



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

**CLUSTERING APLICADO A CONDICIONES DE OPERACIÓN EN DISEÑO
DE ESQUEMAS DE PROTECCIONES ESPECIALES EN SISTEMAS
ELECTRICOS DE POTENCIA**

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL ELÉCTRICO

GERARDO ANDRÉS VARAS SANTANA

PROFESOR GUÍA:
Rigoberto Torres Ávila

MIEMBROS DE LA COMISIÓN:
Ignacio Ñancupil Quirilao
Ricardo Alvarez Malebrán
Francisco Rivera Serrano

SANTIAGO DE CHILE
2024

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGENIERO
CIVIL ELÉCTRICO
POR: GERARDO ANDRÉS VARAS SANTANA
FECHA: 2024
PROF. GUÍA: RIGOBERTO TORRES ÁVILA

CLUSTERING APLICADO A CONDICIONES DE OPERACIÓN EN DISEÑO DE ESQUEMAS DE PROTECCIONES ESPECIALES EN SISTEMAS ELECTRICOS DE POTENCIA

El diseño de Esquemas Especiales de Protección (SPS) es una herramienta para aumentar la capacidad de transmisión de las líneas de un sistema eléctrico de potencia.

Esta memoria propone una implementación que utiliza algoritmos de agrupación para seleccionar puntos de operación representativos. Este enfoque reduce el número de condiciones necesarias para el diseño de Esquemas Especiales de Protección (SPS) en sistemas eléctricos, facilitando la búsqueda de soluciones óptimas y mejorando los tiempos de ejecución.

El objetivo es disminuir la dimensión del problema de optimización mediante técnicas de clusterización, optimizando el costo de tiempo computacional comprometiendo al mínimo la calidad de la solución. Al reducir las condiciones de operación, se busca lograr un equilibrio entre eficiencia y efectividad en el diseño de SPS.

De acuerdo a lo investigado a la fecha, existen trabajos que aplican el clustering a sistemas electricos de potencia para agrupar condiciones de operación similares. Más específico, hay investigaciones relacionadas directamente con la expansión de la capacidad de transmisión. No obstante no hay estudios que se apliquen directamente al sistema electrico chileno en el diseño de SPS. Para medir el desempeño del clustering se aplica las metricas del diseño del SPS en tres términos: costo, tiempo de cálculo, y condiciones no factibles.

Se espera que a mayor tiempo de cálculo menor condiciones no factibles lo cual se obtiene por medio de los gráficos de métricas de resultados. En este trabajo, con el fin de implementar la agrupación se recibe la información de las condiciones de operación de interés proveniente de la herramienta de diseño de SPS. Se le da un formato sobre el cual se calcula la disimilitud en función una dimension de interes que puede ser la generación, transmisión, demanda, costo marginal por barras. Luego se aplica un algoritmo de agrupación que puede ser Jerárquico o K-medoide , y finalmente se da el formato de salida de los datos que puede ingresarse a la herramienta de diseño de SPS con clustering. Se manipula la información mediante el lenguaje Python donde se implementa una interfaz para facilitar su uso. Se lograron los resultados esperados: agrupar condiciones de operación para diseñar SPS alejandose lo mínimo de la solución completa. Asi se muestra en los resultados que para casos reducidos con alrededor del treinta porciento de los datos se logran soluciones que cubren el problema completo. Mientras que para casos con orden de 500 condiciones de operación se llega a menos de 5 porciento de condiciones no factibles.

*A mis padres, por ser mi guía y sostén incondicional
en cada paso de este camino.*

Agradecimientos

Quiero expresar mi profundo agradecimiento a mis profesores guías, Rigoberto Torres Ávila e Ignacio Ñancupil Quirilao , cuya orientación experta, paciencia y apoyo inquebrantable fueron fundamentales en el desarrollo de este trabajo. Sus valiosas enseñanzas y dedicación han sido una fuente constante de inspiración y aprendizaje durante todo este proceso. También quiero agradecer a mi familia por su amor incondicional, comprensión y constante apoyo. Sus palabras de aliento y sacrificios han sido el pilar sobre el cual he construido este logro. A mis amigos y seres queridos, gracias por estar siempre ahí, brindándome momentos de distracción y alegría que han sido fundamentales para mantener el equilibrio durante este desafiante período. Este trabajo no habría sido posible sin la contribución y el apoyo de todos ustedes. A cada uno, mi más sincero agradecimiento.

Tabla de Contenido

1. Introducción	1
1.1. Contexto	1
1.2. Objetivos	2
1.2.1. Objetivo General	2
1.2.2. Objetivos Específicos	2
2. Antecedentes	3
2.1. Diseño SPS	3
2.2. Clustering	4
2.2.1. Métodos de Clustering	5
2.2.2. Disimilitud	6
2.2.3. Algoritmo K-means	7
2.2.4. Algoritmo de Partición alrededor de Medoides(PAM)	8
2.2.5. Algoritmo de agrupamiento de Sustracción(SC)	8
2.2.6. Algoritmo Jerárquico	10
2.2.6.1. Algoritmo Aglomerativo	10
2.2.7. Algoritmo de Agrupamiento Espacial Basado en Densidad de Aplicaciones con Ruido (DBSCAN)	11
2.3. Aplicaciones de Clustering en sistemas eléctricos de potencia	12
2.4. Librería PyClustering	15
3. Clustering aplicado a condiciones de operación en diseño de SPS	17
3.1. Obtención Dataset	18
3.2. Disimilitud	18
3.3. Aplicación Clustering	18
3.4. Formato para herramienta de SPS	19
3.5. Diseño y evaluación de SPS para los puntos representativos obtenidos en la agrupación	19
4. Casos de estudio	20
4.1. Caso 1 : 1 segmento de corredor 500kv zona norte	20
4.1.1. Caso1.1: Dimensión de agrupación transmisión(TUF)	21
4.1.2. Caso1.2: Dimensión de agrupación generación	22
4.1.3. Caso1.3: Dimensión de agrupación demanda	23
4.1.4. Caso1.4: Dimensión de agrupación costo marginal	24
4.1.5. Comentarios Caso1	24
4.2. Caso 2 : 1 segmento de corredor 500kv zona norte	26
4.2.1. Caso2.1: Dimensión de agrupación transmisión(TUF)	27

4.2.2.	Caso2.2: Dimensión de agrupación generación	28
4.2.3.	Caso2.3: Dimensión de agrupación demanda	29
4.2.4.	Caso2.4: Dimensión de agrupación costo marginal	30
4.2.5.	Comentarios Caso 2	31
4.3.	Caso 3 : 1 segmento de corredor 500kv zona norte usando el filtro de Distancia	31
4.3.1.	Comentarios Caso 3	31
5.	 conclusiones	34
	Bibliografía	36
	Anexo	38
A.	Interfaz	38

Índice de Ilustraciones

2.1.	Diagrama metodológico general de diseño de SPS	5
2.2.	Dos tipos de procedimientos de técnicas de clustering	6
2.3.	Clustering en 2 dimensiones (Pmin=3)	12
2.4.	Etapas de agrupación de patrones de carga	12
3.1.	Diagrama Metodologico General clustering aplicado al diseño de SPS	17
4.1.	Comparación de métricas de resultados para diferentes métodos de agrupamiento para la dimensión de transmisión(TUF) Caso 1.	21
4.2.	Comparación de métricas de resultados para diferentes métodos de agrupamiento para la dimensión de generación Caso 1.	22
4.3.	Comparación de métricas de resultados para diferentes métodos de agrupamiento para la dimensión de demanda Caso 1.	23
4.4.	Comparación de métricas de resultados para diferentes métodos de agrupamiento para la dimensión de costo marginal Caso 1.	24
4.5.	Comparación de métricas de resultados para diferentes métodos de agrupamiento para la dimensión de transmisión(TUF) Caso 2.	27
4.6.	Comparación de métricas de resultados para diferentes métodos de agrupamiento para la dimensión de generación Caso 2.	28
4.7.	Comparación de métricas de resultados para diferentes métodos de agrupamiento para la dimensión de demanda Caso 2.	29
4.8.	Comparación de métricas de resultados para diferentes métodos de agrupamiento para la dimensión de costo marginal Caso 2.	30
4.9.	Comparación de métricas de resultados para diferentes filtros de distancia y dos algoritmos de agrupamiento. Caso 3.	32
A.1.	Pestaña Procesamiento de Condiciones	38
A.2.	Pestaña Filtrar por Distancia	39
A.3.	Pestaña Disimilitud	39
A.4.	Pestaña Clustering	40
A.5.	Pestaña Cluster to paper	40

Capítulo 1

Introducción

1.1. Contexto

La integración de energías renovables en el sistema eléctrico está avanzando a un ritmo más rápido que la expansión del sistema de transmisión. Este fenómeno se debe en parte a la significativa disminución de los costos de inversión asociados con la construcción de centrales solares y eólicas. Además, estos tipos de instalaciones presentan tiempos de desarrollo más cortos. Sin embargo, esta rápida expansión puede dar lugar a congestiones en la red y márgenes operativos más ajustados, lo que a su vez incrementa los costos operativos y puede resultar en el vertimiento de energía renovable no utilizada.

Ante la necesidad de aumentar la capacidad de las líneas de transmisión, una alternativa efectiva es la implementación de Esquemas de Protección de la Integridad del Sistema (SIPS). Estos esquemas ofrecen una solución complementaria a la expansión física del sistema de transmisión. Mediante el uso de algoritmos de optimización, los SIPS identifican los agentes del sistema de potencia que permiten relajar los límites operacionales en corredores congestionados, estableciendo nuevos parámetros que superan el criterio N-1.

Una de las principales ventajas de los SIPS es que garantizan la seguridad del sistema frente a fallas en líneas de doble circuito. Además, al gestionar los agentes del sistema, estos esquemas ayudan a reducir los costos operativos y a posponer la necesidad de inversiones adicionales en infraestructura de transmisión, aumentando así la capacidad máxima de operación.

No obstante, el problema de optimización asociado con los SIPS presenta desafíos significativos en términos de costos computacionales. La complejidad del problema crece rápidamente con la inclusión de variables de decisión, ya que se trata de un problema entero mixto. Para abordar esta dificultad, se propone reducir la dimensión del problema de optimización utilizando técnicas de clustering aplicadas a las condiciones de operación. El objetivo es obtener una solución de diseño de SIPS en un menor tiempo de cómputo, minimizando al mismo tiempo el desvío respecto a la solución original.

1.2. Objetivos

1.2.1. Objetivo General

Reducir el número de condiciones de operación para llevar a cabo el diseño de Esquemas Especiales de Protección (SPS) en sistemas eléctricos de potencia, mediante técnicas de clusterización y así disminuir la dimensión del problema de optimización obteniendo una disminución del costo de tiempo computacional comprometiendo al mínimo la calidad de la solución.

