

UCUENCA

Universidad de Cuenca

Facultad de Ingeniería

Carrera de Electricidad

Reducción del espacio de búsqueda aplicada a la planificación óptima de sistemas transmisión de energía eléctrica de gran escala


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

Autor:

Félix Sebastián González Cantos

Director:

Santiago Patricio Torres Contreras

ORCID:  0000-0002-8803-6811

Cuenca, Ecuador

2025-09-24

Resumen

La Planificación de la Expansión del Sistema de Transmisión (TNEP) es una actividad esencial para garantizar que el sistema de energía eléctrica pueda satisfacer la demanda futura al mínimo costo cumpliendo restricciones técnicas y ambientales. El actual desafío para llevar a la práctica cualquier metodología de TNEP es la complejidad computacional derivada del tamaño de los sistemas de transmisión y el alto número de líneas candidatas para la expansión, lo que resulta en espacios de búsqueda grandes, y por ende tiempos de computación inmanejables. Por lo tanto, en este trabajo se propone una metodología basada en la Reducción del Espacio de Búsqueda (SSR) utilizando algoritmos de agrupamiento, para producir un subconjunto o grupo de líneas candidatas en el cual idealmente se encuentre la solución óptima. Se compara el desempeño de los algoritmos no supervisados tales como *Kmeans*, *PAM* y *CLARA*, considerando como método de agrupamiento al criterio de sobrecarga, de costo-beneficio y de mínimo esfuerzo. Para validar que el subespacio de búsqueda contenga la solución óptima se usa como técnica de optimización la metaheurística híbrida *Improved version of DE-PBILc hybrid metaheuristic* (IDE-PBILc). Finalmente, la metodología es aplicada a cuatro sistemas eléctricos de prueba de distintas escalas, el sistema Garver de 6 barras, IEEE de 24 barras, IEEE de 118 barras modificado e IEEE de 300 barras modificado, obteniendo una reducción significativa del tiempo de procesamiento, 12.05%, 98.85%, 78.81% y 95.76% respectivamente, producto de la reducción de la complejidad computacional. También se comprueba que la solución óptima es parte de los subespacios obtenidos para cada uno de los sistemas de prueba.

Palabras clave del autor: reducción de espacio de búsqueda, planificación de la expansión del sistema de transmisión, modelo AC, metaheurística, MINLP no convexa



El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Cuenca ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por la propiedad intelectual y los derechos de autor.

Repositorio Institucional: <https://dspace.ucuenca.edu.ec/>

Abstract

Transmission Network Expansion Planning (TNEP) is an essential activity to ensure that the electric power system can meet future demand at the lowest cost while meeting technical and environmental constraints. The current challenge for implementing any TNEP methodology is the computational complexity derived from the size of transmission systems and the high number of candidate lines, which results in large search spaces and therefore unmanageable computation times. For that reason, in this work, we propose a methodology based on Search Space Reduction (SSR) using clustering algorithms to produce a subset or group of candidate lines in which the optimal solution is ideally found. The performance of unsupervised algorithms such as *Kmeans*, *PAM*, and *CLARA* is compared, considering the overload, cost-benefit, and least effort criteria as clustering methods. To validate that the search subspace contains the optimal solution, the Improved version of DE-PBILc hybrid metaheuristic (IDE-PBILc) is used as an optimization technique. Finally, the methodology is applied to four electrical test systems of different scales: the 6-bus Garver system, the 24-bus IEEE system, the modified 118-bus IEEE system, and the modified 300-bus IEEE system. This results in significant reductions in processing time, 12.05%, 98.85%, 78.81%, and 95.76%, respectively, due to the reduction in computational complexity. It is also verified that the optimal solution is part of the subspace obtained for each of the test systems.

Author Keywords: search space reduction, transmission network expansion planning, AC power flow, metaheuristic, nonconvex MINLP



The content of this work corresponds to the right of expression of the authors and does not compromise the institutional thinking of the University of Cuenca, nor does it release its responsibility before third parties. The authors assume responsibility for the intellectual property and copyrights.

Institutional Repository: <https://dspace.ucuenca.edu.ec/>

Agradecimiento

A mis padres, que a lo largo de mi trayecto han sabido educarme con amor y han fungido como un pilar inquebrantable para mis logros. A todos mis docentes de los cuales me llevo sus cálidas enseñanzas.

Índice de contenido

1. Nomenclatura	9
Índices y conjuntos	9
Funciones.....	9
Parámetros.....	10
Variables	10
2. Introducción	13
3. Revisión de la literatura.....	14
4. Metodología.....	16
4.1. Formulación matemática de TNEP	16
4.1.1. <i>Formulación del problema AC-OPF</i>	16
4.1.2. <i>Formulación del problema de Expansión Óptima de los Sistemas de Transmisión usando el modelo AC</i>	17
4.2. Reducción del espacio de búsqueda	17
4.2.1. <i>Criterios de clasificación de las líneas</i>	18
4.2.1.1. <i>Criterio de sobrecarga</i>	18
4.2.1.2. <i>Criterio de costo-beneficio</i>	18
4.2.1.3. <i>Criterio de mínimo esfuerzo</i>	18
4.3. Algoritmos de agrupamiento	19
4.3.1. <i>Kmeans</i>	19
4.3.2. <i>K-medoides: PAM y CLARA</i>	20
4.3.3. <i>Mecanismos de determinación del número de grupos</i>	21
4.3.4. <i>Mecanismo de inicialización</i>	21
4.4. Implementación y metodología de pruebas	22
4.4.1. <i>Etapas de AC-OPF</i>	22
4.4.2. <i>Etapas de SSR</i>	23
4.4.3. <i>Etapas de determinación de líneas óptimas para TNEP</i>	25
4.4.4. <i>Sistemas de prueba</i>	26
5. Resultados y Discusión.....	26
5.1. <i>Sistema Garver</i>	26
5.1.1. <i>Determinación del grupo óptimo global para el sistema Garver</i> :.....	26
5.1.2. <i>Análisis del grupo óptimo global del sistema Garver</i> :	26
5.1.3. <i>Comparación del rendimiento de la metaheurística con el espacio completo y reducido en el sistema Garver</i> :	27
5.2. <i>Sistema IEEE de 24 barras</i>	28
5.2.1. <i>Determinación del grupo óptimo global para el sistema IEEE de 24 barras</i> :.....	28

5.2.2. Análisis del grupo óptimo global del sistema IEEE de 24 barras:	28
5.2.3. Comparación del rendimiento de la metaheurística con el espacio completo y reducido en el sistema IEEE de 24 barras:	29
5.3. Sistema IEEE modificado de 118 barras.....	30
5.3.1. Determinación del grupo óptimo global para el sistema IEEE de 118 barras:.....	30
5.3.2. Análisis del grupo óptimo global del sistema IEEE de 118 barras:	31
5.3.3. Comparación del rendimiento de la metaheurística con el espacio completo y reducido en el sistema IEEE de 118 barras:	31
5.4. Sistema IEEE modificado de 300 barras.....	32
5.4.1. Determinación del grupo óptimo global para el sistema IEEE de 300 barras:.....	32
5.4.2. Análisis del grupo óptimo global del sistema IEEE de 300 barras:	32
5.4.3. Comparación del rendimiento de la metaheurística con el espacio completo y reducido en el sistema IEEE de 300 barras:	33
5.5. Comparación de resultados de SSR.....	33
6. Conclusiones y Recomendaciones	34
Referencias.....	36
Anexos.....	39

Índice de figuras

Figura 1. Diagrama de flujo de la metodología de TNEP con SSR.....	22
Figura 2. Diagrama de flujo de la etapa de AC-OPF.	23
Figura 3. Diagrama de flujo de la etapa de SSR.	24
Figura 4. Resultados de agrupamiento con el criterio de Mínimo esfuerzo en el sistema Garver de 6 barras.....	27
Figura 5. Resultados de agrupamiento con el criterio bidimensional de Costo beneficio - Mínimo esfuerzo en el sistema IEEE de 24 barras.....	29
Figura 6. Resultados de agrupamiento con el criterio de Mínimo esfuerzo en el sistema IEEE de 118 barras.....	31
Figura 7. Resultados de agrupamiento con el criterio de Mínimo esfuerzo en el sistema IEEE de 300 barras.....	33

Índice de tablas

Tabla 1. Criterio de selección del grupo óptimo para criterios unidimensionales y bidimensionales.	25
Tabla 2. Diez primeros grupos del sistema Garver de 6 barras con menor número de líneas candidatas y tiempo de resolución al aplicar los dos primeros criterios de selección del grupo óptimo global.	26
Tabla 3. Líneas del grupo global optimo del sistema Garver de 6 barras y número de veces que se repiten.	27
Tabla 4. Tabla comparativa de resultados de TNEP con el espacio de búsqueda completo y reducido en el sistema Garver de 6 barras.....	28
Tabla 5. Diez primeros grupos del sistema IEEE de 24 con menor número de líneas candidatas y tiempo de resolución al aplicar los dos primeros criterios de selección del grupo óptimo global.	28
Tabla 6. Líneas del grupo global optimo del sistema IEEE de 24 barras y número de veces que se repiten.	29
Tabla 7. Tabla comparativa de resultados de TNEP con el espacio de búsqueda completo y reducido en el sistema IEEE de 24 barras.	30
Tabla 8. Diez primeros grupos del sistema IEEE de 118 barras con menor número de líneas candidatas y tiempo de resolución al aplicar los dos primeros criterios de selección del grupo óptimo global.	30
Tabla 9. Tabla comparativa de resultados de TNEP con el espacio de búsqueda completo y reducido en el sistema IEEE de 118 barras.	32
Tabla 10. Diez primeros grupos del sistema IEEE de 300 barras con menor número de líneas candidatas y tiempo de resolución al aplicar los dos primeros criterios de selección del grupo óptimo global.....	32
Tabla 11. Tabla comparativa de resultados de TNEP con el espacio de búsqueda completo y reducido en el sistema IEEE de 300 barras.....	33
Tabla 12. Comparación del tamaño de población utilizado en la metaheurística para los cuatro sistemas.....	34
Tabla 13. Tabla de resultados de SSR en los cuatro sistemas.....	34
Tabla 14. Líneas del grupo global optimo del sistema IEEE de 118 barras y número de veces que se repiten.....	39
Tabla 15. Líneas del grupo global optimo del sistema IEEE de 300 barras y número de veces que se repiten.....	39

1. Nomenclatura

Índices y conjuntos

Λ	Conjunto de los nodos de carga.
m	m-ésima barra.
L	L-ésima línea candidata.
\mathbb{Z}	Conjunto de enteros.
j, l	Índices de las barras.
D	Conjunto universo de datos.
G	Grupos de datos.
x_n	n-ésimo dato.
i, j	Índices de elementos pertenecientes a D .
$G_{\text{óptimo}}$	Grupo óptimo global.
G_{2d}	Grupo óptimo local de un criterio bidimensional.
$G_{\text{Lóptimas}}$	Grupo de líneas candidatas óptimas.

Funciones

f	Función de costo polinómico de las inyecciones de potencia activa y reactiva de los generadores del Sistema Eléctrico de Potencia (SEP).
$P(V, \theta)^m$	Potencia activa que sale de la m-ésima barra.
$Q(V, \theta)^m$	Potencia reactiva que sale de la m-ésima barra.
v	Costo de construcción de las líneas de transmisión, operación y compensación.
SC_L	Valor del criterio de sobrecarga para la L-ésima línea candidata.
CB_L	Valor del criterio de costo beneficio para la L-ésima línea candidata.
σ_L	Valor del criterio de mínimo esfuerzo para la L-ésima línea candidata.
SSE	Suma de errores cuadrados.
ED	Ecuación de diferencias.
CH	Ecuación del índice de Calinski-Harabasz.

Gap Ecuación del índice Gap-statistic.

Sil Ecuación del coeficiente de la silueta.

