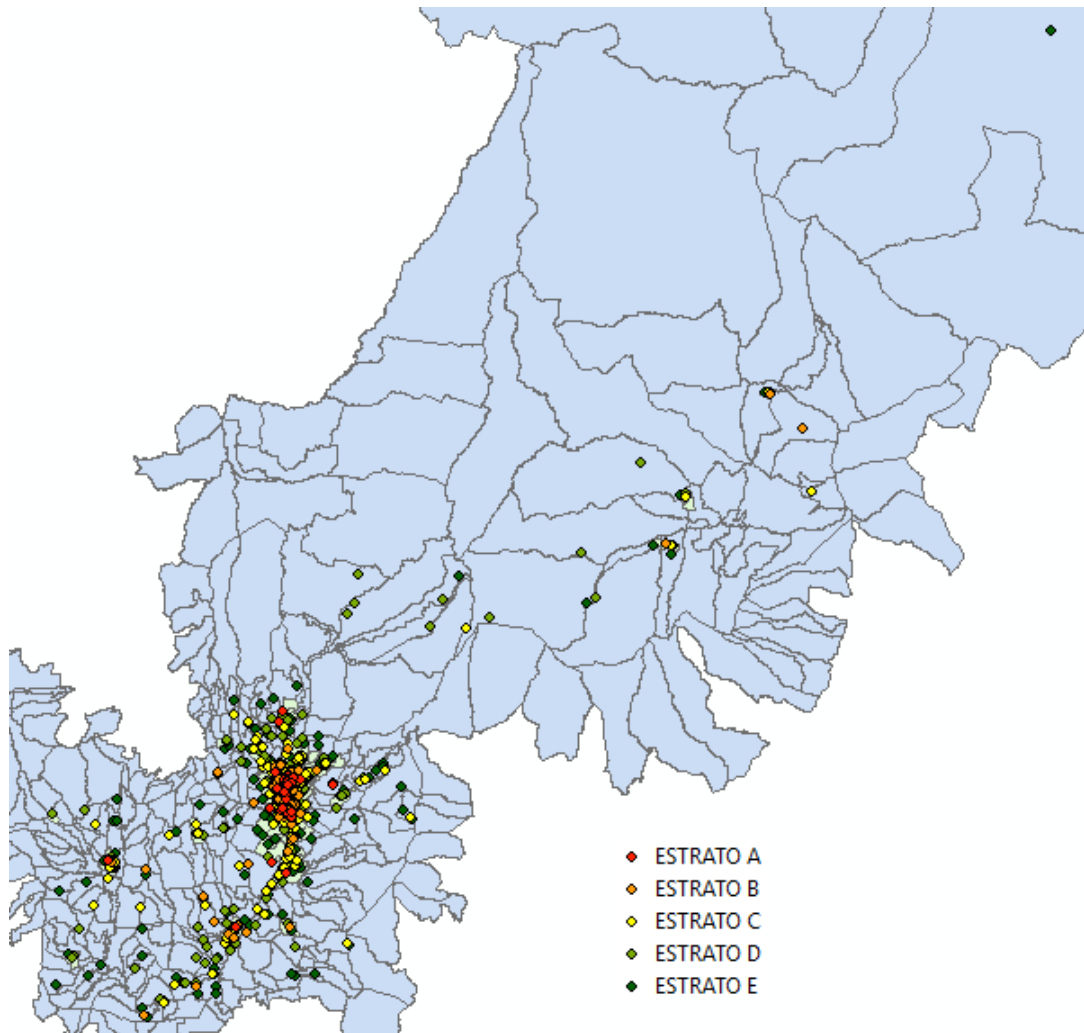


Figura 3.19. Georreferenciación de usuarios con tarifa Comercial.
(Fuente: Propia)

Para el caso de clientes con tarifa comercial se procede de igual manera que en el caso de clientes residenciales, teniendo en cuenta los rangos de estratos presentados en la tabla 3.9. Los resultados obtenidos para este caso se exponen en la figura 3.19. En este caso el número de cargas es evidentemente menor, con respecto a los clientes residenciales, aun así, se puede notar que un gran porcentaje de estas cargas se concentra en el centro de la ciudad de Azogues, siendo muchas de ellas de gran relevancia en cuanto a la demanda de energía promedio mensual (estrato A, B y C). Por otro lado, se puede apreciar que en las partes circundantes a la ciudad de azogues las cargas tienden a ser más dispersas y son dominadas por los estratos D y E.

- **Usuarios con tarifa industrial**

La figura 3.20 muestra que para el caso de los usuarios industriales sucede algo similar al caso de usuarios comerciales, pues un alto porcentaje de usuarios se encuentran conglomerados en el centro de la ciudad de Azogues, siendo dominados en este caso por las cargas pertenecientes a los estratos C, D y E. Por otro lado, las cargas alejadas de este centro urbano se tornan muy dispersas y no contienen un estrato dominante predecible a simple vista.

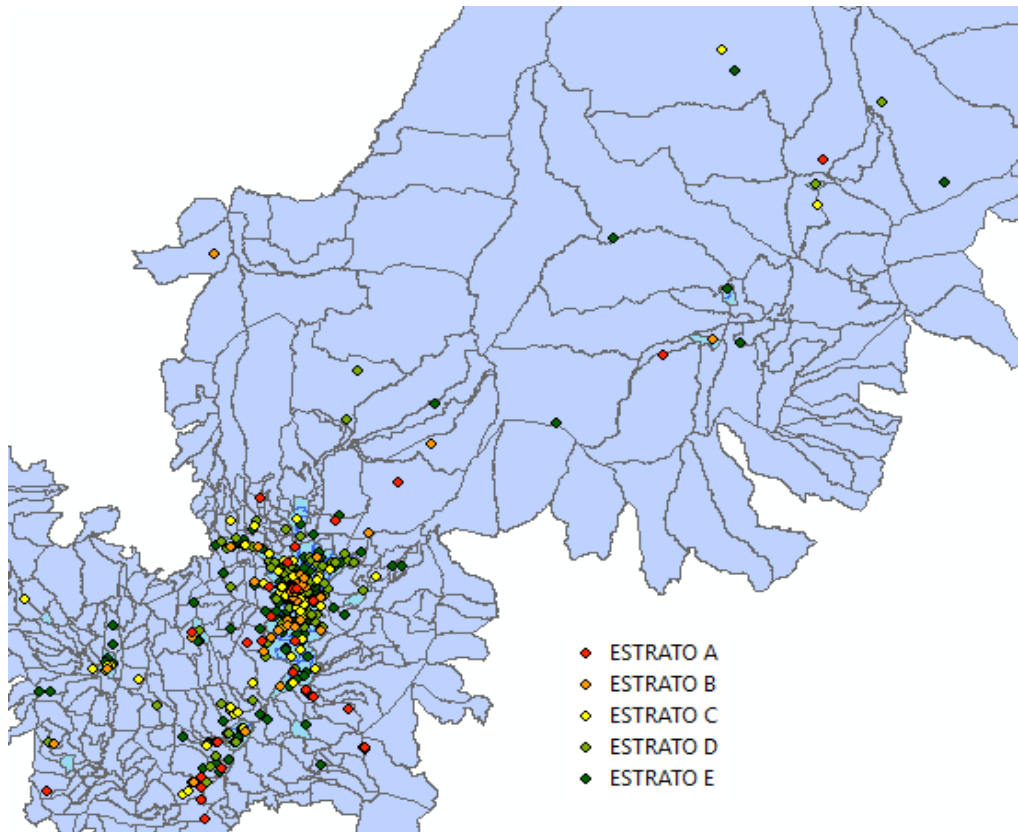


Figura 3.20. Georreferenciación para usuarios con tarifa industrial.
(Fuente: Propia)

3.6.3. Cuadriculación del área de concesión

El segundo paso para la conformación del mapa de estratos consiste en la cuadriculación de la zona de concesión, lo cual básicamente consiste en dividir el área de servicio en recuadros (cuadrículas) de dimensión n para de esta manera caracterizar el área de cada recuadro según las propiedades de los elementos contenidos dentro de cada uno de estos (figura 3.21). La dificultad radica en definir el tamaño de la cuadrícula ideal pues en base a estas dimensiones se define la homogeneidad del mapa de estratos resultante.

La metodología planteada para la definición del tamaño de cada recuadro está basada en el análisis de la redundancia, representatividad y homogeneidad que ofrecen distintas medidas de cuadrícula. Para ello es importante empezar por entender la definición de unidad mínima de análisis.

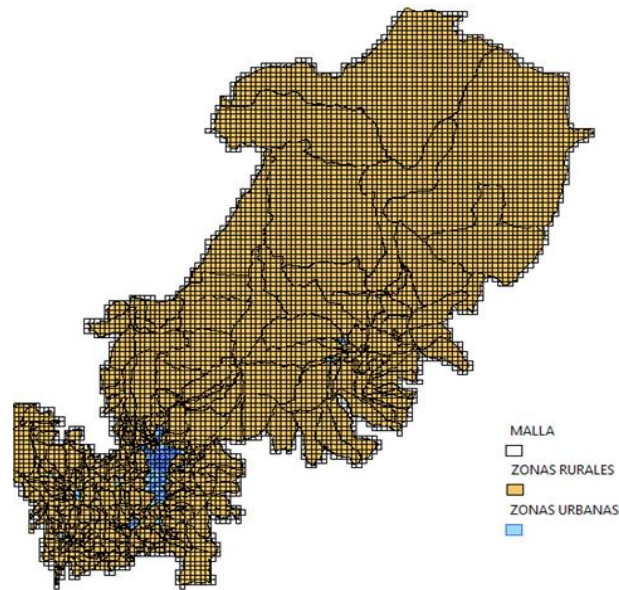


Figura 3.21. Ejemplo de la cuadrícula de la zona de concesión de la EEA.
(Fuente: Propia)

3.6.3.1. Unidad mínima de análisis

El estudio sobre una población de n elementos mediante técnicas estadísticas clásicas puede resultar en ciertos casos muy costosa y técnicamente nada viable, pues por ejemplo, estudiar una muestra aleatoria de k elemento se traduce en la recopilación de información de k puntos esparcidos indistintamente a través de toda la zona de concesión, pudiendo estos estar muy alejados o presentar un difícil acceso para su estudio lo cual representa un trabajo arduo y demorado.

Por esta razón, en muchos estudios es conveniente que la unidad mínima de análisis no sea un elemento puntual y se opta por la conformación de micro agrupaciones que representen las variables o atributos de los elementos que los conforman, para este caso estas agrupaciones se denominan como cuadrículas. La unidad mínima de análisis debe presentar las siguientes características:

- **Redundancia:** Debe existir una cantidad mínima de elementos en cada micro agrupación, de manera que se garantiza que al menos uno elemento tenga los atributos necesarios para la estimación del parámetro estudiado.
- **Representatividad:** El número de cuadrículas especiales (vacías, inconsistentes, etc.) debe ser razonablemente menor al número de cuadrículas totales, en otras palabras las cuadrículas especiales no deben tener un impacto grande en la homogeneidad del cuadrículado general.
- **Homogeneidad:** El área de la cuadrícula debe ser tal que el parámetro estimado no resulte heterogéneo.

El concepto de unidad mínima de análisis se vuelve fundamental a la hora de escoger las dimensiones que deben tener las celdas que dividen a la zona de concesión. Así pues, basándose en este concepto se procede a evaluar distintos tamaños de cuadrícula a fin de escoger el que mejor cumpla los criterios de unidad mínima de análisis.

3.6.3.2. Criterio para la caracterización del estrato de cada cuadrícula

Una vez que se ha determinado las dimensiones de cada cuadrícula se procede, como segundo paso, a caracterizar cada una de ellas en función de las propiedades de los elementos que están dentro de cada celda. El parámetro analizado es el estrato al que pertenece cada usuario, así pues para caracterizar cada cuadrícula se ordenan los datos de los usuarios, teniendo en cuenta la tabla 3.10 para luego extraer el valor de la mediana del grupo de elementos.

El uso de la mediana como medio para la caracterización de la cuadrícula se debe a que tiene como propiedad principal la poca variación con relación a sus datos extremos haciendo de esta una opción muy viable a la hora de caracterizar las cuadrículas pues, además aporta cierto grado de homogeneidad al mapa de estratos.

Tabla 3.10. Equivalencia de estrato.
(Fuente: Propia)

ESTRATO	NÚMERO EQUIVALENTE
A	1
B	2
C	3
D	4
E	5

A manera de ejemplo se presenta la figura 3.22 en la cual se examina una cuadrícula con 6 usuarios de diferentes estratos (E, E, E, E, D y C). Si se considera la tabla 3.10, los estratos de la muestra quedan ordenados como lo indica la ecuación 3.8.

$$x = \{5,5,5,5,4,3\} \quad (3.8)$$

De esta manera, si se calcula la mediana de esta muestra, se tiene como resultado el número 5, por lo tanto la cuadrícula en caracterizada con un estrato tipo E.

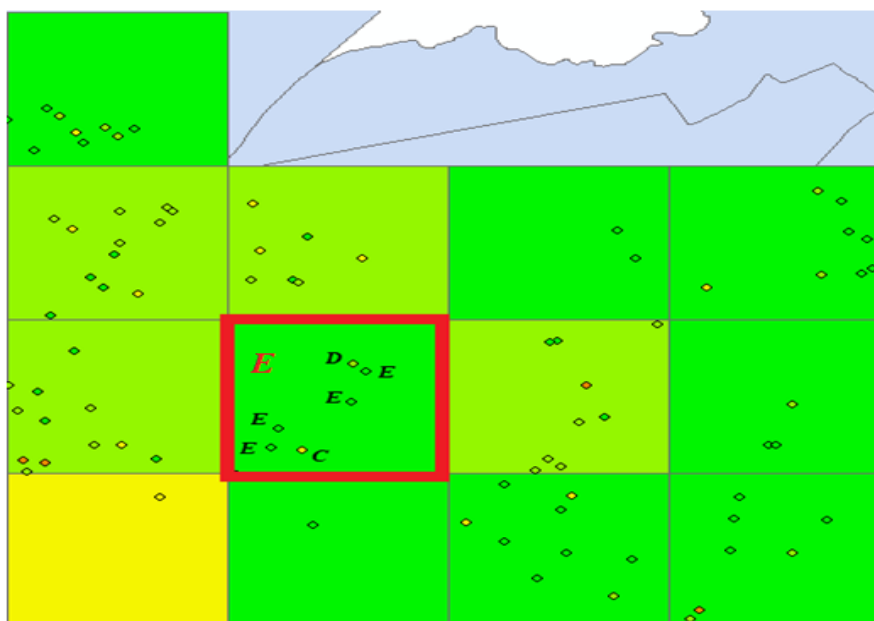


Figura 3.22. Caracterización del estrato de una cuadrícula perteneciente a la zona de concesión de la EEA.
(Fuente: Propia)



3.6.4. Resultados obtenidos

- Usuarios con tarifa residencial.**

Para el caso de usuarios con tarifa residencial, se ha probado con una variedad de tamaños de cuadrícula obteniendo así que las cuadrículas que mejor cumplen con los criterios de unidad mínima de análisis son aquellas cuyas dimensiones están comprendidas entre [400m, 1000m]. Por esta razón se escogen tres tamaños de cuadrícula: 500x500m, 800x800m y 1000x1000m que posteriormente son analizadas con el fin de encontrar el tamaño de cuadrícula que brinde mayor información sin dejar de un lado la homogeneidad del mapa resultante.

*Tabla 3.11. Evaluación de las cuadrículas sin carga y con datos atípicos.
(Fuente: Propia)*

TAMAÑO DE CUADRÍCULA	500x500	800x800	1000x1000
TOTAL DE CUADRÍCULAS	6 830	2 756	1 793
CUADRÍCULAS SIN CARGA	81.3%	76.5%	73.8%
DATOS ATÍPICOS	0	0	0

Para la evaluación de los parámetros de unidad mínima de análisis se toma en cuenta los datos de las cuadrículas presentadas en la tabla 3.11, además según el análisis visual realizado en cada caso, se califican dichos parámetros sobre 10 puntos. Cabe destacar que los valores presentados en la tabla 3.11 están dados en base a las gráficas obtenidas para los diferentes tamaños de cuadrículas. Así pues, en la tabla 3.12 se presentan los resultados de la evaluación de cada una de las dimensiones escogidas.

*Tabla 3.12. Evaluación de criterios de la Unidad Mínima.
(Fuente: Propia)*

TAMAÑO DE CUADRÍCULA	500x500	800x800	1000x1000
REPRESENTATIVIDAD	9	6	5
REDUNDANCIA	8	8	7
HOMOGENEIDAD	7	8	8
PROMEDIO	8	7.33333333	6.66666667

Según el análisis descrito, la cuadrícula que mejor cumple los criterios de representatividad, redundancia y homogeneidad es la cuadrícula de 500x500m. El mapa resultante con el uso de esta medida de cuadrícula se muestra en la figura 3.23. Se debe mencionar que, dada la gran densidad de usuarios residenciales, al escoger tamaños de cuadrícula menores a 400x400m, la homogeneidad del mapa de estratos tiende a reducirse mucho en las zonas rurales, aunque en las zonas urbanas no es tan apreciable. Por el contrario, si se escoge tamaños de cuadrícula mayores a 1000x1000m el mapa de estratos tiende a volverse demasiado heterogéneo.

Dado a ciertos criterios técnicos es aconsejable que en la parte urbana; las cuadrículas sean reducidas, por tal motivo se puede subdividir a las cuadrículas de 500x500m de las zonas urbanas en celdas de 250x250m, obteniéndose el resultado mostrado por la figura 3.24, en la cual se tiene una pequeña pérdida de homogeneidad en la parte urbana, con la ventaja de obtener mayor representatividad.

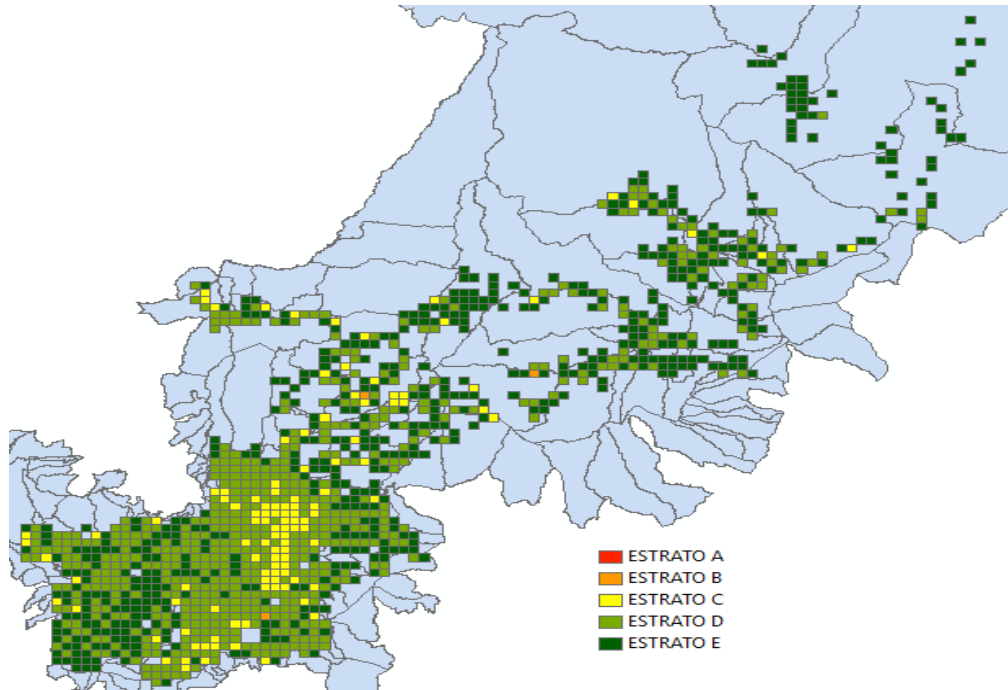


Figura 3.23. Mapa de estratos para los usuarios con tarifa residencial.
(Fuente: Propia)

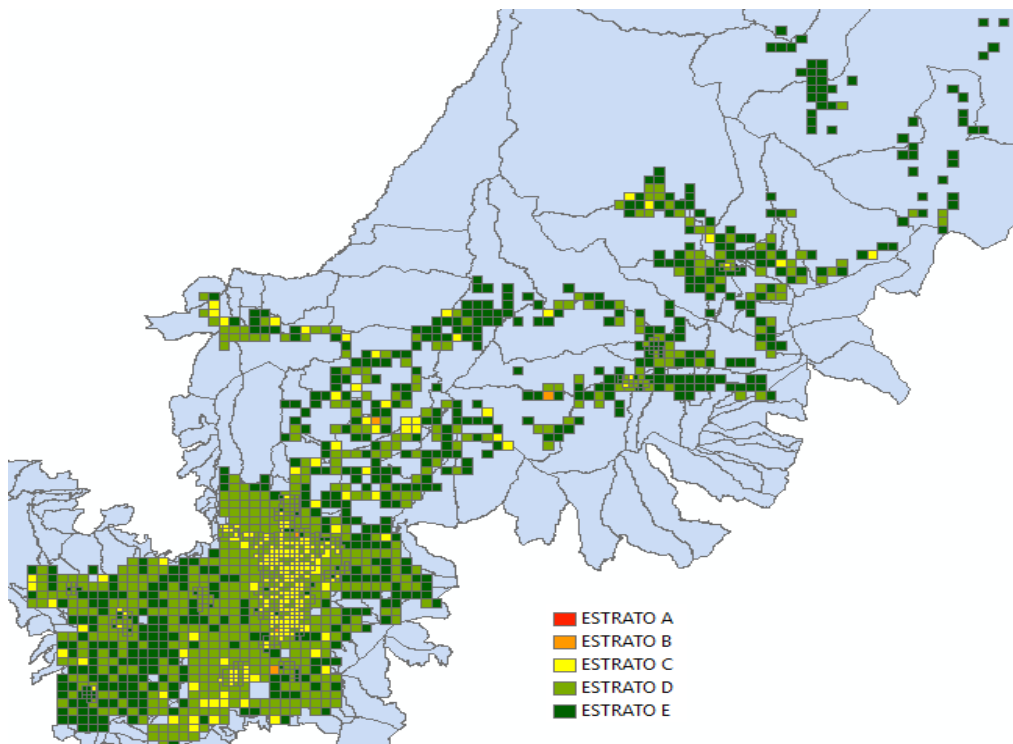


Figura 3.24. Variación del mapa de estratos para usuarios residenciales.
(Fuente: Propia)



- **Usuarios con tarifa comercial.**

Para los usuarios con tarifa comercial se siguen los mismos pasos que en el caso de usuarios residenciales, así pues, luego de examinar distintos tamaños de cuadrícula se concluye que las celdas que mejor cumplen con los criterios de unidad mínima están en el rango de [250m, 700m] de esta manera se escoge tres dimensiones de cuadrícula: 300x300m, 500x500m y 700x700m. En la tabla 3.13 se muestra la evaluación de las cuadrículas sin carga y con datos atípicos de las cuadrículas escogidas.