1.2.2. Objetivos Específicos

Para alcanzar este objetivo general se definen los siguientes objetivos específicos:

- Comparar las diferentes técnicas de clustering existentes y su aplicación a condiciones de operación en sistemas eléctricos de potencia.
- Proponer una modelación de clustering para ser integrada en un modelo matemático que permita su identificación y diseño preliminar para un sistema eléctrico dado. Definiendo criterios tanto para escoger la dimensión y métrica para la agrupación como para escoger el punto representativo de cada cluster.
- Proponer un modelo capaz de identificar de un conjunto de opciones de diseño de clustering para un sistema eléctrico real la mejor opción que entregue el mejor resultado de diseño de SPS.
- Analizar el desempeño mediante la comparación de los resultados obtenidos con y sin el clustering, aplicado a un caso en el Sistema Eléctrico de Chile.

Capítulo 2

Antecedentes

2.1. Diseño SPS

Este trabajo se enmarca en una de las propuestas de investigación de la tesis de Magister de Ignacio Nancupil, "MARCO PARA LA IDENTIFICACIÓN Y DISEÑO DE ESQUEMAS ESPECIALES DE PROTECCIÓN PARA EL MANEJO DE CONGESTIONES DEL SISTEMA DE TRANSMISIÓN"[1].

El diseño de SIPS permite solucionar problemas de congestión del sistema de transmisión. Por medio la detección de condiciones particulares de las cuales se conoce que causan un estrés inusual del sistema de potencia, toma algún tipo de acción predeterminada para contrarrestar la condición observada de forma controlada. Todo esto mediante un algoritmo de optimización.

Se busca un relajo del criterio N-1 sin perder condiciones de seguridad. Es decir, se obtiene un aplazamiento de la expansión de la infraestructura de transmisión. Como resultado se genera un beneficio económico asociado a la implementación del SPS ya que permite puntos de operación a menor costo marginal.

No obstante, este algoritmo de optimización es costoso en términos de tiempo de ejecución ya que intervienen numerosas variables de decisión asociadas a la demanda y a la generación tanto continua como binarias, junto con una serie de restricciones de asociadas a los agentes del sistema eléctrico.

En la figura 2.1 , se muestra el diagrama metodológico general de diseño de SPS. Antes de hacer la optimización, en el paso "selección de corredores de interés", se obtiene un conjunto (numeroso) en que se propone realizar clustering para reducir el número de condiciones de operación para el algoritmo de diseño.

Esta herramienta de diseño de SPS usa información proveniente del Modelo de Planificación a largo plazo(PLP), en que a partir de la curva de carga se caracteriza la demanda del mes en 10 condiciones representativas(Block). De esta forma el año, en terminos de la demanda, queda representado por 120 bloques. La Hidrología(Hydro) corresponde a escenarios de caudales de centrales hidroeléctricas con un pronóstico de por medio. No hay un número fijo como en el caso de la demanda, en general son 59 escenarios hidrológicos. De esta forma

1 año está caracterizado por un total de $120 \times 59 = 7080$ condiciones de operación.

Como resultado del diseño de Esquemas de Protección Especial (SPS), se identifican los agentes del sistema eléctrico que desempeñan un papel crucial en la implementación de las soluciones propuestas. Estos agentes incluyen:

- **Generación:** Las unidades generadoras pueden ser ajustadas para aumentar o reducir su producción de energía. La gestión de la generación implica modificar la potencia generada para equilibrar el sistema y asegurar que los límites operacionales se mantengan dentro de los parámetros seguros.
- **Demanda:** La demanda de energía también puede ser ajustada a través de medidas de control de carga. Esto puede incluir la reducción temporal de la demanda en ciertas áreas para aliviar la congestión en el sistema o para adaptarse a cambios en la oferta de generación.
- **Acciones Correctivas:** Estas acciones incluyen:
 - *Rampas de Subida y Bajada:* Se refiere al ajuste gradual de la generación o la demanda. Las rampas de subida permiten que la generación aumente suavemente para evitar sobrecargas, mientras que las rampas de bajada ayudan a reducir la generación o la demanda de manera controlada para evitar caídas abruptas en el suministro.
 - *Conexión y Desconexión de Equipos:* Implica el control de la conexión o desconexión de líneas de transmisión, transformadores, y otras infraestructuras clave. Esta acción puede ayudar a redirigir el flujo de energía y gestionar la carga en diferentes partes del sistema.

Estas acciones correctivas son esenciales para mantener la estabilidad del sistema eléctrico y garantizar que los límites operacionales se respeten, minimizando así el riesgo de fallos o desconexiones. La identificación y gestión efectiva de estos agentes contribuye a la eficacia del diseño de SPS y a la optimización del rendimiento del sistema de transmisión.

2.2. Clustering

El clustering pertenece al ámbito del “aprendizaje automático no supervisado”, que se define como la tarea de descubrir una estructura oculta a partir de datos “sin etiquetar”. Los elementos de datos no están preclasificados o etiquetados, lo que hace que la evaluación de algoritmos de aprendizaje no supervisado sea difícil.

Clustering es el proceso de agrupar un conjunto de objetos en subconjuntos conocidos como clusters basándose en alguna medida de disimilitud entre los objetos. El objetivo es agrupar objetos similares y asegurarse de que los objetos ubicados en grupos distintos sean diferentes.

Por una parte, el aprendizaje automático supervisado comienza con una muestra de datos sobre la cual tenemos conocimiento previo y trata de extrapolar este conocimiento a volúmenes más grandes de datos. Por otro lado, el aprendizaje no supervisado no tiene conocimiento previo, por lo que su objetivo es inferir la estructura "natural" presente en los datos.

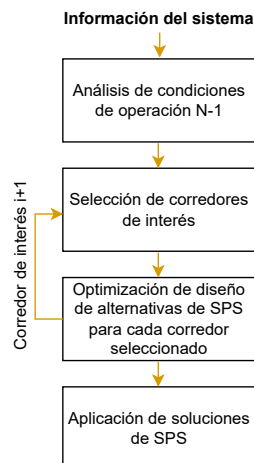


Figura 2.1: Diagrama metodológico general de diseño de SPS

En general, el aprendizaje no supervisado, aprovechando las grandes cantidades de datos sin procesar disponibles, es ampliamente reconocido como uno de los desafíos más importantes que enfrenta el aprendizaje automático en la actualidad.

El dominio del aprendizaje automático no supervisado, y en particular el clustering, está en constante desarrollo. Diferentes algoritmos de clustering aplicados al mismo conjunto de datos pueden producir tipos distintos de agrupaciones, y no existen principios generales que guíen la selección de algoritmos.

2.2.1. Métodos de Clustering

Muchos algoritmos de agrupamiento requieren que se proporcione el número de grupos como parámetro de entrada, lo que obliga a estos algoritmos a combinar o dividir clústeres naturales, o a generar clústeres que no existen naturalmente en los datos. La búsqueda de agrupamientos con un número predeterminado de clústeres es un problema mal definido porque un conjunto de puntos puede agruparse de muchas maneras. Incluso si un conjunto de datos no tiene una estructura significativa, un algoritmo de agrupamiento puede encontrar alguna partición de los datos.

Existen muchos tipos de algoritmos de clustering. Los más importantes son:

- Algoritmos de partición, representados por el algoritmo k-means y sus variantes.
- Algoritmos jerárquicos, que incluyen algoritmos aglomerativos y divisivos.
- Otras clases que incluyen el clustering basado en densidad, el clustering basado en

cuadrícula, el clustering espectral.

- Se han desarrollado algoritmos especializados para el clustering de datos categóricos, datos en flujo continuo, colecciones de documentos y datos multimedia, series temporales, entre otros.

Los algoritmos se pueden clasificar en dos categorías, de acuerdo con parámetros dados por el usuario. Una es en base al número de clusters N_c obteniendo el número de clusters deseados en base al grado de similitud entre cada par de datos. La otra de acuerdo con un radio que usa el grado deseado más alto de disimilitud entre cada par de datos, siendo el número de clusters determinado por el algoritmo. De esta forma los algoritmos dividen el conjunto de datos bajo distintos enfoques. En la figura 2.2 se muestra un ejemplo de estas dos categorías.

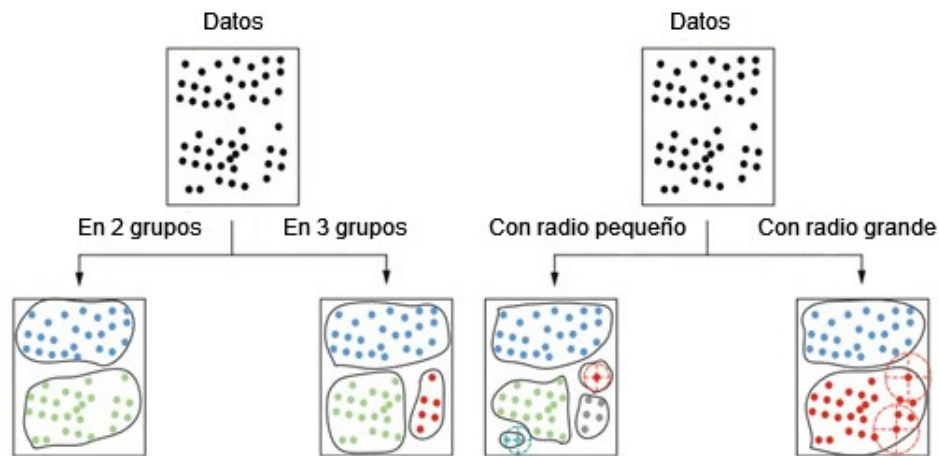


Figura 2.2: Dos tipos de procedimientos de técnicas de clustering

2.2.2. Disimilitud

La disimilitud es una medida fundamental en el análisis de datos y clustering, ya que cuantifica las diferencias entre elementos de un conjunto. Esta medida es crucial porque constituye el dato de entrada para cualquier algoritmo de clustering, influenciando directamente la forma en que los datos se agrupan. Las principales métricas de disimilitud incluyen:

- **Disimilitud Euclidiana:** Calcula la distancia entre dos puntos en un espacio n -dimensional. La fórmula general es:

$$d_{eucl}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

donde x y y son los puntos en el espacio, y x_i y y_i son las coordenadas en la dimensión i . Esta medida proporciona una forma intuitiva de cuantificar la separación entre puntos, siendo adecuada para datos numéricos en espacios geométricos.

- **Disimilitud de Fréchet:** Se emplea para comparar curvas o secuencias, considerando la distancia mínima necesaria para transformar una curva en otra mediante deformaciones

continuas. Esta métrica es útil para evaluar similitudes y diferencias en la forma de las curvas, teniendo en cuenta tanto la longitud como la forma de estas.

Elegir la medida de disimilitud adecuada es esencial para los algoritmos de clustering, ya que impacta en la identificación de patrones y en la interpretación de los resultados. La disimilitud determina cómo se agrupan los datos, facilitando el análisis y la clasificación de grandes conjuntos de información.