Parámetros

α_1	Costo de la potencia activa no suministrada
α_2	Costo de la potencia reactiva no suministrada
φ_1	Constante de penalización de uso de generadores ficticios (10^7)
c_{jl}	Costo de inversión de construcción de la línea entre los nodos j y l .
c_{op}	Costo operativo.
\bar{V}/\underline{V}	Magnitud de voltaje máxima y mínima.
$\bar{\theta}/\underline{\theta}$	Ángulo de voltaje máxima y mínima.
S	Potencia compleja.
\bar{S}	Límite de flujo de potencia aparente.
\bar{x}_{jl}	Máximo número de circuitos entre las barras j y l .
c_L	Costo de inversión de construcción de L-ésima línea.
X_L	Impedancia en serie de la L-ésima línea.
R_L	Resistencia en serie de la de L-ésima línea.
d	Dimensiones de agrupamiento de datos.
N	Número de datos en el conjunto universo.
N_L	Número de líneas candidatas.
N_ϵ	Número de ensayos de AC-OPF realizados.
η_1/η_2	Valor inicial de carga e incrementos de carga en los N_ϵ ensayos de AC-OPF
N_a	Número de algoritmos de agrupamiento
N_B	Número de criterios de agrupamiento de líneas candidatas
N_c	Número de mecanismos de determinación del número de grupos candidatos

Variables

r_P	Potencia activa no suministrada
r_Q	Potencia reactiva no suministrada
V	Magnitud de voltaje.
θ	Ángulo de fase de voltaje.
S_{desde} / S_{hacia}	Potencia aparente por las ramas en ambos terminales.
P_G	Potencia activa generada en la barra.
Q_G	Potencia reactiva generada en la barra.
P_D	Potencia activa demandada en la barra.
Q_D	Potencia reactiva demandada en la barra.
x_{jl}	Línea candidata entre las barras j y l .
S_L	Potencia compleja por la L -ésima línea candidata.
γ_L	Susceptancia en serie de la L -ésima línea.
K	Número de grupos.
g_k	Centroide del k -ésimo grupo.
t_{ij}	Disimilaridad entre los elementos i y j .
g'_{ij}	Variable booleana que indica si los elementos i y j pertenecen al mismo grupo o no, 1 y 0 respectivamente.
B	Suma de cuadrados entre grupos.
W	Suma de cuadrados dentro de cada grupo.
W_b^*	Suma de cuadrados dentro del b -ésimo conjunto de datos.
S_{k+1}	Estimado de la desviación estándar entre grupos.
a_i	Distancia promedio entre el elemento x_n y los elementos del grupo que lo contiene.
b_i	Distancia promedio entre el grupo que no contiene al elemento x_n y este.
aux_1 / aux_2	Variables auxiliares.
$T_{T.agr}$	Tiempo de total de agrupamiento de datos.
T_{agr}	Tiempo de agrupamiento de datos.
T_{AC-OPF}	Tiempo que tarda realizar las N_ϵ pruebas de AC-OPF.

T_{mec}	Tiempo que tarda el mecanismo de determinación de número de grupos.
T_{λ}	Tiempo de procesamiento de la heurística.
T_{Total}	Tiempo total de procesamiento.

2. Introducción

El sistema de transmisión de energía eléctrica es fundamental para el aprovisionamiento del servicio eléctrico, su principal objetivo es el de interconectar al sistema de distribución con las centrales de generación eléctrica [1]. Debido al incesante incremento de la carga es necesario planificar la expansión del sistema de transmisión con miras a satisfacer la demanda a largo plazo y al menor costo posible, esto se logra a través de la adición y/o repotenciación de líneas en el sistema [2], [3]. La técnica de obtener la cantidad, momento en el que construir y la ubicación óptima de las líneas a añadir se denomina TNEP [4]. Esta planificación debe considerar la de confiabilidad y seguridad del suministro eléctrico, sin embargo, esto es un gran reto por la incertidumbre en el crecimiento de la carga, la variabilidad de la generación [5] y la complejidad computacional derivada del incremento de tamaño de los sistemas de transmisión [4]. En este documento se considera a la solución óptima de la TNEP como el conjunto de líneas que minimiza los costos totales de inversión garantizando el aspecto técnico y económico de la operación presente y futura del SEP.

Para obtener de forma sistemática la solución óptima a la TNEP se han desarrollado modelos matemáticos que consideran subgrupos de variables del problema y que en conjunto con el modelamiento de flujos de potencia dictan la complejidad de su resolución [4]. Entre las variables se consideran variables determinísticas y estocásticas, cuyo modelamiento presenta un alto grado de incertidumbre. Entre las variables determinísticas se hallan los parámetros eléctricos de las líneas candidatas y de las centrales de generación eléctrica operativas y futuras. En cambio, entre las variables estocásticas se encuentran la inserción de energías renovables a la red; el costo y disponibilidad de recursos primarios para la generación; la demanda energética y la distribución del flujo de potencia por el sistema [4]. El modelamiento de los flujos de potencia es de gran importancia ya que permite analizar el comportamiento del sistema y su operación ante condiciones específicas de generación y demanda.

Existen diversas formas de modelar los flujos de potencia, en [6] se indica que existen dos categorías flujos AC y DC, siendo el modelo DC más rápido y sencillo pero menos preciso. Por lo que la selección del modelo dependerá de la precisión, capacidad computacional y parámetros relevantes para el estudio [7]. Al utilizar el modelo AC este puede ser resuelto mediante algoritmos clásicos como la programación lineal [8], Programación Lineal Entera Mixta (MILP) [9], [10], [11], [12] y Programación No Lineal Entera Mixta (MINLP) [13] y mediante algoritmos no clásicos como la descomposición de Benders [14] y la teoría de juegos [15]. En caso de utilizar MINLP la resolución es un problema de tipo Polinómico No Determinista Completo (NP-Hard) que presenta la particularidad de que su tiempo de cómputo crece exponencialmente en función del número de líneas candidatas [2], [5].

Para solventar el problema computacional de la TNEP en [9] y [10] se aplica SSR que según [16] es una técnica que analiza las características del espacio de posibles soluciones, empleando muestras y minería de datos, con la finalidad de agrupar las posibles soluciones en subespacios, a partir de dicho proceso se encuentra un grupo dentro de un subespacio que contiene la solución óptima simplificando así el proceso de optimización, ya que se lo realiza a un subconjunto del espacio en vez de a todo el conjunto de soluciones posibles [17], [18]. Entre las herramientas que pueden ser empleadas para reducir el espacio de búsqueda se encuentran los algoritmos de aprendizaje automático. Que se dividen en dos categorías

principales los de aprendizaje supervisado, que requieren la categorización manual de datos; y los de aprendizaje no supervisado, donde el algoritmo automáticamente detecta patrones en el conjunto de datos [19]. Entre los algoritmos de aprendizaje no supervisado se encuentra el algoritmo *Kmeans*, que está diseñado para categorizar datos en k grupos empleando muestreo aleatorio mediante la asignación fortuita de centroides [20].

Por lo que en este trabajo se presenta una metodología computacional para reducir el tiempo de solución de la TNEP considerando un modelo de MINLP no convexo mediante el agrupamiento de las líneas en base a los criterios de sobrecarga, costo-beneficio y mínimo esfuerzo. Se abordará al problema de TNEP sin considerar la incertidumbre de las variables estocásticas al suponer que dichos valores son determinísticos. Finalmente, para evaluar la validez de los algoritmos de agrupamiento se los aplicará en sistemas de gran escala como los sistemas de prueba IEEE de 300 barras e IEEE de 118 barras y sistemas de pequeña y mediana escala como el sistema Garver de 6 barras e IEEE de 24 barras. La aplicación de varios algoritmos tiene el fin de comparar su rendimiento en función de la escala del sistema de prueba, de esta manera se aguarda reducir el tiempo computacional y encontrar un algoritmo y criterio de agrupamiento óptimo para cada sistema analizado.

3. Revisión de la literatura

En la literatura se encontraron distintos enfoques para la solución del problema de TNEP entre ellos se encuentra la aplicación de algoritmos de aprendizaje no supervisado a partir de modelos metaheurísticos como [14] y [15] que utilizan el algoritmo de descomposición de Benders y teoría de juegos respectivamente. En [14] se propone un modelo de expansión de la transmisión multi-anual considerando la congestión, se utiliza el algoritmo de descomposición para dividir al problema de TNEP en dos subproblemas, uno referente a la seguridad y otro a la congestión, a partir de los resultados se establece que el método presenta dificultades de convergencia debido a la naturaleza no convexa del TNEP pero presenta buen rendimiento al resolver el problema modelando el flujo de potencia del SEP mediante flujos DC. Finalmente, en [15] se propone la aplicación de la teoría de juegos para el problema de planificación de la expansión de generación y transmisión (GTEP) en un mercado eléctrico descentralizado, se evalúa el problema mediante dos modelos de dos niveles, uno para el operador del sistema de transmisión y otro para las empresas generadoras, donde en el segundo nivel se plantea un juego de Stackelberg obteniendo resultados positivos en un modelo de tres barras, aun así en la literatura no se presenta un caso de aplicación a sistemas de gran escala.

También es posible aplicar modelos matemáticos clásicos para la resolución del problema de TNEP entre estos se encuentran la programación lineal [8], programación no lineal [21], MILP [9], [10], [11], [12] y MINLP [13]. En [8] se presenta un algoritmo de solución al problema de TNEP en base a un algoritmo genético y programación lineal para un sistema con mercado eléctrico completamente descentralizado donde la solución óptima involucra la división justa de costos de operación e inversión para la construcción de las líneas, a partir de los resultados se determina que la inversión total es menor que con métodos tradicionales pero no se consideran factores técnicos como la confiabilidad de la respuesta y el modelo de flujos de potencia es el tipo DC. En [21] se plantea el uso de optimización no lineal, en base al modelo DC de flujos de potencia, para resolver el problema de TNEP considerando variables estocásticas, el modelo se aplica al sistema de transmisión de Alemania mostrando que esta

metodología puede ser aplicada a problemas de TNEP de gran escala, pero los requisitos computacionales son altos debido a la no linealidad del problema.

En [11] se utilizan flujos AC y se modela al problema de GTEP mediante MILP considerando múltiples perfiles de demanda con resolución horaria, para solventar los retos computacionales se aplica la descomposición de Benders obteniendo como resultado que el algoritmo puede resolver el problema en 15 días de cómputo para la planificación de los próximos 20 años del caso de estudio planteado por el Comité de Confiabilidad Eléctrica de Texas (ERCOT). En [13] se propone un modelo MINLP no convexo de dos etapas para la TNEP, la primera consiste en la selección de líneas que minimicen el costo operacional y de inversión y es resuelto por una metaheurística. La segunda etapa se encarga principalmente del problema operativo que es resuelto mediante la formulación AC de flujo de potencia óptimo considerando contingencias. Para la primera etapa se comparan cuatro metaheurísticas donde IDE-PBILc obtiene los mejores resultados, sin embargo, su tiempo de resolución se incrementa en función del número de líneas candidatas, pasando de 13 minutos en el sistema IEEE de 118 barras a 1.9 horas en el sistema IEEE de 300 barras.

Es por esto que para solventar el problema computacional y reducir el tiempo de procesamiento se propone el uso de técnicas de SSR que en [9], [10] y [12] han brindado buenos resultados. En [9] se utiliza un modelo AC de flujos de potencia aplicando SSR para reducir el número de líneas candidatas en un problema multi etapa que considera contingencias para la planificación, obteniendo resultados satisfactorios como que el tiempo computacional para la obtención de las líneas óptimas del modelo de 24 barras evaluado se redujo de 13 días a 10 minutos, en comparación con el tiempo de cálculo computacional de analizar el comportamiento de todo el espacio de búsqueda. En [10], se aplica el algoritmo *Kmeans* para el agrupamiento de las líneas candidatas en base a tres criterios de clasificación determinando que el criterio de agrupamiento óptimo depende de la escala del sistema. Finalmente, en [12] se consideran líneas candidatas de alto voltaje en DC y AC para el problema de TNEP en el sistema eléctrico del sur de Brasil, concluyendo que la aplicación de SSR permitió obtener la solución óptima y reducir el tiempo de cálculo en un 85.9%, con respecto a considerar el espacio completo.