*Tabla 3.13. Evaluación de las cuadrículas sin carga y con datos atípicos.
(Fuente: Propia)*

TAMAÑO DE CUADRÍCULA	300x300	500x500	700x7000
TOTAL DE CUADRÍCULAS	18586	6830	3565
CUADRÍCULAS SIN CARGA	98.5%	97.07%	73.8%
DATOS ATÍPICOS	1%	0.4%	0.14%

Los resultados de la evaluación realizada se muestran en la tabla 3.14, en la cual se puede ver que la cuadrícula que mejor cumple con los parámetros de representatividad, redundancia y homogeneidad es aquella de 500x500m (figura 3.25) sin embargo, dado a ciertos criterios técnicos se prefiere cuadrículas de reducida dimensión para el área urbana y ya que un gran porcentaje de usuarios se encuentran agrupados en esta zona se ha decidido optar por la cuadrícula de 300x300m (figura 3.26) la misma que presenta un promedio ligeramente menor al obtenido por la cuadrícula de 500x500m.

*Tabla 3.14. Evaluación de criterios de la Unidad Mínima.
(Fuente: Propia)*

	300x300	500x500	600x600
REPRESENTATIVIDAD	9	9	7
REDUNDANCIA	7	8	8
HOMOGENEIDAD	8	8	8
PROMEDIO	8	8.33333333	7.66666667

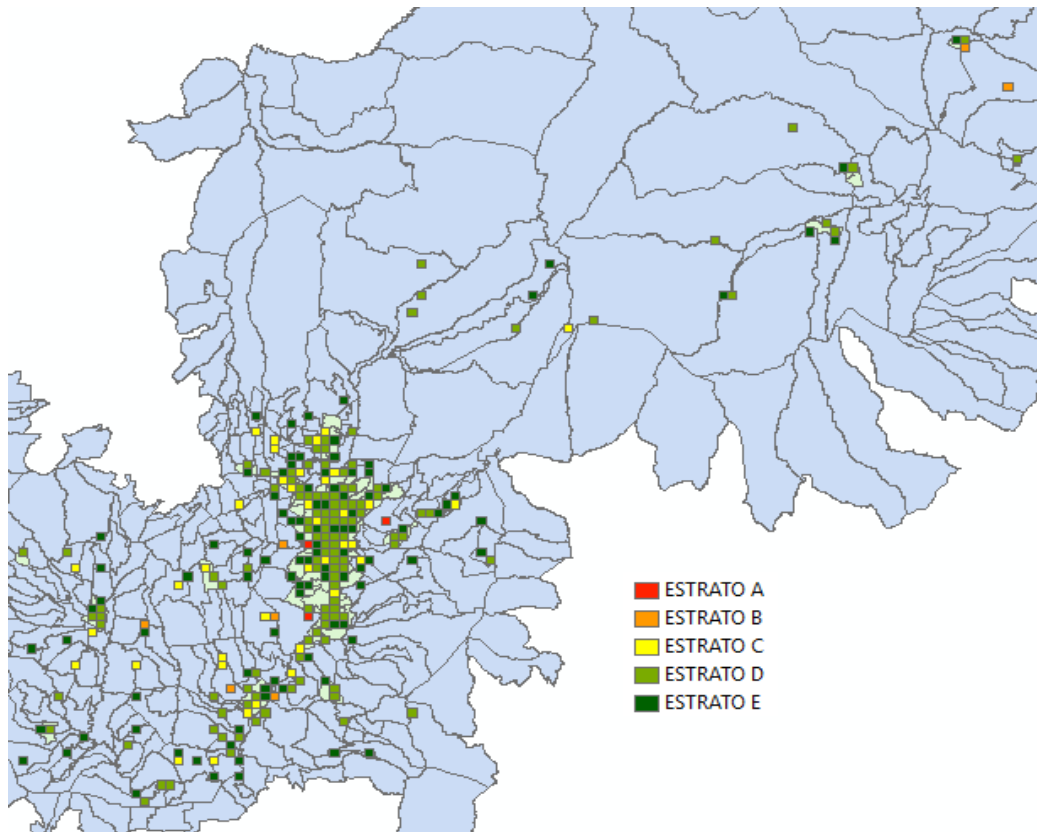


Figura 3.25. Mapa de estratos usando cuadrícula de 300x300m para clientes con tarifa comercial.
(Fuente: Propia)

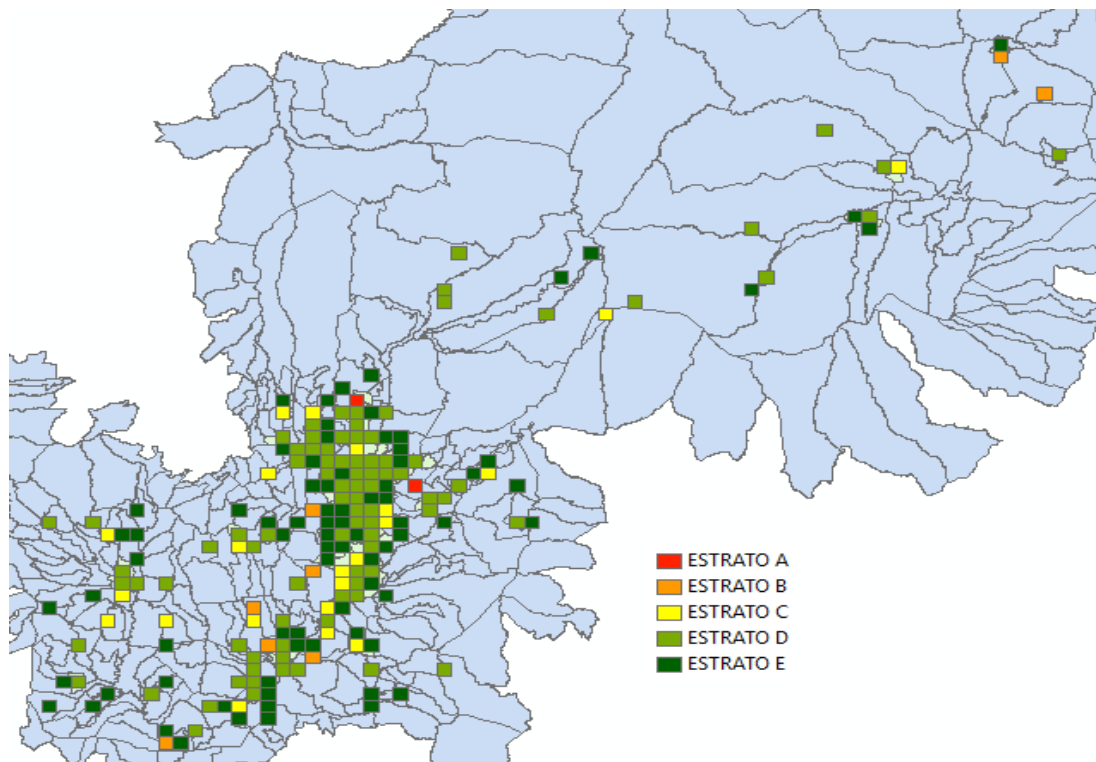


Figura 3.26. Mapa de estratos usando cuadrícula de 500x500m para clientes con tarifa comercial.
(Fuente: Propia)

- **Usuarios con tarifa industrial.**

Los usuarios con tarifa industrial son un caso especial en este análisis pues, si se toma en cuenta el bajo número de usuarios y se le añade el hecho de que un gran porcentaje está concentrado en las zonas urbanas (se requiere cuadrículas pequeñas) mientras que el restante de usuarios se encuentran muy dispersos en las áreas rurales de la zona de concesión (se requiere cuadrículas grandes), se obtiene un panorama en el que se hace difícil encontrar una cuadrícula idónea que satisfaga de una manera eficaz los criterios de la unidad mínima de análisis sobre todo el parámetro de homogeneidad.

Puede parecer una solución el diferir las cuadrículas de la zona urbana y la zona rural, para de esta manera tener cuadrículas de mayor dimensión en la zona rural y de menor tamaño en la zona urbana, sin embargo dada la gran dispersión que tienen los usuarios fuera del área urbana, muchas de las cuadrículas tienden a ser caracterizadas por una o dos cargas, tal como se muestra en la figura 3.27, por tal motivo se vuelve contra productivo el aumentar las dimensiones en las cuadrículas para las zonas rurales pues se tiende a caracterizar grandes zonas con un número reducido de datos.

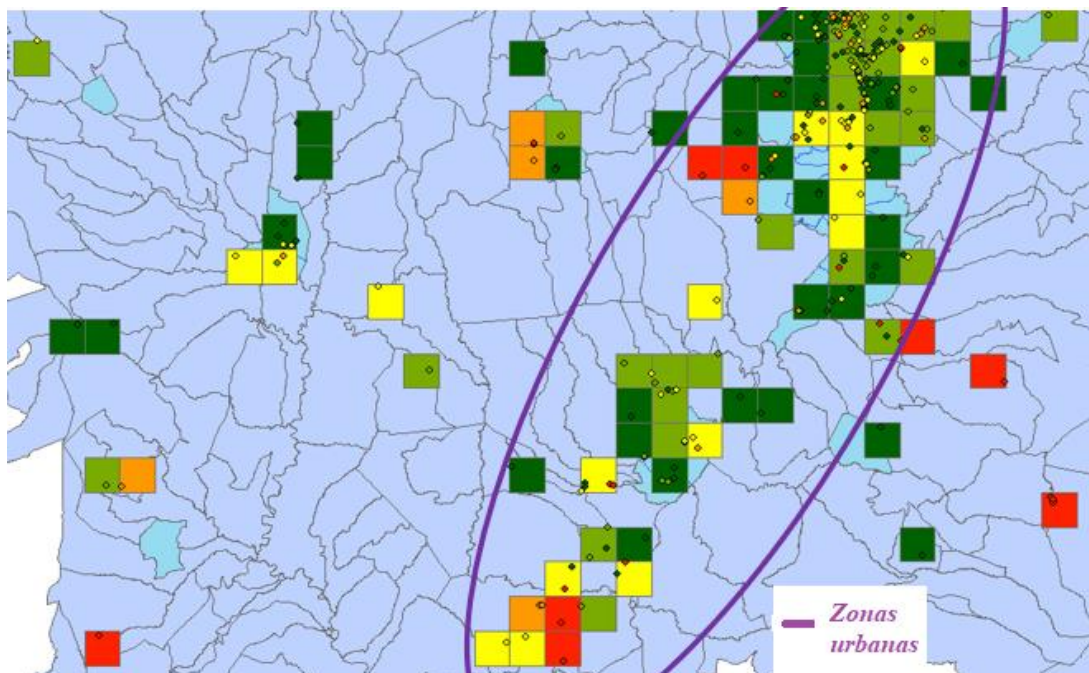


Figura 3.27. Contraste entre la densidad de usuarios presentes en zonas urbanas y zonas rurales para clientes con tarifa industrial.
(Fuente: Propia)

Tomando en cuenta los anteriores aspectos, se procede a evaluar distintos tamaños de celdas obteniendo de esta manera que las cuadrículas que mejor describen las características de unidad mínima de análisis se encuentran en un rango de [100m, 500m] por lo que se escoge tres tamaños de cuadrícula: 100x100m, 300x300m y 500x500m para realizar el análisis. Los resultados de la evaluación realizada se muestran en la tabla 3.15. Así pues, se tiene que la cuadrícula que mejor cumple con los criterios de representatividad, homogeneidad y redundancia es la de 500x500m (figura 3.28)

Tabla 3.15. Evaluación de los parámetros de unidad mínima para clientes industriales.
(Fuente: Propia)

TAMAÑO DE CUADRÍCULA	100x100	300x300	500x500
REPRESENTATIVIDAD	7	8	7
REDUNDANCIA	5	6	8
HOMOGENEIDAD	4	6	7
PROMEDIO	5.33333	6.66666	7.33333

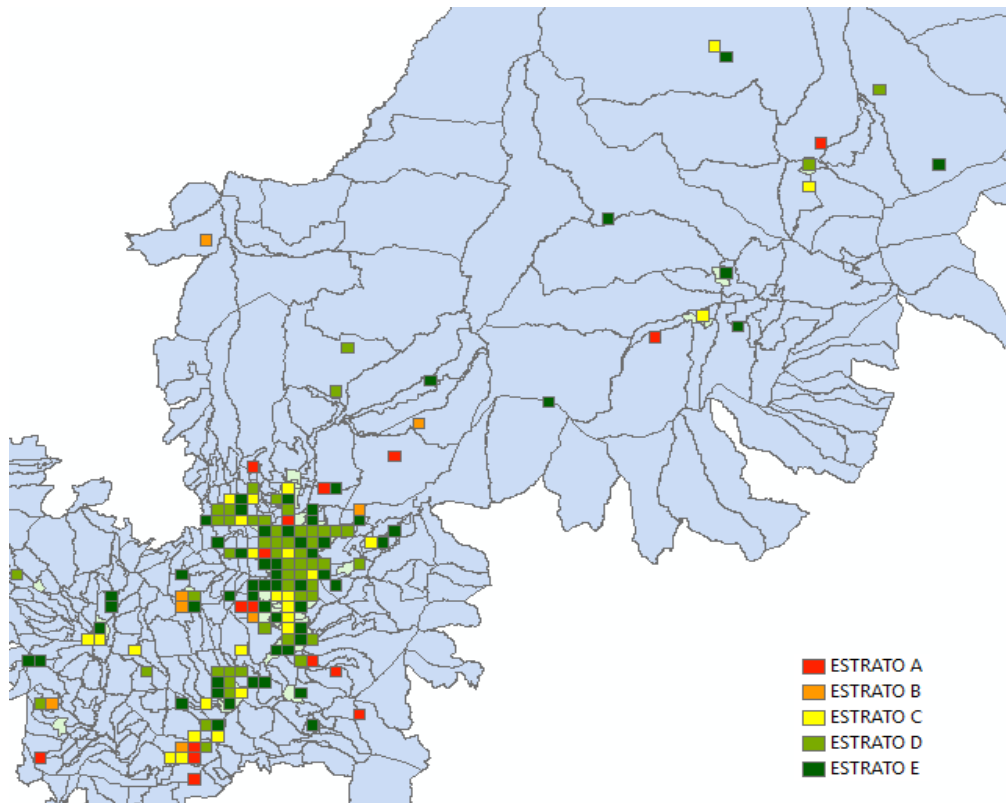


Figura 3.28. Mapa de estratos para usuarios con tarifa industrial.
(Fuente: Propia)

3.7. Demanda máxima diversificada para usuarios residenciales.

Dado el alto porcentaje de usuarios residenciales (90.5%) y con el fin de poner en un contexto aplicativo la zonificación realizada previamente, se plantea calcular la demanda máxima diversificada para distintos números de usuarios, pertenecientes a cada uno de los estratos en los que se los ha dividido el sector residencial. Esto a su vez sirve para comprobar los resultados obtenidos por el método planteado en este estudio.

Para el cálculo de la demanda máxima diversificada, se utiliza el método propuesto por la REA (Rural Electrification Administration) el mismo que dada su gran base histórica y al número de estudios realizados, brindan una buena aproximación de la demanda máxima requerida para los usuarios con una tarifa residencial. Los datos obtenidos con este método posteriormente son comparados con datos de mediciones efectuadas a usuarios al azar y con los resultados de esta comparación es posible evaluar si la demanda predicha por cálculos se cumple en realidad.



3.7.1. Método REA (Rural Electrification Administration)

La REA a través de sus estudios establece que la demanda residencial de sus usuarios puede ser calculada mediante el producto de dos factores (Vest, 1957), tal como lo indica la ecuación 3.9:

$$D. \text{coinc}_{max} = \text{factor } A * \text{factor } B \quad (3.9)$$

En donde:

$D. \text{coinc}_{max}$ = Demanda coincidente para un grupo de usuarios

$\text{factor } A$ = Factor de coincidencia para un grupo de usuarios.

$\text{factor } B$ = Energía consumida mensualmente por cada cliente.

Los valores de los factores A y B están basadas en gráficas, sin embargo las ecuaciones 3.10 y 3.11 brindan una buena aproximación de los valores de estos factores, para cualquier número de usuarios.

$$\text{factor } A = C \left[1 - 0.4C + 0.4(C^2 + 40)^{\frac{1}{2}} \right] \quad (3.10)$$

$$\text{factor } B = 0.005925 \left(\frac{\frac{kWh}{mes}}{\text{consumidor}} \right)^{0.885} \quad (3.11)$$

En donde:

C = Número de usuarios.

$\frac{\frac{kWh}{mes}}{\text{consumidor}}$ = Energía promedio mensual consumida por cada usuario.

Además, se puede relacionar el factor de coincidencia de “n” usuarios tal como se indica en la ecuación 3.12.

$$f_{coinc} = \frac{\text{factor } A}{3.29(\text{consumidores})} \quad (3.12)$$

En donde

3.29 = factor A para un solo consumidor

En la tabla 3.16 se muestra el factor A para usuarios agrupados de cinco a quinientos miembros.

Cabe resaltar que este método fue diseñado para un grupo de cinco o más clientes (≥ 5), por lo cual si se desea aplicar este método para un grupo menor se debe seguir los siguientes pasos (Pilamunga Ubidia, 2009):

- iv. Determinar la demanda coincidente máxima de 5 clientes.
- v. Despejar la suma de demandas individuales partiendo de la ecuación 2.20.
- vi. Determinar la demanda máxima individual despejándola de la ecuación 3.10.

$$D. \text{individual}_{max} = \frac{\sum D_{individuales}}{5} \quad (3.10)$$

- vii. Calcular la demanda coincidente máxima del grupo de clientes que se desee tal como se indica en la ecuación 3.11.

$$D. \text{coinc}_{max} = D. \text{individual}_{max} * n * f_{coinc} \quad (3.11)$$

Tabla 3.16. Factor B para “n” número de usuarios.
Fuente: (Vest, 1957)

# DE USUARIOS	FACTOR A
5	9.49
6	10.8
7	12.1
8	13.5
9	14.8
10	16.1
11	17.4
12	18.7
13	20.1
14	21.4
15	22.7
16	24
17	25.3
18	26.6
19	27.8
20	29.2
21	30.4
22	31.7
23	32.8
24	33.9
25	34.9
26	36
27	37.2
28	38.9
29	39.5
30	40.7
31	41.9
32	43.1
33	44.3
34	45.4
35	46.6
36	47.7

# DE USUARIOS	FACTOR A
37	48.9
38	50
39	51.25
40	52.3
41	43.4
42	54.5
43	55.4
44	56.7
45	57.9
46	59
47	60.2
48	61.4
49	62.4
50	63.5
51	64.7
52	65.7
53	66.7
54	68
55	69
56	70.2
57	71.2
58	72.3
59	73.6
60	74.5
61	75.6
62	76.7
63	77.8
64	78.9
65	80
66	81.1
67	82.15
68	83.2
69	84.3
70	85.4
71	86.5
72	87.6

# DE USUARIOS	FACTOR A
73	88.65
74	89.7
75	90.75
76	91.8
77	92.85
78	93.9
79	94.95
80	96
81	97.15
82	98.3
83	99.15
84	100
85	101
86	102
87	103
88	104
89	105.5
90	107
91	108
92	109
93	110
94	111
95	112
96	113
97	114
98	115
99	116
100	117
150	168
200	218
250	267
300	315
350	364
400	412
450	462
500	512

En la figura 3.29 se muestra el nomograma característico del método REA. Este es un método gráfico por el cual se puede conocer la demanda máxima diversificada de un grupo de clientes tan solo con

trazar una línea desde el eje perteneciente al número de clientes, hasta el eje del promedio de consumo de cada usuario, siendo la respuesta el punto de intersección entre la línea trazada y el eje de la demanda. En esta figura se puede concluir que la demanda máxima para un grupo de 8 clientes con un consumo promedio de 700kWh es de 20kW.

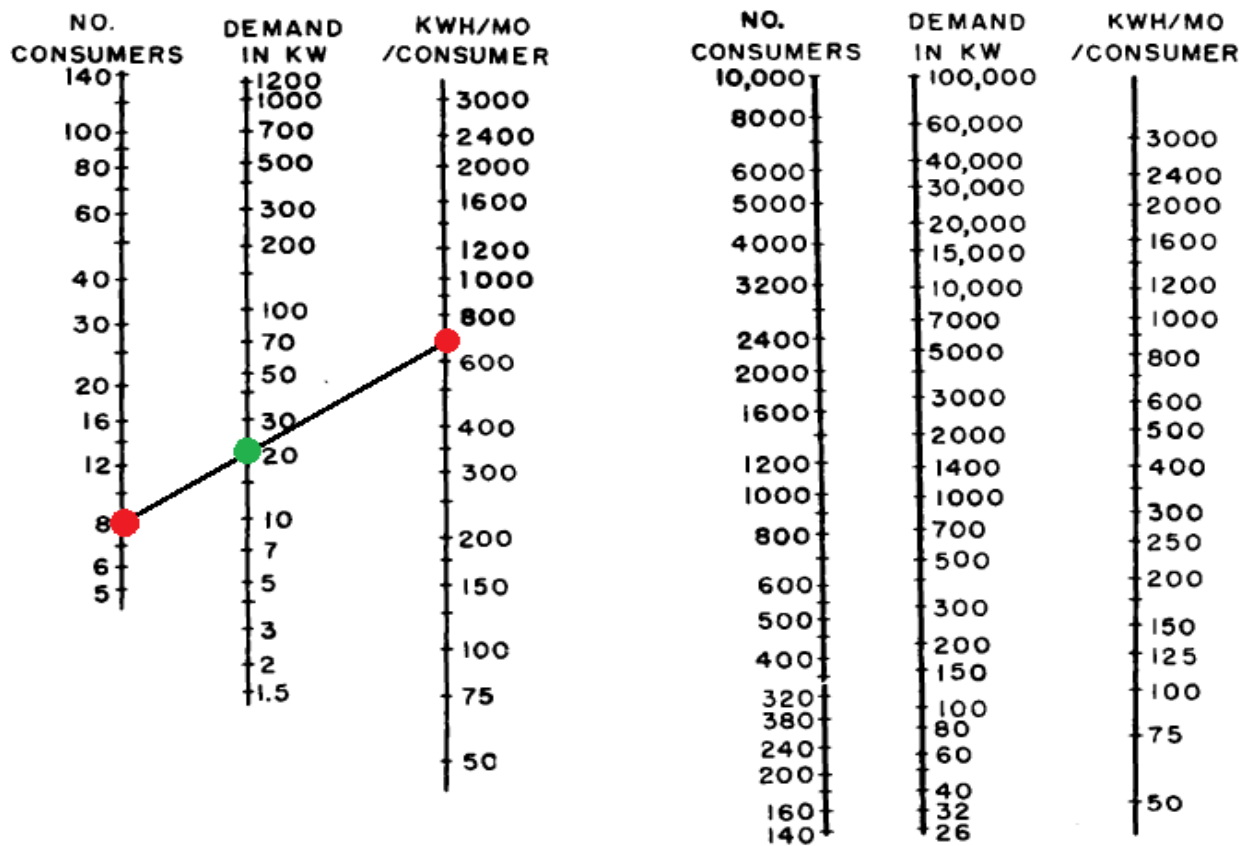


Figura 3.29. Nomograma característico del método REA.
Fuente: (Vest, 1957)

3.7.2. Criterios adicionales para el cálculo de la demanda máxima unitaria.

Para estimar de mejor manera la demanda máxima unitaria se debe subdividir a los usuarios en aquellos que disponen de cocina de inducción y aquellos que no cuentan con este dispositivo, así pues se tiene que:

- **Usuarios con cocina de inducción:**

La demanda máxima diversificada de los usuarios con cocina de inducción está dada por la ecuación 3.12:

$$DMD = factorA * factorB + DMD_{cocinas\ de\ induccion} \quad (3.12)$$

En donde:

DMD = demanda máxima diversificada

$factor\ A$ =energía consumida mensualmente por cada cliente.