2.2.3. Algoritmo K-means

Es un algoritmo popular que utiliza un procedimiento iterativo para agrupar los datos en un número predefinido de grupos. Se dice que es un algoritmo rígido debido a que cada dato pertenece a un único cluster. Busca los grupos en el conjunto de datos de manera que se minimice una función de costo.

Se define la función de costo como la suma de las disimilitudes entre los miembros de los N_c grupos, siendo su expresión:

$$J = \sum_{j=1}^{N_c} \left(\sum_{i \in G_j} d_{i,j}^2 \right) \quad (2.1)$$

De acuerdo con esto, un punto i se considera que pertenece al $j^{\text{ésimo}}$ grupo si entre todos los centros, el $j^{\text{ésimo}}$ centro tiene la menor disimilitud con el punto i .

Se define una función de pertenencia dada por:

$$u_{i,j} = \begin{cases} 1 & \text{si el } j^{\text{ésimo}} \text{ dato pertenece al } i^{\text{ésimo}} \text{ cluster,} \\ 0 & \text{si no.} \end{cases} \quad (2.2)$$

En cada iteración se actualizan los centros de los clusters con la expresión:

$$z_j = \frac{1}{|G_j|} \sum_{x_i \in G_j} x_i \quad (2.3)$$

La expresión z_j representa el centro del j -ésimo cluster, donde G_j denota el conjunto de puntos del j -ésimo cluster y $|G_j|$ su tamaño. Por otro lado, x_i representa el i -ésimo dato perteneciente a G_j .

El proceso iterativo de este algoritmo es:

- Paso 1: Inicializar N_c número de centros de clúster seleccionados del conjunto de datos.
- Paso 2: Encontrar los valores de pertenencia para todos los puntos de datos usando la ecuación 2.2.
- Paso 3: Calcular la función de costo definida en la ecuación 2.1.
- Paso 4: Actualizar los centros de clúster utilizando la ecuación 2.3 y regresar al Paso 2.

Se pueden usar dos criterios de detención. En el primero, el algoritmo se detendrá si la diferencia entre el valor de la función de costo obtenido en dos iteraciones consecutivas es menor que un umbral, mientras que en el segundo, el algoritmo se detendrá si se ha ejecutado

durante un número suficiente de veces.

No obstante, se identifican dos inconvenientes. En primer lugar, KM necesita que el número de clústeres esté predefinido. En segundo lugar, su rendimiento depende en gran medida de la selección aleatoria de los centros de clúster iniciales. Por lo tanto, puede suceder que KM encuentre una solución incorrecta, especialmente para un conjunto de datos en el que existan clústeres tanto muy pequeños como muy grandes. Para superar este desafío, un enfoque consiste en ejecutar KM varias veces para encontrar soluciones mejores o utilizar un método de sustracción para seleccionar los centros de clúster iniciales de manera apropiada.

2.2.4. Algoritmo de Partición alrededor de Medoides(PAM)

El algoritmo PAM, también conocido como k-medoides, difiere del k-means en que los centros de los grupos (medoides) se seleccionan en cada iteración a partir de los puntos de datos en el conjunto de datos. Un medoide se define como un punto de datos cuya disimilitud promedio con respecto a otros puntos de datos en el grupo es mínima.

Requiere la selección inicial tanto del número deseado de clusters como de los medoides. Para hacerlo, se incorpora una fase de construcción que consiste en un procedimiento iterativo para seleccionar sabiamente los medoides iniciales.

El proceso iterativo de este algoritmo se describe a continuación:

1. Paso 1: Inicializar N_c número de medoides (centros de clúster) seleccionados del conjunto de datos.
2. Paso 2: Encontrar los valores de pertenencia para todos los puntos de datos.
3. Paso 3: Actualizar los medoides de acuerdo con las disimilitudes de los puntos de datos con respecto a los medoides seleccionados previamente.
4. Paso 4: Detener la actualización de los medoides si se cumple el criterio de detención. De lo contrario, regresar al paso 2.
5. Paso 5: Asignar cada punto de datos al clúster centrado por el medoide más cercano.

2.2.5. Algoritmo de agrupamiento de Sustracción(SC)

SC se introdujo principalmente como una herramienta de preprocesamiento para otros algoritmos de agrupamiento como KM y FCMD. SC tiene como objetivo principal encontrar centros de clúster mediante un procedimiento iterativo. Sin embargo, también puede utilizarse como algoritmo de agrupamiento por sí mismo.

En este algoritmo, se asigna primero un valor de densidad D_i a un punto de datos x_i utilizando la medida de densidad definida en la ecuación (2.4) :

$$D_i = \sum_{j=1}^{N_B} e^{-\left(\frac{d_{i,j}^2}{(ra/2)^2}\right)} \quad (2.4)$$

Donde $d_{i,j}$ es la disimilitud entre los puntos de datos i y j , y r_a es una constante positiva utilizada para representar el radio de vecindad. Según la ecuación (2.4), cuanto menor sea el valor de $d_{i,j}$ (más similar es el punto de datos j al punto de datos i), mayor será el valor de D_i . En otras palabras, dado que la medida de densidad en la ecuación (2.4) se define en función de la suma de similitudes entre todos los puntos de datos, el punto de datos con el valor de densidad máximo se elegirá como el primer centro de clúster, ya que tiene la mayor similitud con el mayor número de puntos de datos.

En el siguiente paso del algoritmo, todos los valores de densidad se revisarán con respecto al valor de densidad del centro de clúster determinado en el paso anterior. Esta revisión sustractiva se realiza utilizando la siguiente ecuación:

$$D_i^p = D_i^{(p-1)} - D_c^{(p-1)} e^{-\left(\frac{d_{i,c}^2}{(r_b/2)^2}\right)} \quad \text{para } i = 1, 2, \dots, N_B \quad (2.5)$$

Donde:

$$D_c^{(p-1)} = \max_i \{D_i^{(p-1)}\} \quad (2.6)$$

En la ecuación (2.5), r_b es una constante positiva destinada a definir un vecindario alrededor del centro anterior cuyos miembros experimentan una reducción medible en sus valores de densidad. Además, $D_c^{(p-1)}$ es el valor de densidad del centro de clúster seleccionado en la iteración $p - 1$, $d_{i,c}^{(p-1)}$ es la disimilitud entre el punto de datos i y el centro determinado en la iteración $p - 1$.

Ahora, el centro de clúster en la iteración p se seleccionará como aquel que tenga el valor de densidad más alto entre los valores revisados obtenidos de la Ecuación (2.5). El proceso de encontrar centros de clúster se repetirá hasta que el criterio de rechazo rechace el centro candidato. Se deben considerar varios criterios de rechazo, desde uno simple hasta uno más complicado.

Una vez que se determinan los centros de clúster, se debe evaluar la similitud de cada punto de datos con cada centro de clúster. En consecuencia, se formará una matriz con dimensiones $N_C \times N_B$ de la siguiente manera:

$$R = \begin{bmatrix} r_{1,1} & \cdots & r_{1,N_B} \\ \vdots & \ddots & \vdots \\ r_{N_C,1} & \cdots & r_{N_C,N_B} \end{bmatrix} \quad (2.7)$$

Donde:

$$r_{i,j} = D_j^1 - D_{c_i}^1 e^{-\left(\frac{d_{j,c_i}^2}{(r_b/2)^2}\right)} \quad (2.8)$$

Finalmente, suponiendo que el elemento i -ésimo de la columna j -ésima de R tiene el valor más bajo en la columna, se puede afirmar que el punto de datos j -ésimo debe asignarse a un clúster cuyo centro sea el i -ésimo centro de clúster determinado por el algoritmo.

2.2.6. Algoritmo Jerárquico

Los algoritmos de agrupamiento jerárquico evitan una de las principales limitaciones de los algoritmos de partición, que es la necesidad de especificar de antemano el número deseado de clústeres. En su lugar, construyen sistemas jerárquicos de clústeres descritos por estructuras tipo árbol conocidas como dendrogramas. Cortar estos árboles en diferentes niveles genera agrupamientos distintos, proporcionando una mayor flexibilidad.

Los clusters jerárquicos pueden construirse utilizando métodos aglomerativos o métodos divisivos. El primer enfoque comienza con las hojas de los dendrogramas, que son conjuntos unitarios, y construye el árbol de clústeres fusionando clústeres en un nivel determinado para formar clústeres más grandes en el siguiente nivel. El segundo enfoque comienza con el conjunto completo de objetos y divide sucesivamente conjuntos de objetos para producir un árbol de clústeres.

2.2.6.1. Algoritmo Aglomerativo

Se emplea un algoritmo llamado 'The Single-Link Algorithm' (Algoritmo de enlace único). El clustering de enlace único comienza con un agrupamiento inicial $\pi_1 = \alpha S$, la partición que consiste en objetos individuales.

El algoritmo produce una secuencia creciente de particiones π_1, π_2, \dots . Se mantiene una lista actual de clústeres L . Su valor inicial es $L = \{\{x_1\}, \dots, \{x_n\}\}$. Si el agrupamiento actual es π , el algoritmo busca el par de clústeres C_1, C_2 en L que tiene la distancia mínima entre sí (medida entre sus objetos más cercanos). La fusión de clústeres tiene un comportamiento local: las regiones donde los clústeres están más cercanos tienen una mayor influencia en este proceso que la estructura global del grafo de disimilaridad de los objetos. El efecto es producir clústeres alargados.

La condición de parada más común, que adoptamos a menos que se especifique lo contrario, es que $\pi_i = \omega S$, es decir, el agrupamiento se detiene solo cuando existe un clúster.

Es importante notar que la lista L puede contener tanto clústeres singleton como clústeres de varios elementos.

En cada fase del clustering jerárquico, se fusionan dos de los clústeres "más cercanos". La noción de clústeres cercanos depende de la disimilitud específica entre clústeres considerada en cada variante del algoritmo de clustering.

Definición 2.1 *Si U y V son dos clústeres, la disimilitud entre ellos se define utilizando una de las siguientes funciones de dos argumentos reales definidas en el conjunto de subconjuntos*

de S :

$$\begin{aligned}
 sl(U, V) &= \min\{d(u, v) \mid u \in U, v \in V\}; \\
 cl(U, V) &= \max\{d(u, v) \mid u \in U, v \in V\}; \\
 gav(U, V) &= \frac{\sum\{d(u, v) \mid u \in U, v \in V\}}{|U| \cdot |V|}; \\
 cen(U, V) &= \|c_U - c_V\|_2; \\
 ward(U, V) &= \frac{|U| \cdot |V|}{|U| + |V|} \|c_V - c_U\|_2.
 \end{aligned}$$

Los nombres de las funciones sl , cl , gav y cen definidas anteriormente son acrónimos de los términos "single link", "complete link", "group average centroid", respectivamente. Están asociados con variantes de los algoritmos de clustering jerárquico que discutimos más adelante. Observa que en el caso de la función $ward$, el valor equivale al aumento en la suma de los errores cuadráticos cuando los clústeres U , V son reemplazados por su unión.