Se emplearon tres algoritmos de agrupamiento para la SSR, entre ellos *Kmeans* fue aplicado en todos los sistemas ya que en [10] fue implementado de manera exitosa para la reducción de espacio de búsqueda en el problema de TNEP, sin embargo se utilizó el mecanismo de inicialización tradicional, que incrementa el tiempo de convergencia del algoritmo [22] y para la determinación del número de grupos se empleó el método del codo, que presenta cierta subjetividad en su interpretación. Debido a la presencia de valores extremos en algunos criterios de agrupamiento unidimensionales y todos los criterios bidimensionales se propone la implementación de los algoritmos *Clustering LARge Application (CLARA)* y *Partitioning Around Medoids (PAM)* para el agrupamiento de datos en los sistemas Garver e IEEE de 24 barras y para los en los sistemas de IEEE de 118 e IEEE de 300 barras se propone el uso de *CLARA*. En [23] se comparan los tres algoritmos de agrupamiento propuestos, en conjuntos con un gran número de datos, a partir de los resultados se indica que *CLARA* y *Kmeans* producen mejores o iguales resultados y en un menor tiempo de procesamiento que *PAM*, mientras que *CLARA* obtiene mejores indicadores que *Kmeans*, sin embargo lo hace en un mayor tiempo de procesamiento. Resultados similares se presentan en [24] donde en bases de datos de similares proporciones a las de los sistemas de IEEE de 300 y 118 barras *CLARA*

produce mejores resultados que *PAM* y *Kmeans* en un tiempo comparable, siendo *Kmeans* la que peores resultados produce, sin embargo a escalas de datos proporcionales a las encontradas en los sistemas de 6 y 24 barras *PAM* produce mejores resultados que *Kmeans* y *CLARA* y en un tiempo menor a *CLARA* pero superior a *Kmeans*. Finalmente, en [25] se concluye que a mayor cantidad de datos a clasificar y mayor número de grupos, el tiempo de convergencia de *PAM* presenta un crecimiento exponencial, a diferencia de *CLARA* que es lineal.

En base a la revisión de la literatura se ha determinado que actualmente se carece de artículos académicos que evalúen el rendimiento de varios algoritmos de agrupamiento para reducir el espacio de búsqueda en el problema de TNEP y que comparen los criterios de agrupamiento en sistemas de diversas escalas, utilizando el modelo AC-OPF y resuelto mediante MINLP no convexo para el modelamiento del sistema y para la selección de las líneas candidatas óptimas para el TNEP la metaheurística IDE-PBILc, por lo que en este artículo se propone:

- Adaptar los sistemas Garver 6, IEEE de 24 barras, IEEE de 118 barras e IEEE de 300 barras para la aplicación de SSR.
- Comparar el tiempo de resolución de la metaheurística IDE-PBILc con el espacio de búsqueda completo y con el espacio de búsqueda reducido.
- Proponer un mecanismo, criterio y algoritmo de agrupamiento óptimo para la SSR de acuerdo con el tamaño del sistema.

De esta forma se espera que el espacio de búsqueda reducido disminuya el tiempo de obtención de una respuesta de la metaheurística, debido a la menor cantidad de líneas candidatas.

4. Metodología

4.1. Formulación matemática de TNEP

4.1.1. Formulación del problema AC-OPF

Para modelar el comportamiento del SEP se utiliza el modelo de flujo óptimo de potencia AC (AC-OPF) [13], [26] con MNILP no convexa, su resolución consiste en minimizar la función de costo polinómico de las inyecciones de potencia activa y reactiva de todos los generadores del sistema, ecuación (1), esta función también considera penalizaciones en la topología por el uso de generadores ficticios. La función objetivo está sometida a restricciones de igualdad, ecuaciones (2) y (3), que son ecuaciones no lineales que deben cumplirse para el balance de potencia activa y reactiva en cada barra del sistema. Y restricciones de desigualdad que indican que la magnitud y ángulo del voltaje en todas las barras deben establecerse dentro de ciertos límites, ecuación (4) y (5), que la potencia compleja en cada línea debe permanecer dentro de sus límites operativos ecuaciones (6) y (7) y que las potencias activa y reactiva no suministradas deben encontrarse dentro de un rango definido, ecuaciones (8) y (9).

$$\min f^m = \sum_{(j \in \Lambda)} (\alpha_1 \cdot r_{Pj}^m \cdot \varphi_1) + (\alpha_2 \cdot r_{Qj}^m) \quad (1)$$

$$P(V, \theta)^m - P_G^m + P_D^m - r_P^m = 0 \quad (2)$$

$$Q(V, \theta)^m - Q_G^m + Q_D^m - Q_P^m = 0 \quad (3)$$

$$\underline{V} \leq V^m \leq \bar{V} \quad (4)$$

$$\underline{\theta} \leq \theta^m \leq \bar{\theta} \quad (5)$$

$$S^{desde,m} \leq \bar{S} \quad (6)$$

$$S^{hacia,m} \leq \bar{S} \quad (7)$$

$$\underline{r}_P \leq r_P^m \leq \bar{r}_P \quad (8)$$

$$\underline{r}_Q \leq r_Q^m \leq \bar{r}_Q \quad (9)$$

4.1.2. Formulación del problema de Expansión Óptima de los Sistemas de Transmisión usando el modelo AC.

Para la selección de las líneas óptimas para la expansión del sistema se utiliza el planteamiento presentado en [13] cuya función objetivo se encuentra en la ecuación (10). Sin embargo, únicamente se emplea la metaheurística IDE-PBILc y no se consideran contingencias ni el costo de pérdidas en las líneas. La ecuación (11) limita el número de líneas que pueden colocarse entre las barras j y l . Se utiliza la metaheurística IDE-PBILc ya que entre todas las metaheurísticas analizadas presenta el mayor porcentaje de éxito, esta metaheurística es la combinación de Evolución Diferencial (DE) y Aprendizaje Incremental Basado en la Población (PBILc) en adición a dos técnicas que incrementan su rendimiento.

$$\min v = \sum_{(j,l) \in \Omega} (c_{jl} \cdot x_{jl}) + c_{op} \quad (10)$$

sujeto a

$$0 \leq x_{jl} \leq \bar{x}_{jl}; \quad x_{jl} \in \mathbb{Z} \quad (11)$$

La primera técnica es la modificación de la topología inicial con nuevas líneas, que en caso de presentar una mayor función objetivo que la topología inicial se la considera como solución óptima, sin embargo, si al adicionar nuevas líneas la función objetivo no mejora estas líneas se eliminan iterativamente reduciendo así el espacio de búsqueda. La segunda técnica identifica el número de líneas que se pueden remover de la topología óptima actual y las elimina de forma iterativa, una vez terminado dicho proceso se adiciona un nuevo circuito que tiene un costo de inversión menor que las líneas eliminadas, si durante este proceso mejora la función objetivo, entonces la topología obtenida reemplaza a la previa como mejor solución.

4.2. Reducción del espacio de búsqueda

Dentro del espacio de búsqueda existen dos conceptos clave que requieren diferenciación: grupos y subespacios. Los grupos son conjuntos de elementos de las posibles soluciones mientras que los subespacios son conjuntos de atributos dentro del espacio de datos, ambos elementos son interdependientes [17], en consecuencia su búsqueda no puede ser

estrictamente secuencial, por lo tanto, los algoritmos de búsqueda deben considerar ambos conceptos de forma simultánea. De acuerdo con [20] existen varios tipos de subespacios y dependiendo de esto deben aplicarse distintos criterios para los algoritmos de agrupamiento. En el caso de análisis se cuenta con un espacio de ejes paralelos, es decir un espacio euclidiano donde la norma L_p puede ser utilizada como medida de distancia. Este tipo de espacio tiene la característica principal de que solo un subgrupo de atributos es considerado para calcular la norma y así formar los grupos. Para definir los subespacios pueden ser aplicadas dos técnicas básicas; búsqueda de arriba hacia abajo [27] y de abajo hacia arriba [28]. La búsqueda de subespacios de arriba hacia abajo puede utilizar la asunción de la localidad o el muestreo aleatorio para la conformación de grupos preliminares y la posterior selección del subespacio óptimo a partir del espacio de dimensiones completas. En el caso de análisis los subespacios son los tres criterios de clasificación de líneas propuestos y su combinación bidimensional, por lo que se tienen seis subespacios de caracterización de los datos a los cuales se aplicará la técnica de búsqueda de arriba hacia abajo, por lo que se inicia con la creación de grupos en función de los criterios de clasificación de las líneas candidatas para la posterior selección del subespacio óptimo para cada sistema.

4.2.1. Criterios de clasificación de las líneas

De acuerdo con [20] para la correcta selección de subespacios debe iniciarse mediante una fase de preprocesamiento para determinar la validez de sus características y así mejorar la calidad de los grupos, es por esto que para el agrupamiento se consideraron tres criterios que han sido evaluados en [10], [29], [30] y han brindado buenos resultados para la selección de líneas candidatas para la TNEP.

4.2.1.1. Criterio de sobrecarga

El criterio se presenta en la ecuación (12) y está conformado por la relación entre la potencia compleja que transporta la línea candidata y su cargabilidad [10], [29]. Este criterio evalúa el grado de utilización de la línea en la nueva topología de red, donde que su valor sea cercano a la unidad indica que la línea candidata juega un rol clave en la conducción de potencia hacia las cargas, en [10] se concluye que este criterio presenta buenos resultados para sistemas de pequeña escala.

$$SC_L = \frac{S_L}{\bar{S}_L} \quad (12)$$

4.2.1.2. Criterio de costo-beneficio

Este criterio fue utilizado en [10] siendo recomendado como característica de agrupamiento para el sistema de mediana escala IEEE de 24 y el sistema de gran escala IEEE de 118 barras. El criterio está conformado por la relación entre la inversión para la construcción de la línea y la potencia compleja transportada por esta, ecuación (13), por lo que es expresado en $[\$/MW]$.

$$CB_L = \frac{c_L}{S_L} \quad (13)$$

4.2.1.3. Criterio de mínimo esfuerzo

El criterio de mínimo esfuerzo [10], [29], [30] considera la susceptancia en serie de la línea candidata, ya que por el camino con mayor susceptancia o menor esfuerzo se logra la mejor distribución de los flujos de potencia. El criterio se presenta en la ecuación (14) mientras que en la ecuación (15) se presenta el cálculo de la susceptancia en serie de la línea.

$$\sigma_L = -\frac{1}{2}\gamma_L\theta_L^2 \quad (14)$$

sujeto a

$$\gamma_L = -\frac{X_L}{R_L^2 + X_L^2} \quad (15)$$

4.3. Algoritmos de agrupamiento

Para la clasificar a las líneas de acuerdo con su magnitud de cada criterio se emplean algoritmos de agrupamiento. Cuya finalidad es particionar, sin supervisión humana, un conjunto de datos sin etiquetar. Los elementos que conforman los grupos deben guardar similitud en la mayor cantidad de características posible, mientras que, entre datos de distintos grupos debe existir la mayor cantidad de diferencias posibles [17], [20], [31]. Para lograr esto se cuantifican las similitudes y diferencias entre los datos mediante la utilización de medidas de proximidad entre elementos y el punto central asignado a cada grupo. Según [17] el funcionamiento general de estos algoritmos asume que se tiene un grupo de datos D , que contiene N instancias datos x_1, x_2, \dots, x_n donde cada instancia tiene d dimensiones, el proceso de agrupamiento resulta en un conjunto de grupos G y un número de grupos k , de forma que $G = G_1, G_2, \dots, G_k$, donde se satisfacen las restricciones:

- La combinación de todos los grupos es igual al dominio.
- La intersección de dos grupos cualquiera es un conjunto vacío.
- La cardinalidad de cualquier grupo es un conjunto no vacío.

Existen distintos tipos de algoritmos de agrupamiento, en este caso serán utilizados los algoritmos de agrupamiento particional que agrupan los datos mediante la optimización de una función objetivo cuyo valor se mejora iterativamente incrementado la calidad de las particiones [20]. Estos métodos requieren de un conjunto de valores iniciales para la definición de centroides o medoides y del número de grupos k en el que clasificar los datos, estos valores iniciales presentan un gran impacto en el rendimiento de los algoritmos [20]. Y a partir de ellos se inicia un proceso iterativo de mejora de la ubicación de los centroides o medoides y asignación de grupo a cada uno de los elementos hasta la convergencia, que generalmente ocurre cuando un porcentaje específico de los datos no cambia de grupo entre iteraciones. Sin embargo en [32] se propone un algoritmo de agrupamiento particional que no requiere de la ubicación inicial de centroides ni de k , denominado *U-Kmeans* cuyos resultados son prometedores pero su implementación es compleja, por lo que no será utilizado. A continuación, se presentan los algoritmos de agrupamiento, mecanismos de determinación de número de grupos óptimo y el mecanismo de inicialización implementados.