$factor\ B$ = factor de coincidencia para un grupo de usuarios.

$DMD_{cocinas\ de\ induccion}$ = demanda máxima diversificada de “n” cocinas de inducción



La demanda de ‘n’ cocinas de inducción ($DMD_{cocinas\ de\ induccion}$) esta dada a su vez por la ecuación 3.13 y 3.14:

$$DMD_{cocinas\ de\ induccion} = 0.6 * N * FC_{ci} + DMU_{ci} \quad (3.13)$$

$$FC_{ci} = 0.465 * N_{ci}^{-0.128443} + 0.037 \quad (3.14)$$

En donde:

N = Número de cocinas de inducción.

FC_{ci} = Factor de carga de “n” cocinas de inducción (dada por la ecuación 3.14)

N_{ci} = Número de cocinas de inducción

DMU_{ci} = Demanda máxima unitaria de una cocina de inducción.

Para el cálculo de la DMU_{ci} se toma la ecuación 3.15, en la cual la demanda promedio de una cocina de inducción es de 4kW.

$$DMU_{ci} = 0.8(4kW) = 3.2kW \quad (3.15)$$

- **Usuarios sin cocina de inducción:**

Para el caso de los usuarios sin cocina de inducción se toma la ecuación 3.16 para el cálculo de la demanda máxima diversificada.

$$DMD = factorA * factorB \quad (3.16)$$

En donde:

DMD = Demanda máxima diversificada

$factor A$ = Energía consumida mensualmente por cada cliente.

$factor B$ = Factor de coincidencia para un grupo de usuarios.

3.7.3. Resultados obtenidos

La tabla 3.17 y 3.18 muestran la demanda máxima diversificada para diferentes números de usuarios con cocina de inducción y sin cocina de inducción, respectivamente.

Tabla 3.17. Tabla de demandas máximas diversificadas para usuarios con cocina de inducción.
(Fuente: Propia)

NÚMERO DE USUARIOS	ESTRATOS			
	0– 48 kWh	48 - 94 kWh	94 - 160 kWh	160 -390 kWh
1	1.84 kW	2.26 kW	2.84 kW	4.03 kW
2	3.2 kW	3.93 kW	4.92 kW	6.99 kW
3	4.37 kW	5.37 kW	6.73 kW	9.55 kW
4	5.36 kW	6.6 kW	8.26 kW	11.72 kW
5	6.23 kW	7.67 kW	9.6 kW	13.62 kW
6	7.02 kW	8.63 kW	10.81 kW	15.33 kW
7	7.74 kW	9.5276 kW	11.92 kW	16.92 kW
8	8.43 kW	10.36 kW	12.98 kW	18.41 kW
9	9.084 kW	11.17 kW	13.98 kW	19.84 kW
10	9.71 kW	11.95 kW	14.96 kW	21.22 kW

Tabla 3.18. Tabla de demandas máximas diversificadas para usuarios sin cocina de inducción.
(Fuente: Propia)

NÚMERO DE USUARIOS	ESTRATOS			
	0 – 48 kWh	48 – 94 kWh	94 – 160 kWh	160 -390 kWh
1	0.51 kW	0.93 kW	1.5 kW	2.69 kW
2	0.88 kW	1.62 kW	2.61 kW	4.67 kW
3	1.21 kW	2.21 kW	3.57 kW	6.39 kW
4	1.48 kW	2.72 kW	4.38 kW	7.84 kW
5	1.73 kW	3.16 kW	5.09 kW	9.11 kW
6	1.94 kW	3.56 kW	5.73 kW	10.26 kW
7	2.15 kW	3.93 kW	6.33 kW	11.32 kW
8	2.33 kW	4.27 kW	6.88 kW	12.32 kW
9	2.52 kW	4.6 kW	7.42 kW	13.28 kW
10	2.69 kW	4.92 kW	7.94 kW	14.2 kW

Para la elaboración de las tablas 3.17 y 3.18 se usa el factor de coincidencia propuesto por Vest (ecuación 3.12) sin embargo, se puede hacer uso de otros factores de coincidencia, en otras palabras, los valores de las tablas 3.17 y 3.18 pueden variar según el factor de coincidencia que se considere para el cálculo.

3.7.4. Verificación de los resultados

Para verificar los resultados calculados se obtuvo los datos de las mediciones realizadas durante 24 horas de 38 usuarios con tarifa residencial escogidos al azar. A partir de estos datos se elabora la tabla 3.19 en la que se registra el estrato del cliente y de la zona en donde este reside, además, de la demanda promedio mensual de energía y del pico de demanda máxima de cada usuario, de esta manera se puede comparar la demanda máxima unitaria de cada cliente con la demanda máxima unitaria calculada mediante el método REA para cada uno de los estratos.

Los datos entregados al examinar la tabla es que del 100% de usuarios residenciales medidos el 5.02% tiene un pico de demanda irregular mientras que el resto de usuarios pueden ser caracterizados ya sea por la demanda máxima calculada para usuarios con cocina de inducción o sin cocina de inducción respetivamente.

Esto a su vez lleva a la conclusión de que los estratos escogidos representan de una manera efectiva los distintos niveles de consumo presentes en el área de concesión, lo cual se deduce por el nivel de exactitud al predecir las distintas demandas de los usuarios.



Tabla 3.19. Tabla de comparativa entre la demanda máxima medida y la demanda máxima unitaria calculada.
(Fuente: Propia)

Código de Cliente	Demanda máxima [kW]	Energía mensual [kWh]	Categoría del cliente	Categoría de cuadrícula	Demanda máxima (cocina de inducción) [kW]	Demanda máxima (sin cocina de inducción) [kW]
4259	2.103	484.83	A	C	2.84	1.5
13433	0.396	124.75	C	C	2.84	1.5
13524	4.516	216.25	B	C	2.84	1.5
16014	0.508	198.5	B	C	2.84	1.5
18515	1.526	269.83	B	C	2.84	1.5
38471	0.153	25.25	E	C	2.84	1.5
46292	0.597	104.91	C	D	2.26	0.93
46466	0.3	70.33	D	D	2.26	0.93
59220	0.254	136.41	C	C	2.84	1.5
59568	0.948	448.91	A	C	2.84	1.5
60376	0.784	262.66	B	C	2.84	1.5
61689	1.198	224.08	B	E	1.84	0.51
62877	0.234	62.45	D	D	2.26	0.93
72207	0.297	89.16	D	C	2.84	1.5
72769	0.272	105.75	C	C	2.84	1.5
75978	0.806	247.41	B	C	2.84	1.5
80176	0.893	319.83	A	C	2.84	1.5
87353	2.296	291.25	B	C	2.84	1.5
89607	1.058	341.5	A	C	2.84	1.5
91918	0.917	233.01	B	C	2.84	1.5
94433	0.61	72.75	D	C	2.84	1.5
100636	1.187	121.25	C	D	2.26	0.93
101006	3.626	246.625	B	C	2.84	1.5
101667	1.234	664.25	A	C	2.84	1.5
102350	0.953	223.75	B	D	2.26	0.93
107870	0.878	163.08	C	D	2.26	0.93
111054	1.7	162.25	C	C	2.84	1.5
117317	1.222	445.41	A	D	2.26	0.93
118299	0.734	171.83	B	C	2.84	1.5
127696	1.56	271.58	B	E	1.84	0.51
133280	1.132	405.16	A	C	2.84	1.5
133629	0.435	140.25	C	D	2.26	0.93
135129	0.496	126.16	C	C	2.84	1.5
135418	0.372	145.75	C	C	2.84	1.5
141440	0.374	131.66	C	C	2.84	1.5
145243	0.716	213.25	B	C	2.84	1.5
147371	0.466	158.91	C	C	2.84	1.5
152702	2.6	252.66	B	C	2.84	1.5



CAPÍTULO 4: MAPA DE DENSIDAD

Por cuestiones económicas y técnicas, en ciertos estudios la evaluación puntual de elementos no es un proceso óptimo por lo cual se recurre a la conformación de micro agrupaciones que representan ciertos parámetros importantes de los elementos que los conforman. Para el caso de los mapas de densidades, estos micro conglomerados son designados con el nombre de bloques de consumo.

Lo que se busca es que a partir de la información con la que cuentan las empresas de comercialización y distribución de energía eléctrica, caracterizar cada bloque de consumo siguiendo un parámetro eléctrico el cuál permita estimar variables o atributos para distintas dimensiones de zonas geográficas.

Existe una gran variedad de parámetros eléctricos que pueden ser relacionados con los bloques de consumo, pero la demanda coincidente es uno de los mejores parámetros de estudio ya que permite la agregación de otros bloques y es fácilmente calculable. Así pues, de manera general la elaboración de un mapa de densidades se basa en dividir la zona de concesión en cuadrículas, también denominadas como bloques de consumo, las cuales están diferenciadas generalmente por colores que representan la densidad de demanda existente en cada área. Todo con el fin de poder representar gráficamente la concentración de carga en las distintas zonas del área de concesión de la EEA.

Los pasos básicos que se siguen para la elaboración del mapa de densidades son los siguientes:

- Caracterizar la curva de carga típica de cada tipo de tarifa
- Determinar la demanda de cada cliente en base a la hora de máxima demanda en el sistema.
- Georreferenciar los usuarios en el mapa.
- Cuadricular el área de concesión.
- Estimar la demanda coincidente en cada cuadrícula.
- Determinar los rangos de densidades de demanda

4.1. Caracterización de las curvas de consumo.

El objetivo general de un sistema eléctrico es cubrir la demanda de energía de los clientes, sin embargo dado que la demanda varía con el tiempo y se encuentra fuertemente ligada a la ubicación de los usuarios, se vuelve imprescindible caracterizar el consumo de los clientes para de esta manera asegurar la robustez del sistema frente a la demanda de energía requerida.

Una de las maneras más efectivas de conocer el comportamiento de la carga es tomar mediciones y elaborar una curva de carga que describa las necesidades de demanda con relación al tiempo. Este es un procedimiento muy usado por las compañías de distribución de energía, sin embargo solo se lo lleva a cabo hasta el segmento de medio voltaje pues se vuelve extremadamente costoso seguir este procedimiento en sistemas primarios y secundarios dado el volumen de estos. Para estos casos se ejecutan campañas de medición en puntos aleatorios de los circuitos primarios y secundarios, de esta manera según la información obtenida se estima las curvas de carga típicas para los diversos tipos de clientes. A continuación, se describe los parámetros a considerar para la estimación de este tipo de curvas de carga.



4.1.1. Campañas de medición.

Antes de realizar un estudio de los parámetros de una población es necesario tener claro los datos que se necesitan para el estudio. Estos datos pueden estar contenidos en una base histórica o puede que no se cuente con los datos necesarios, en cuyo caso se debe proceder a obtener la información necesaria mediante campañas de medición.

Una campaña de medición es una actividad en la que se recoge información con el fin de establecer las características de consumo de una carga determinada. Básicamente se cuenta con dos tipos de campañas:

- **Formal:** Este tipo de campaña se basa en un proceso apegado a bases estadísticas, técnicas de muestreo y clasificación que dictan los puntos más óptimos en donde se deben realizar las mediciones.
- **Informal:** Este tipo de medición se basa en el almacenamiento de información recogida en el tiempo con el ideal de que en algún momento esta información sea relevante. En muchos casos este tipo de información es cuestionada basándose en su coherencia y representatividad, sin embargo es usada frecuentemente.

4.1.2. Técnicas de muestreo

Para el caso de una campaña formal de medición, previamente, se debe identificar el universo de la población a la que se le realizará el estudio para posteriormente obtener una muestra de la misma. Una muestra es un subconjunto de la población y tiene como principal característica el ser estadísticamente representativo a la población, con un cierto grado de confiabilidad.

Existen básicamente dos tipos de muestreo:

- **Muestreo no probabilístico:** La muestra se escoge en base a ciertos criterios subjetivos.
- **Muestreo probabilístico:** En este caso la muestra se elige al azar. Se puede distinguir algunos tipos:
 - **Muestreo Aleatorio simple:** En este caso cada elemento de la población tiene la misma probabilidad de ser escogido.
 - **Muestreo Sistemático:** Un elemento es escogido al azar y con referencia a ese elemento se eligen los demás usando intervalos constantes hasta completar la muestra requerida.
 - **Muestreo Estratificado:** En este caso se divide a la población en estratos, los cuales representan subconjuntos de una población con características homogéneas con respecto a un parámetro en común. Luego se escoge al azar un número “ n ” de individuos de cada estrato basándose en la proporción del estrato o con ciertos criterios que consideren la varianza.
 - **Muestreo por Conglomerados:** Este caso es especialmente útil cuando no se dispone de información de los elementos de una población. Se inicia separando a población en conglomerados, los cuales corresponden a subconjuntos de la población con características heterogéneas. Posteriormente se escoge la cantidad de estratos y finalmente se analizan los elementos de cada conglomerado o se toma una muestra aleatoria simple de cada uno de ellos.

La elección del tipo de muestreo que debe ser utilizado depende del conocimiento que se tenga de la población (Gomez Sagner , 2007)

4.1.3. Características de las mediciones.

Una vez determinado el número y lugar de las mediciones (campaña formal) o se ha definido y organizado la información con la que se cuenta a priori (campaña informal), es necesario evaluar los criterios en los que van a estar fundamentados los datos necesarios.

- i. **Especificaciones técnicas:** Esta referido a las características de los equipos utilizados tales como precisión, funciones básicas, parámetros de medición, etc.
- ii. **Periodo de análisis:** consiste en el tiempo en el que un sistema o componente es evaluado. El periodo de análisis puede ser continuo o discreto, según lo demande la aplicación a la cuál responden los datos.
- iii. **Periodo de medición:** Se refiere a la magnitud de tiempo en el que el equipo evalúa los parámetros de interés de un elemento. Este periodo esta a su vez influenciado por tres parámetros:
 - **Intervalo de demanda:** intervalo de tiempo definido en el que la demanda promedio es calculada por el equipo.
 - **Simultaneidad:** Parámetro usado generalmente en mediciones informales, el cuál muestra la diferencia entre el periodo de análisis de los datos y el periodo de medición necesario.
 - **Periodicidad:** Se refiere a la regularidad con la que se efectúan las mediciones, mostrando así la repetitividad del periodo de medición.
 - **Estacionalidad:** parámetro referido a todos aquellos factores no eléctricos, tales como factores socioeconómicos, estructurales, urbanísticos, que son afectados por las estaciones climatológicas y que tienden a marcar comportamientos de consumo particulares en ciertas épocas del año.

4.1.4. Generación y normalización de las curvas de demanda.

La generación y normalización de las curvas de demanda típicas son realizadas con el objetivo de reproducir la demanda con relación al tiempo de cualquier usuario, conociendo su demanda de energía. Para poder realizar esto, se siguen los siguientes pasos básicos:

- **Elaboración de la base de datos:** Una vez obtenido todos los datos de la muestra examinada se procede a recopilar toda la información y de esta manera armar la base de datos.
- **Filtrado de datos:** Cuando se recopila la información, esta generalmente contiene algunos errores tales como mediciones que no son registradas o información desproporcionada con relación a valores esperados (outliers) los cuáles necesariamente deben ser examinados y filtrados de ser el caso.
- **Determinación del periodo de las curvas:** Si se realiza una campaña de medición formal, este punto queda definido a priori, por lo que las mediciones son tomadas en base al tipo de curva que se quiere obtener (diaria, semanal o mensual). Por otro lado, para el caso de una campaña informal se depende del periodo de medición, simultaneidad, periodicidad y estacionalidad de

los datos recolectados para posteriormente definir si se consideran curvas de demanda típicas diarias, semanales o mensuales.

- **Clasificación:** Una vez depurado y puesta a punto la información de la base de datos, se crea una matriz de datos que sirve de base para un algoritmo o proceso de clasificación. De esta manera se obtienen grupos de cargas con comportamientos similares. Generalmente los grupos obtenidos son clasificados en base al tipo de tarifa al que pertenecen (residenciales, comerciales o industriales), pudiendo crearse a su vez subdivisiones dentro de los grupos para de esta manera obtener curvas de demanda más específicas.
- **Normalización:** El proceso de normalización consiste en obtener una curva representativa en forma (más no en magnitud) para cada grupo en por unidad, para lo cual se define la demanda máxima de la curva y luego se divide cada dato para este valor.

En la tabla 4.1 se muestra las demandas típicas de los usuarios residenciales para cada tipo de día tanto en kW como en valores en por unidad. Además, en las figuras 4.1, 4.2 y 4.3 se muestra las curvas de demanda típicas para días laborables, sábados, domingos y feriados, respectivamente. Las figuras y tablas para los usuarios comerciales e industriales se presentan en el Anexo 7.

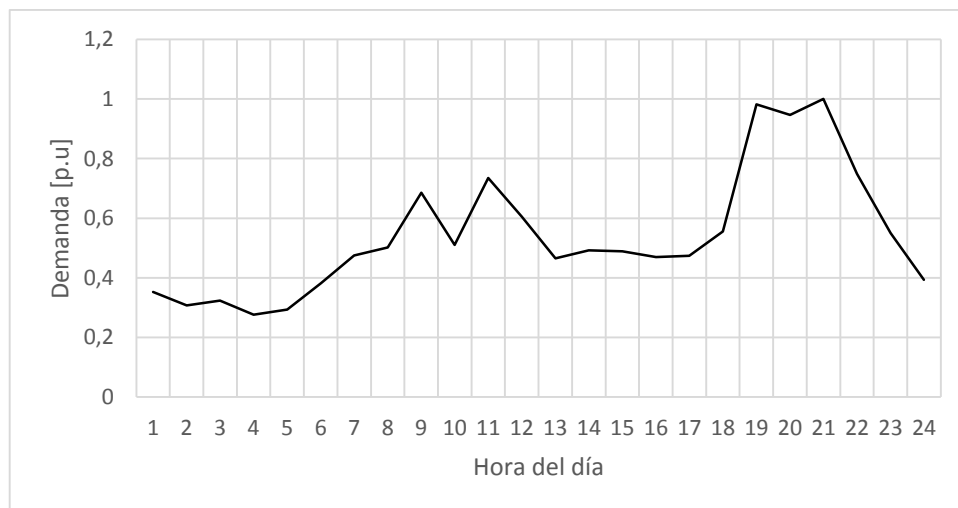


Figura 4.1. Curva típica de demanda para días laborables de usuarios residenciales.
(Fuente: Empresa Eléctrica Azogues)

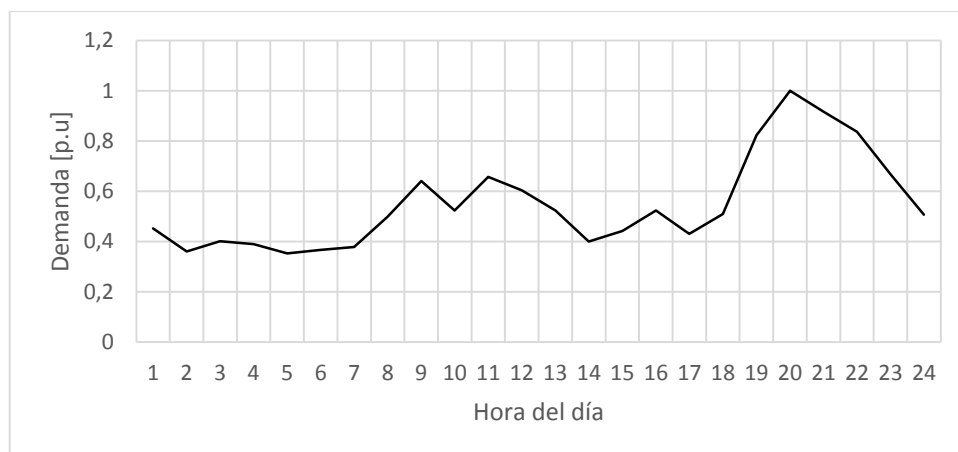


Figura 4.2. Curva típica de demanda para días sábados de usuarios residenciales.
(Fuente: Empresa Eléctrica Azogues)



Figura 4.3. Curva típica de demanda para domingos y feriados de usuarios residenciales
(Fuente: Empresa Eléctrica Azogues)

Tabla 4.1. Demanda horaria para días laborables, sábados, domingos y feriados de usuarios con tarifa residencial.
(Fuente: Empresa Eléctrica Azogues)

HORAS	Demanda (KW)			Demanda (p.u.)		
	Días Laborables (KW)	Sábados (KW)	Domingos y Feriados (KW)	Días Laborables (p.u.)	Sábados (p.u.)	Domingos y Feriados (p.u.)
01:00:00	0.19781854	0.22435939	0.17708845	0.35280062	0.45161793	0.34253094
02:00:00	0.17258326	0.17865891	0.17160574	0.30779461	0.35962644	0.33192607
03:00:00	0.18120742	0.19909277	0.17027143	0.32317542	0.4007582	0.3293452
04:00:00	0.15480988	0.19335054	0.15677727	0.27609657	0.38919953	0.3032443
05:00:00	0.16489901	0.17521305	0.18213191	0.29409009	0.35269019	0.35228617
06:00:00	0.21380262	0.18206219	0.16608108	0.38130752	0.36647696	0.32124008
07:00:00	0.26623875	0.18771105	0.17380098	0.47482505	0.37784768	0.33617219
08:00:00	0.28123444	0.24810537	0.20434518	0.50156918	0.49941674	0.39525189
09:00:00	0.38452525	0.31834748	0.23304081	0.68578379	0.6408086	0.45075602
10:00:00	0.28606922	0.25991078	0.35154884	0.5101918	0.5231801	0.67997858
11:00:00	0.41200414	0.32641049	0.29737312	0.73479117	0.65703881	0.57518993
12:00:00	0.33927851	0.29955644	0.30651371	0.60508822	0.60298371	0.59286999
13:00:00	0.26097872	0.25996127	0.36024701	0.46544401	0.52328173	0.69680289
14:00:00	0.27617245	0.19828243	0.3090714	0.49254135	0.39912704	0.59781716
15:00:00	0.27416737	0.21958558	0.2649593	0.48896537	0.44200862	0.51249393
16:00:00	0.26318611	0.25984806	0.23086598	0.46938079	0.52305385	0.44654938
17:00:00	0.26586496	0.2135199	0.2734351	0.47415841	0.42979888	0.52888814
18:00:00	0.31174438	0.25303217	0.35153436	0.55598231	0.50933399	0.67995058
19:00:00	0.55066317	0.40875807	0.44287822	0.98208342	0.82279806	0.85663119
20:00:00	0.53052518	0.49679027	0.51176704	0.9461682	1	0.98987845
21:00:00	0.56070916	0.45539906	0.51699988	1	0.91668274	1
22:00:00	0.41990703	0.41537075	0.40695779	0.74888562	0.83610887	0.78715259
23:00:00	0.30826101	0.33114212	0.28059085	0.54976988	0.66656322	0.54272904
00:00:00	0.22085893	0.25138148	0.23520712	0.39389213	0.50601127	0.45494619

Como se puede notar en la tabla 4.1, y en las figuras 4.1, 4.2 y 4.3, la demanda máxima de un usuario residencial se da en los días laborables a las 21h00 teniendo un pico de demanda muy similar al que se tiene los días domingos y feriados, los cuáles suceden a la misma hora. En cuanto a los días sábados se puede inspeccionar la demanda máxima se da a las 20h00 y que este día es típicamente cuando menos consumo tienen los usuarios residenciales.