El criterio específico de selección para fusionar bloques define el algoritmo de agrupamiento. A medida que se crean nuevos clústeres fusionando dos clústeres existentes, la matriz de distancias debe ajustarse para reflejar las disimilaridades entre el nuevo clúster y los clústeres existentes.

2.2.7. Algoritmo de Agrupamiento Espacial Basado en Densidad de Aplicaciones con Ruido (DBSCAN)

DBSCAN es otra técnica de agrupamiento basada en la densidad que ha sido utilizada en diferentes áreas de la ciencia. De manera similar al agrupamiento por sustracción, los fundamentos de DBSCAN se basan en el concepto de densidad en un espacio n -dimensional. Considerando un espacio 2-D mostrado en la Figura 2.3, se cumplen las siguientes definiciones:

- Un punto de datos p_r es un punto central si al menos P_{\min} puntos de datos, incluido p_r , están dentro de una distancia ϵ del punto p_r .
- Si el punto de datos q_r está dentro de la distancia ϵ del punto de datos p_r , se dice que el punto de datos q_r es directamente alcanzable desde el punto de datos p_r .
- Un punto de datos p_r es un ruido si no es directamente alcanzable desde ningún punto central.
- Un punto de datos q_r es alcanzable desde el punto de datos p_r si hay un camino p_{r1}, \dots, p_{rn} con $p_{r1} = p_r$ y $p_{rn} = q_r$, donde cada $p_{r(i+1)}$ es directamente alcanzable desde p_{ri} .

En este caso, un cluster es un conjunto de al menos P_{\min} datos en el que cada par es directamente alcanzable. Por ejemplo, en la Figura 2.3, si $P_{\min} = 3$, se formará un cluster con ocho elementos, mientras que los otros tres datos serán ruido.

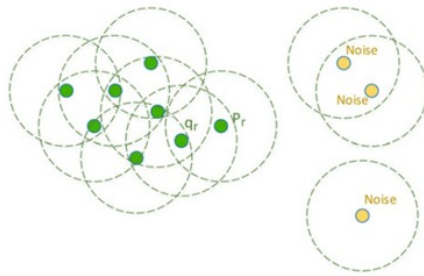


Figura 2.3: Clustering en 2 dimensiones ($P_{min}=3$)

2.3. Aplicaciones de Clustering en sistemas eléctricos de potencia

Lo primero es el formato a partir del cual se manejan los datos. En [2] se da un formato en que cada punto de operación se caracteriza por generación, flujo por las líneas, demandas. La generación la divide en convencional y renovables. la demanda la divide en demanda carga del sistema y carga reducida.

En [3] se muestran las etapas del clustering de patrones de carga, como lo muestra la figura 2.4. Que se usa en general como metodología para procesar los datos.

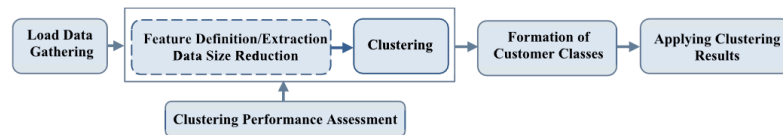


Figura 2.4: Etapas de agrupación de patrones de carga

En cuanto a la precisión de los métodos de reducción de dimensiones. Se puede medir tanto en términos del procesamiento de datos crudos como en las métricas de salida del modelo, donde estas últimas revelan qué tan bien se ajusta una técnica de muestreo a una instancia particular del modelo. En [4], se cuantifica la robustez de varias técnicas de muestreo y clustering con diferentes instancias del modelo, variando de forma independiente los parámetros del modelo, como por ejemplo el costo marginal de generación. Dado que los hallazgos obtenidos indican que el rendimiento de las técnicas consideradas es, de hecho, dependiente del modelo, se proporciona una mayor comprensión sobre el rendimiento de las técnicas comunes de reducción de dimensiones en aplicaciones de planificación de sistemas de energía.

Se comprenden diversos métodos para planificación usando clustering. En [5] se aborda con dos ideas principales: condiciones de operación se caracterizan por sus patrones del Flujo Óptimo de Potencia (OPF) en vez de patrones de generación-demanda. Esto recae en que la expansión de la red es el problema, y las pérdidas y congestiones son las que llevan a la inversión en la infraestructura. La segunda idea es clasificar los patrones de OPF usando una técnica de 'momentos' un enfoque bien conocido para abordar problemas de Reconocimiento Óptico de Patrones.

En [6], se proporciona una reducción óptima de datos de alta dimensionalidad para las curvas de demanda representativas que alimentan el problema de TEP (transmission expansion planning). La idea clave detrás de esta estrategia es extraer patrones de demanda a partir de los datos de demanda del sistema eléctrico mediante la implementación de un algoritmo de clustering jerárquico aglomerativo (HACA) basado en la regla del codo y un criterio de enlace, como la varianza de Ward.

En [7] propone otro tipo de cluster, denominado cluster espectral. Propone un método de selección de días representativos basado en clustering espectral orientado a la demanda (SC-DOR), el cual extrae la curva de duración de carga neta diaria (NLDC) y la curva de duración de rampa de carga neta (NLRDC) como un vector de características para capturar con precisión la demanda en términos de adecuación y flexibilidad. Además, los vectores de características se consideran como vértices conectados en un grafo, y los días representativos se obtienen a través de la partición del grafo mediante clustering espectral, el cual es adaptable a diferentes distribuciones de datos

En [8] se presenta un algoritmo híbrido de agregación de series temporales basado en clustering, diseñado para capturar tanto los valores extremos como las dinámicas temporales de los datos de entrada mediante la selección de representantes extraídos. El objetivo principal de este enfoque es preservar tanto los valores extremos como la cronología temporal en los representantes finales. Los resultados mostraron que el método propuesto puede reducir el tiempo de cálculo del modelo de co-planificación en hasta un 99 %, con un error del 2 % en el costo total de co-planificación. Este rendimiento se logra utilizando únicamente catorce días representativos, lo que equivale al 3.8 % del conjunto de datos.

Debido a restricciones computacionales, los modelos de optimización de sistemas de energía (ESOMs) y los modelos de planificación de expansión de generación (GEPMs) frecuentemente representan las variaciones intraanuales en la demanda y el suministro utilizando los datos de un número limitado de días históricos representativos. En [9] se contribuye a:

- proponer criterios y métricas para evaluar la representatividad
- proporcionar un enfoque novedoso basado en optimización para seleccionar un conjunto representativo de días
- evaluar y comparar el enfoque desarrollado con múltiples enfoques disponibles en la literatura.

Se demuestra que el enfoque basado en optimización logra resultados más precisos que los enfoques disponibles en la literatura. Como consecuencia, al aplicar este enfoque para seleccionar un conjunto representativo de días, se puede mejorar la precisión de los ESOMs/GEPMs sin aumentar el costo computacional. La principal desventaja es que el enfoque es costoso desde el punto de vista computacional y requiere un esfuerzo de implementación.

La gran mayoría de trabajos en el campo de clustering de sistemas de energía se realiza mediante enfoques de aprendizaje no supervisado. Se han propuesto varios métodos de clustering para la agrupación de sistemas de energía. Sin embargo, estos métodos tienen sus ventajas y desventajas, las cuales deben ser examinadas y abordadas antes de ser utilizadas.

Se presenta un enfoque para la selección de días representativos que combina dos estrategias: una basada en la entrada y otra en el costo, utilizando los algoritmos de clustering k-means y k-medoides. Se analizan las propiedades matemáticas de estos algoritmos, incluyendo un método para calcular la "brecha de optimalidad" entre las decisiones de inversión obtenidas con el modelo de días representativos y el modelo completo, así como la relación entre el error de clustering y la brecha de optimalidad. Además, se proponen dos enfoques innovadores para capturar las condiciones operativas extremas: el costo de reducción de carga el costo más alto", que ayudan a identificar los "días extremos"[10].

Por ejemplo, se pueden seleccionar horas operativas representativas al discretizar la curva de duración de carga, lo cual captura el efecto de los niveles de carga en los costos operativos del sistema. Sin embargo, este enfoque puede no ser adecuado si los costos operativos dependen de parámetros distintos a la carga, como la disponibilidad de recursos renovables, o si existen restricciones operativas intertemporales significativas, como los límites en el aumento de capacidad de los generadores [11]. En respuesta a estas limitaciones, se proponen dos técnicas de clustering jerárquico diseñadas para capturar las características estadísticas importantes de los parámetros, como la carga y la disponibilidad de recursos renovables, en la selección de días representativos. Estas técnicas abarcan autocorrelaciones temporales y correlaciones entre diferentes ubicaciones.

Existen métodos adicionales a los ya desarrollados, como el clustering algebraico multi-rejilla (AMG), que emplea un enfoque multinivel para agrupar datos, capturando mejor las relaciones jerárquicas y dependencias en comparación con los métodos tradicionales [12]. Para las 8,760 condiciones de un año, el método AMG logra un error de planificación inferior al 1 % para más de 200 clústeres, mientras que el error más bajo alcanzado por k-means++ es del 2.51 %. Además, el tiempo necesario para resolver el caso completo horario supera las 24 horas, lo que resalta la capacidad de las técnicas de clustering, independientemente del método utilizado, para acelerar significativamente el proceso de solución. El tiempo de ejecución se reduce a menos de una hora para más de 300 clústeres, mostrando una mejora notable respecto al caso horario.

En relación a la recopilación presentada en [13] se muestra que el algoritmo de K-means ha sido aplicado para agrupar zonas de sistemas eléctricos. Un ejemplo es que definen un vector fila para cada generador que representa la respuesta modal del generador, y luego utilizan el ángulo entre los vectores de dos generadores como criterio para obtener disimilitudes. Otra aplicación que menciona es una mejora al aplicar el algoritmo a las muestras del rotor de los generadores en conjunto con el algoritmo de optimización por enjambre de partículas (PSO) para idear un método PSO-KM con el objetivo de lograr una agrupación mejorada.

Otro ejemplo, se aplica para agrupar barras de un sistema eléctrico. Para ello, se utiliza primero el coeficiente de Pearson para evaluar la similitud de las variaciones de ángulo de fase de los voltajes. Luego, la distancia entre el valor del coeficiente de Pearson entre dos barras y el valor de unidad se considera como la disimilitud entre las barras. Finalmente, el algoritmo KM utiliza estos valores de disimilitud para agrupar las barras.

Para PAM, se expone que se ha aplicado a las disimilitudes entre las variaciones del ángulo de rotor de los generadores para agruparlos con el fin de realizar una equivalencia dinámica.

En el caso de prueba ha mostrado mejor desempeño que k-means logrando en un test 274 sobre 1000 casos de éxitos encontrando los grupos correctos.

Para SC, se obtiene mejores resultados estableciendo su eficacia en comparación con los algoritmos FCMd y KM. En particular, se ha mostrado la efectividad del algoritmo SC en la identificación de áreas. Es importante destacar que SC se distingue principalmente por no tener una operación de selección aleatoria, por lo que solo se ejecuta una vez y sus resultados son deterministas.