4.3.1. Kmeans

El algoritmo de agrupamiento *Kmeans* es el algoritmo particional más utilizado [33], [32], [22], en esencia es un problema de optimización que busca minimizar la Suma de Errores Cuadrados (SSE), ecuación (16). Su funcionamiento comienza con la selección de un número k de centroides iniciales, a cada elemento se le asigna un grupo de acuerdo a su distancia con los centroides, una vez todos los grupos son conformados se actualiza la posición de los centroides, estos dos pasos se repiten de forma iterativa hasta que los centroides no cambian de posición o hasta que un porcentaje definido de los datos no cambie de grupo, siendo el segundo caso el motivo de convergencia más común, ya que el primer motivo es un problema del tipo NP-Hard [35].

$$\min SSE(G) = \sum_{k=1}^K \sum_{x_n \in G_k} \|x_n - g_k\|^2 \quad (16)$$

sujeto a

$$g_k = \frac{\sum_{x_n \in G_k} x_n}{|G_k|} \quad (17)$$

La medición de distancia entre los elementos del conjunto de datos y los centroides puede realizarse de múltiples formas, esto puede afectar significativamente la posición de los centroides y la calidad de la solución final [20]. Entre los tipos de medición se encuentran la distancia Manhattan, la distancia Euclidiana y la similitud coseno. En el caso de análisis debido a la naturaleza de los datos se utilizará la distancia Euclidiana.

4.3.2. *K-medoides: PAM y CLARA*

A diferencia de *Kmeans* los métodos basados en *Kmedoides* presentan resultados más uniformes dentro de cada grupo ya que son menos sensibles a elementos con valores extremos, por lo que se recomienda su uso con bases de datos con gran cantidad de ruido o grandes bases de datos con valores heterogéneos [33], [34]. Estos algoritmos buscan minimizar la función objetivo de error absoluto y utilizan un elemento real como centro del grupo, a diferencia de *Kmeans* que minimiza la SSE y utiliza centroides. El algoritmo inicia con la selección de un elemento como centro de cada grupo para posterior al agrupamiento ser remplazado por un punto arbitrario no representativo, la función objetivo es evaluada en ambos casos y se calcula el costo del cambio, que en caso de ser menor a 0 hace que se utilice el punto no representativo como centroide para formar k nuevos grupos, repitiéndose el proceso hasta cumplir con las condiciones de convergencia, que son iguales que las de *Kmeans*.

Para mejorar el rendimiento computacional del algoritmo *Kmedoides* en [35] se propone el algoritmo *PAM* que se basa en el análisis de disimilaridades entre los elementos, intercambiando iterativamente los medoides y centroides a partir de la matriz de disimilaridades hasta la convergencia. La ecuación de disimilaridades se encuentra en (18). Sin embargo este método no es recomendable para grandes bases de datos [20] por lo que en [36] se propone *CLARA* que aplica *PAM* a un número menor de datos los cuales selecciona mediante un método de muestreo, reduciendo así el tiempo de cálculo y volviendo lineal el número de cálculos y almacenamiento necesario con respecto al número de elementos a

procesar. En [36] recomienda utilizar este método con bases de datos con por lo menos 120 elementos, por lo que puede ser empleado en cualquiera de los sistemas de prueba.

$$\min ED = \sum_{i=1}^N \sum_{j=1}^N t_{ij} g'_{ij} \quad (18)$$

4.3.3. Mecanismos de determinación del número de grupos

Para definir el número k de grupos óptimo se han desarrollado mecanismos de estimación del número de grupos entre los que se encuentran el método del codo [19], [37], [38], [39] que es un método gráfico que calcula el valor de SSE, ecuación (16), para distintos valores de k , siendo seleccionado el valor de k donde el decremento de SSE es marginal para un mayor número de grupos, sin embargo esto no es óptimo ya que el punto de codo es subjetivo. El método de inicialización a partir del índice de Calinski-Harabasz [20], [40], [41] es un problema de optimización que consiste en escoger el número k que maximice la función objetivo de la ecuación (19), que está constituida por la división entre la razón de la suma de cuadrados entre grupos y el número de grupos menos uno y la razón de suma de cuadrados dentro de los grupos y el número de elementos totales N menos el número de grupos k . En [40] se aplica este mecanismo en distintos conjuntos de datos presentando mejores resultados que otros mecanismos en conjuntos de pequeña, mediana y gran escala.

$$CH(K) = \frac{\left[\frac{B(k)}{(k-1)} \right]}{\left[\frac{W(k)}{N-k} \right]} \quad (19)$$

También se emplearon los mecanismos de inicialización *Gap-Statistic* y el coeficiente de silueta. El primero utiliza la ecuación (20) para distintos valores de k siendo seleccionado el número de grupos que sea mayor o igual a la diferencia del coeficiente de $k + 1$ y la desviación estándar de este, representada en la ecuación (21). Mientras tanto el coeficiente de la silueta utiliza las distancias inter-grupales e intra-grupales las cuales son promediadas primero para cada grupo y luego para todos los grupos, siendo seleccionado como número de grupos óptimo el k con mayor valor de *Sil*.

$$Gap(k) = \frac{1}{B} \times \sum_b \log(W_b^*(k)) - \log(W(k)) \quad (20)$$

$$Gap(k) \geq Gap(k + 1) - s_{k+1} \quad (21)$$

$$Sil = \frac{\sum_{i=1}^N \frac{b_i - a_i}{\max(a_i, b_i)}}{N} \quad (22)$$

4.3.4. Mecanismo de inicialización

En [22] se propone el mecanismo de inicialización tradicional utilizado por *Kmeans* que consiste en la ubicación aleatoria de los k centroides, sin embargo este método presenta resultados de calidad variable, debido a la ubicación aleatoria de centroides, y causa que el algoritmo tarde más tiempo en encontrar una solución que con otros métodos de inicialización. Mientras tanto en [42] se propone el algoritmo de inicialización *Kmeans++* que para un grupo

de datos de 1024 elementos tarda cerca de la mitad de tiempo y produce resultados 20% mejores que el método de inicialización tradicional. *Kmeans++* utiliza un enfoque probabilístico donde el primer centroide se escoge de forma aleatoria, el siguiente centroide seleccionado es el que se encuentra a mayor distancia del último centroide definido, siendo escogido a partir de una puntuación de probabilidad con pesos, y así hasta tener un número k de centroides. Por lo que se decidió utilizar para los tres algoritmos a *Kmeans++* como mecanismo de inicialización.

4.4. Implementación y metodología de pruebas

Los cálculos fueron realizados en un sistema Intel i5, 3.1GHZ, 8GB RAM. Se utilizó la plataforma de cálculo científico Matlab para la implementación de los mecanismos de SSR y la metaheurística IDE-PBILc, en adición se empleó Matpower en Matlab para la resolución de AC-OPF. En función de los resultados de [13] se empleó *Parallel Computing Toolbox* de Matlab para reducir el tiempo de procesamiento de la metaheurística, utilizando 4 procesos en paralelo, adicionalmente se realizaron pruebas de esta herramienta para el agrupamiento de datos en el sistema Garver determinándose que no es conveniente ya incrementa el tiempo de resolución, por lo que no fue implementada. La Figura 1 presenta el diagrama de flujo general de la implementación donde se divide al problema en tres etapas: Etapa de AC-OPF, etapa de SSR y etapa de determinación de líneas óptimas para la TNEP.

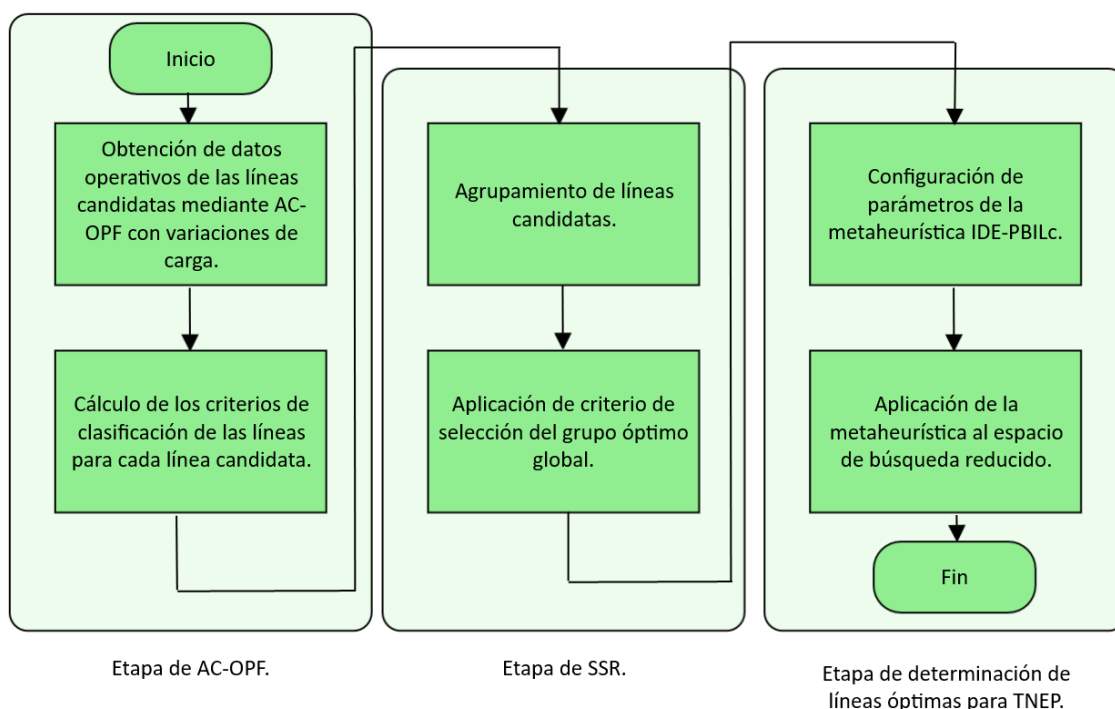


Figura 1. Diagrama de flujo de la metodología de TNEP con SSR.

4.4.1. Etapa de AC-OPF

En la etapa de AC-OPF se simula el comportamiento de la implementación de las líneas candidatas, en la Figura 2 se encuentra su diagrama de flujo. En esta etapa se calculan los tres criterios para diferentes niveles de demanda en cada uno de los sistemas, donde K_1 es la demanda inicial, K_2 es el aumento de demanda entre pasos y N_e es el número de pasos,

siendo diez en todos los sistemas. En los sistemas Garver, IEEE de 24 barras e IEEE de 118 barras las constantes son $K_1 = 0.5$ y $K_2 = 0.07$ lo que permite analizar el comportamiento de cada sistema con un incremento y decremento de demanda de hasta el 50 por ciento de su valor inicial. Mientras tanto en el caso del sistema IEEE de 300 barras por motivos de convergencia las constantes son $K_1 = 1.13$ y $K_2 = 0.01$ por lo que se analiza el funcionamiento del sistema para un incremento de demanda de hasta el 25 por ciento. Los valores de cada criterio obtenidos en cada paso son normalizados para evitar sesgos al clasificarlos.

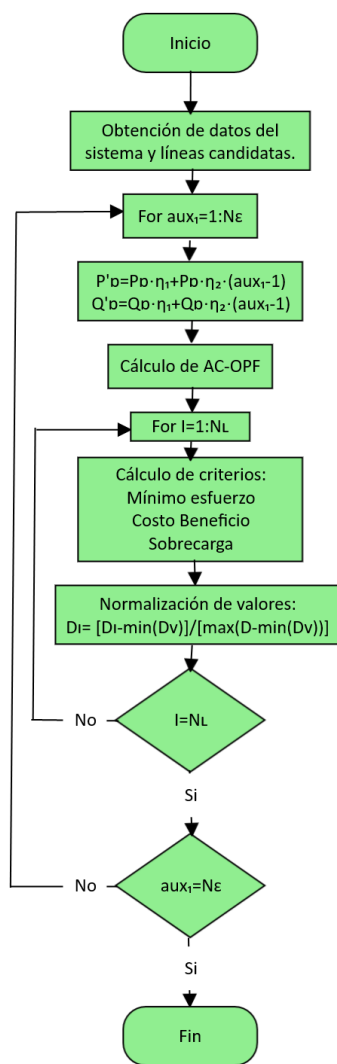


Figura 2. Diagrama de flujo de la etapa de AC-OPF.