4.2. Determinación de la potencia de cada cliente en base a la hora de máxima demanda.

4.2.1. Metodología para la estimación de la demanda de cada cliente

La demanda de cada usuario, a la hora de máxima demanda en el sistema, es el dato principal con el cuál se elabora el mapa de densidades. Para obtener dicho dato se siguen los siguientes pasos:

- i. Relacionar cada usuario con su grupo tarifario
- ii. Calcular la demanda promedio partiendo de la energía promedio mensual, para lo cual se debe tener muy en cuenta el periodo de medición de la curva asociada. Matemáticamente se puede expresar la demanda promedio con la relación entre la energía promedio mensual y el periodo de medición de la curva típica asociada, tal como se muestra en la ecuación 4.1.

$$D_{prom} = \frac{E_{mensual}}{T_{CDTN}} \quad (4.1)$$

En donde:

D_{prom} : Demanda promedio

$E_{mensual}$: Energía mensual facturada

T_{CDTN} : Periodo de medición referente a la curva de demanda típica normalizada

- iii. Obtener la demanda máxima, para lo cual se debe relacionar la demanda promedio con el factor de carga. Esta relación matemática se muestra en la ecuación 4.2.

$$D_{max} = \frac{D_{prom}}{F_{carga}} = \frac{E_{mensual}}{T_{CDTN} * F_{carga}} \quad (4.2)$$

En donde:

D_{max} : Demanda máxima

F_{carga} : Factor de carga

- iv. Obtener la demanda a la hora especificada, para lo cual se multiplica el valor de la demanda máxima y el valor de la curva típica normalizada a la hora designada tal como se presenta en la ecuación 4.3.

$$D(h) = D_{max} * CDTN(h) \quad (4.3)$$

En donde:

$D(h)$: demanda a la hora "h"

$CDTN(h)$: curva de demanda típica normalizada a la hora "h".



4.2.2. Ejemplo para la determinación de la demanda de un cliente.

Para determinar la carga de un cliente a la hora de máxima demanda en el sistema antes que nada se debe conocer la hora a la cual se da el pico de demanda en las subestaciones. Para el caso de la EEA, esta se da a las 19:00 (dato proporcionado por el departamento de planificación de la EEA).

Como segundo paso es necesario conocer tanto el grupo tarifario como la demanda de energía promedio mensual del cliente escogido para realizar el cálculo. En la tabla 4.2 se presentan los datos del usuario que son usados para este ejemplo.

Tabla 4.2. Datos del usuario usado para el ejemplo.
(Fuente: Empresa Eléctrica Azogues)

Número de medidor	Tensión	Tarifa	Demanda Promedio [kWh]
124138	Baja	Residencial	118.33

En este caso, se trata de un usuario de tarifa residencial por lo que se lo relaciona con su respectiva curva de demanda típica normalizada. Dado que en los días laborables es donde se tiene una mayor demanda, se toma la curva de demanda típica perteneciente a este tipo de día (figura 4.1) para calcular la demanda del usuario.

Como se puede apreciar en la figura 4.1 y en la tabla 4.1, a las 19h00 el consumo es de 0.98 p.u. Con este dato, y la información proporcionada por la tabla 4.2 se procede a calcular la demanda máxima de este usuario, tal como se muestra en la ecuación 4.4.

$$D_{max} = \frac{E_{mensual}}{T_{CDTN} * F_{carga}} = \frac{118.33}{720 * 0.542} = 0.29kW \quad (4.4)$$

Ya que se relaciona la energía mensual consumida por el cliente con las curvas típicas de demanda normalizadas, las mismas que son curvas diarias, es necesario dividir la energía mensual obtenida para 720 que representa la cantidad promedio de horas en un mes. De esta manera se obtiene la energía consumida en un día.

Como paso final se evalúa la demanda del usuario a la hora de máxima demanda en el sistema (19h00) tomando la ecuación 4.3 en la que se relaciona la demanda máxima del cliente con su curva de demanda típica normalizada. El resultado obtenido se muestra en la ecuación 4.5.

$$D(19h00) = D_{max} * CDTN(19h00) = 0.29 * 0.98 = 0.2842 kW \quad (4.5)$$

4.3. Georreferenciación de los usuarios en el mapa

Como tercer paso para la creación del mapa de densidades se exporta la base de datos al programa ArcGIS, en donde deben estar registradas las coordenadas (longitud y latitud) del medidor de cada usuario, además del tipo de cliente y de la demanda a la hora de máximo consumo. Así pues, en las coordenadas de cada usuario se coloca un punto de referencia tal como se muestra en la figura 4.4.

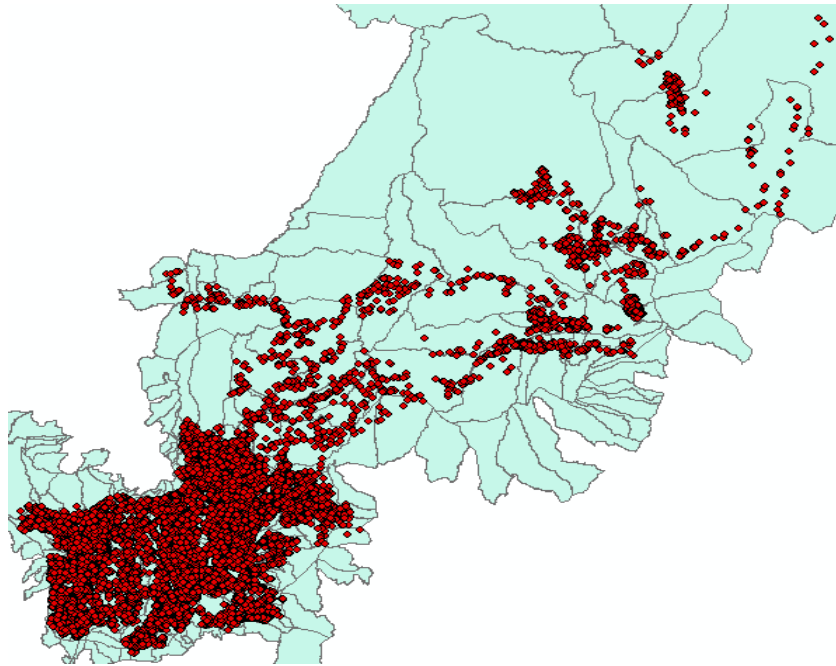


Figura 4.4. Georreferenciación de los usuarios de la EEA.
(Fuente: Propia)

Para este caso se ha optado por agrupar a los usuarios con características similares en estratos con la finalidad de examinar, de forma visual, su demanda y distribución a través del área de concesión, prestando singular atención a los usuarios con mayor magnitud de demanda. Cabe resaltar que este no es un paso obligatorio a la hora de realizar el mapa de densidades, sin embargo, al realizarlo se puede obtener mayor comprensión acerca de los parámetros estudiados.

Los pasos y conceptos usados para la agrupación de los clientes son las mismas presentadas en el capítulo 3 para la elaboración del mapa de estratos.

4.3.1. Filtrado de datos atípicos de la muestra.

Como se puede notar en la figura 4.5, la distribución de los datos tiene una asimetría notable por tal motivo se procede con el método propuesto por Hubert & Vandervieren para determinar los posibles outliers presentes en la base de datos.

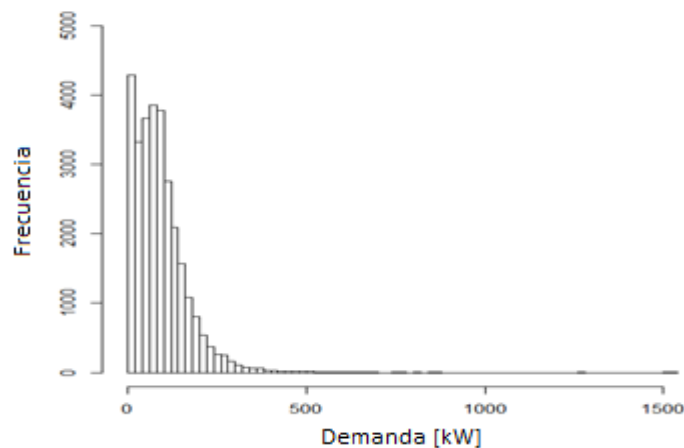


Figura 4.5. Distribución de datos sin la determinación de outliers.
(Fuente: Propia)

Los resultados obtenidos, al usar el método antes mencionado para la detección de datos atípico, son los límites $[-0.082, 0.839]$, lo que significa que cualquier usuario con una demanda mayor a 0.839 kW o menor a -0.082 kW es considerado como outlier. Así pues, se obtuvo que, de un total de $31\ 121$ usuarios, 641 (2%) de ellos superan los 0.839 kW a la hora de máxima demanda del sistema por lo tanto se filtran estos datos quedando un total de $30\ 480$ (98%) usuarios que componen la base de datos definitiva. El límite inferior (-0.082 kW) no es tomado en cuenta pues ningún usuario contiene una demanda negativa.

El resumen de los datos obtenidos en este proceso se muestra en la tabla 4.3, mientras que la figura 4.6 muestra la distribución de los datos una vez se han filtrados los elementos considerados como atípicos.

Tabla 4.3. Resumen de datos obtenidos en el proceso de filtrado de outliers.
(Fuente: Propia)

Resumen mapa de densidad	
Muestra inicial	31121
Muestra Final	30 480 (98%)
Datos atípicos	641 (2%)
Valor extremo 1	0.839
Valor extremo 2	-0.082
Medcouple	0.1527
Cuartil 1	0.096
Cuartil 3	0.317

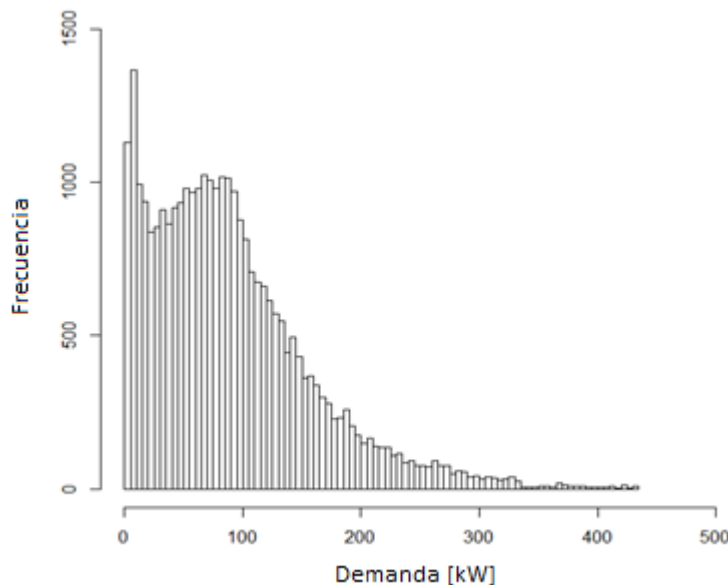


Figura 4.6. Distribución de datos una vez filtrados los outliers.
(Fuente: Propia)

4.3.2. Definición del número de grupos

Una vez filtrados los datos atípicos y puesto a punto la base de datos de los clientes se procede a evaluar cuál es el número ideal de grupos en los que se debe dividir a los usuarios. Para tal propósito, se usa el método k-means, con el cual se conforma distintos números de agrupaciones para posteriormente evaluar la cohesión intra-grupal de los elementos y la separación inter-grupal. Este procedimiento se encuentra descrito en el capítulo 3.

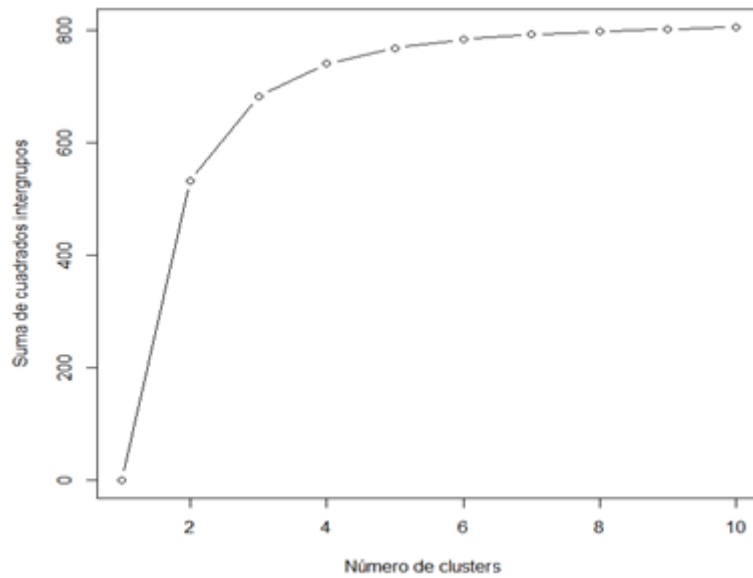


Figura 4.7. Variación de la distancia inter-cluster con relación al número de clusters en los que se agrupan los usuarios para la elaboración del mapa de densidades.
(Fuente: Propia)

La figura 4.7 muestra la variación de la distancia entre grupos con relación al número de clusters en los que se divide a los usuarios. Se nota que a partir de cuatro grupos se tiene un coeficiente de separación muy cercano a su límite máximo.

La figura 4.8 muestra la variación de la distancia (similitud) de los elementos que componen cada grupo con relación al número de grupos en los que se divide a los usuarios. Se puede observar que a partir de 4 grupos la distancia entre elementos del mismo cluster tiende a ser muy próxima a su mínimo. De esta manera se puede concluir que cuatro es el número ideal de grupos en los que se debe dividir a los usuarios para obtener índices de cohesión y separación aceptables, con el uso del menor número de grupos posible.

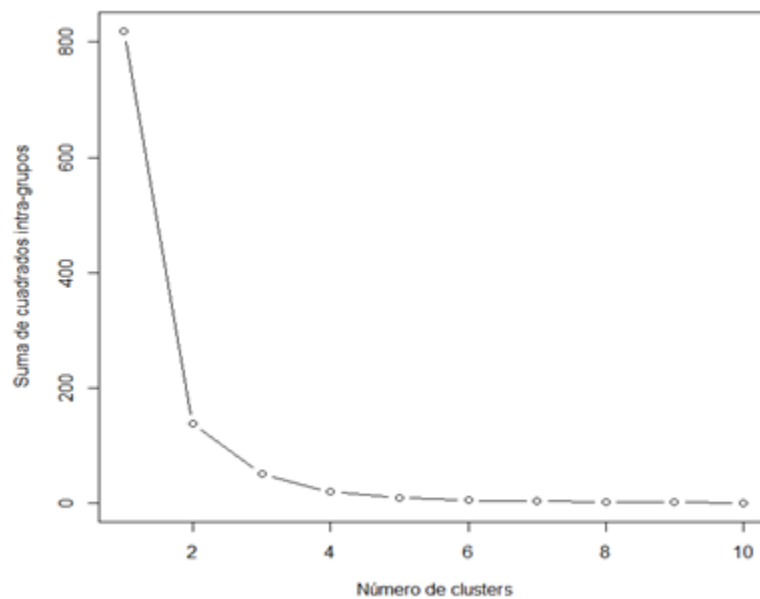


Figura 4.8. Variación de la distancia intra-cluster con relación al número de clusters en los que se agrupan los usuarios para la elaboración del mapa de densidades.
(Fuente: Propia)

4.3.3. Límites de cada grupo

Para determinar los límites de cada agrupación se utiliza el método propuesto por Dalenius & Hodges, con lo cual se obtienen los resultados expuestos en la tabla 4.4. Al igual que en el capítulo 3, los datos considerados como outliers en este punto forman el grupo A considerado de alta demanda.

Tabla 4.4. Límites de demanda para los grupos de cargas del mapa de densidades.
(Fuente: Propia)

Rango de estratos para			
Grupo	Límites [kW]	Número de usuarios dentro del estrato	Porcentaje [%]
A	> 1.84	657	1.7
B	(0.43, 1.84]	3848	11.3
C	(0.25, 0.43]	7327	23.5
D	(0.13, 0.25]	9414	30.2
E	[0.000742, 0.13]	10375	33.3

4.3.4. Resultados obtenidos

La figura 4.9 muestra a los usuarios georreferenciados y separados por colores según su demanda a la hora de máxima carga del sistema. En este caso se puede examinar que un gran porcentaje de la demanda se encuentra entre los límites [0.000742; 0.43) (estratos C, D y E). Por otra parte, las demandas cuyos límites son [0.43; 15.93] (estratos A y B) se encuentran agrupadas principalmente en la ciudad de Azogues y sus alrededores.

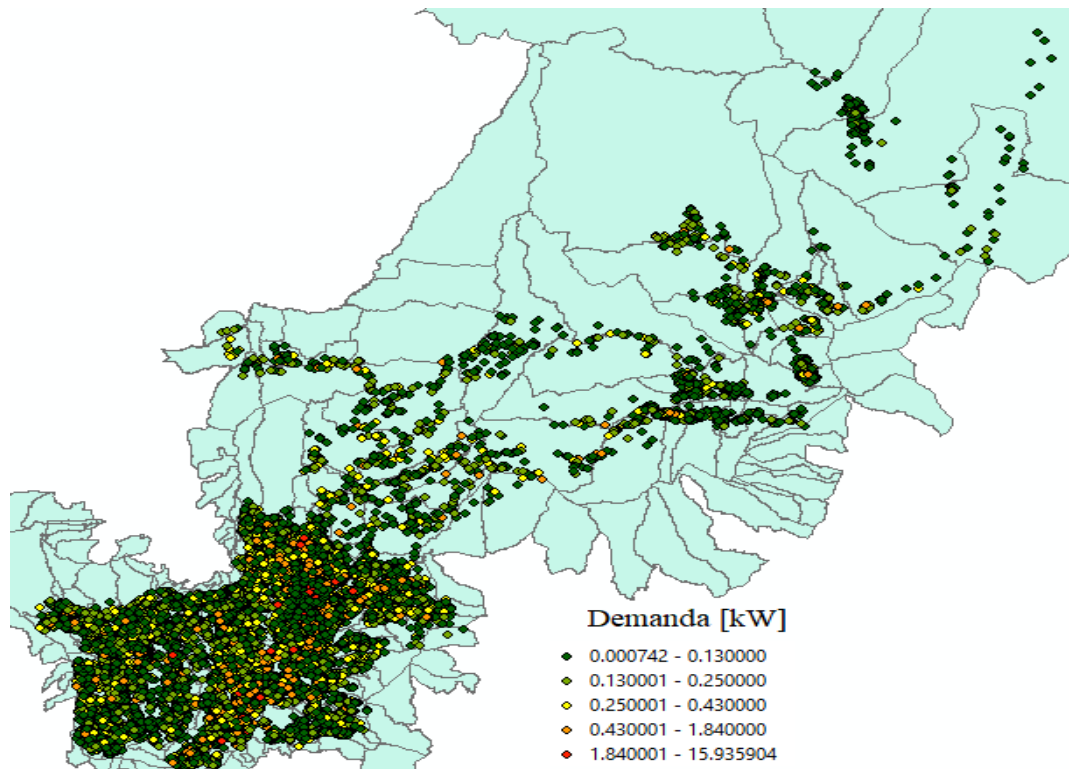


Figura 4.9. Usuarios georreferenciados divididos en grupos de consumo.
(Fuente: Propia)



4.4. Cuadrícula del área de concesión

En este punto se divide el área de concesión de la EEA en cuadrículas de dimensión “n” con el fin de caracterizar el área de cada cuadrícula en función de parámetros representativos de los elementos que estén en su interior. El parámetro con el que se examina cada cuadrícula es la sumatoria de las demandas de los usuarios.

Para determinar la dimensión ideal de las cuadrículas se examinan diferentes tamaños para posteriormente evaluar las celdas en base a los criterios de unidad mínima de análisis (Representatividad, Homogeneidad y Redundancia), de manera similar a como se lo realiza en el capítulo 3. Así pues, con la ayuda del software ArcGIS se ha cuadrículado la zona de concesión de la EEA con tres tamaños de cuadrícula diferentes: 400x400m, 600x600m y 800x800m. Las figuras 4.10, 4.11 y 4.12 muestran la cuadrícula de la zona de concesión con los diferentes tamaños de celda seleccionados.

Cabe aclarar que los diferentes estratos que se pueden apreciar en las figuras antes citadas fueron realizados con el método *Jenks* que trae por defecto el software ArcGIS, esto con el fin de tener una mejor guía para la elección del tamaño de cuadrícula. Más adelante se analiza el número de grupos ideales que se deben usar, así como los límites que deben tener cada uno de ellos.

Para la evaluación la dimensión de las celdas se ha tomado en cuenta la tabla 4.5 que indica el porcentaje de cuadrículas vacías y el porcentaje de celdas que son consideradas de alta demanda o atípicas para cada una de las dimensiones escogidas. Así pues, los resultados de la evaluación son mostrados en la tabla 4.6 en donde se puede notar que la cuadrícula que mejor cumple los criterios de unidad mínima es la que tiene una dimensión de 400x400m.