El algoritmo DBSCAN es más simple que el SC y se dice que es más adecuado para sistemas eléctricos de potencia debido a los siguientes puntos:

- No necesita el número de clusters como dato de entrada. Esto es importante ya que en distintas condiciones no se comportan los elementos de la misma forma.
- Necesita menos parámetros como entrada.
- No emplea operaciones aleatorias de selección y por tanto su solución es determinística.
- Funciona bien para conjuntos de datos desequilibrados. Puede encontrar fácilmente pequeños clusters y ruidos.

No obstante, en las aplicaciones usualmente se emplea para etiquetas el sistema eléctrico en áreas de operación más que para agrupar condiciones de operación similares.

La aplicación de técnicas de clustering para la selección de puntos de operación críticos en estudios de estabilidad de tensión ha sido objeto de investigación reciente. Esta metodología se enfoca en agrupar puntos de operación que están relacionados con la estabilidad de tensión, considerando aspectos como la longitud de las líneas y la capacidad de los generadores síncronos para participar en la regulación del voltaje. Por ejemplo, al definir los grupos mediante clustering, se puede seleccionar un punto representativo para cada clúster, ya sea el centroide de los datos o el punto de operación de mayor demanda. Esta selección permite identificar los escenarios más críticos dentro de cada grupo y facilita el análisis detallado de la estabilidad de tensión en diversas condiciones operativas. Esta técnica ha demostrado ser eficaz para mejorar la precisión en la identificación de condiciones operativas extremas y para optimizar los estudios de estabilidad en redes eléctricas [14].

Otra aplicación descrita en [15] utiliza k-means para la clusterización en múltiples dimensiones, considerando perfiles de demanda, generación, así como perfiles de sol y viento. Además, se incluye la hidrología en un estudio proyectado para los próximos ocho años. Este enfoque es significativo ya que ofrece una visión sobre las dimensiones de interés en las que se puede aplicar el clustering, destacando su utilidad en la modelización y análisis de sistemas complejos.

2.4. Librería PyClustering

La biblioteca `pyclustering` [16] es una herramienta robusta para el análisis de datos y clustering en Python. Esta biblioteca proporciona una amplia gama de algoritmos de agrupamiento, tales como K-Means, DBSCAN y Clustering Jerárquico, que permiten realizar

análisis complejos sobre conjuntos de datos. El algoritmo K-Means de **pyclustering** es conocido por su capacidad para agrupar datos en torno a centroides, minimizando la variabilidad dentro de cada grupo. Por otro lado, DBSCAN se destaca por su habilidad para identificar clusters basados en la densidad de puntos, lo cual es útil para datos con formas no lineales y presencia de ruido. Además, el Clustering Jerárquico ofrecido por la biblioteca construye una jerarquía de clusters, lo que facilita el análisis a diferentes niveles de granularidad. **pyclustering** también incluye herramientas para evaluar la calidad de los clusters, como el análisis de siluetas, que permite medir la cohesión y separación de los grupos formados. La flexibilidad y el conjunto de características ofrecidas por **pyclustering** hacen de esta biblioteca una opción valiosa para la optimización de procesos de clustering, proporcionando soluciones eficientes y efectivas para el análisis de grandes volúmenes de datos.

Capítulo 3

Clustering aplicado a condiciones de operación en diseño de SPS

El presente trabajo propone un marco metodológico para la agrupación de condiciones de operación del sistema eléctrico de potencia, por medio de la aplicación de técnicas de clustering. En la figura 3.1 se muestra en general los pasos para la ejecución de este estudio.

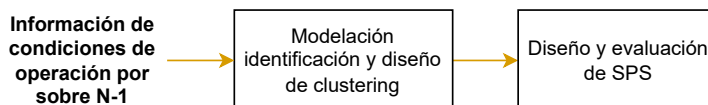


Figura 3.1: Diagrama Metodológico General clustering aplicado al diseño de SPS

El primer bloque consiste en los siguientes pasos:

1. Emplear un código en Python para dar formato a los datos de entrada, caracterizar los puntos de operación en función de los valores del flujo por las líneas, costo marginal de las barras, potencia de la demanda y potencia de los generadores. Tiene información del sistema eléctrico completo.
2. Filtrar por distancia. Dado que el problema de diseño de un Esquema Especial de Protección (SPS) se considera de naturaleza "local", es decir, su impacto se limita principalmente a las líneas específicas para las cuales se está diseñando el SPS, las condiciones en áreas distantes no deberían influir significativamente en la solución. Por lo tanto, se realiza un filtrado basado en la distancia, considerando únicamente las barras que se encuentran dentro de una proximidad específica respecto a las barras de las líneas en las que se está diseñando el SPS. Este enfoque asegura que el análisis se concentre en las zonas relevantes, optimizando así la precisión y eficiencia del diseño del SPS.
3. Lograr el valor de disimilitud de los puntos de operación mediante la manipulación de la información contenida en el dataset anterior. Se tiene la opción de la disimilitud euclidiana o de frechet, para las dimensiones de generación, transmisión, demanda, y costo marginal por barra.
4. Aplicar algoritmo de clustering para agrupar datos (k-medoide, jerárquico).

5. Aplicar criterio para seleccionar el punto representativo de cada clustering. En este caso se usa el punto que minimiza la disimilitud a los demas puntos.
6. Un ultimo paso de este bloque es darle el formato adecuado para ingresar a la herramienta de diseño de SPS la información de los puntos representativos de cada cluster, junto con la información necesaria para el manejo de los archivos de resultados. Se obtiene un archivo excel.

El segundo bloque consiste en usar el archivo excel para ejecutar la herramienta de diseño de SPS en el software DeepEdit. Como resultado se obtiene las métricas: Costo del SPS, tiempo de diseño, condiciones no factibles.

3.1. Obtención Dataset

La herramienta de diseño de SPS en DeepEdit genera un archivo con la información de las condiciones de operación por sobre el criterio N-1. En la figura A.1 se ve la pestaña de la interfaz que recibe el archivo, denominado '*Cluster_Conditions.csv*', que tiene las siguientes columnas: Block, Hydro, Type, Name, Param y Value.

- Las columnas Block y Hydro etiquetan las condiciones de operación.
- La columna Type especifica el tipo de elemento, que puede ser línea, generador, barra o demanda.
- Name indica el nombre del elemento.
- Param representa diversos parámetros, como el flujo en las líneas, la potencia de los generadores, la demanda, el costo marginal por barra, así como la potencia máxima y mínima de los generadores y el flujo máximo en las líneas.
- Value proporciona el valor correspondiente a cada variable en cada fila.

Este formato permite representar no solo variables operativas específicas para cada condición, sino también parámetros como los valores máximos o mínimos, los cuales pueden indicar situaciones como la desconexión o el mantenimiento de unidades.

3.2. Disimilitud

Antes de aplicar la disimilitud se filtra por distancia mediante la pestaña que se observa en la figura A.2. Luego se tiene la opción de buscar diferencias entre condiciones de operación mediante la disimilitud euclideana y de frechet, de acuerdo a una dimensión de interes, como se observa en la figura A.3

3.3. Aplicación Clustering

Con la matriz de disimilitud que relaciona dos condiciones de operación se procede a calcular los centros de clustering , teniendo como opción los algoritmos kmedoide o jerárquico, como se ve en la figura A.4. Se obtiene los centros segun la etiqueta que se ha venido usando que son pares de la forma (block,hydro).

3.4. Formato para herramienta de SPS

Finalmente, es necesario preparar los datos en el formato requerido para utilizar la herramienta de diseño de SPS en DeepEdit. Este proceso se lleva a cabo en la pestaña mostrada en la figura A.5. En esencia, se debe completar una plantilla en Excel con la información de los centros obtenidos en el paso anterior. Además, se configuran otras opciones necesarias para identificar y gestionar los archivos de salida generados por el diseño en DeepEdit.

3.5. Diseño y evaluación de SPS para los puntos representativos obtenidos en la agrupación

Como resultado, se obtienen tres métricas clave: los costos en millones de dólares, el tiempo de cálculo utilizando clustering, y la cantidad de condiciones no factibles. Las condiciones no factibles se refieren a la situación en la que, al no considerar todas las condiciones posibles, el SPS diseñado con un conjunto reducido de agentes no logra manejar adecuadamente todas las condiciones de operación. En otras palabras, se prueba si los agentes seleccionados pueden implementar acciones correctivas efectivas para cada condición de operación que no se encuentra entre los puntos representativos utilizados en el diseño. Como era de esperarse, algunos de estos puntos no cumplen con la factibilidad, y esos son los puntos identificados como no factibles.

Capítulo 4

Casos de estudio

4.1. Caso 1 : 1 segmento de corredor 500kv zona norte

Se prueba para 1 línea del corredor Polpaico500->NvaPAzucar500->NvaMaitencillo500->NvaCardones500->Cumbre500. Con la intención de disminuir los tiempos de cálculo, se calcula un sips para sobrecargas sólo en la línea NvaMaitencillo500->NvaCardones500. se configura el parámetro basecapsip con el valor 450. Como resultado se obtiene un número de condiciones de operación para diseño de SIPS de 52 puntos de operación.

El Resultado sin Clustering se puede obtener ya que es un caso con pocos puntos de operación. En este caso es:

- CostSPS: 12 [MUSD]
- Total SPS Time: 5901.655[seg]

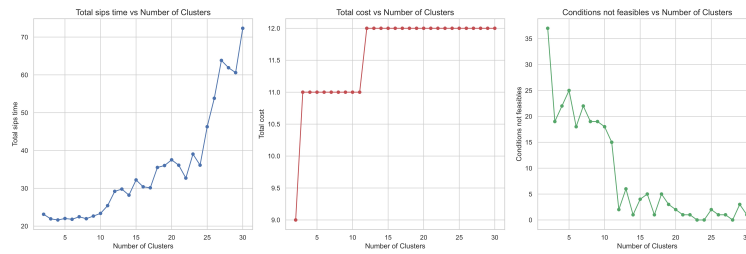
Se varía sobre las siguientes variables de decisión en el clustering:

- **Dimensión de Agrupación:** Generación, Transmisión(TUF), Demanda, Costo Marginal.
- **Disimilitud:** Euclidiana, Fréchet.
- **Algoritmo:** K-medoide, Jerárquico.

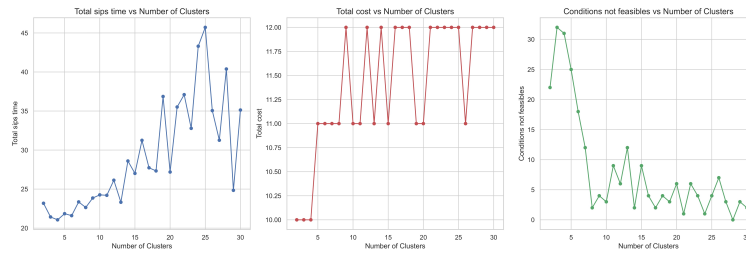
De esta forma se obtienen $4 \times 2 \times 2 = 16$ casos de resultados.