4.4.2. Etapa de SSR

Para cada criterio se obtienen N_ϵ datos por línea candidata, por lo tanto, la ecuación (23) indica el número de datos a clasificar para cada criterio. En la Figura 3 se presenta el diagrama de flujo de la etapa de agrupamiento y selección de un grupo óptimo global siendo este el espacio de búsqueda reducido. Los algoritmos de agrupamiento utilizados para todos los sistemas fueron *Kmeans* y *CLARA*, mientras que *PAM* fue empleado en el sistema Garver

e IEEE de 24 barras ya que en [25] se recomienda su uso para bases de datos con menos de 1000 elementos y en las pruebas realizadas el tiempo de resolución de este algoritmo en sistemas de gran escala es injustificablemente elevado. Mientras tanto los mecanismos de determinación del número de grupos óptimo utilizados son *Gap-Statistic*, *Calinski-Harabasz* y el coeficiente de la silueta para todos los algoritmos de agrupamiento, sin embargo, para *Kmeans* también se utilizó el método del codo. Los algoritmos de agrupamiento fueron probados con dos límites máximos de número de grupos, 6 y 12, entregando resultados variados.

$$N = N_L \times N_e \tag{23}$$

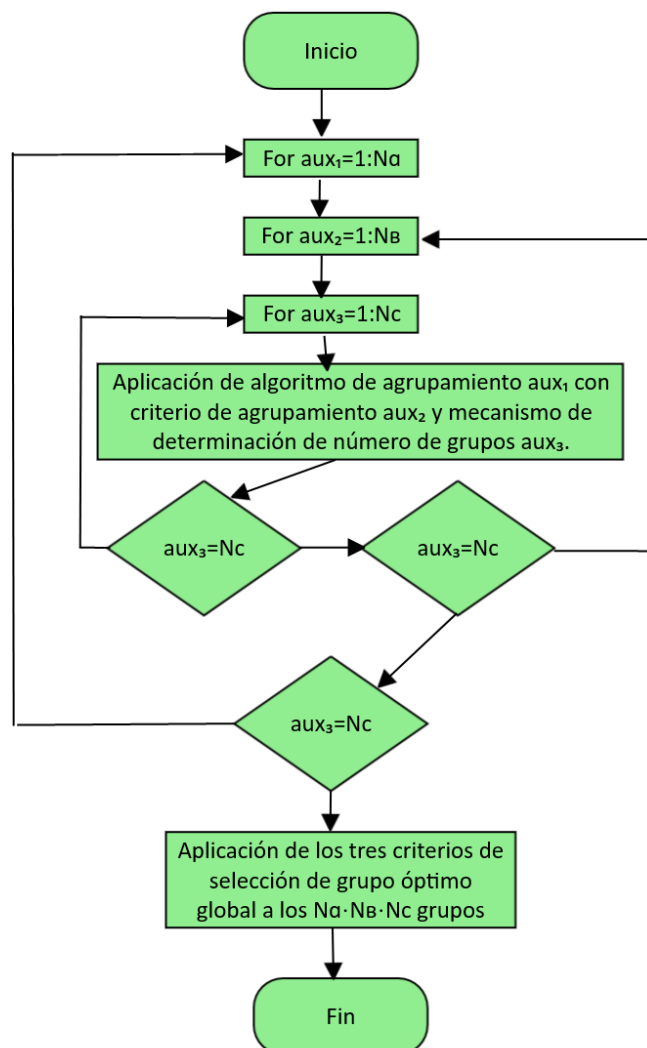


Figura 3. Diagrama de flujo de la etapa de SSR.

La selección del grupo óptimo depende del criterio de clasificación de las líneas, donde para los criterios unidimensionales se selecciona un único conjunto, mientras que en el caso de la combinación de dos criterios el grupo óptimo está conformado por la unión de los conjuntos óptimos de cada criterio, ecuación (24). Los criterios de selección de grupo óptimo para cada criterio unidimensional y bidimensional de clasificación de líneas se presentan en la Tabla 1, siendo determinados por el valor extremo normalizado de la magnitud del criterio.

$$G_{2d} = G_1 \cup G_2 \tag{24}$$

Tabla 1. Criterio de selección del grupo óptimo para criterios unidimensionales y bidimensionales.

Criterio de clasificación de líneas.	G_1	G_2
Costo-beneficio	Mínimo	-
Sobrecarga	Máximo	-
Mínimo esfuerzo	Máximo	-
Costo-beneficio y sobrecarga	Mínimo	Máximo
Sobrecarga y mínimo esfuerzo	Máximo	Máximo
Costo-beneficio y mínimo esfuerzo	Mínimo	Máximo

Finalmente, para reducir el espacio de búsqueda a un único conjunto de líneas candidatas y determinar un algoritmo de agrupamiento y parámetros de agrupamiento óptimos para cada sistema se selecciona un grupo óptimo global, los criterios de selección del grupo óptimo global se aplican secuencialmente y se presentan a continuación:

- El grupo óptimo global debe contener al grupo de líneas óptimas para la expansión del SEP, este conjunto es obtenido mediante la aplicación de la metaheurística al espacio de búsqueda completo. Es decir, debe cumplirse la ecuación (25).
- Se selecciona el grupo que está conformado por el menor número de líneas que se repitan por lo menos cinco veces dentro del grupo a lo largo de los 10 ensayos realizados. De esta forma se garantiza que el resultado sea repetible pese a la variabilidad inicial de los centroides o medoides.
- Por último, se determina como grupo óptimo al que tenga el menor tiempo total, siendo este dado por la ecuación (26).

$$G_{LOptimas} \in G_{\text{optimo}} \tag{25}$$

$$T_{T.agr} = T_{AC-OPF} + T_{mec} + T_{agr} \tag{26}$$

4.4.3. Etapa de determinación de líneas óptimas para TNEP

En cuanto a los parámetros de la metaheurística se utilizan los valores de número de iteraciones y valor máximo de la función objetivo recomendados en [13] que son dependientes del tamaño y escenario del sistema. No obstante, debido a la variabilidad de los resultados se analizan los valores promedio de 10 pruebas para todos los sistemas menos el IEEE de 300 barras, en dicho caso se realizan 5 pruebas debido a su tiempo de resolución elevado con el espacio de búsqueda completo. Para determinar el tamaño de la población para cada sistema se aplica a la ecuación (27), siendo n el número de líneas candidatas en el espacio reducido. En [13] se sugiere que el tamaño de la población se encuentre dentro del intervalo recomendado $[y, 1.25 \cdot y]$, sin embargo para el sistema IEEE de 118 barras fue necesario superarlo para asegurar un porcentaje de éxito adecuado. Finalmente, para determinar el tiempo de procesamiento total se empleó la ecuación (28), donde $T_{T.agr} = 0$ al no aplicar SSR.

$$y = 0.0003 \cdot n^2 + 0.3448 \cdot n + 25.252 \tag{27}$$

$$T_{Total} = T_{T.agr} + T_{\lambda} \tag{28}$$

4.4.4. Sistemas de prueba

Para el sistema Garver, IEEE 24 barras e IEEE 118 barras se plantea el escenario sin compensación en paralelo donde se optimiza únicamente el costo de las líneas que se añadirán al sistema. Sin embargo, el sistema IEEE de 118 barras fue modificado como en [13], incrementándose el costo de las líneas de 338kV, ya que su costo de inversión inicial era muy próximo a las de menor voltaje y se redujo la capacidad nominal de las líneas iniciales para generar congestión, ya que el sistema original no requiere de la adición de líneas. Mientras tanto en el sistema IEEE de 300 barras se consideró compensación de reactivos en paralelo con límites de -1000 y 1000 MVAR y un costo de 10k\$/MVAR. Este sistema fue modificado eliminando la red de distribución de 66kV, se fijó la carga nominal de las líneas entre 198 MVA y 300 MVA y se definió una demanda reactiva de 250 MVAR en las barras 118 y 218 para asegurar la convergencia de AC-OPF ya que los valores originales eran elevados.

5. Resultados y Discusión

5.1. Sistema Garver

5.1.1. Determinación del grupo óptimo global para el sistema Garver:

En este sistema se clasifican $N = 150$ datos, ya que en el espacio completo se tienen 15 líneas candidatas y se realizan 10 ensayos. De las pruebas realizadas 82 de los 114 grupos óptimos cumplen los dos primeros criterios para ser seleccionados como grupo óptimo global, los 10 primeros con el menor tiempo de resolución se presentan en la Tabla 2. Al aplicar el tercer criterio se determina que la combinación óptima de algoritmo, criterio de agrupamiento de líneas y mecanismo de determinación del número de grupos es *PAM*, criterio de mínimo esfuerzo y Calinski-Harabasz con máximo 6 grupos, respectivamente. Adicionalmente se observa que los 10 grupos óptimos más rápidos utilizan Calinski-Harabasz y el criterio de selección más repetido es el criterio de mínimo esfuerzo, repitiéndose 6 veces, seguido por los criterios bidimensionales costo-beneficio y mínimo esfuerzo, 3 veces sobrecarga y mínimo esfuerzo, 1 vez.

Tabla 2. Diez primeros grupos del sistema Garver de 6 barras con menor número de líneas candidatas y tiempo de resolución al aplicar los dos primeros criterios de selección del grupo óptimo global.

Algoritmo de agrupamiento	Criterio de determinación del número de grupos	Máximo número de grupos	Criterio de clasificación de líneas	Número de líneas	Tiempo (s)
<i>PAM</i>	Calinski-Harabasz	6	σ	3	2.63
<i>KMEANS</i>	Calinski-Harabasz	6	CB- σ	3	2.69
<i>KMEANS</i>	Calinski-Harabasz	6	SC- σ	3	2.7
<i>KMEANS</i>	Calinski-Harabasz	6	σ	3	2.70
<i>PAM</i>	Calinski-Harabasz	12	σ	3	2.71
<i>PAM</i>	Calinski-Harabasz	12	CB- σ	3	2.71
<i>KMEANS</i>	Calinski-Harabasz	12	CB- σ	3	2.73
<i>KMEANS</i>	Calinski-Harabasz	12	σ	3	2.73
<i>CLARA</i>	Calinski-Harabasz	6	σ	3	2.76
<i>CLARA</i>	Calinski-Harabasz	12	σ	3	2.98

5.1.2. Análisis del grupo óptimo global del sistema Garver:

Las líneas candidatas que conforman el grupo óptimo global y el número de veces que se repiten a lo largo de los ensayos se presenta en la Tabla 3. Se determina que el criterio de agrupamiento óptimo coincide con los resultados de [23], [25], [24], que indican que a la escala de la base de datos se recomienda PAM. En la Figura 4 se presentan los grupos y sus valores del coeficiente de menor esfuerzo donde el grupo 2 es el global óptimo.

Tabla 3. Líneas del grupo global óptimo del sistema Garver de 6 barras y número de veces que se repiten.

Línea		Número de veces que se repite
Desde (Barra)	Hacia (Barra)	
3	5	9
4	6	7
1	3	7

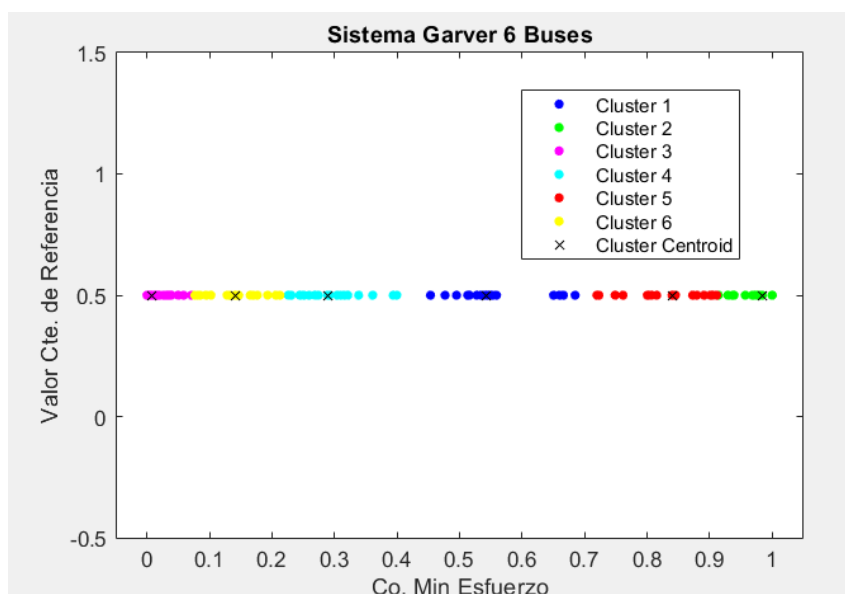


Figura 4. Resultados de agrupamiento con el criterio de Mínimo esfuerzo en el sistema Garver de 6 barras.