Tabla 4.5. Evaluación del porcentaje de celdas vacías y celdas con datos atípicos según las dimensiones de cada cuadrícula. (Fuente: Propia)

	400	600	800
TOTAL CELDAS	18586	4795	2756
CELDAS VACIAS (%)	86.34456042	79.0823775	76.3425254
DATOS DE ALTA DEMANDA (%)	0.021521575	0.08342023	0.14513788

Tabla 4.6. Resultados de la evaluación de cada cuadrícula con base en los criterios de unidad mínima. (Fuente: Propia)

TAMAÑO DE CUADRÍCULA	400x400	600x600	800x800
REPRESENTATIVIDAD	9	7	6
REDUNDANCIA	9	9	9
HOMOGENEIDAD	9	9	9
PROMEDIO	9	8.33333333	8

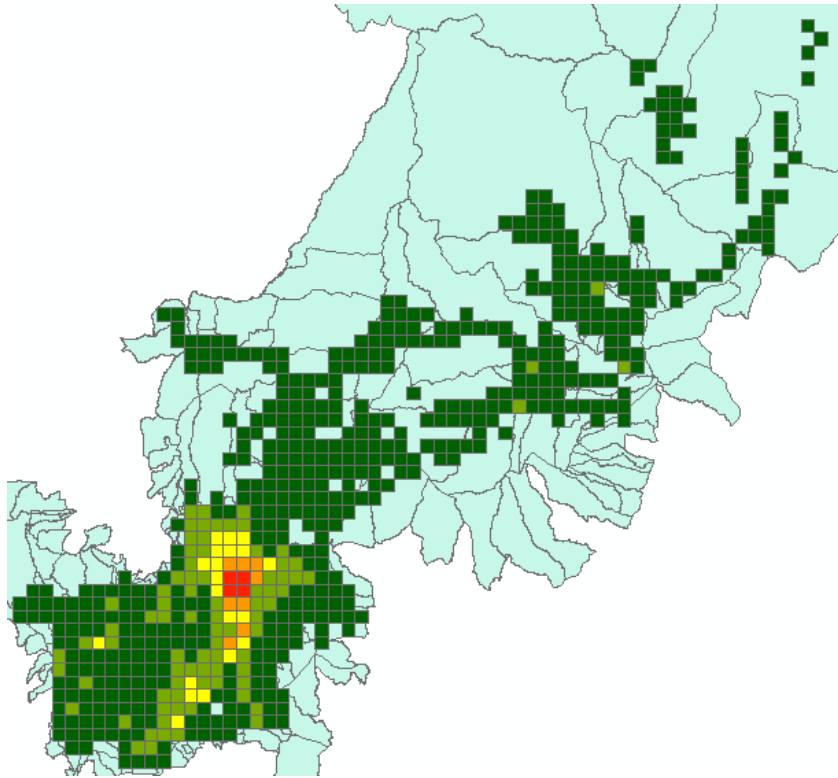


Figura 4.10. División del área de concesión de la EEA mediante cuadrículas de 800x800m.
(Fuente: Propia)

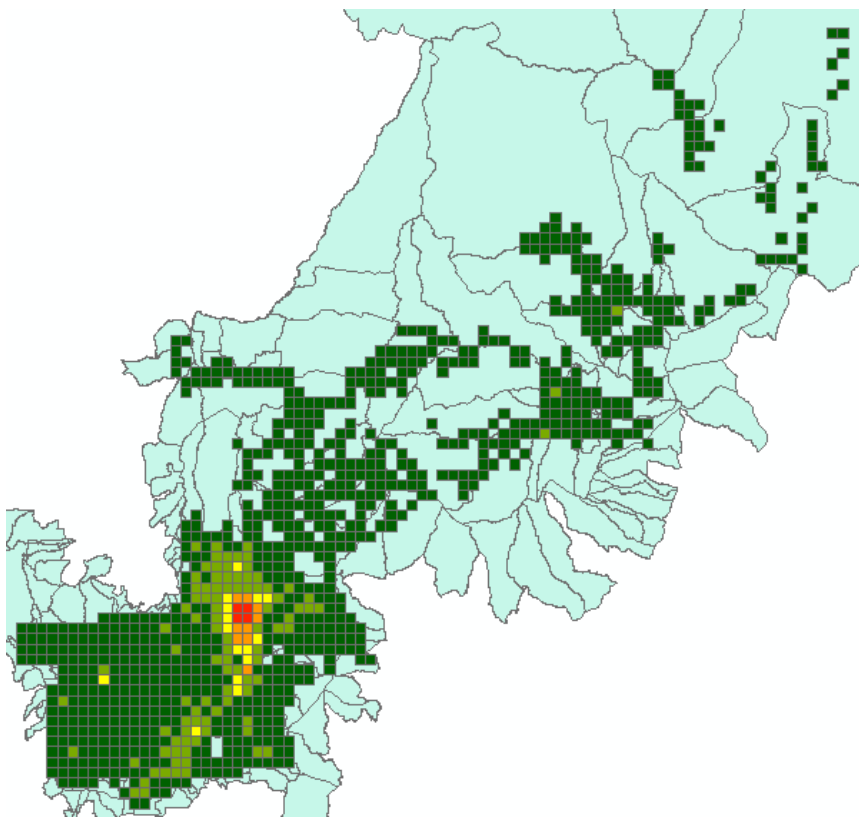


Figura 4.11. División del área de concesión de la EEA mediante cuadrículas de 600x600m.
(Fuente: Propia)

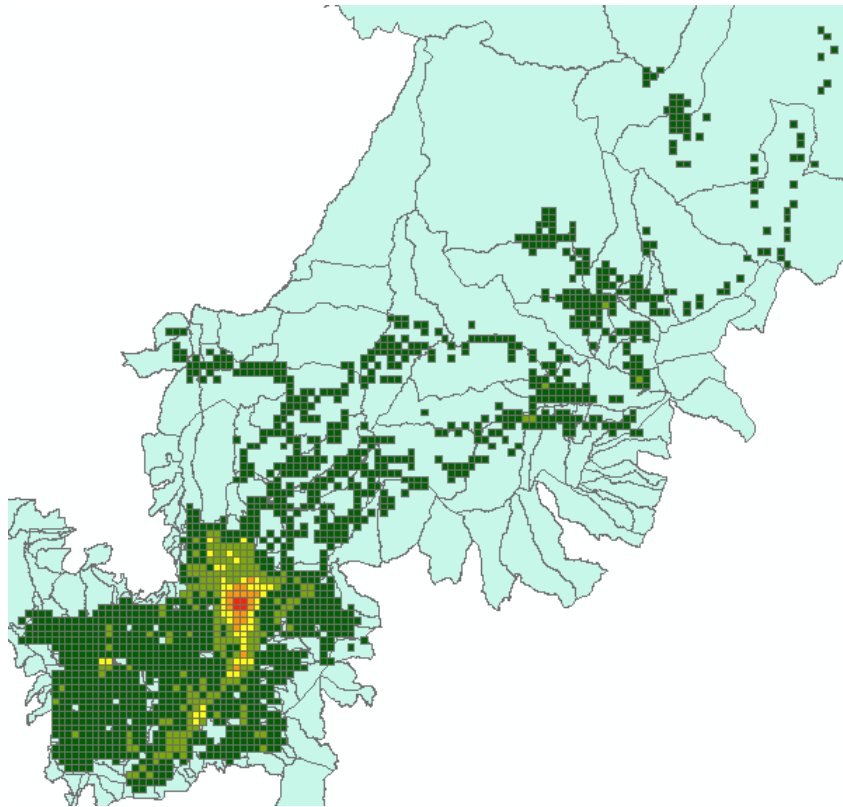


Figura 4.12. División del área de concesión de la EEA mediante cuadrículas de 400x400m.
(Fuente: Propia)

4.5. Clasificación de cuadrículas

Como último punto en la elaboración del mapa de densidades, se debe clasificar las diferentes cuadrículas según la demanda que estas presentan. Para lo cual se siguen los siguientes pasos:

1. Determinación de la demanda de cada cuadrícula.
2. Determinar la cantidad de grupos en los que se dividirán las cuadrículas.
3. Definir los intervalos de cada uno de los grupos.

4.5.1. Determinación de la demanda de cada cuadrícula.

Para determinar la demanda de cada cuadrícula se debe realizar una sumatoria de todas las demandas contenidas dentro de la celda, para ello ArcGIS ofrece una herramienta que facilita este trabajo. Los pasos a seguir para determinar la demanda de cada cuadrícula mediante el uso de las herramientas de ArcGIS se detallan en el Anexo 4.

4.5.2. Definición del número de grupos.

Para la definición del número de grupos en los que se van a dividir las cuadrículas se procede de una manera similar a lo expuesto en el capítulo 3. Así pues, se usa el método k-means para agrupar las celdas cuyas demandas sean similares en k grupos. Posteriormente se varía la cantidad de grupos (k) y se calculan los índices de cohesión y separación obtenidos en cada caso para de esta manera formar una gráfica que muestre las variaciones de los índices según el número de grupos que se tenga.

La figura 4.13 muestra una curva exponencial decreciente perteneciente a la variación de la distancia intra-cluster según la cantidad de grupos que se tenga. Se puede notar claramente que la distancia entre elementos tiende a acercarse a su mínimo si se clasifica los datos en 3 grupos o más.

Por otro lado, la figura 4.14. muestra un comportamiento logarítmico perteneciente a la variación de la distancia inter-grupal con relación al número de grupos que se disponga. En esta gráfica se verifica que la separación entre grupos es muy cercana a su valor máximo siempre y cuando se tengan 4 grupos o más.

De esta manera se concluye que el número óptimo de grupos en los que se puede clasificar a las diferentes cuadrículas debe ser de por lo menos cuatro grupos para de esta manera conseguir índices de cohesión y separación muy próximos a sus valores ideales.

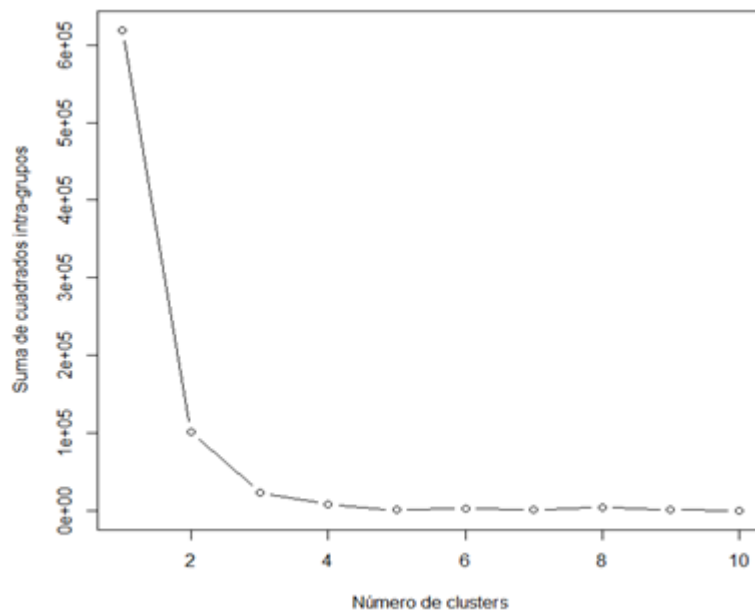


Figura 4.13. Variación de la cohesión entre elementos según el número de grupos. (Fuente: Propia)

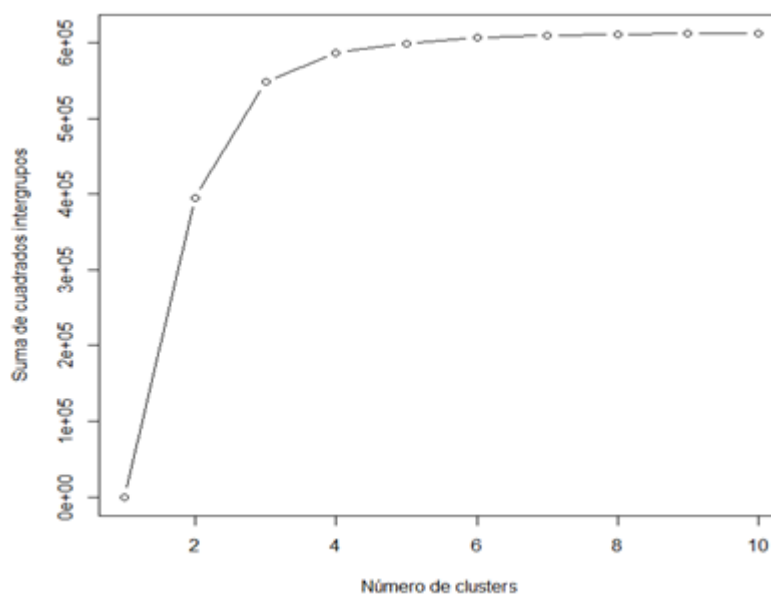


Figura 4.14. Variación de la separación entre clusters según la cantidad de grupos que se disponga. (Fuente: Propia)

4.5.3. Definición de los intervalos de los grupos.

En esta parte se hace uso del método propuesto por Dalenius & Hodges para determinar los límites (rango) que debe tener cada uno de los grupos. Los resultados se pueden ver en la tabla 4.7.

Tabla 4.7. Límites obtenidos por el método de Dalenius & Hodges.
(Fuente: Propia)

GRUPO	LÍMITES [kW]
1	[0; 5.53]
2	(5.53; 22.11]
3	(22.11; 99.47]
4	(99.47; 332.56]

4.6. Mapa de densidades.

Una vez realizado todo el proceso para obtener el número de grupos (y los límites) en los que se debe clasificar las distintas celdas, el resultado obtenido es mostrado en la figura 4.15 en la cual se puede notar claramente que las cuadrículas de 400x400m, seleccionadas en el punto 4.4 del presente capítulo, brindan una buena resolución mientras que la cantidad de grupos escogidos hacen al mapa muy homogéneo. Se debe destacar también que la mayor densidad de carga se encuentra en la zona central de la ciudad de Azogues, de esta manera mientras más se aleja del centro de esta ciudad la densidad de carga va disminuyendo, a excepción de cuando se tiene otros centro urbanos tales como Pindilig o Guapán.

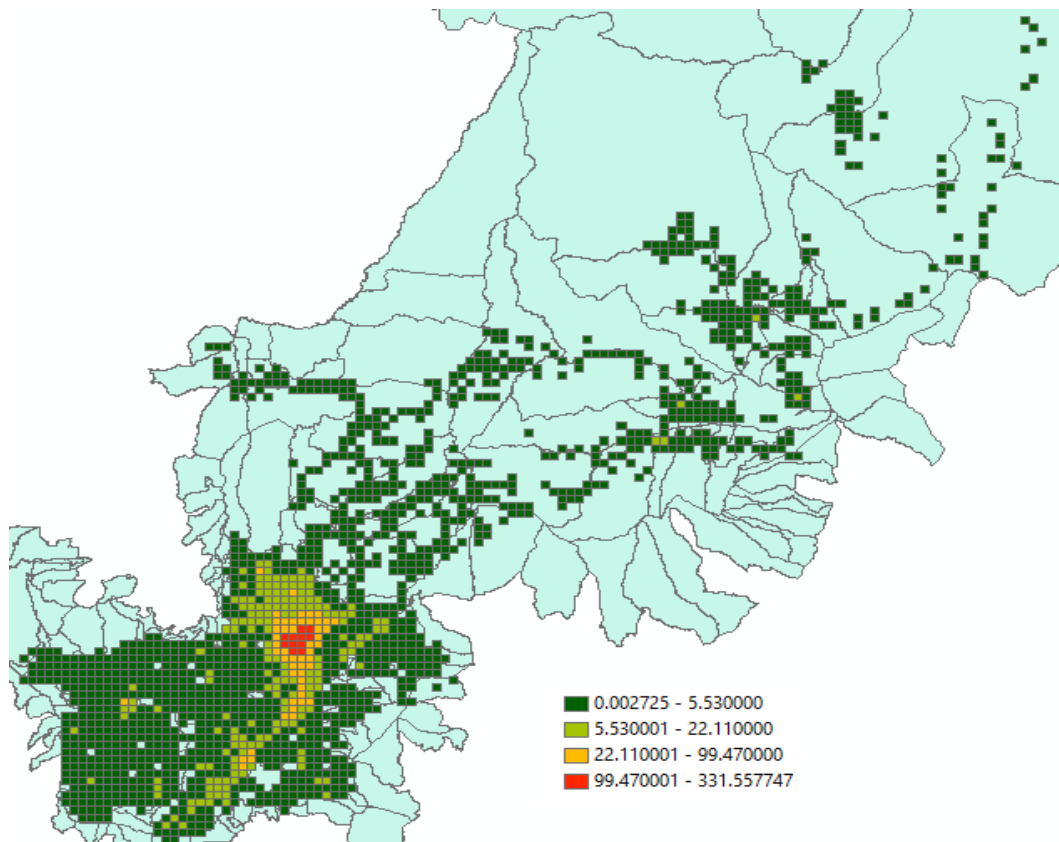


Figura 4.15. Mapa de densidades.
(Fuente: Propia)

CAPÍTULO 5: CONCLUSIONES Y RECOMENDACIONES

5.1. CONCLUSIONES

- Se ha demostrado, que la mayoría de conceptos y metodologías tales como la detección de datos atípicos, la definición de las características de los estratos (número de estratos y rango) así también como los criterios de unidad mínima de análisis han podido ser usados tanto para la conformación del mapa de estratos como para la elaboración del mapa de densidades, demostrando de esta manera la gran versatilidad de los elementos usados para este estudio
- La detección de datos atípicos, pese a ser muy necesaria para este estudio pues muchos de los procesos se basan en la media aritmética y por esta razón son susceptibles a datos atípicos, también brinda la posibilidad de examinar a aquellos usuarios con consumos fuera de lo común y en base a ello descubrir diversos problemas tales como errores de facturación o inconvenientes con el tipo de tarifa al que pertenece alguno de estos usuarios.
- La validez de los resultados obtenidos con el método planteado para este estudio pueden ser corroborados cuando son llevados al campo practico, para este caso, se ha calculado la demanda máxima diversificada para cada estrato del grupo residencial, el cual abarca el mayor porcentaje de usuarios, y se han comparado los datos obtenidos con mediciones realizadas a clientes al azar del mismo grupo tarifario obteniéndose que los datos calculados predicen con gran exactitud a la mayoría de usuarios medidos.
- Los resultados obtenidos con el método planteado en el presente documento han sido muy satisfactorios, sobre todo para el caso del mapa de densidades y el mapa de estratos de clientes residenciales. Estos casos en particular comparten la característica de poseer muchos datos para su análisis lo cual conlleva a que los mapas resultantes sean muy homogéneos y representativos.
- Dado el bajo número y la gran dispersión geográfica presente en los usuarios con tarifa industrial se ha comprobado que su estratificación no es viable pues el mapa resultante no cumple de una manera cabal con el concepto de unidad mínima de análisis, además se debe tener en cuenta que este tipo de usuarios declara su demanda y en la mayoría de casos las readecuaciones y costos de nuevos servicios son asumidos por los mismos consumidores, por esta razón en muchos casos se vuelve innecesaria la zonificación de este tipo de clientes.
- Mediante los mapas de estratos y de densidad de carga se ha podido concluir que la ciudad de Azogues es el punto central de carga para la EEA pues, como lo indican los mapas de estratos realizados para cada grupo de usuarios, el mayor consumo de energía promedio mensual se da en el centro de esta ciudad para cualquier grupo tarifario. Por otro lado, el mapa de densidad de carga indica que la demanda se encuentra concentrada en el centro de Azogues y el punto de mayor demanda está en el sector del hospital Homero Castanier Crespo y sus alrededores.
- En el capítulo 4 se puede comprobar que la metodología propuesta en este documento da resultados similares a los obtenidos por el método Jenks, el cual es el algoritmo standard de clasificación usado por el software ArcGIS.



5.2. RECOMENDACIONES

- Uno de los puntos más conflictivos al realizar este estudio fue a la hora de elaborar la base de datos inicial, dada la transición de sistema en el que se veía envuelta la compañía, por esta razón la base de datos contaba con ciertos problemas tales como errores de facturación o errores en el registro de datos. Por consiguiente, se aconseja pulir los datos en el registro y verificar los consumos, ya sea de energía o de potencia, que estén fuera de lo común mediante mediciones para veracidad de los datos en cuanto a los casos atípicos.
- Se recomienda realizar la zonificación del mercado eléctrico de la empresa por lo menos para un periodo de diez años, con lo cual se podrá analizar la evolución en la demanda de cada sector y variación en la densidad de carga de cada zona. De este modo se puede adquirir un mejor entendimiento de las características y la evolución que ha venido teniendo la carga en los últimos años y en base a ello se puede aproximar una tendencia para el comportamiento futuro de los usuarios.
- Simular el sistema con los datos obtenidos de la demanda máxima diversificada para de esta manera evaluar las características del sistema, tomando especial énfasis en las pérdidas producidas en los transformadores, para lo cual se debe tener muy en cuenta que la base de datos GIS debe estar actualizada.
- Dadas las características (consumo y número de usuarios) presentes en los clientes con tarifa residencial y comercial, se recomienda la unión de estos dos grupos tarifarios en un mismo mapa de estratos lo cual ayudará a una mejor caracterización del área de concesión de la empresa sin repercutir de una manera significativa en las características de los estratos.



REFERENCIAS

- Aggarwal, C., & Reddy, C. (2014). *Data Clustering Algorithms and Applications*. New York: CRC Press.
- Arosemena, G. (2008). Ecuador en la segunda mitad del siglo XXI. *Escuela de Postgrado y Administración de Empresas (ESPAE)*. Guayaquil - Ecuador.
- Brys, G., Hubert, M., & Struyf, A. (2003). A robust measure of skewness. 3-6.
- Cochran, W. G. (1980). *Técnicas de Muestreo*. John Wiley & Sons.
- Conelec. (2013). Plan Maestro de Electrificación 2013 - 2022. Quito, Ecuador.
- Cowan, G. (1998). *Statistical data analysis*. New York: Oxford University Press.
- Dalenius, T., & Hodges, J. L. (1959). Minimum Variance Stratification. *American Statistical Association*.
- Flores Taípe, F. (2011). El sector eléctrico ecuatoriano en los últimos 20 años: estrategias para alcanzar la seguridad energética.
- Forgy, E. (1965). Cluster analysis of multivariate data: efficiency vs. interpretability of classification. *Biometrics*.
- Gomez Sagner, D. I. (2007). Análisis de estacionalidad de la curva de demanda para clientes BT1. *Universidad de Chile*.
- Gönen, T. (2014). *Electric power distribution engineering* (3 ed.). CRC Press.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis*. Pearson.
- Hans-Hermann, B. (2008). Origins and extensions of the k-means algorithm in cluster analysis. *Journ@l Electronique d'Histoire des Probabilités et de la Statistique Vol. 4*.
- Hatigan, J., & Wong, M. (1979). Algorithm AS136 A K-means Clustering Algorithm. *Applied Statistics*.
- Hidiroglou, M., & Lavallée, P. (1988). On the stratification of skewed populations. *Survey Methodology*.
- Hubert, M., & Vandervieren, E. (2007). An adjusted boxplot for skewed distributions. *ScienceDirect*, 5186-5201.
- Jain, A., Diun, R., & Mao, J. (2000). Statistical pattern recognition: A review. *IEEE Trans. Pattern Anal. Machine Intelligence*.
- Lee Willis, H. (n.d.). *Spatial Electric Load Forecasting*. New York: Marcel Dekker.
- Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information*.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium On Mathematical Statistics and Probabilities*.
- Moncada, J. (1974). La evolución de la planificación en el Ecuador. *Nueva Sociedad No13*.
- Morissette, L., & Chartier, S. (2013). The k-means clustering technique: General considerations and implementation in Mathematica. *Tutorials in Quantitative Methods for Psychology*, 15-24.



Pérez Miñaca, C. A., & Salazar Yépez, G. (2014). Zonificación del Mercado de Consumo Eléctrico. *IEEE*.

Pilamunga Ubidia, J. M. (2009). Análisis del factor de carga y el factor de diversidad para el diseño eléctrico de centros comerciales populares del municipio de Quito. .

Sebnem, E. (2011). Computational methods for optimum stratification: A review. *World Statistical Congress*, 3305-3306.

Sethi, V. (1963). A note on optimum stratification of populations for estimating the population means. *Australian Journal of Statistics*.

Vest, S. J. (1957). Estimating kW demanda for future loads on rural distribution systems. *AIEE*.