Se aplica las pruebas desde 2 a 30 agrupaciones dando los siguientes resultados.

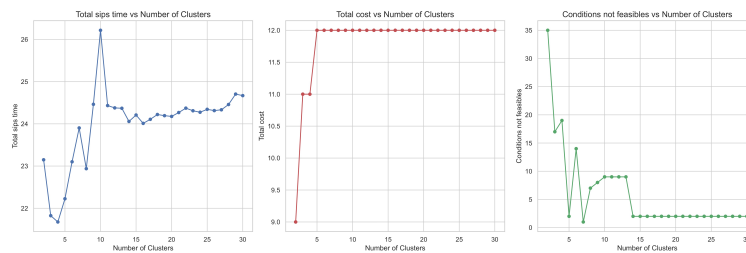
4.1.1. Caso1.1: Dimensión de agrupación transmisión(TUF)



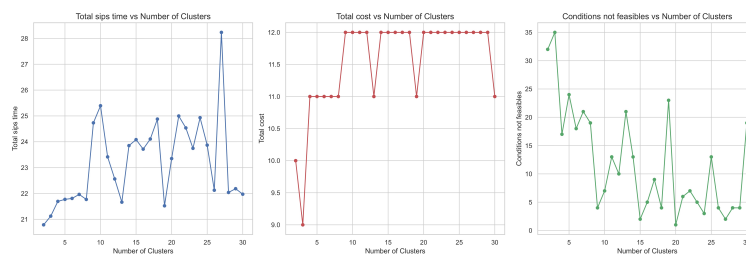
(a) Métricas de resultados para distancia euclideana - clúster Jerárquico



(b) Métricas de resultados para distancia euclideana - clúster K-medoide



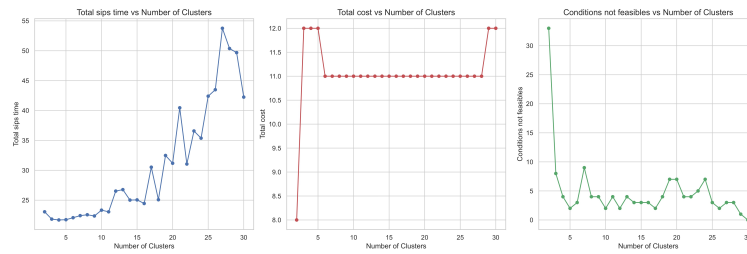
(c) Métricas de resultados para distancia frechet - clúster Jerárquico



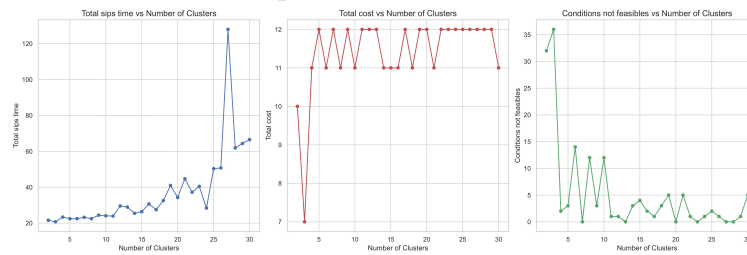
(d) Métricas de resultados para distancia frechet - clúster K-medoide

Figura 4.1: Comparación de métricas de resultados para diferentes métodos de agrupamiento para la dimensión de transmisión(TUF) Caso 1.

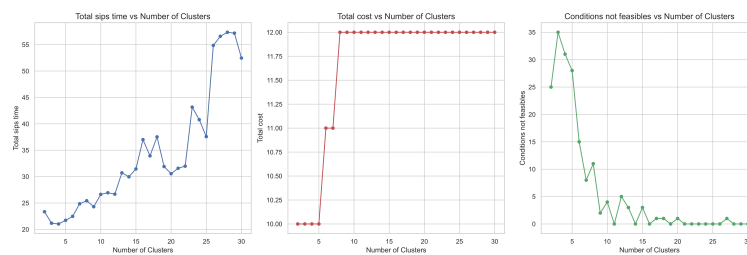
4.1.2. Caso1.2: Dimensión de agrupación generación



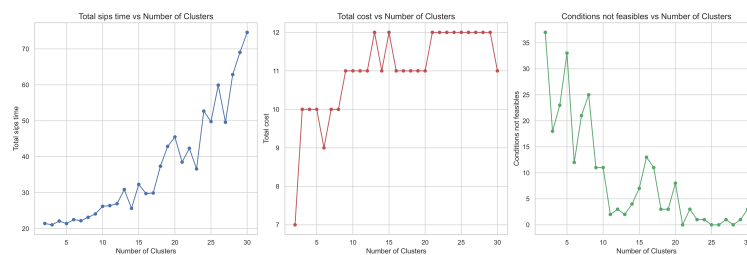
(a) Métricas de resultados para distancia euclideana - clúster Jerárquico



(b) Métricas de resultados para distancia euclideana - clúster K-medoide



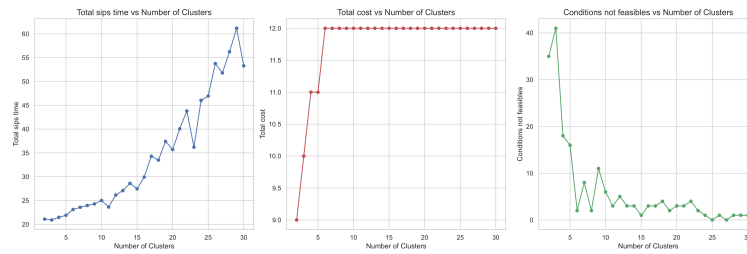
(c) Métricas de resultados para distancia frechet - clúster Jerárquico



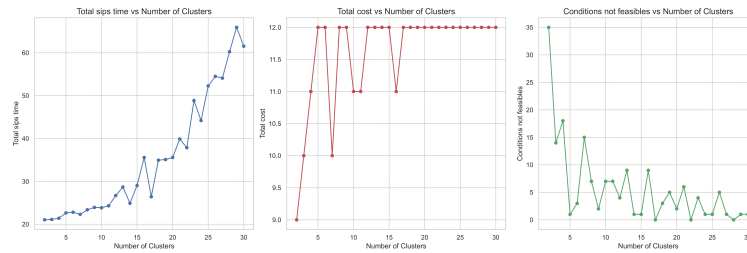
(d) Métricas de resultados para distancia frechet - clúster K-medoide

Figura 4.2: Comparación de métricas de resultados para diferentes métodos de agrupamiento para la dimensión de generación Caso 1.

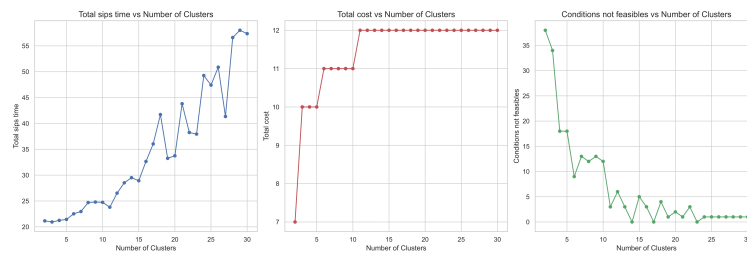
4.1.3. Caso1.3: Dimensión de agrupación demanda



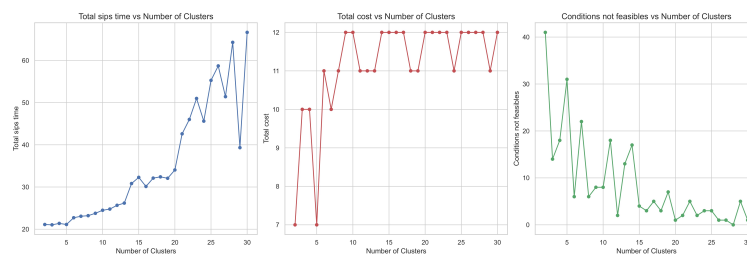
(a) Métricas de resultados para distancia euclideana - clúster Jerárquico



(b) Métricas de resultados para distancia euclideana - clúster K-medoide



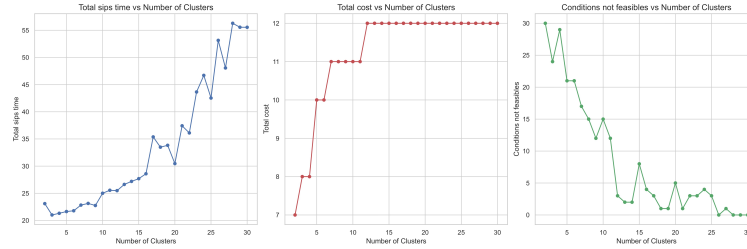
(c) Métricas de resultados para distancia frechet - clúster Jerárquico



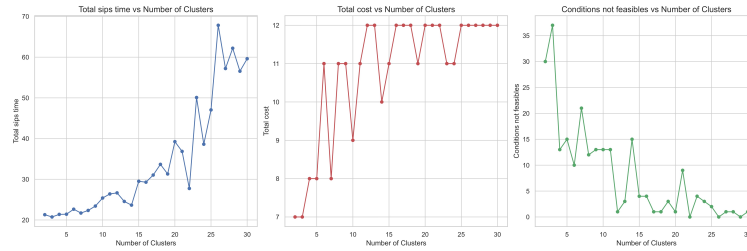
(d) Métricas de resultados para distancia frechet - clúster K-medoide

Figura 4.3: Comparación de métricas de resultados para diferentes métodos de agrupamiento para la dimensión de demanda Caso 1.

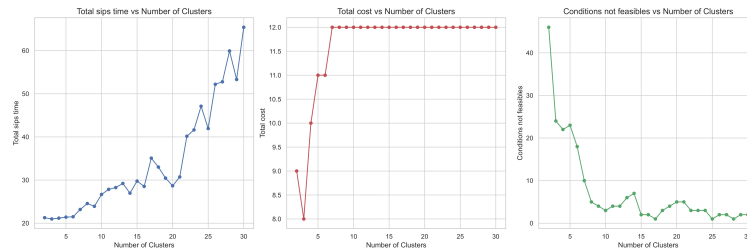
4.1.4. Caso1.4: Dimensión de agrupación costo marginal



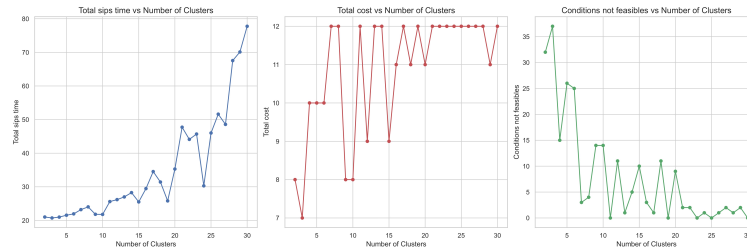
(a) Métricas de resultados para distancia euclidea - clúster Jerárquico



(b) Métricas de resultados para distancia euclidea - clúster K-medoide



(c) Métricas de resultados para distancia frechet - clúster Jerárquico



(d) Métricas de resultados para distancia frechet - clúster K-medoide

Figura 4.4: Comparación de métricas de resultados para diferentes métodos de agrupamiento para la dimensión de costo marginal Caso 1.