5.1.3. Comparación del rendimiento de la metaheurística con el espacio completo y reducido en el sistema Garver:

En la Tabla 4 se compara la aplicación de la metaheurística al espacio completo y al grupo óptimo global, donde se determina que la aplicación de SSR permite una reducción del espacio de búsqueda del 80% y del tiempo de procesamiento del 12.05%, manteniendo el costo de la topología óptima y el porcentaje de éxito.

Tabla 4. Tabla comparativa de resultados de TNEP con el espacio de búsqueda completo y reducido en el sistema Garver de 6 barras.

	Sin SSR	Con SSR	Diferencia (%)
Líneas óptimas		$l_{3-5} = 1$ $l_{4-6} = 1$ $l_{1-3} = 1$	-
Número de líneas candidatas	15	3	80
Costo de la topología óptima	160	160	0
Porcentaje de éxito	100	100	0
Tamaño población	30	27	10
Tiempo Total (s)	14	12.3131	12.05

5.2. Sistema IEEE de 24 barras

5.2.1. Determinación del grupo óptimo global para el sistema IEEE de 24 barras:

En este sistema se clasifican $N = 410$ datos, ya que en el espacio completo se tienen 41 líneas candidatas y se realizan 10 ensayos. De las 114 pruebas realizadas en este sistema 33 de los grupos cumplen los dos primeros criterios para ser el grupo óptimo global. En la Tabla 5 se presentan los 10 primeros con el menor espacio de búsqueda. Al aplicar el tercer criterio se establece que, el mecanismo Calinski-Harabasz con un máximo de 12 grupos, el criterio bidimensional costo-beneficio y mínimo esfuerzo y el algoritmo de agrupamiento PAM brindan el mejor resultado. Además de los 10 resultados 6 de ellos utilizan el algoritmo de agrupamiento óptimo y 8 de ellos el criterio bidimensional costo-beneficio y mínimo esfuerzo, demostrando la eficacia de dichos parámetros para este sistema.

Tabla 5. Diez primeros grupos del sistema IEEE de 24 con menor número de líneas candidatas y tiempo de resolución al aplicar los dos primeros criterios de selección del grupo óptimo global.

Algoritmo agrupamiento	Mecanismo número de grupos	Grupos máximos	Criterio de clasificación de líneas	Número de líneas	Tiempo (s)
PAM	Calinski-Harabasz	12	CB- σ	12	5.45
KMEANS	Calinski-Harabasz	12	CB- σ	13	4.90
CLARA	Gap	12	CB- σ	13	43.31
CLARA	Calinski-Harabasz	12	CB- σ	14	5.19
PAM	Gap	12	CB- σ	19	55.95
PAM	Gap	12	CB	20	57.92
PAM	Gap	6	CB- σ	21	17.31
CLARA	Calinski-Harabasz	6	CB- σ	22	4.90
PAM	Gap	6	SC- σ	22	17.13
PAM	Calinski-Harabasz	6	CB- σ	23	4.97

5.2.2. Análisis del grupo óptimo global del sistema IEEE de 24 barras:

Las líneas candidatas que conforman el grupo óptimo global y el número de veces que se repiten a lo largo de los ensayos se presenta en la Tabla 6. El criterio de agrupamiento óptimo coincide con los resultados de [23], [25], [24] donde para un grupo de 410 datos se

recomienda *PAM* o *Kmeans*. En la Figura 5 se presentan los resultados del agrupamiento empleando el criterio bidimensional costo-beneficio y mínimo esfuerzo, siendo el grupo óptimo global la unión de los grupos 1 y 6.

Tabla 6. Líneas del grupo global óptimo del sistema IEEE de 24 barras y número de veces que se repiten.

Línea			Número de veces que se repite
Desde (Barra)	Hacia (Barra)	Número de línea	
1	1	1	10
16	19	28	9
2	4	4	7
6	10	10	7
7	8	11	7
12	13	20	7
1	5	3	6
12	23	21	6
15	16	24	6
13	14	38	6
14	16	23	5
15	21	25	5

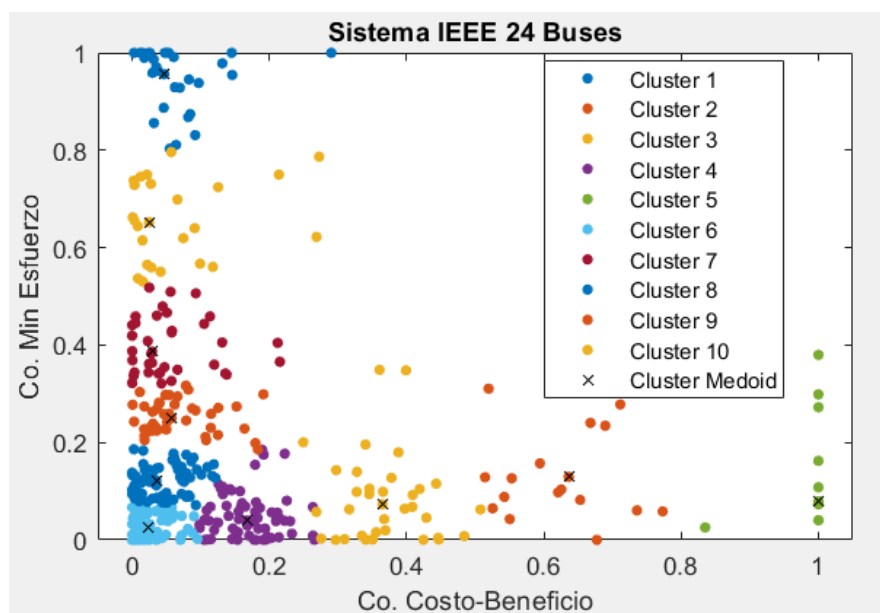


Figura 5. Resultados de agrupamiento con el criterio bidimensional de Costo beneficio - Mínimo esfuerzo en el sistema IEEE de 24 barras.

5.2.3. Comparación del rendimiento de la metaheurística con el espacio completo y reducido en el sistema IEEE de 24 barras:

En la Tabla 7 se compara el resultado de aplicar la metaheurística al espacio completo y al espacio reducido, al comparar los resultados con los obtenidos en el sistema Garver se determina que debido al mayor número de líneas candidatas se tiene una mayor reducción del espacio de búsqueda y tiempo de resolución.

Tabla 7. Tabla comparativa de resultados de TNEP con el espacio de búsqueda completo y reducido en el sistema IEEE de 24 barras.

	Sin SSR	Con SSR	Diferencia (%)
Líneas óptimas	$l_{6-10}=1$ $l_{7-8}=1$ $l_{14-16}=1$		-
Número de líneas candidatas	41	12	70.73
Costo de la topología óptima	86	86	0
Porcentaje de éxito	100	100	0
Tamaño población	50	36	28
Tiempo Total (s)	2700	31.03	98.85

5.3. Sistema IEEE modificado de 118 barras

5.3.1. Determinación del grupo óptimo global para el sistema IEEE de 118 barras:

En este sistema se clasifican $N = 1860$ datos, ya que en el espacio completo se tienen 186 líneas candidatas y se realizan 10 ensayos. De las 78 pruebas realizadas 23 cumplen con los dos criterios iniciales para ser seleccionados como grupo óptimo global. Los diez primeros grupos óptimos con menor número de líneas se presentan en la Tabla 8, donde se establece que el criterio de clasificación de líneas más repetido es el de sobrecarga con 4 grupos, seguido por el criterio de costo-beneficio y el criterio bidimensional sobrecarga y costo-beneficio, ambos con 3 grupos. El algoritmo de agrupamiento más repetido es *Kmeans* con 6 grupos y el mecanismo el método del coeficiente de silueta con 7 grupos. Al aplicar el tercer criterio se determina que el criterio bidimensional sobrecarga y costo-beneficio, el algoritmo *Kmeans* y el método del coeficiente de silueta con máximo 6 grupos en conjunto brindan el grupo óptimo global.

Tabla 8. Diez primeros grupos del sistema IEEE de 118 barras con menor número de líneas candidatas y tiempo de resolución al aplicar los dos primeros criterios de selección del grupo óptimo global.

Algoritmo agrupamiento	Mecanismo número de grupos	Grupos máximos	Criterio de clasificación de líneas	Número de líneas	Tiempo (s)
<i>KMEANS</i>	Silueta	6	CB-SC	31	3.06
<i>KMEANS</i>	Silueta	12	SC	31	3.32
<i>KMEANS</i>	Silueta	12	CB-SC	31	3.43
<i>KMEANS</i>	Silueta	6	SC	32	3.02
<i>CLARA</i>	Silueta	6	CB-SC	44	3.28
<i>CLARA</i>	Silueta	12	SC	53	3.77
<i>CLARA</i>	Silueta	6	SC	59	3.30
<i>KMEANS</i>	Calinski-Harabasz	12	CB	65	3.15
<i>CLARA</i>	Calinski-Harabasz	6	CB	69	3.64
<i>KMEANS</i>	Calinski-Harabasz	6	CB	97	3.15

5.3.2. Análisis del grupo óptimo global del sistema IEEE de 118 barras:

Las líneas candidatas que conforman el grupo óptimo global y el número de veces que se repiten a lo largo de los ensayos se presenta en el Tabla 14. De donde se determina que el criterio de agrupamiento óptimo coincide con los resultados de [23], [25], [24], de acuerdo con el número N de datos clasificados, ya que se indica que *Kmeans* puede aplicarse a conjuntos de datos de gran y pequeña escala. Mientras que en la Figura 6 se presentan los grupos obtenidos con la combinación de mecanismo, algoritmo de agrupamiento y criterio óptimos, siendo el grupo 1 el óptimo global.

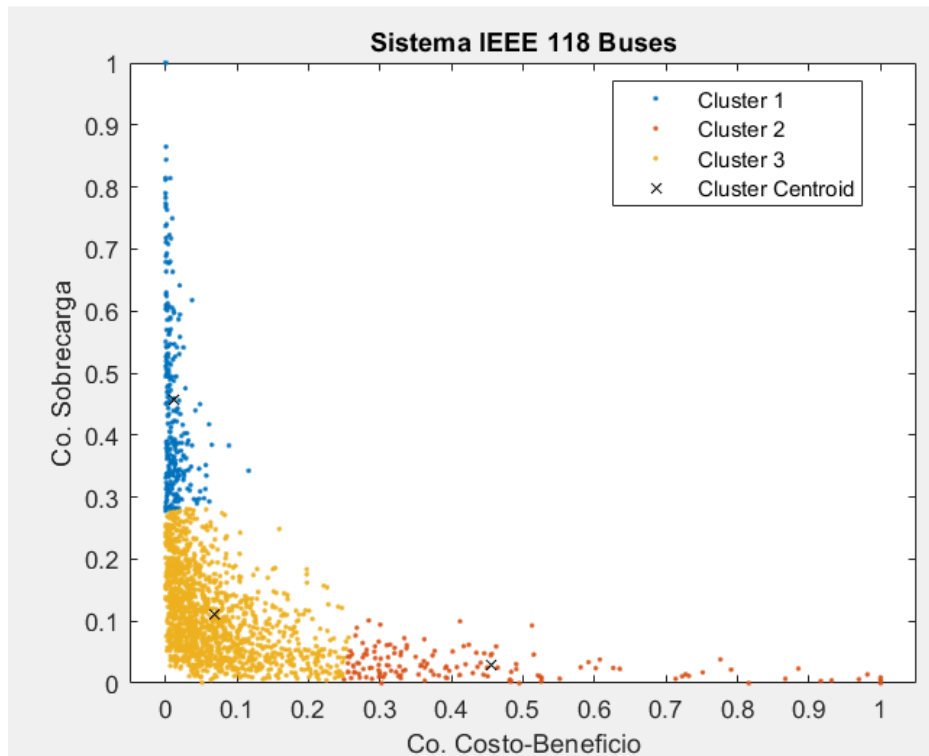


Figura 6. Resultados de agrupamiento con el criterio de Mínimo esfuerzo en el sistema IEEE de 118 barras.

5.3.3. Comparación del rendimiento de la metaheurística con el espacio completo y reducido en el sistema IEEE de 118 barras:

En la Tabla 9 se compara el resultado de aplicar la metaheurística al espacio completo y al grupo óptimo global, siendo la reducción de espacio de búsqueda mayor que en el sistema IEEE de 24 barras y manteniendo el costo de la topología y porcentaje de éxito del espacio completo.

Tabla 9. Tabla comparativa de resultados de TNEP con el espacio de búsqueda completo y reducido en el sistema IEEE de 118 barras.