Walpole, R. E., Myers, R. H., & Myers, S. L. (2012). *Probabilidad y estadística para ingeniería y ciencias* (9 ed.). Mexico: Pearson.

Westinghouse Electric Corporation. (1959). *Electric utility engineering reference book: distribution system* (Vol. 3). Pennsylvania.

Zimmer, S., Kim, J.-k., & Nusser, S. (2012). A Hierarchical Clustering Algorithm for Multivariate Stratification in Stratified Sampling. *Survey Research Methods*.



ANEXOS

ANEXO 1: IMPORTACIÓN DE BASE DE DATOS A ArcGIS

Uno de los primeros pasos a realizarse en el software ArcGIS, para el proceso de zonificación, es la importación de la base de datos depurada, la misma que necesariamente debe tener una extensión .xls. Los pasos necesarios que se deben realizar son los siguientes:

1. Abrir ArcCatalog. Para ello es necesario ir a la pestaña Windows > Catalog lo cual hará que la ventana Catalog se abra en la parte derecha del espacio de trabajo.
2. Vincular el software a la dirección del proyecto. Para ello, en la ventana Catalog, se debe identificar el icono de *Connect To Folder* (Figura A.1) en el cual se dará clic.

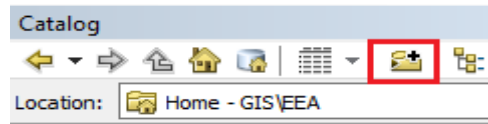


Figura 6.1. Icono "Connect To Folder".
(Fuente: Propia)

3. Se abrirá la ventana mostrada en la Figura A.2 en la cual se debe buscar la dirección del proyecto que se desea vincular. Una vez se ha vinculado la carpeta del proyecto, esta aparecerá en la ventana Cataloga.

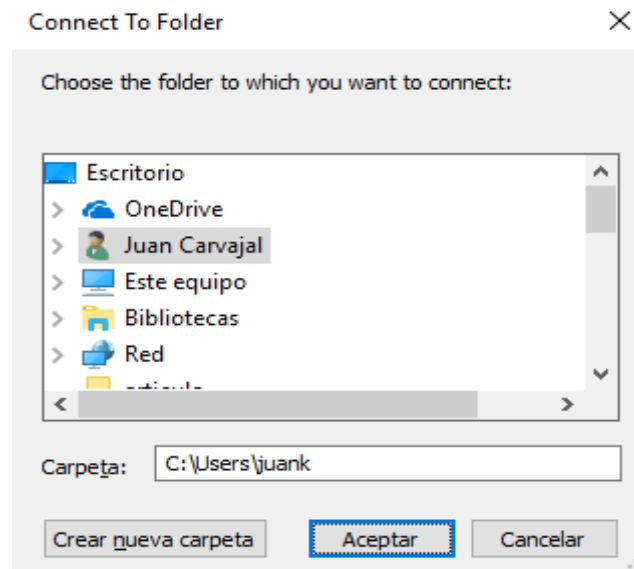


Figure 6.2. Ventana "Connect To Folder".
(Fuente: Propia)

4. Buscar el archivo que contiene la base de datos. Una vez encontrado, se debe dar clic en el recuadro identificado con "+" que está al lado del nombre del archivo, esto con la finalidad de que se abran las subsecciones del documento (Figura A.3).

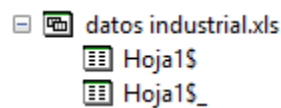


Figura 6.3. Archivo de donde se extrae la base de datos.
(Fuente: Propia)

5. En la primera hoja se debe dar clic derecho para que aparezca el menú presentado en la Figura A.4. Aquí se debe escoger la opción Create Feature Class > From XY Table.

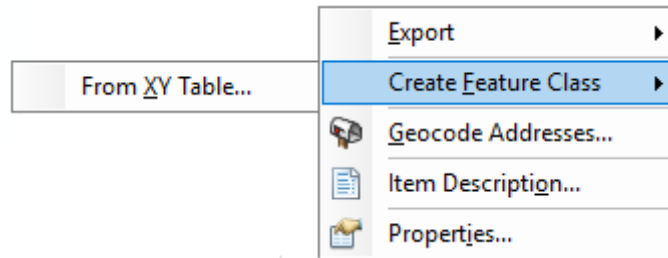


Figura 6.4. Menú para la creación de una tabla con la información de la base de datos.
(Fuente: Propia)

- En la ventana mostrada por la figura A.5 se debe definir el lugar en donde se guardará la tabla a crearse y el sistema de coordenadas en las que están basados los datos de los clientes.

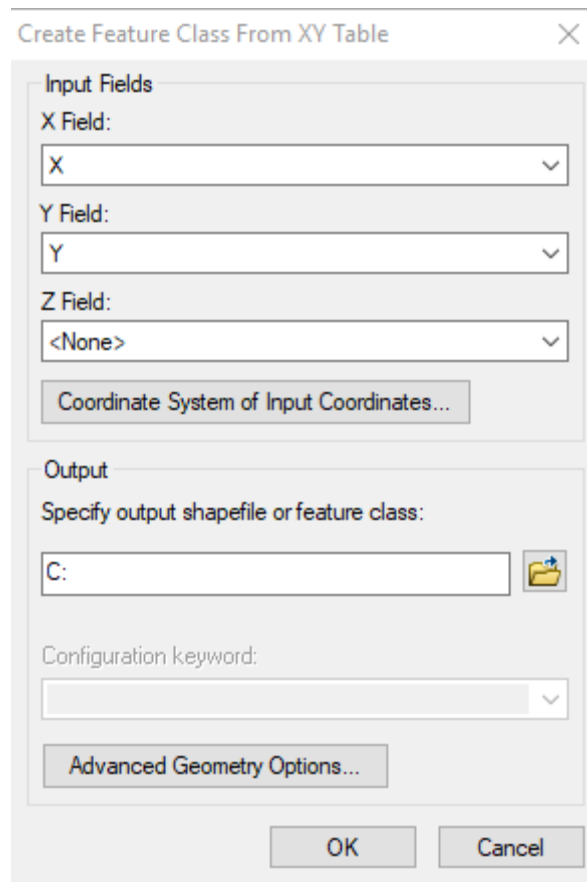


Figura 6.5. Ventana Create Feature Class From XY Table.
(Fuente: Propia)

- Para la definición del sistema de coordenadas se debe oprimir el botón “Coordinate System of Input Coordinates...” (Figura A.5) con lo que se abrirá la ventana presentada en la figura A.6 y en donde se debe escoger el sistema de coordenadas con que se trabajará. En este caso se tomó el sistema WGS 1984 UTM Zone 17S.

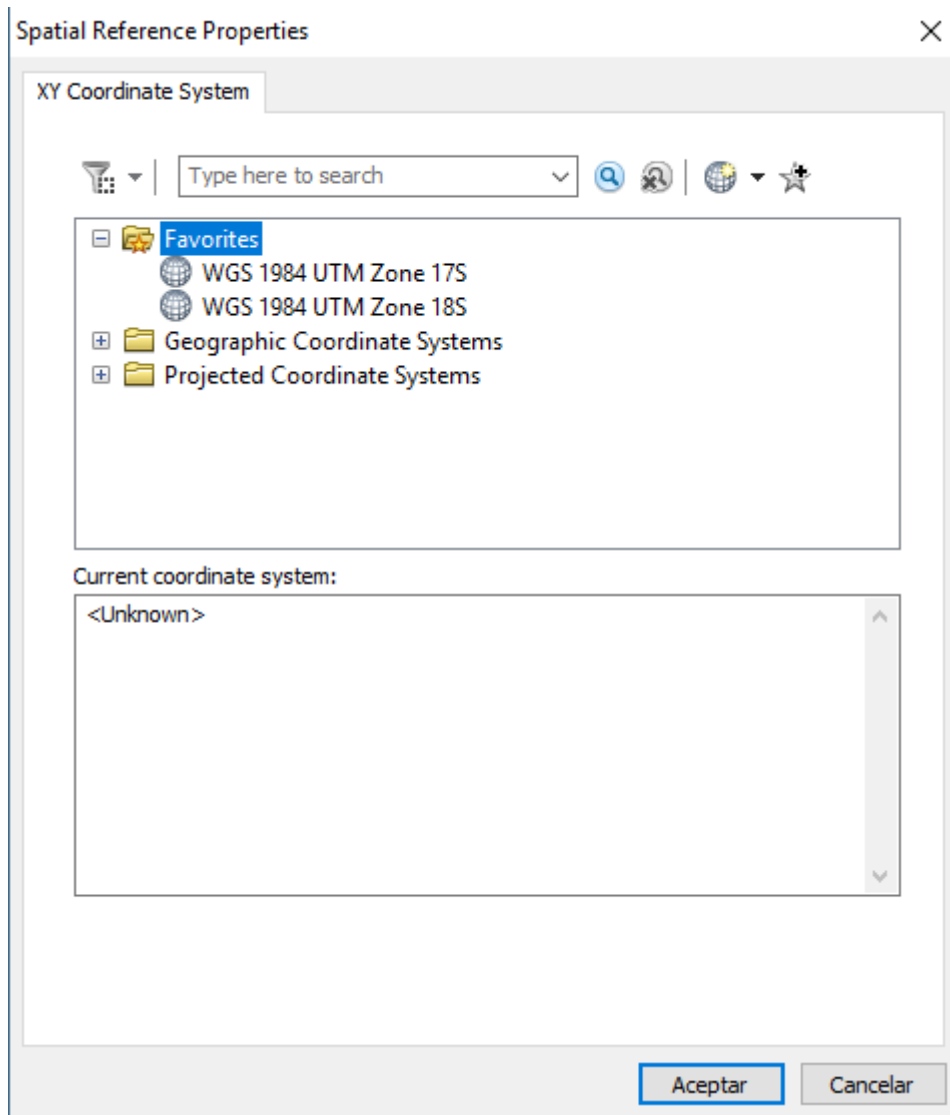


Figura 6.6. Ventana para la elección del sistema de coordenadas en las que estará fundamentada la base de datos.
(Fuente: Propia)

8. Una vez se han definido los parámetros necesarios se procede a la creación de la tabla de datos, la misma que estará guardada en la carpeta que se definió previamente, por lo cual es necesario dirigirse a esta carpeta y arrastrar el archivo creado hacia la parte de *Table of Contents* quedando así una capa similar a la mostrada en la figura A.7.

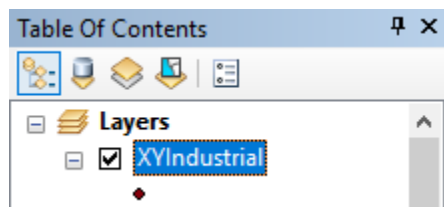


Figura 6.7. Capa de puntos generada a partir de la base de datos.
(Fuente: Propia)

ANEXO 2: CLASIFICACIÓN DE ELEMENTOS

Muchas veces se desea agrupar a ciertos elemento según un rango de valores. ArcGIS da la opción de clasificar un grupo de objetos por colores según los valores de cada objeto. Así pues, por ejemplo, se puede agrupar los elementos de la figura A.8a a fin de obtener un resultado como el mostrado en la figura A.8b. Para ello se siguen los siguientes pasos:

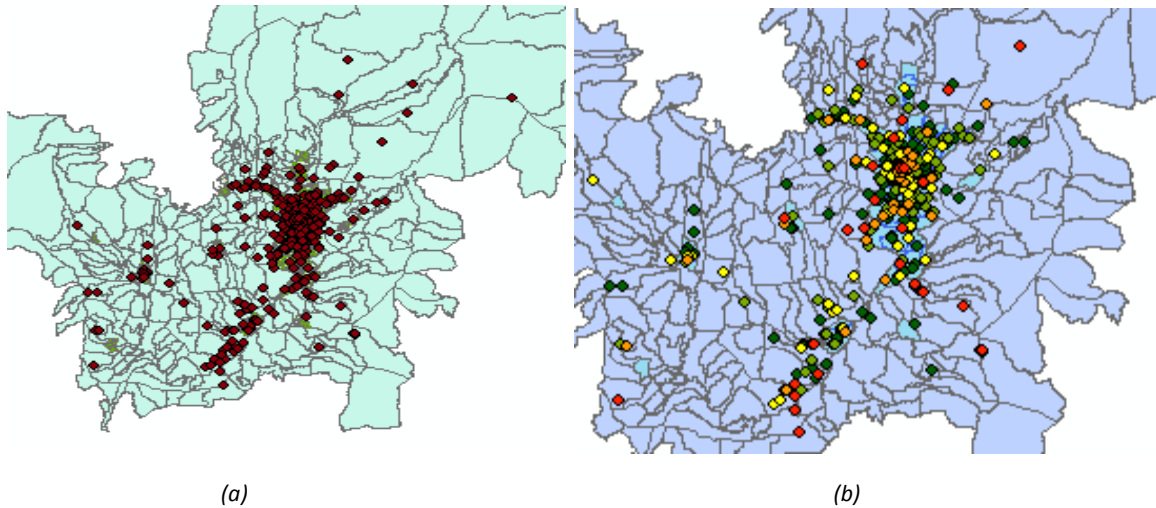


Figura 6.8 Ubicación de medidores para el sector Industrial de la EEA (a) sin agrupación (b) con agrupación. (Fuente: Propia)

1. Dar doble clic en la capa que se desea tratar. Se abrirá una ventana llamada *Layer Properties*.
2. Dirigirse a *Symbology > Quantities > Graduated Colors*. Aquí se podrá escoger el parámetro en el que se basará la clasificación, tal como lo indica la figura A.9.

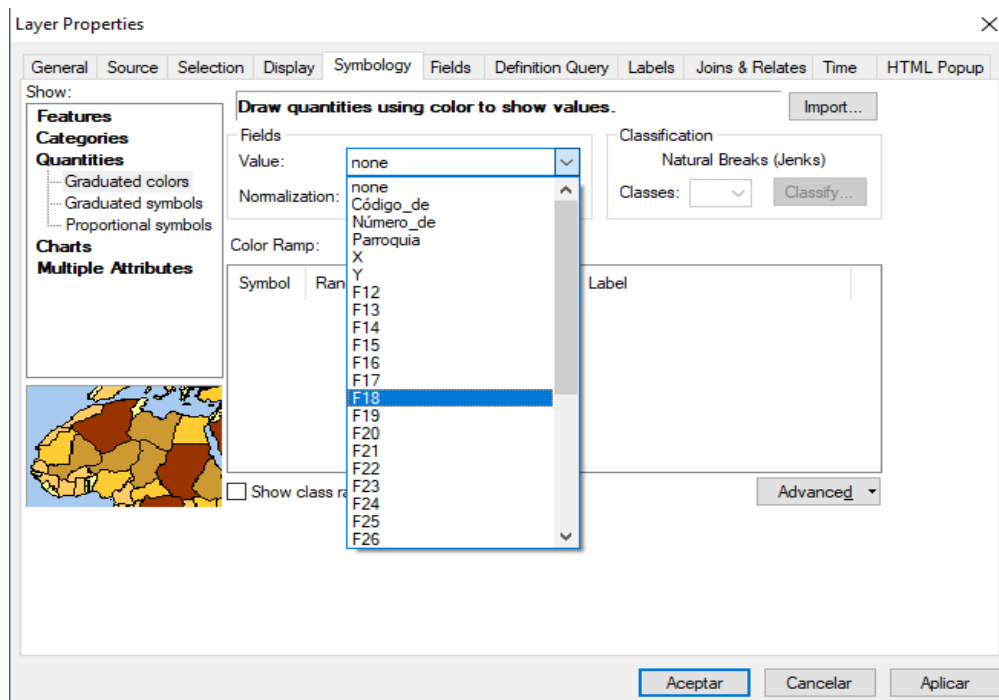


Figura 6.9. Propiedades de capa. (Fuente: Propia)

3. Seleccionar el número de clases en las que se agruparán los datos y la gama de colores que estos tendrán, tal como se muestra en la figura A.10.

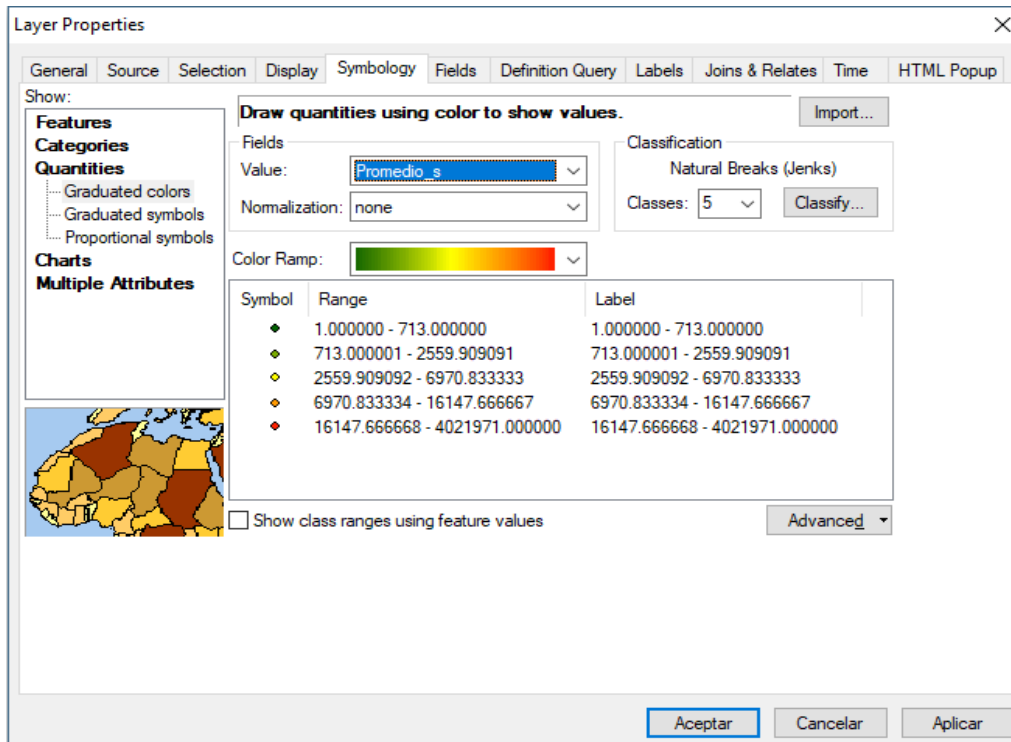


Figura 6.10. Número de clases y colores que se tendrá en la clasificación. (Fuente: Propia)

2. Finalmente, si se desea, se puede optar para que el software genere el número de agrupaciones y sus límites adecuados de cada uno de ellos. Para ello se debe dar clic en el botón *Classify*, mostrado en la figura A.10, a fin de ingresar a la ventana *Classification* (figura A.11) en la cual se puede seleccionar el método que se desea usar para la clasificación y la cantidad de grupos que se requiere (si se escoge el método manual).

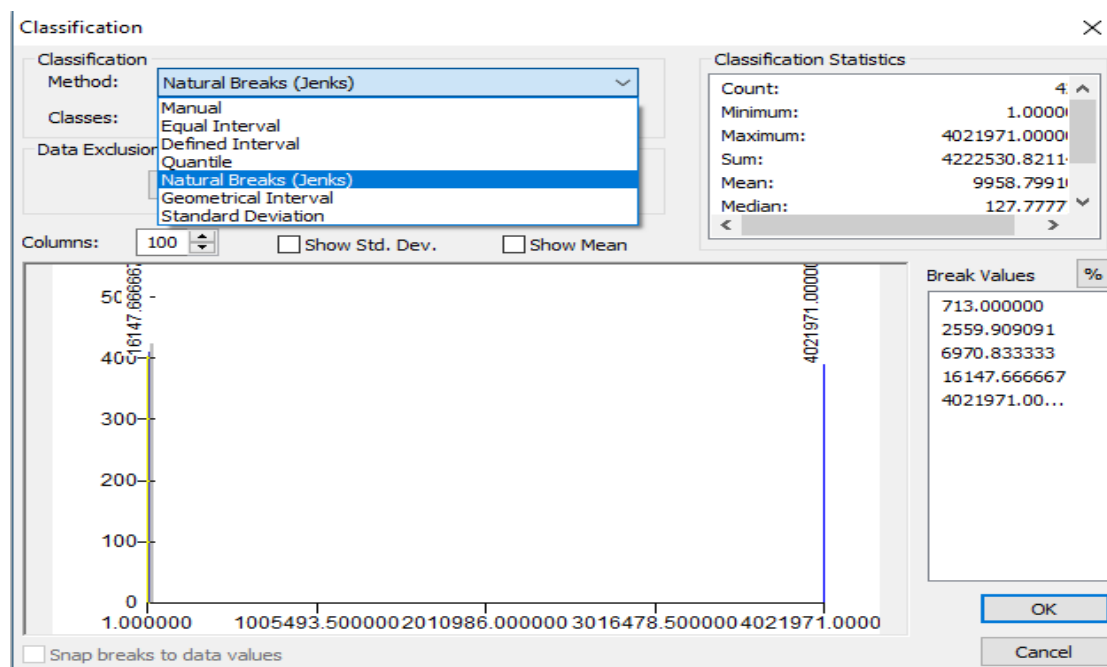


Figura 6.11. Parámetros de Clasificación. (Fuente: Propia)

ANEXO 3: CUADRICULACIÓN DE UNA ZONA

Para generar una cuadrícula de dimensión “n” en una zona específica se siguen los siguientes pasos:

1. En la ventana Catalog, dirigirse a *System Toolboxes > Cartography Tools > Data Driven Pages > Grid Index Features* (Figura A.12)

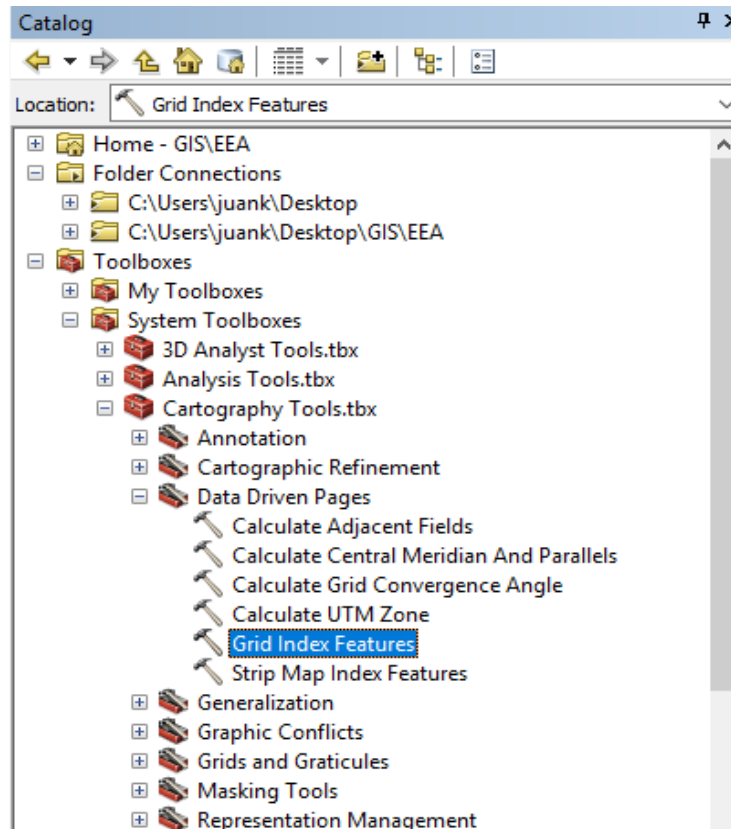


Figura 6.12. Herramienta para la creación de una cuadrícula en una zona determinada. (Fuente: Propia)

2. Aparecerá una ventana similar a la mostrada en la figura A.13, en la que se debe llenar los siguientes campos:
 - a. *Output feature class*: designa el lugar en donde se desea guardar la capa de que se generará.
 - b. *Input feature*: en este apartado se debe escoger la capa que servira de base para la cuadrículacion. En la figura A.13 se ha tomado la capa de “ZONAS URBANAS” como base.
 - c. *Map Scale*: si se desea se puede escoger una escala en la capa resultante.
 - d. *Polygon Width*: Aquí se establece el ancho que tendra cada cuadrícula.
 - e. *Polygon Height*: En este apartado se establece la altura de la cuadrícula.
 - f. *Polygon Grid Origin Coordinate*: En este apartado se establecen las coordenadas del punto de inicio para la cuadrícula. Este campo viene establecido por defecto en concordancia con las coordenadas de la capa base seleccionada.
3. Tras aceptar todos los datos introducidos en el paso 2. Se generará una cuadrícula como la mostrada en la figura A.14.