4.1.5. Comentarios Caso1

En comparación con las casi dos hora de resolución del problema completo este caso reducido muestra un gran desempeño en cuando al tiempo que tarda en diseñar el SPS con las condiciones representativas para todos los casos, ya que en el los resultados el peor desempeño ronda menos de dos minutos. Esto anterior referido a los diseños que alcanzan soluciones no factibles igual a cero.

Se observa que el algoritmo Kmedoide es mucho más oscilatorio que el Jerárquico, y es un resultado esperable, debido a la naturaleza con que se van armando los clusters a medida que se agregan más elementos. El jerárquico tiende a generar grupos más alargados y por eso su orden, mientras que el kmedoide aveces entrega estructuras no tan intuitivas.

La dimensión TUF no fue de buen desempeño ya que solo con la disimilitud euclideana logra una solución sin soluciones infactibles.

El Costo marginal es de mejor desempeño pero presenta mucha oscilación.

La demanda y la generación entregan buenos resultados, pero la generación sobresale ya que posee mayor suavidad sobretodo con la combinación de la disimilitud de frechet y algoritmo jerárquico.

Se busca reducir las oscilaciones para permitir un proceso de búsqueda más eficiente, basado en una lógica que supere la simple búsqueda exhaustiva.

4.2. Caso 2 : 1 segmento de corredor 500kv zona norte

Se prueba para 1 línea del corredor Polpaico500->NvaPAzucar500->NvaMaitencillo500->NvaCardones500->Cumbre500. Con la intención de buscar solución a un problema con mayor número de condiciones de operación, se calcula un sips nuevamente para sobrecargas en la línea NvaMaitencillo500->NvaCardones500, pero esta vez se configura el parámetro basecapsip con el valor 0. Como resultado se obtiene un número de condiciones de operación para diseño de SIPS de 592 puntos de operación.

El resultado completo no se pudo obtener ya que la solución no convergía por lo que se analiza fundamentalmente cuánto disminuyen las condiciones no factibles.

Se aplica como conjunto de interés las variables de generación, demanda, costo marginal, y transmisión (TUF) en base a las cuales se calcula la disimilitud en base a la distancia de Fréchet y distancia euclidiana. Se aplica cluster jerárquico y k-medoide escogiendo el punto representativo como el que tiene la menor suma de disimilitud con los demás elementos del cluster.

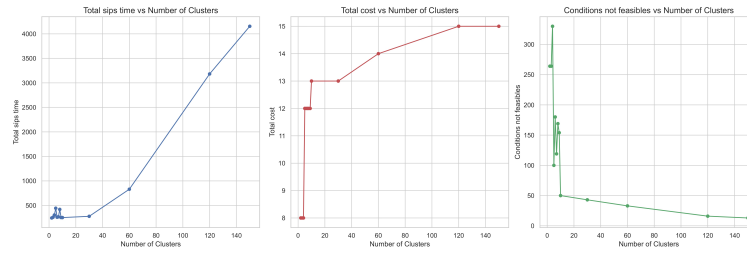
Se varía sobre las siguientes variables de decisión en el clustering:

- **Dimensión de Agrupación:** Generación, Transmisión (TUF), Demanda, Costo Marginal.
- **Disimilitud:** Euclidiana, Distancia de Fréchet.
- **Algoritmo:** K-medoide, Jerárquico.

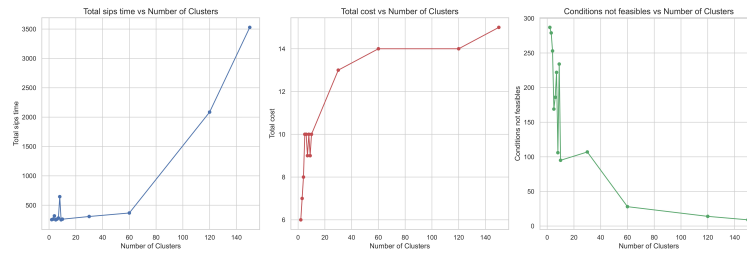
De esta forma se obtienen $4 \times 2 \times 2 = 16$ casos de resultados.

Se aplica las pruebas desde 2 a 10, luego se prueba con 30, 60, 90, 120 y finalmente con 150 agrupaciones, dando los resultados expuestos en las siguientes secciones.

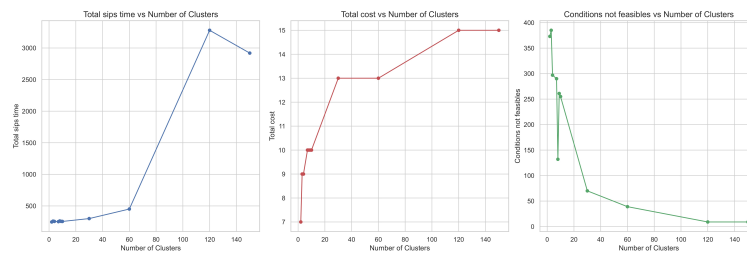
4.2.1. Caso2.1: Dimensión de agrupación transmisión(TUF)



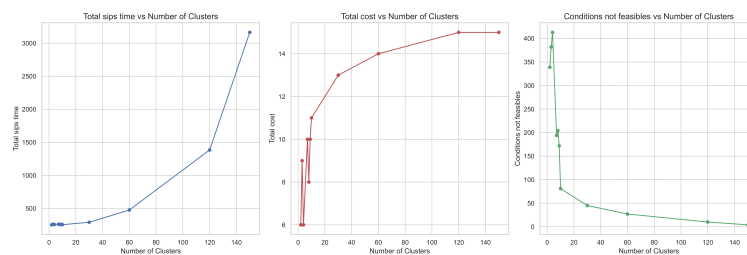
(a) Métricas de resultados para distancia euclideana - clúster Jerárquico



(b) Métricas de resultados para distancia euclideana - clúster K-medoide



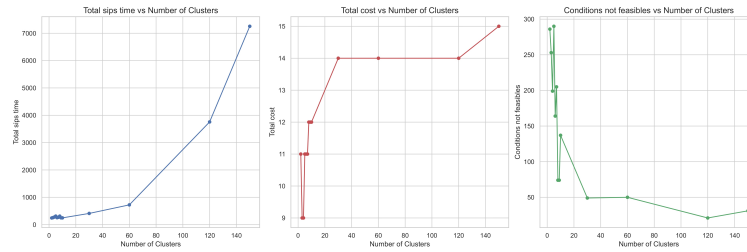
(c) Métricas de resultados para distancia frechet - clúster Jerárquico



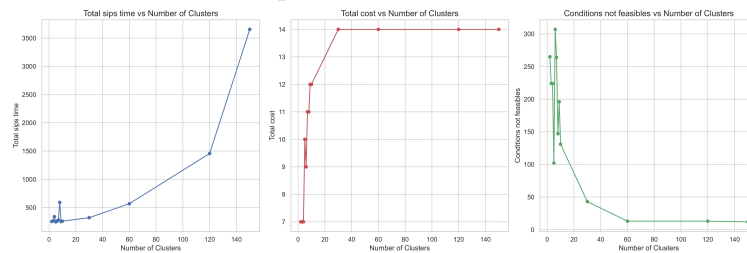
(d) Métricas de resultados para distancia frechet - clúster K-medoide

Figura 4.5: Comparación de métricas de resultados para diferentes métodos de agrupamiento para la dimensión de transmisión(TUF) Caso 2.

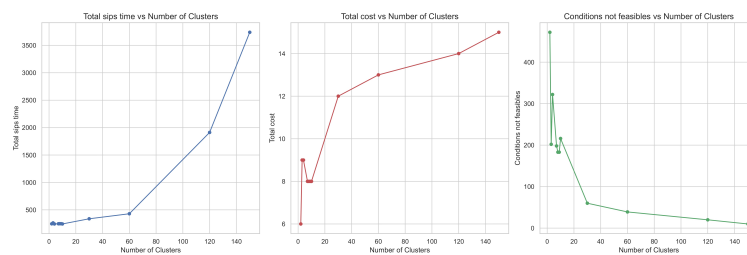
4.2.2. Caso2.2: Dimensión de agrupación generación



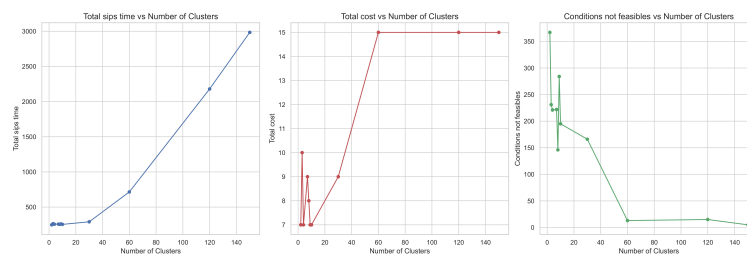
(a) Métricas de resultados para distancia euclideana - clúster Jerárquico



(b) Métricas de resultados para distancia euclideana - clúster K-medoide



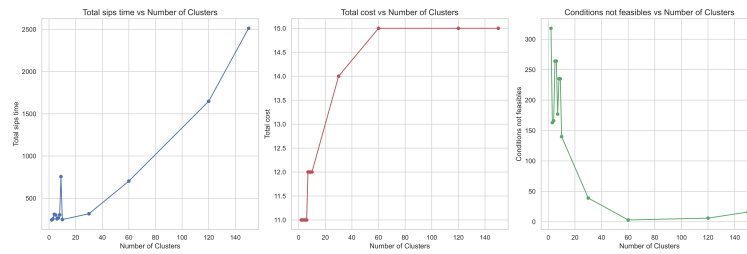
(c) Métricas de resultados para distancia frechet - clúster Jerárquico



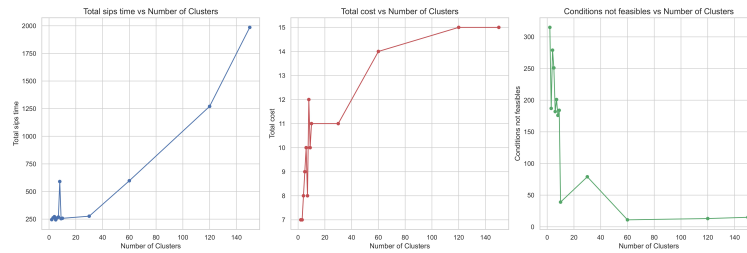
(d) Métricas de resultados para distancia frechet - clúster K-medoide

Figura 4.6: Comparación de métricas de resultados para diferentes métodos de agrupamiento para la dimensión de generación Caso 2.