	Sin SSR	Con SSR	Diferencia (%)
Líneas óptimas	$l_{8-9} = 1, l_{8-5} = 1$ $l_{9-10} = 1, l_{15-17} = 1$ $l_{23-25} = 1, l_{25-27} = 1$ $l_{94-100} = 1, l_{17-113} = 1$		-
Número de líneas candidatas	186	31	83.33
Costo de la topología óptima	181.6	181.6	0
Porcentaje de éxito	100	100	0
Tamaño población	100	60	40
Tiempo Total (s)	780	164.49	78.91

5.4. Sistema IEEE modificado de 300 barras

5.4.1. Determinación del grupo óptimo global para el sistema IEEE de 300 barras:

En este sistema se clasifican $N = 3750$ datos, ya que en el espacio completo se tienen 375 líneas candidatas y se realizan 10 ensayos. De las 78 pruebas realizadas tan solo 7 cumplen los dos criterios iniciales de selección del grupo óptimo global, es decir tan solo el 9% de los grupos pueden ser seleccionados. Estos grupos son presentados en la Tabla 10 siendo los más repetidos el criterio de Costo beneficio con 6 grupos, el mecanismo de Calinski-Harabasz con 4 grupos y el algoritmo CLARA con 5 grupos. Al aplicar el tercer criterio la unión de los parámetros más repetidos coincide con las características del grupo óptimo global, considerando 12 grupos como valor máximo del mecanismo de selección del número de grupos óptimo.

Tabla 10. Diez primeros grupos del sistema IEEE de 300 barras con menor número de líneas candidatas y tiempo de resolución al aplicar los dos primeros criterios de selección del grupo óptimo global.

Algoritmo agrupamiento	Mecanismo número de grupos	Grupos máximos	Criterio de clasificación de líneas	Número de líneas	Tiempo (s)
CLARA	Calinski-Harabasz	12	CB	89	3.86
CLARA	Gap	12	CB- σ	91	50.67
KMEANS	Calinski-Harabasz	12	CB	148	3.02
CLARA	Silueta	6	CB	161	3.81
CLARA	Silueta	12	CB	161	4.38
CLARA	Calinski-Harabasz	6	CB	200	3.87
KMEANS	Calinski-Harabasz	6	CB	249	3.11

5.4.2. Análisis del grupo óptimo global del sistema IEEE de 300 barras:

Las líneas candidatas que conforman el grupo óptimo global y el número de veces que se repiten a lo largo de los ensayos se presenta en el Tabla 15. De donde se determina que el criterio de agrupamiento óptimo coincide con los resultados de [23], [25], [24] donde se determina que CLARA es idóneo para el agrupamiento de un gran número de datos. Mientras

que en la Figura 7 se presenta el resultado de agrupamiento de los parámetros óptimos para el sistema, donde el grupo 6 es el óptimo global.

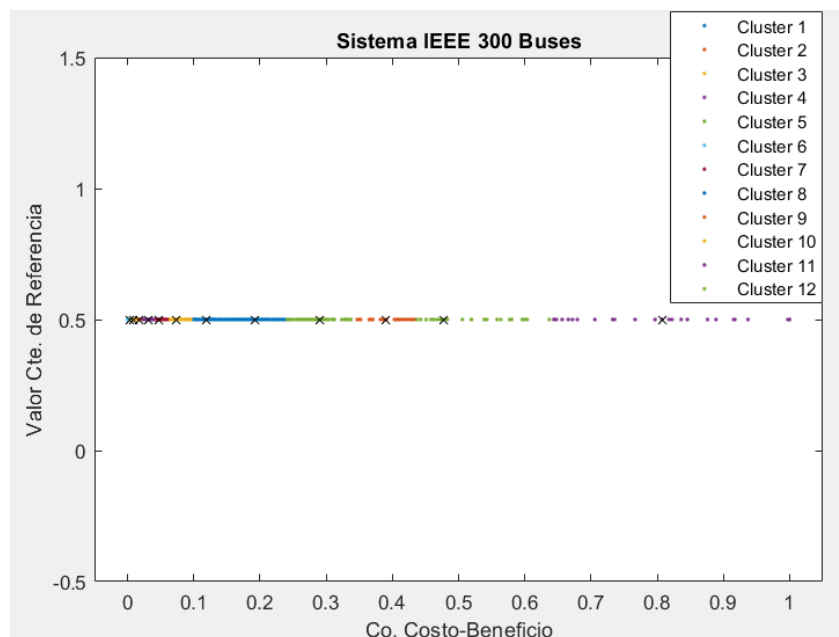


Figura 7. Resultados de agrupamiento con el criterio de Mínimo esfuerzo en el sistema IEEE de 300 barras.

5.4.3. Comparación del rendimiento de la metaheurística con el espacio completo y reducido en el sistema IEEE de 300 barras:

En la Tabla 11 se compara el resultado de aplicar la metaheurística al espacio completo y al grupo óptimo global. Donde se obtiene una reducción de tiempo porcentualmente mayor que en el sistema IEEE de 118 barras modificado y se mantiene el porcentaje de éxito y costo de la topología del espacio de búsqueda completo.

Tabla 11. Tabla comparativa de resultados de TNEP con el espacio de búsqueda completo y reducido en el sistema IEEE de 300 barras.

	Sin SSR	Con SSR	Diferencia (%)
Líneas óptimas	$l_{119-120} = 1$ $l_{119-121} = 1$ $l_{133-171} = 1$ $l_{221-223} = 1$		-
Número de líneas candidatas	375	89	77.87
Costo de la topología óptima	42.7	42.7	0
Porcentaje de éxito	100	100	0
Tamaño población	200	60	70
Tiempo Total (s)	6840	289.96	95.76

5.5. Comparación de resultados de SSR

Finalmente, en la Tabla 12 se presentan el intervalo de población recomendado y el seleccionado para los sistemas con espacio reducido. En la Tabla 13 se presentan los parámetros del grupo óptimo global, el porcentaje de grupos óptimos que cumplen los dos primeros criterios para ser seleccionados como grupo óptimo global con respecto al número

de pruebas realizadas, el número N de datos clasificados, el porcentaje de reducción del espacio de búsqueda y del tiempo de resolución en cada sistema. De donde se determina que la aplicación de SSR con los criterios, algoritmos y mecanismos postulados brinda excelentes resultados en la reducción de tiempo de procesamiento computacional para el problema de TNEP.

Tabla 12. Comparación del tamaño de población utilizado en la metaheurística para los cuatro sistemas.

Sistema	Garver 6 barras	IEEE de 24 barras	IEEE de 118 barras	IEEE de 300 barras
# líneas candidatas	3	12	31	89
Tamaño de población (utilizado)	27	36	60	60
Tamaño de población (recomendado)	26-33	29-37	36-45	58-73

Tabla 13. Tabla de resultados de SSR en los cuatro sistemas.

Sistema	Garver 6 barras	IEEE de 24 barras	IEEE de 118 barras	IEEE de 300 barras
Algoritmo de agrupamiento	<i>PAM</i>	<i>PAM</i>	<i>Kmeans</i>	<i>CLARA</i>
Criterio de clasificación de líneas	Mínimo Esfuerzo	Costo Beneficio - Mínimo Esfuerzo	Costo Beneficio - Sobrecarga	Costo Beneficio
Mecanismo de determinación de # grupos	Calinski-Harabasz	Calinski-Harabasz	Silueta	Calinski-Harabasz
# Máximo de grupos	6	12	6	12
Número de datos N	150	410	1860	3750
Reducción de espacio de búsqueda (%)	80	70.73	83.33	77.87
Reducción de tiempo de procesamiento con SSR (%)	12.05	98.85	78.91	95.76
Grupos que cumplen los dos primeros criterios globales (%)	71.93	28.95	29.49	8.97

6. Conclusiones y Recomendaciones

A partir de los resultados obtenidos se determina que la aplicación de SSR en el modelo de TNEP planteado brinda excelentes resultados, reduciendo el tiempo de procesamiento en todos los sistemas analizados con respecto al espacio de búsqueda completo. Aun así, se determina que la metodología brinda mejores resultados en los sistemas de mayor tamaño donde la reducción de tiempo de procesamiento mínima ocurre en el sistema IEEE de 118 barras modificado, siendo del 78.91%, mientras que la máxima ocurre en el sistema IEEE de 24 barras siendo del 98.85%. En el sistema de mayor escala, IEEE de 300 barras modificado, la reducción temporal es del 95.67%. Sin embargo, en el sistema Garver de 6 barras la reducción del tiempo de procesamiento es tan solo del 12.05% pese a presentar una reducción del espacio de búsqueda del 80%, esto se debe a que el espacio inicial cuenta con un número pequeño de líneas candidatas.

A partir del análisis se determina que la combinación óptima de algoritmo de agrupamiento, criterio de selección de líneas y mecanismo de determinación de número de grupos para sistemas de pequeña escala, como el sistema Garver de 6 barras, es *PAM*, criterio de mínimo esfuerzo y Calinski-Harabasz; para sistemas de mediana escala, como el sistema IEEE de 24 barras es *PAM*, el criterio bidimensional costo-beneficio y sobrecarga y Calinski-Harabasz; mientras que para sistemas de gran escala como el sistema IEEE de 118 barras, se

recomienda *Kmeans*, el criterio bidimensional costo-beneficio y sobrecarga y el mecanismo del coeficiente de silueta; finalmente, para sistemas de gran escala como el sistema IEEE de 300 barras es *CLARA*, el criterio costo-beneficio y Calinski-Harabasz. Adicionalmente, se determinó que los criterios de agrupamiento óptimos coinciden con los recomendados en la literatura que dependen del número de datos a clasificar, sustentando así la validez del planteamiento.

Para futuros trabajos se recomienda la implementación de algoritmos como [23], [25], [32] ya que brindan mejores resultados de agrupamiento, sin embargo se desconoce su tiempo computacional y la complejidad de su implementación es alta. También se sugiere aplicar la metodología planteada a sistemas donde se consideren contingencias y el despacho de energías renovables, para evaluar su validez en dichos escenarios.

Referencias

- [1] H. W. Beaty y D. G. Fink, *Standard handbook for electrical engineers*, 14. ed. London: McGraw-Hill, 2000.
- [2] R. A. R. Lázaro y G. Rendón, "PLANEAMIENTO DE LA EXPANSIÓN DE SISTEMAS DE TRANSMISIÓN CONSIDERANDO INCERTIDUMBRE EN LA DEMANDA Y LA GENERACION", *Sci. Tech. Año XIV*, núm. 38, pp. 37–41, 2008.
- [3] I. M. D. Mendonça, I. C. Silva Junior, B. H. Dias, y A. L. M. Marcato, "Identification of relevant routes for static expansion planning of electric power transmission systems", *Electr. Power Syst. Res.*, vol. 140, pp. 769–775, nov. 2016, doi: 10.1016/j.epsr.2016.05.011.
- [4] N. Gideon Ude, H. Yskandar, y R. Coneth Graham, "A Comprehensive State-of-the-Art Survey on the Transmission Network Expansion Planning Optimization Algorithms", *IEEE Access*, vol. 7, pp. 123158–123181, 2019, doi: 10.1109/ACCESS.2019.2936682.
- [5] S. Lumbreras y A. Ramos, "The new challenges to transmission expansion planning. Survey of recent practice and literature review", *Electr. Power Syst. Res.*, vol. 134, pp. 19–29, may 2016, doi: 10.1016/j.epsr.2015.10.013.
- [6] S. L. Gbadamosi y N. I. Nwulu, "A comparative analysis of generation and transmission expansion planning models for power loss minimization", *Sustain. Energy Grids Netw.*, vol. 26, p. 100456, jun. 2021, doi: 10.1016/j.segan.2021.100456.
- [7] J. D. Glover, T. J. Overbye, A. B. Birchfield, y M. S. Sarma, *Power system analysis & design*, Seventh edition. Australia ; Boston [MA]: Cengage, 2023.
- [8] R.-C. Leou, "A multi-year transmission planning under a deregulated market", *Int. J. Electr. Power Energy Syst.*, vol. 33, núm. 3, pp. 708–714, mar. 2011, doi: 10.1016/j.ijepes.2010.11.020.
- [9] M. Torres R., C. A. Castro, y M. J. Rider, "A Search Space Reduction Strategy within the MILP Branch Flow Model for Concurrent AC Multistage Transmission Expansion and Reactive Power Planning", en *2018 North American Power Symposium (NAPS)*, Fargo, ND: IEEE, 2018, pp. 1–6. doi: 10.1109/NAPS.2018.8600545.
- [10] W. Minchala, "Reducción de espacio de búsqueda usando algoritmos de aprendizaje no supervisado aplicado al problema de la expansión del sistema de transmisión de energía eléctrica", Universidad de Cuenca, Cuenca, 2023.
- [11] C. Li, A. J. Conejo, P. Liu, B. P. Omell, J. D. Sirola, y I. E. Grossmann, "Mixed-integer linear programming models and algorithms for generation and transmission expansion planning of power systems", *Eur. J. Oper. Res.*, vol. 297, núm. 3, pp. 1071–1082, mar. 2022, doi: 10.1016/j.ejor.2021.06.024.
- [12] A. H. Domínguez, A. H. Escobar, y R. A. Gallego, "An MILP model for the static transmission expansion planning problem including HVAC/HVDC links, security constraints and power losses with a reduced search space", *Electr. Power Syst. Res.*, vol. 143, pp. 611–623, feb. 2017, doi: 10.1016/j.epsr.2016.10.055.
- [13] E. G. Morquecho *et al.*, "Comparison of an Improved Metaheuristic and Mathematical Optimization Based Methods to Solve the Static AC TNEP Problem", *IEEE Trans. Power Syst.*, vol. 39, núm. 2, pp. 3240–3256, mar. 2024, doi: 10.1109/TPWRS.2023.3305431.
- [14] W. Yao, J. Zhao, F. Wen, Y. Xue, y G. Ledwich, "A Hierarchical Decomposition Approach for Coordinated Dispatch of Plug-in Electric Vehicles", *IEEE Trans. Power Syst.*, vol. 28, núm. 3, pp. 2768–2778, ago. 2013, doi: 10.1109/TPWRS.2013.2256937.
- [15] M. Jenabi, S. M. T. Fatemi Ghomi, y Y. Smeers, "Bi-Level Game Approaches for Coordination of Generation and Transmission Expansion Planning Within a Market Environment", *IEEE Trans. Power Syst.*, vol. 28, núm. 3, pp. 2639–2650, ago. 2013, doi: 10.1109/TPWRS.2012.2236110.
- [16] Q. Zheng, B.-W. Feng, H.-C. Chang, y Z.-Y. Liu, "Dynamic space reduction optimization framework and its application in hull form optimization", *Appl. Ocean Res.*, vol. 114, p. 102812, sep. 2021, doi: 10.1016/j.apor.2021.102812.