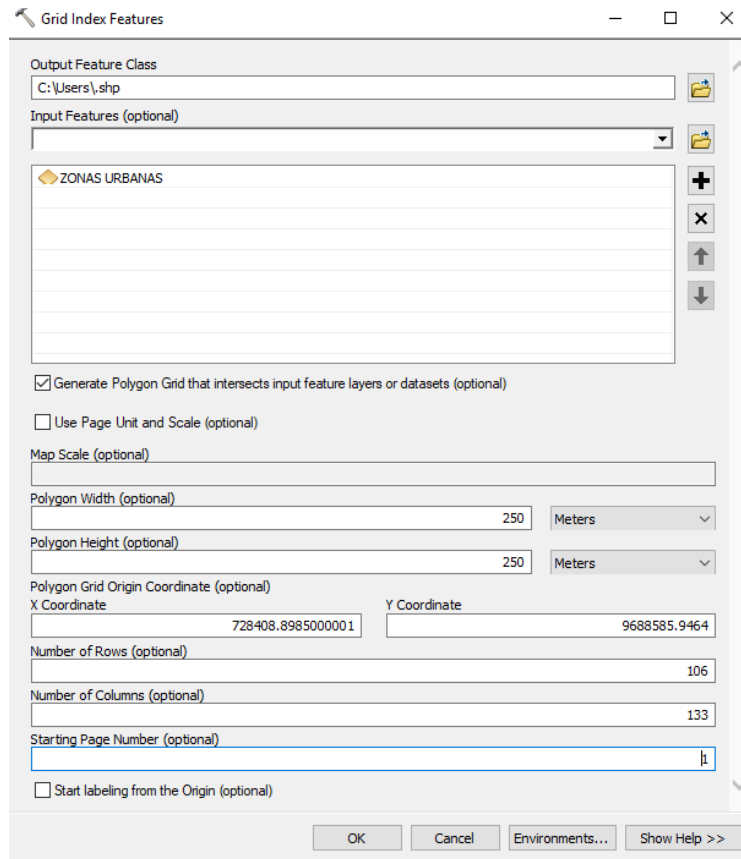


Figura 6.13. Ventana de creación de cuadrícula. (Fuente: Propia)

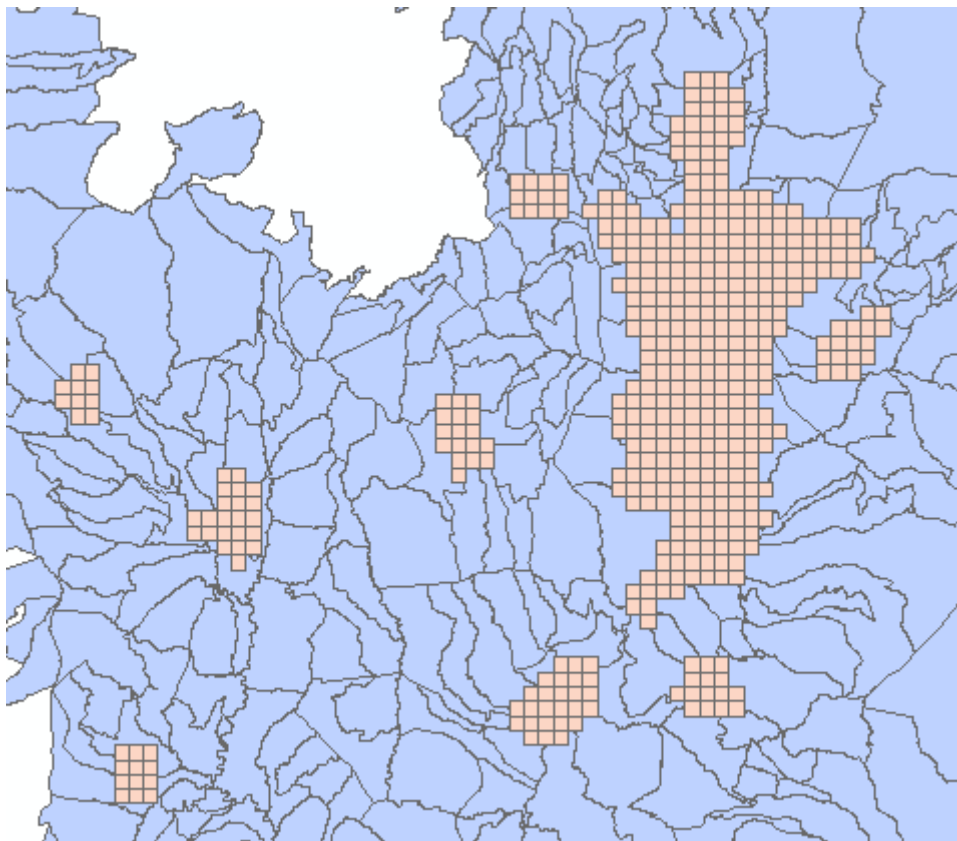


Figura 6.14. Cuadrícula generada para la zona urbana de la EEA. (Fuente: Propia)

ANEXO 4: CARACTERIZACIÓN DE CUADRÍCULAS

Para la caracterización de las cuadrículas es necesario examinar los datos que están dentro de cada una de ellas en base a un parámetro específico, para ello ArcGIS ofrece la herramienta *Spatial Join* cuya función es unir las características de dos capas en base a su posición espacial y, además, permite efectuar ciertas operaciones matemáticas y estadísticas con los datos obtenidos. Así pues, para la caracterización de las cuadrículas se siguen los siguientes pasos.

1. Dirigirse a *Catalog > Toolboxes > System Toolboxes > Analysis Tool > Overlay > Spatial Join* (Figura A.15)

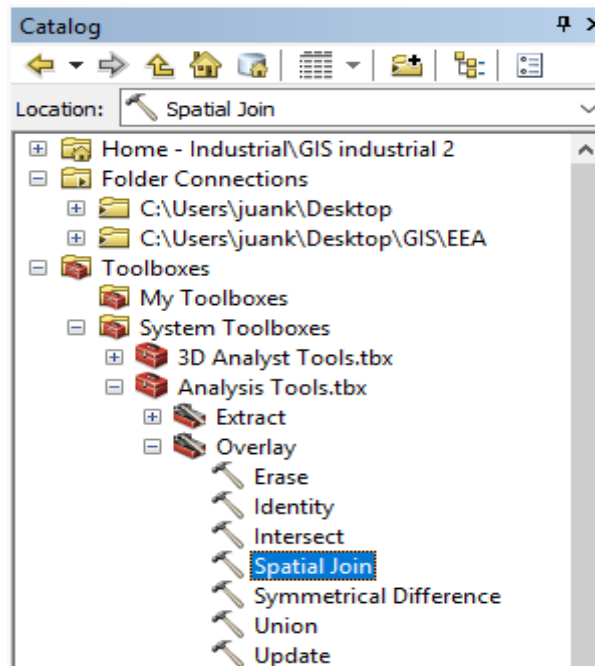


Figura 6.15. Dirección de la herramienta *Spatial Join*. (Fuente: Propia)

2. Aparecerá una ventana similar a la mostrada en la figura A.16, en ella se deben especificar los siguientes parámetros:
 - a. *Target Features*: Se define cual es la capa base, a la cual se quiere integrar nuevas características, para el caso de la zonificación, en este apartado se coloca la capa de cuadrículas.
 - b. *Join Features*: En este apartado se especifica la capa cuyas características serán unidas a integradas a otra. Para el caso de la zonificación, en este apartado se debe ubicar la capa de puntos con las coordenadas de cada usuario.
 - c. *Output Feature Class*: Se define el lugar en donde se almacenará la capa que se va a crear.
 - d. *Join Operation*: Para el caso de la zonificación debe estar configurada como JOIN_ONE_TO_ONE.
 - e. *Match Operation*: Debe estar configurada en INTERSECT.

3. Especificar los atributos de los datos resultantes. Como se menciono anteriormente, esta herramienta permite realizar operaciones matemáticas y estadísticas en los datos, para ello se debe:

- 6.1.1. Identificar la variable deseada en el campo *Field Map of JoinFeatures*.
 - 6.1.2. Abrir el menú de opciones y seleccionar *Merge Rule*.
 - 6.1.3. Para el mapa de estratos se selecciona la opción *Median* o *Mean*, mientras que para el mapa de densidades se escoje la opción *Sum* (Figura A.17).
4. Caracterizar cada cuadrícula. La resultante de esta operación será una capa cuadriculada que contiene entre sus propiedades los atributos seleccionados en el paso 3 pero que no son diferenciables a simple vista por lo que se procede a caracterizar cada cuadrícula tal como se realizó en el Anexo 1 dando como resultado la figura A.18.
 5. Eliminar las celdas con datos nulos. Para ello se da doble clic sobre la capa creada con la herramienta *Spatial Join* y en la pestaña *Definition Query* se selecciona el botón *Query Builder*, se mostrará una ventana similar a la presentada en la figura A.19 en la cual se debe definir una condición para eliminar celdas, para el caso de la zonificación la condición dada es que la demanda debe ser mayor a 0, por lo cual las celdas cuya demanda sea cero (no hay consumidores) se eliminarán dando como resultado la figura A.20.

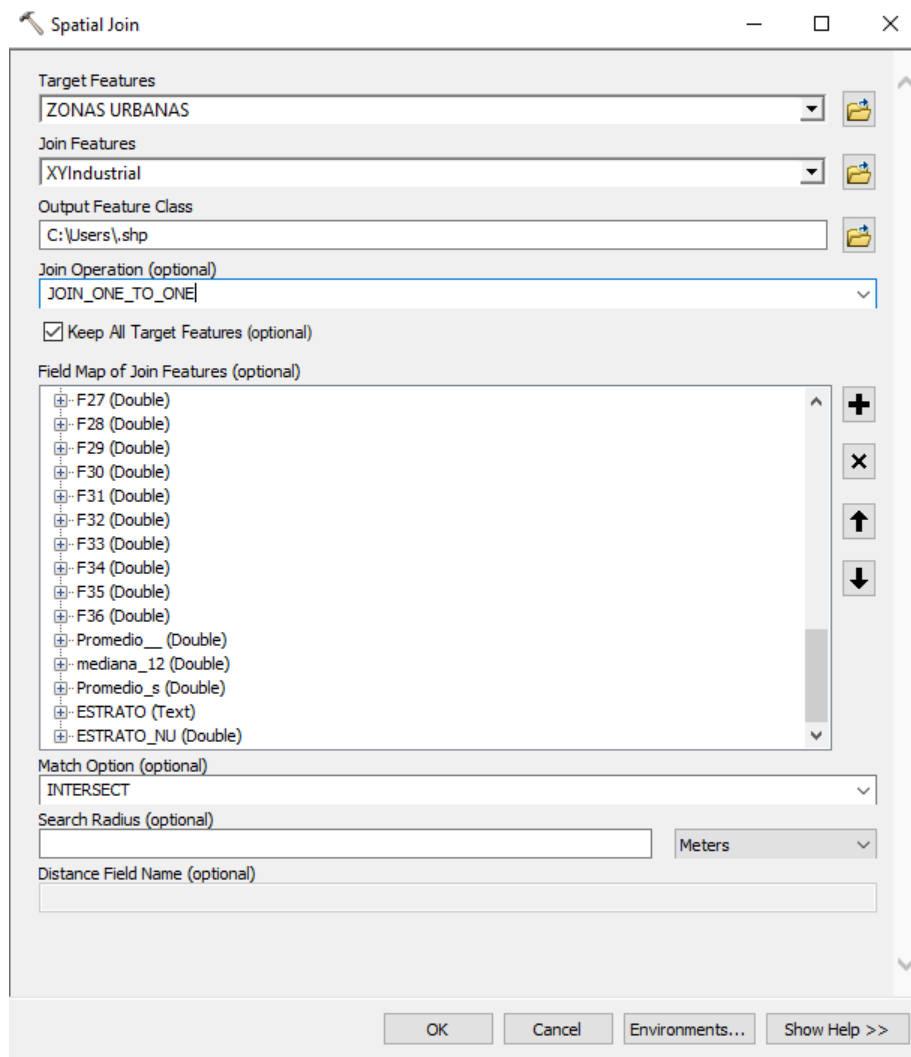


Figura 6.16. Ventana de parámetros de la herramienta *Spatial Join*. (Fuente: Propia)

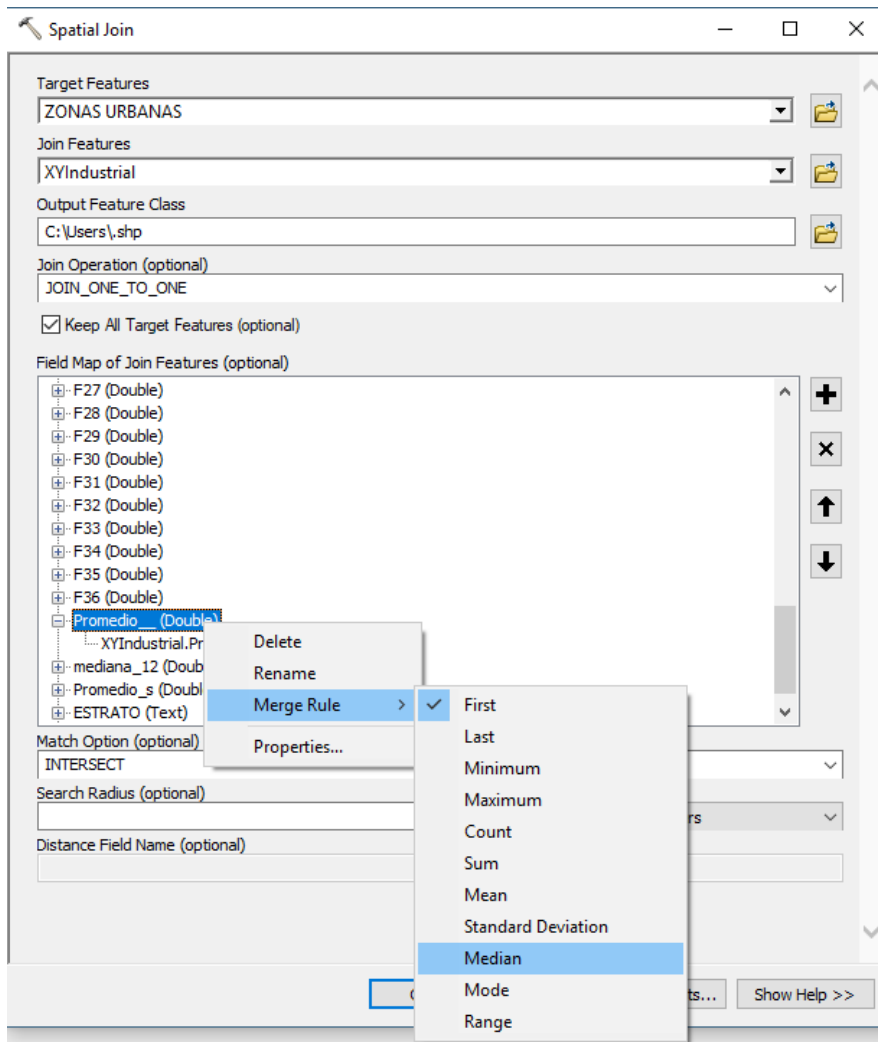


Figura 6.17. Elección de atributos para el Spatial Join. (Fuente: Propia)

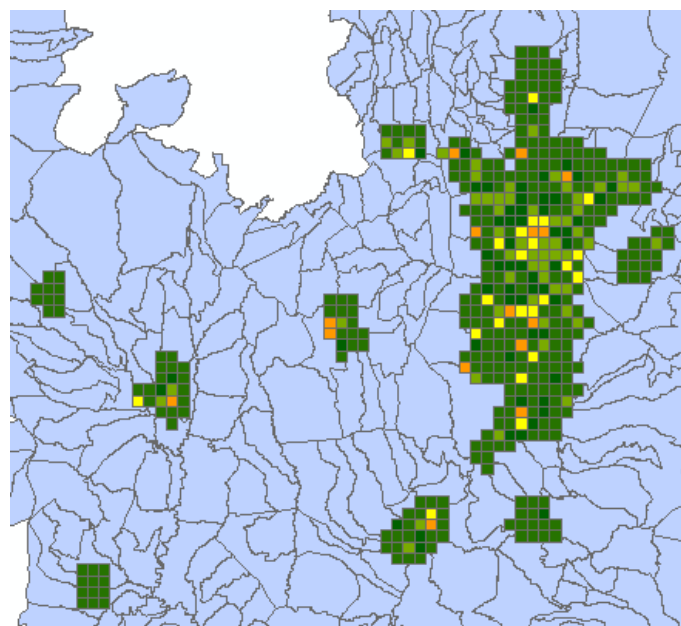


Figura 6.18. Resultado del Spatial Join. (Fuente: Propia)

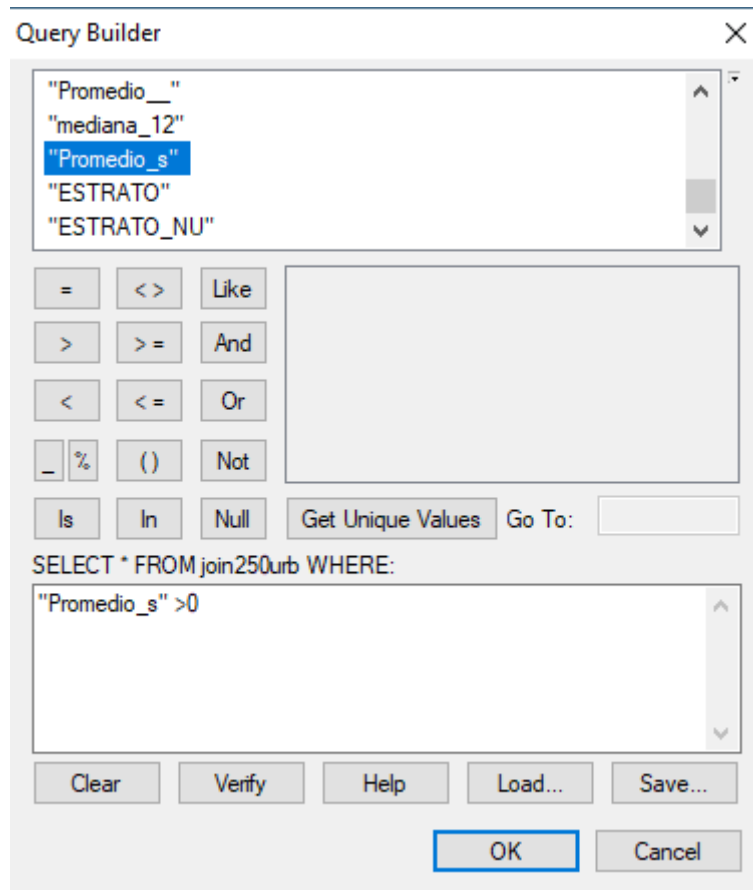


Figura 6.19. condición para eliminar celdas innecesarias. (Fuente: Propia)

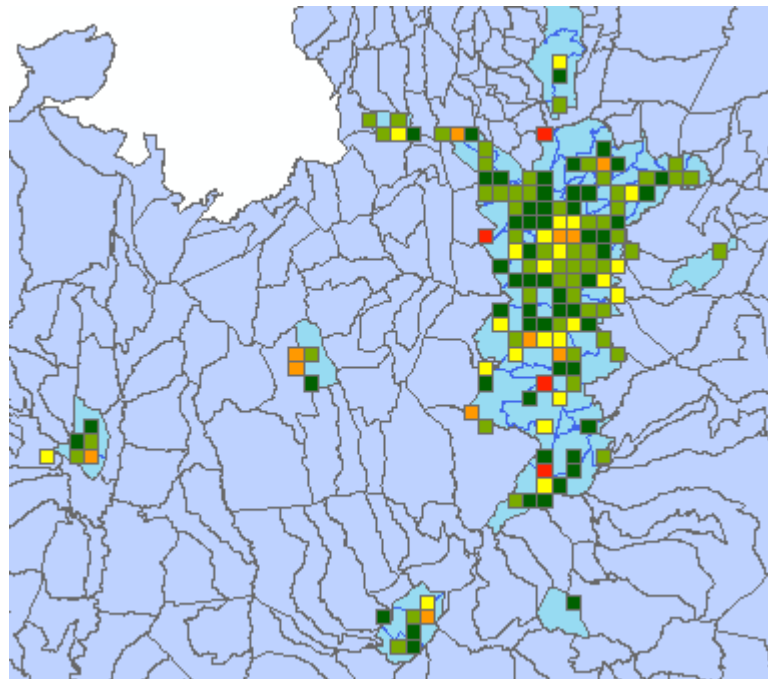


Figura 6.20. Resultado del Spatial Join tomando en cuenta el condicional impuesto en la ventana Query Builder. (Fuente: Propia)

ANEXO 5: PROGRAMA PARA CALCULAR LAS CARÁCTERÍSTICAS DE LOS ESTRATOS Y LA DEMANDA MÁXIMA UNITARIA

El siguiente algoritmo calcula de una manera rápida la cantidad de estratos ideal en los que dividir a los usuarios además de los límites de cada estrato, también, examina los outliers de la base de datos y calcula la demanda máxima unitaria de los usuarios residenciales. Este algoritmo fue desarrollado en el software RStudio Versión 1.1.423.

```
#-----
# FILTRADO DE DATOS
#-----
library(robustbase)

datgen = datosResidencial          ## Documento desde donde se extraen todos los datos

columint= datosResidencial$PROMEDIO  ## Variable de interés

hist(columint,breaks = 100,freq=T,xlim = c(0,1800),ylim = c(0,5000))

medcoup=mc(columint)

q1 = quantile(columint,0.25)
q3 = quantile(columint,0.75)
interq = IQR(columint)

if (medcoup>=0){
  lext1 = q3+1.5*exp(3*medcoup)*interq
  lext2 = q1-1.5*exp(-4*medcoup)*interq
}
else{
  lext1 = q3+1.5*exp(4*medcoup)*interq
  lext2 = q1-1.5*exp(-3*medcoup)*interq
}

}

hist(datfilt$PROMEDIO,breaks = 100,freq=T,xlim = c(0,400),ylim = c(0,1500))

#-----
# DEFINICION DE NUMERO DE ESTRATOS (k-MEANS)
#-----

datestrat=datfilt$PROMEDIO  ##columna desde la que se extraen los datos

## Determinar numero de clusters óptimos

sumbt = kmeans(datestrat, centers=1)$betweenss
for (i in 2:10) {sumbt[i] = kmeans(datestrat, centers=i)$betweenss}
plot(1:10, sumbt, type="b", xlab="Numero de clusters", ylab="Suma de cuadrados intergrupos")

sumbt = kmeans(datestrat, centers=1)$withinss
for (i in 2:10) {sumbt[i] = kmeans(datestrat, centers=i)$withinss}
plot(1:10, sumbt, type="b", xlab="Numero de clusters", ylab="Suma de cuadrados intra-grupos")
```

```

#-----
# DETERMINAR RANGO DE ESTRATOS (DALENIUS-HODGES)
#-----

library(stratification)
x = datfilt$PROMEDIO
estratos= strata.cumrootf(x, CV=0.5, Ls=4)
estratos
#-----
# DEMANDA MAXIMA DIVERSIFICADA (cocinas de inducción)
#-----

limestrat=c(estratos$bh, max(datfilt$PROMEDIO))
numfil=10
tablaDMD=matrix(0,numfil+1,length(limestrat))
DMUusu=c()
for (i in 1:length(limestrat)) {
  usuar=5                                ## Número de usuarios hasta donde se quiere hacer la tabla
  consum = limestrat[i]
  N=usuar
  DMUci=3.2                              ## Demanda máxima unitaria de cada cocina
  a=9.49                                 ## Factor a método REA
  b=0.005925*(consum^0.885)             ## Factor b método REA
  FCci=0.465*(N^-0.128443)+0.037
  DMCci=0.6*N*FCci+DMUci
  Dcoinc=(a*b)+DMCci
  DMUusu=c(DMUusu,(Dcoinc)/(0.667*5))
}

for (i in 1:nrow(tablaDMD)) {

  a=(i-1)*(1-0.4*(i-1)+0.4*((i-1)^2+40)^0.5);
  fcoinc=a/(3.29*(i-1))

  for (j in 1:length(limestrat)) {
    if(i==1){
      tablaDMD[i,j]=limestrat[j]
    }
    if(i==2){
      tablaDMD[i,j]=DMUusu[j]
    }
    if(i>2){
      tablaDMD[i,j]= (i-1)*fcoinc*DMUusu[j]
    }
  }

}

}
tablaDMD

```

```

#-----
# DEMANDA MAXIMA DIVERSIFICADA (sin cocinas de inducción)
#-----

limestrat=c(estratos$bh, max(datfilt$PROMEDIO))
#limestrat=c(60,110,180,310,500)
numfil=10
tablaDMD=matrix(0,numfil+1,length(limestrat))
DMUusu=c()
for (i in 1:length(limestrat)) {
  usuar=5
  consum = limestrat[i]

  a=9.49
  b=0.005925*(consum^0.885)

  Dcoinc=(a*b)
  DMUusu=c(DMUusu,(Dcoinc)/(0.667*5))
}

for (i in 1:nrow(tablaDMD)) {

  a=(i-1)*(1-0.4*(i-1)+0.4*((i-1)^2+40)^0.5);
  fcoinc=a/(3.29*(i-1))

  for (j in 1:length(limestrat)) {
    if(i==1){
      tablaDMD[i,j]=limestrat[j]
    }
    if(i==2){
      tablaDMD[i,j]=DMUusu[j]
    }
    if(i>2){
      tablaDMD[i,j]= (i-1)*fcoinc*DMUusu[j]
    }
  }

}

}
tablaDMD

```

ANEXO 6: ELABORACIÓN DE DENDOGRAMAS

Para la elaboración de dendogramas se ha usado el programa IBM SPSS Statistics versión 25. Los pasos que se siguen para la elaboración de dendogramas son los siguientes:

- 6.1.3.1. Importar los datos hacia el programa para lo cual se debe recurrir a la pestaña Archivo > Importar Datos.
- 6.1.3.2. Configurar las variables, para lo cual se debe recurrir a la pestaña “Vista de variables” y configurar los campos de cada variable usada tal como se muestra en la figura A.21.

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	promedio	N Numérico	13	2		Ninguno	Ninguno	13	Derecha	Nominal	Entrada

Figura 6.21. Ejemplo configuración de variables en SPSS. (Fuente: Propia)

- 6.1.3.3. Clasificar los datos , para lo cual se recurre a la pestaña Analizar > Clasificar > Clúster Jerárquico. (Figura A.22)

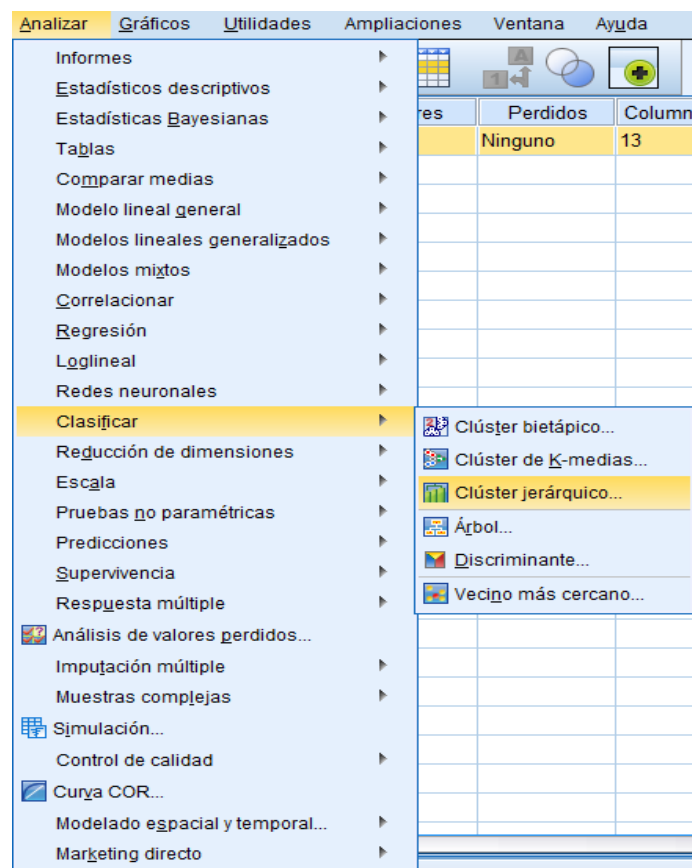


Figura 6.22. Proceso seguido para clasificar los datos. (Fuente: Propia)

- 6.1.3.4. Una vez que se ha accedido a la opción “ Clúster Jerárquico” aparecerá una ventana similar a la presentada en la figura A.23, en donde se deben configurar los siguientes campos:

- 6.1.3.4.1. **Gráficos:** En este apartado se debe seleccionar la opción “Dendograma” y configurar el sentido del gráfico tal como se muestra en la figura A.23 y A.24.
- 6.1.3.4.2. **Método:** En este punto se selecciona el método por el cual se clasificarán los datos y la medida que se usará para evaluarlos. Para este caso se ha escogido el método de Ward y la distancia Euclidiana cuadrada como parámetros de evaluación. (Figura A.25).

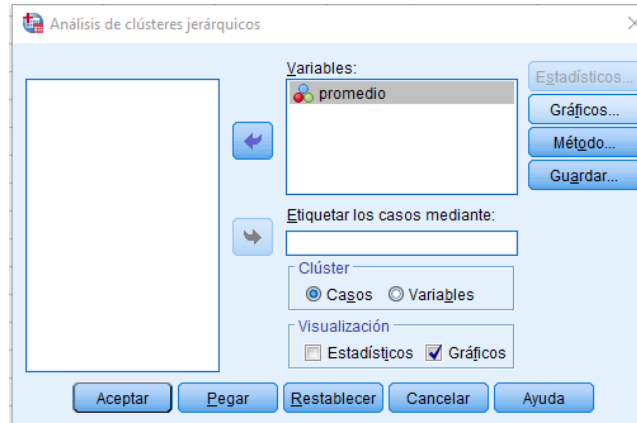


Figura 6.23. Ventana de configuración para análisis cluster. (Fuente: Propia)

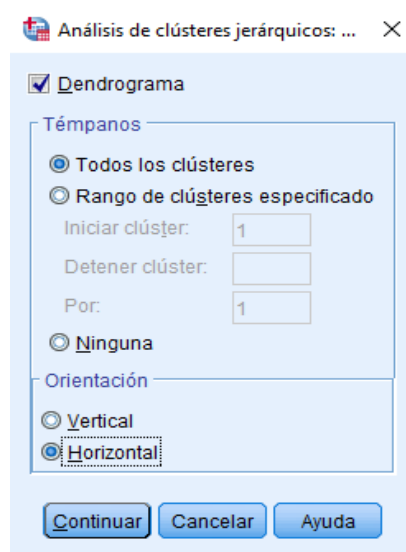


Figura 6.24. Parámetros de la sección de Gráficos para la clasificación de datos en SPSS. (Fuente: Propia)

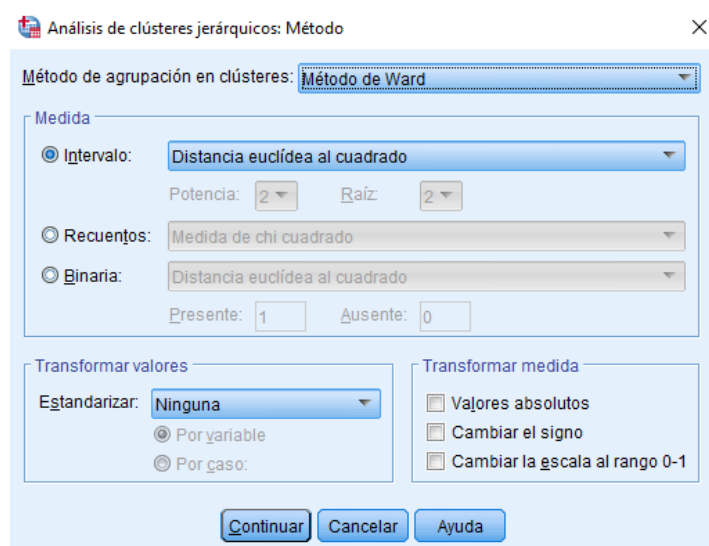


Figura 6.25. Parámetros de la sección de Métodos para la clasificación de datos en SPSS. (Fuente: Propia)

Los resultados obtenidos con este método e pueden inspeccionar en la figura A.26

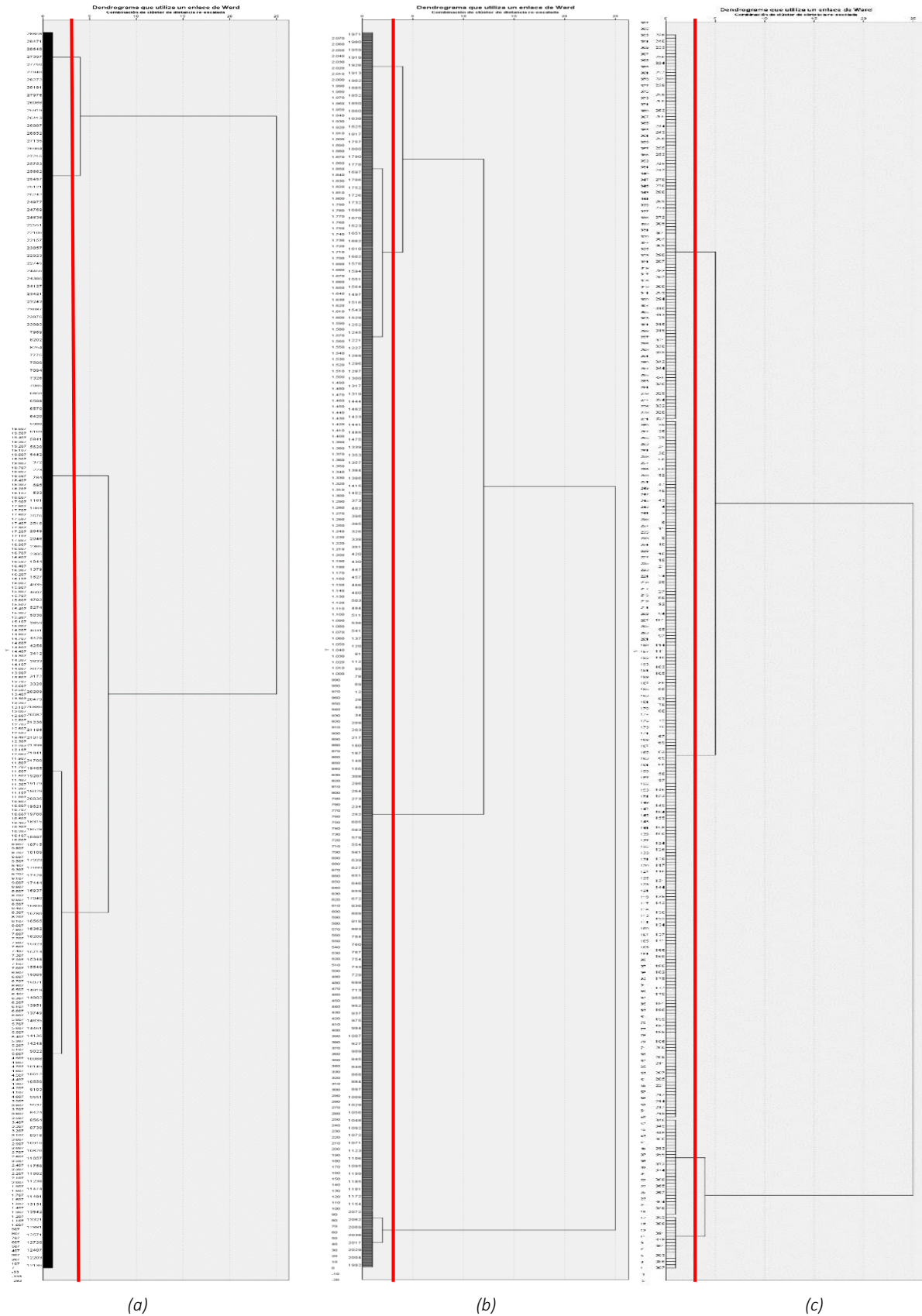


Figura A. 6.26. Resultados obtenidos por el método de Ward. (Fuente: Propia)

ANEXO 7: GRÁFICAS Y TABLAS PARA USUARIOS COMERCIALES E INDUSTRIALES

Tabla 6.1. Consumos típicos para usuarios con tarifa comercial.

HORAS	Demanda (KW)			Demanda (p.u.)		
	Días Laborables (KW)	Sábados (KW)	Domingos y Feriados (KW)	Días Laborables (p.u.)	Sábados (p.u.)	Domingos y Feriados (p.u.)
01:00:00	3.44136651	3.63600006	3.1589	0.4442738	0.56236976	0.5080722
02:00:00	3.32740571	3.3786666	3.1695608	0.42956168	0.52256873	0.50978687
03:00:00	3.31827919	3.38407927	3.1695608	0.42838346	0.52340589	0.50978687
04:00:00	3.31458943	3.12960062	3.1695608	0.42790712	0.4840464	0.50978687
05:00:00	3.52181355	3.23492091	3.25880146	0.45465935	0.50033599	0.52414018
06:00:00	3.69731308	3.25379212	3.30022586	0.477316	0.50325475	0.53080282
07:00:00	4.07781379	3.63863274	3.50106502	0.52643791	0.56277695	0.56310545
08:00:00	5.29949523	3.95186943	3.78270724	0.68415464	0.61122438	0.60840432
09:00:00	6.97165464	5.61737732	5.02700145	0.9000272	0.86882373	0.80853452
10:00:00	7.16268361	5.78107435	5.03777	0.92468867	0.89414228	0.81026651
11:00:00	7.31916031	5.90376394	4.97022275	0.94488951	0.91311833	0.79940233
12:00:00	7.23546013	6.06455996	5.01097958	0.93408398	0.93798819	0.80595759
13:00:00	6.96887049	5.86629961	5.08945614	0.89966777	0.90732383	0.81857963
14:00:00	6.76485075	5.60060632	4.94086671	0.87332922	0.86622981	0.79468075
15:00:00	7.29159397	5.62462435	4.99139925	0.94133075	0.86994461	0.80280832
16:00:00	7.275802	5.7222559	5.11638225	0.93929204	0.885045	0.82291038
17:00:00	7.54184015	5.64204279	5.24647839	0.97363705	0.87263867	0.84383482
18:00:00	7.47622574	5.73465043	5.35391652	0.96516636	0.88696203	0.86111499
19:00:00	7.74604882	6.46549711	5.29080229	1	1	0.8509638
20:00:00	6.76756553	6.03396938	6.21742342	0.87367969	0.93325684	1
21:00:00	5.64351178	4.86892136	5.90817382	0.72856651	0.75306218	0.95026081
22:00:00	4.50026225	4.50727337	3.96583775	0.5809752	0.69712712	0.63785872
23:00:00	3.7577851	4.03713194	3.56639553	0.48512283	0.62441168	0.5736131
0:00:00	3.51800008	3.6807978	3.805446	0.45416704	0.5692985	0.61206158

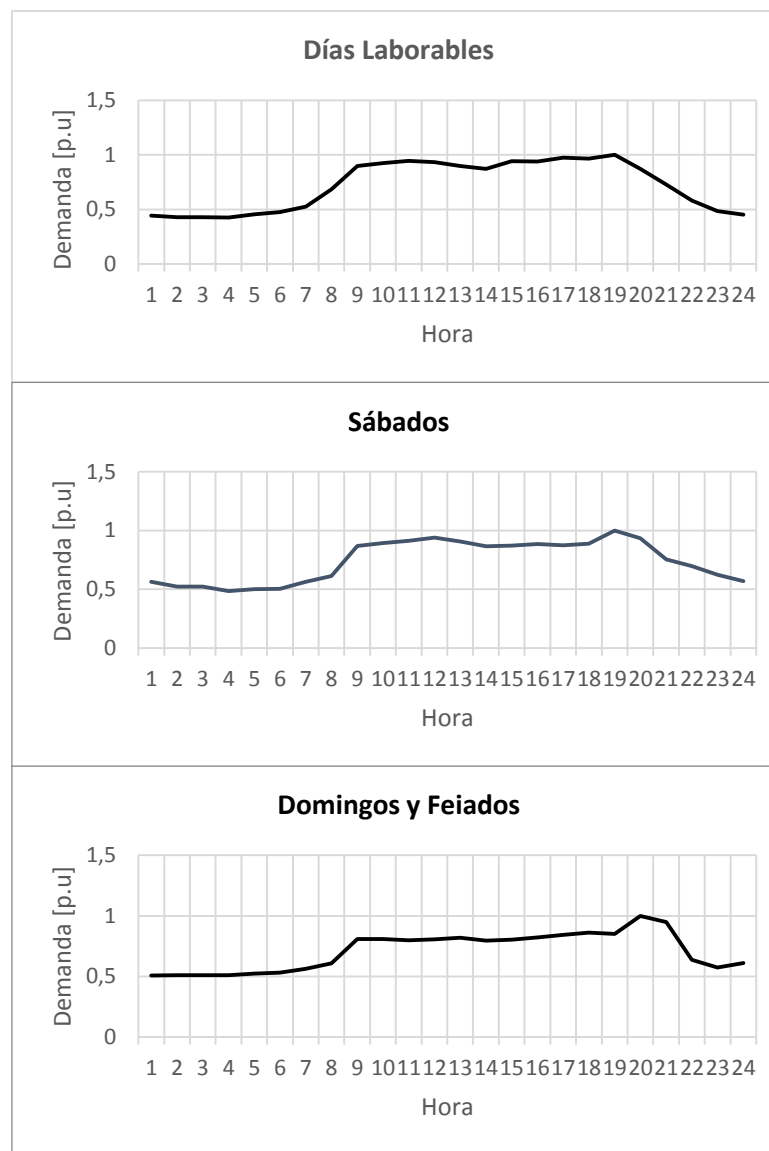


Tabla 6.2. Consumos típicos para usuarios con tarifa industrial.

HORAS	Demanda (KW)			Demanda (p.u.)		
	Días Laborables (KW)	Sábados (KW)	Domingos y Feriados (KW)	Días Laborables (p.u.)	Sábados (p.u.)	Domingos y Feriados (p.u.)
01:00:00	2.51082609	2.64767942	2.40115594	0.23443765	0.41141935	0.41307883
02:00:00	2.53596002	2.56442104	2.21377925	0.23678442	0.39848194	0.3808438
03:00:00	2.60490884	2.37616145	2.47965316	0.24322222	0.36922854	0.42658297
04:00:00	2.55328066	2.39784997	2.31097374	0.23840166	0.37259869	0.39756449
05:00:00	2.66425056	2.64074409	2.40896805	0.248763	0.41034168	0.41442278
06:00:00	2.64140448	2.4102508	2.3325466	0.24662985	0.37452564	0.40127574
07:00:00	3.34257426	3.21545862	3.16929946	0.31209858	0.4996458	0.54522512
08:00:00	7.30065635	4.82749214	3.77251782	0.68166757	0.75013753	0.64899878
09:00:00	8.85037003	5.66405411	3.7470728	0.82636545	0.88012977	0.64462139
10:00:00	9.89381814	5.60430988	3.88834752	0.92379296	0.87084619	0.66892534
11:00:00	10.435732	6.43547613	5.81282732	0.97439185	1	1
12:00:00	9.6799773	5.863101	4.40098151	0.90382649	0.9110594	0.75711547
13:00:00	6.42798372	5.10846963	5.02259631	0.60018549	0.79379824	0.86405393
14:00:00	10.0392685	6.18268487	5.41133812	0.93737377	0.96071911	0.93093048
15:00:00	10.7099952	6.0392724	4.5230472	1	0.93843443	0.77811484
16:00:00	9.48615206	4.55000547	3.03269443	0.88572888	0.70701924	0.5217245
17:00:00	5.15041771	3.17362591	2.04943493	0.48089823	0.49314547	0.3525711
18:00:00	3.02140142	3.02795597	2.30485303	0.28211044	0.47051001	0.39651153
19:00:00	2.79877376	2.58284352	2.14906526	0.26132353	0.40134459	0.36971084
20:00:00	2.56047616	2.81691885	2.55611846	0.23907351	0.43771724	0.43973755
21:00:00	2.58814441	2.73846667	2.30974705	0.24165692	0.42552666	0.39735346
22:00:00	2.72483945	2.42839536	2.11496715	0.25442023	0.3773451	0.36384483
23:00:00	2.62753572	2.69106505	2.25775062	0.24533491	0.41816099	0.38840834
00:00:00	2.60100695	2.5825803	2.68878029	0.2428579	0.40130369	0.46255981