- [17] I. Aljarah, H. Faris, y S. Mirjalili, Eds., *Evolutionary Data Clustering: Algorithms and Applications*. en *Algorithms for Intelligent Systems*. Singapore: Springer Singapore, 2021. doi: 10.1007/978-981-33-4191-3.
- [18] A. S. S. M. Barkat Ullah, R. Sarker, y D. Cornforth, “Search space reduction technique for constrained optimization with tiny feasible space”, en *Proceedings of the 10th annual conference on Genetic and evolutionary computation*, Atlanta GA USA: ACM, jul. 2008, pp. 881–888. doi: 10.1145/1389095.1389268.
- [19] E. Hossain, *Machine Learning Crash Course for Engineers-Springer*, Springer. Boise, 2024.
- [20] C. Reddy, *Data Clustering Algorithms and Applications*. Minneapolis: Chapman & Hall, 2021.
- [21] D. S. Stock, Y. Harms, D. Mende, y L. Hofmann, “Robust nonlinear mathematical transmission expansion planning based on German electricity market simulation”, *Electr. Power Syst. Res.*, vol. 189, p. 106685, dic. 2020, doi: 10.1016/j.epsr.2020.106685.
- [22] S. Lloyd, “Least squares quantization in PCM”, *IEEE Trans. Inf. Theory*, vol. 28, núm. 2, pp. 129–137, mar. 1982, doi: 10.1109/TIT.1982.1056489.
- [23] H. Drias, A. Kechid, y N. Fodil-Cherif, “A hybrid clustering algorithm and web information foraging”, *Int. J. Hybrid Intell. Syst.*, vol. 13, núm. 3–4, pp. 137–149, feb. 2017, doi: 10.3233/HIS-160231.
- [24] P. O. Olukanmi, F. Nelwamondo, y T. Marwala, “Performance evaluation of sampling-based large-scale clustering algorithms”, en *2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA)*, Bloemfontein, South Africa: IEEE, ene. 2019, pp. 194–199. doi: 10.1109/RoboMech.2019.8704854.
- [25] E. Schubert y P. J. Rousseeuw, “Faster k-Medoids Clustering: Improving the PAM, CLARA, and CLARANS Algorithms”, 2018, doi: 10.48550/ARXIV.1810.05691.
- [26] R. D. Zimmerman, C. E. Murillo-Sánchez, y R. J. Thomas, “MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education”, *IEEE Trans. Power Syst.*, vol. 26, núm. 1, pp. 12–19, feb. 2011, doi: 10.1109/TPWRS.2010.2051168.
- [27] C. C. Aggarwal, J. L. Wolf, P. S. Yu, C. Procopiuc, y J. S. Park, “Fast algorithms for projected clustering”, *ACM SIGMOD Rec.*, vol. 28, núm. 2, pp. 61–72, jun. 1999, doi: 10.1145/304181.304188.
- [28] R. Agrawal, J. Gehrke, D. Gunopulos, y P. Raghavan, “Automatic subspace clustering of high dimensional data for data mining applications”, *ACM SIGMOD Rec.*, vol. 27, núm. 2, pp. 94–105, jun. 1998, doi: 10.1145/276305.276314.
- [29] N. E. Matute, S. P. Torres, E. G. Morquecho, F. Astudillo-Salinas, J. C. Lopez, y W. C. Flores, “Improving the AC Transmission Expansion Planning by Using Initial Solutions Algorithms”, en *2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, The Hague, Netherlands: IEEE, oct. 2020, pp. 494–498. doi: 10.1109/ISGT-Europe47291.2020.9248778.
- [30] A. Monticelli, A. Santos, M. V. F. Pereira, S. H. Cunha, B. J. Parker, y J. C. G. Praca, “Interactive Transmission Network Planning Using a Least-Effort Criterion”, *IEEE Trans. Power Appar. Syst.*, vol. PAS-101, núm. 10, pp. 3919–3925, oct. 1982, doi: 10.1109/TPAS.1982.317043.
- [31] S. S. Jacob y R. Vijayakumar, “Modern Techniques used for Big Data Clustering: A Review”, 6, vol. 7, pp. 1–5, 2018.
- [32] K. P. Sinaga y M.-S. Yang, “Unsupervised K-Means Clustering Algorithm”, *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
- [33] B. Mirkin y B. Mirkin, *Clustering for Data Mining: A Data Recovery Approach*, 0 ed. Chapman and Hall/CRC, 2005. doi: 10.1201/9781420034912.

- [34] E. Herman, K.-E. Zsido, y V. Fenyves, “Cluster Analysis with K-Mean versus K-Medoid in Financial Performance Evaluation”, *Appl. Sci.*, vol. 12, núm. 16, Art. núm. 16, ene. 2022, doi: 10.3390/app12167985.
- [35] J. E. Gentle, L. Kaufman, y P. J. Rousseeuw, “Finding Groups in Data: An Introduction to Cluster Analysis.”, *Biometrics*, vol. 47, núm. 2, p. 788, jun. 1991, doi: 10.2307/2532178.
- [36] L. Kaufman y P. J. Rousseeuw, “CLUSTERING LARGE DATA SETS”, en *Pattern Recognition in Practice*, Elsevier, 1986, pp. 425–437. doi: 10.1016/B978-0-444-87877-9.50039-X.
- [37] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, y B. D. Satoto, “Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster”, *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 336, núm. 1, p. 012017, abr. 2018, doi: 10.1088/1757-899X/336/1/012017.
- [38] R. Nainggolan, R. Perangin-angin, E. Simarmata, y A. F. Tarigan, “Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method”, *J. Phys. Conf. Ser.*, vol. 1361, núm. 1, p. 012015, nov. 2019, doi: 10.1088/1742-6596/1361/1/012015.
- [39] D. Marutho, S. Hendra Handaka, E. Wijaya, y Muljono, “The Determination of Cluster Number at k-Mean Using Elbow Method and Purity Evaluation on Headline News”, en *2018 International Seminar on Application for Technology of Information and Communication*, Semarang: IEEE, sep. 2018, pp. 533–538. doi: 10.1109/ISEMANTIC.2018.8549751.
- [40] Z. Syahputri, S. Sutarman, y M. A. P. Siregar, “Determining The Optimal Number of K-Means Clusters Using The Calinski Harabasz Index and Krzanowski and Lai Index Methods for Grouping Flood Prone Areas In North Sumatra”, *Sinkron*, vol. 9, núm. 1, pp. 571–580, ene. 2024, doi: 10.33395/sinkron.v9i1.13246.
- [41] T. Calinski y J. Harabasz, “A dendrite method for cluster analysis”, *Commun. Stat. - Theory Methods*, vol. 3, núm. 1, pp. 1–27, 1974, doi: 10.1080/03610927408827101.
- [42] D. Arthur y S. Vassilvitskii, “k-means++: the advantages of careful seeding”, presentado en ACM-SIAM Symposium on Discrete Algorithms, New Orleans, ene. 2007, pp. 1027–1035. Consultado: el 28 de mayo de 2025. [En línea]. Disponible en: <https://www.semanticscholar.org/paper/k-means%2B%2B%3A-the-advantages-of-careful-seeding-Arthur-Vassilvitskii/ab25e57c716c34c02fe8f78738ffbd44fe7732fa#citing-papers>

Anexos

Tabla 14. Líneas del grupo global óptimo del sistema IEEE de 118 barras y número de veces que se repiten.

Línea			Número de veces que se repite
Desde (Barra)	Hacia (Barra)	Número de línea	
8	9	7	10
68	69	107	10
8	5	8	9
26	30	38	9
65	68	104	9
9	10	9	8
38	37	51	8
30	38	54	8
69	77	119	8
23	24	30	7
63	59	93	7
63	64	94	7
64	65	97	7
77	78	121	7
86	87	134	7
80	99	153	7
94	100	155	7
99	100	159	7
5	6	5	6
23	25	31	6
30	17	36	6
8	30	37	6
23	32	41	6
38	65	96	6
82	83	129	6
85	86	133	6
17	113	178	6
3	5	4	5
15	17	21	5
25	27	33	5
77	82	128	5

Tabla 15. Líneas del grupo global óptimo del sistema IEEE de 300 barras y número de veces que se repiten.

Línea			Número de veces que se repite
Desde (Barra)	Hacia (Barra)	Número de línea	
37	9001	1	10
2	8	5	10
5	9	10	10
8	14	14	10
14	15	19	10
41	42	54	10
44	47	61	10
70	71	85	10
78	84	101	10
117	118	140	10
118	119	141	10
119	120	145	10
119	121	146	10
125	126	151	10
133	137	169	10
133	171	172	10
181	187	223	10
186	188	225	10
190	231	230	10
213	214	257	10

221	223	269	10
231	232	280	10
241	237	286	10
3	4	301	10
16	15	307	10
21	20	308	10
45	44	311	10
45	46	312	10
63	64	314	10
73	74	315	10
122	157	323	10
145	180	331	10
159	117	335	10
182	139	339	10
229	230	351	10
234	236	352	10
7002	2	357	10
7003	3	358	10
7061	61	359	10
7062	62	360	10
7166	166	361	10
7024	24	362	10
7001	1	363	10
7130	130	364	10
7011	11	365	10
7023	23	366	10
7139	139	368	10
7012	12	369	10
7017	17	370	10
7039	39	371	10
7057	57	372	10
7071	71	375	10
226	231	274	9
238	239	353	9
2	6	4	8
33	34	35	8
35	77	42	8
77	78	96	8
137	186	180	8
181	138	222	8
217	219	263	8
227	231	275	8
231	237	281	8
132	170	326	8
142	175	328	8
151	170	332	8
155	156	334	8
7049	49	367	8
4	16	9	7
137	188	181	7
162	164	206	7
162	165	207	7
188	138	227	7
220	238	268	7
85	99	317	7
223	224	350	7
1	5	3	6
12	21	17	6
187	188	226	6
224	225	271	6
120	1200	356	6
126	132	154	5
165	166	209	5

190	240	231	5
191	192	232	5
36	35	310	5
62	61	313	5
141	174	327	5
119	1190	355	5