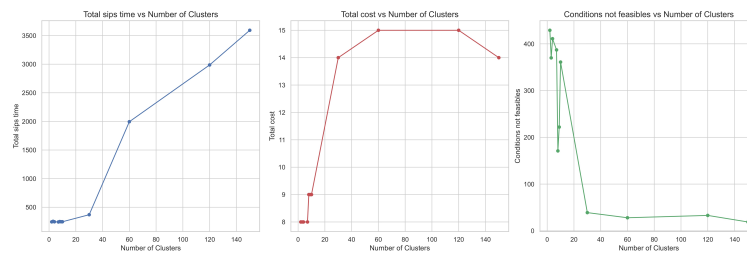
4.2.3. Caso2.3: Dimensión de agrupación demanda



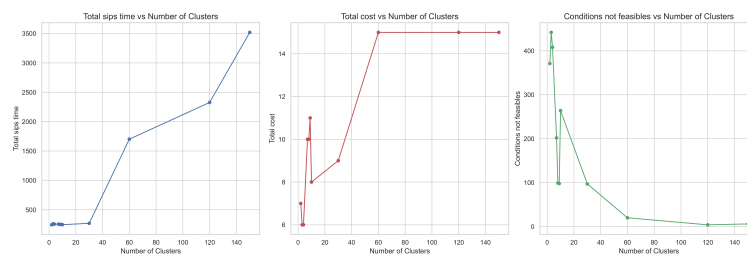
(a) Métricas de resultados para distancia euclideana - clúster Jerárquico



(b) Métricas de resultados para distancia euclideana - clúster K-medoide



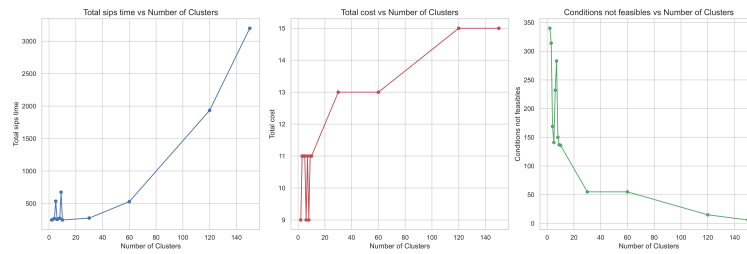
(c) Métricas de resultados para distancia frechet - clúster Jerárquico



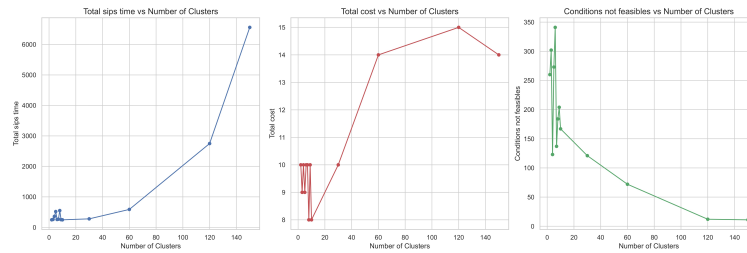
(d) Métricas de resultados para distancia frechet - clúster K-medoide

Figura 4.7: Comparación de métricas de resultados para diferentes métodos de agrupamiento para la dimensión de demanda Caso 2.

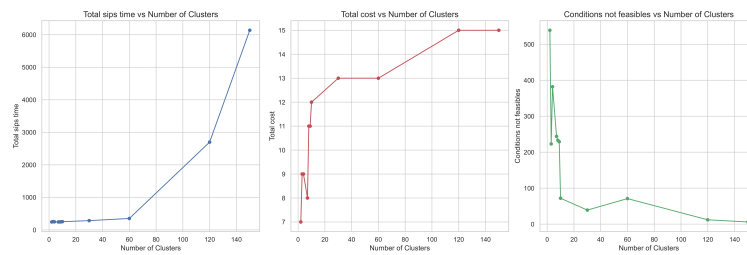
4.2.4. Caso2.4: Dimensión de agrupación costo marginal



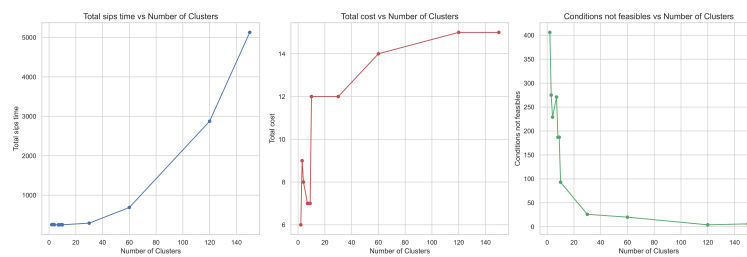
(a) Métricas de resultados para distancia euclideana - clúster Jerárquico



(b) Métricas de resultados para distancia euclideana - clúster K-medoide



(c) Métricas de resultados para distancia frechet - clúster Jerárquico



(d) Métricas de resultados para distancia frechet - clúster K-medoide

Figura 4.8: Comparación de métricas de resultados para diferentes métodos de agrupamiento para la dimensión de costo marginal Caso 2.

4.2.5. Comentarios Caso 2

Este Caso presenta un desafío mayor ya que si bien aun se está enfocando en una de las 4 líneas del corredor, ya el problema completo no es abordable computacionalmente.

Esta vez la dimensión de transmisión entrega un resultado bastante bueno para la combinación disimilitud de frechet y kmedoide. Se ve favorecido por la gran oscilación de kmedoide.

La generación presenta buen comportamiento menos en la combinación disimilitud euclideana con jerárquico ya que comienza a aumentar las condiciones no factibles en el ultimo punto con 150 condiciones representativas.

La demanda presenta resultados con condiciones no factibles muy bajas para 60 puntos representativos para la combinación cluster jerárquico disimilitud euclideana.

El costo marginal presenta menos oscilación que las otras dimensiones, y sus mejores resultados se dan con 150 clusters para la combinación cluster jerárquico disimilitud de frechet.

En realidad las 4 dimensiones presentan un comportamiento poco oscilatorio ya que los pasos de aumento de clusters siendo 30 permite una curva más suave. En cuanto a los resultados, en los puntos de operación no factible son del orden del 1 % del total de puntos de operación de interés, en cuanto al tiempo es del orden de 1 hora. Y es costo tiende a 15 Millones de dólares.

Con los 590 puntos de operación, se obtienen resultados favorables entre 60 y 150 clusters. Es decir, con menos de 30 % de los datos se puede diseñar un sip con un bajo margen de error.

4.3. Caso 3 : 1 segmento de corredor 500kv zona norte usando el filtro de Distancia

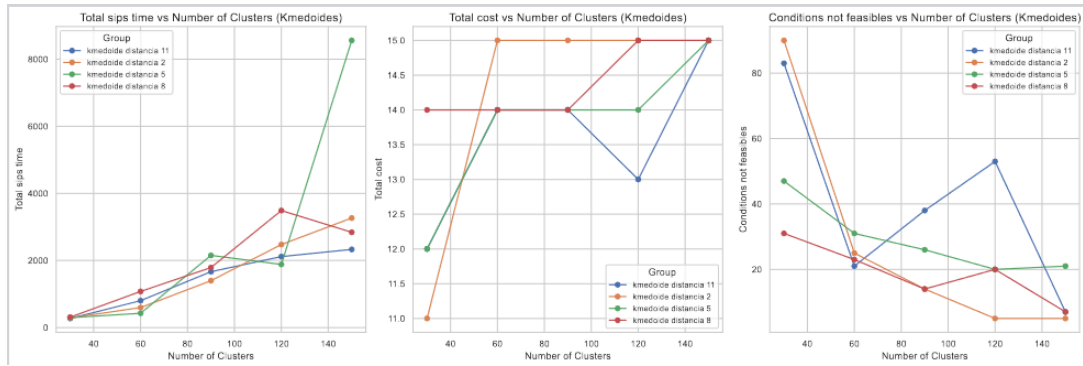
Se realiza una prueba para las condiciones del caso 2, variando el filtro de distancia para evaluar su influencia en los resultados. Se espera que los resultados reflejen la característica local del problema.

Se aplica como conjunto de interés la variable de generación, en base a la cual se calcula la disimilitud en base a la distancia de Frechet . Se aplica cluster jerárquico y k-medoide escogiendo el punto representativo como el que tiene la menor suma de disimilitud con los demás elementos del cluster.

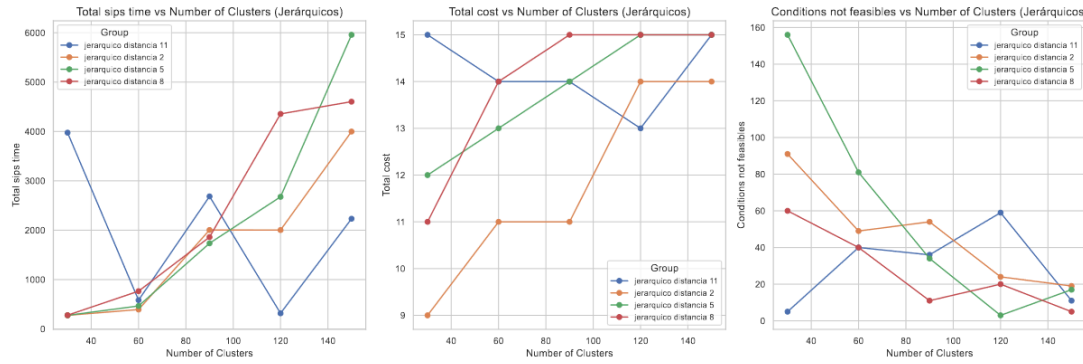
Se aplica las pruebas desde 30, 60,90,120 y finalmente con 150 agrupaciones, dando los resultados expuestos en las figura 4.9.

4.3.1. Comentarios Caso 3

Los resultados de aplicar al filtro expresan que con un filtro 2 no se obtienen buenos resultados para el cluster jerárquico. No obstante para el algoritmo k-medoide es el que menos condiciones no factibles logra.



(a) Métricas de resultados del clúster k-medoide con filtro de distancias



(b) Métricas de resultados para clúster jerárquico con filtro de distancias

Figura 4.9: Comparación de métricas de resultados para diferentes filtros de distancia y dos algoritmos de agrupamiento. Caso 3.

Con filtro 5, presenta un mal desempeño en cuanto al tiempo para ambos algoritmos de agrupación. Y tampoco son buenos para las condiciones no factibles.

Para el filtro 8, el tiempo no es muy buen desempeño pero en cuanto a las condiciones no factibles es el que mejores resultados logra para ambos algoritmos.

Y el filtro 11, el mayor que se probó, ofrece mejoras en cuanto al tiempo de diseño pero no tiene mejoras en las condiciones no factibles.

En cuanto al costo la que tiene un mejor comportamiento en relación a la suavidad de la curva es el filtro de distancia 8. Mientras que el filtro 11 tiene un comportamiento no intuitivo oscilando y no siendo un costo creciente.

La interpretación de estos resultados es que si se filtra con una distancia pequeña se pierde información siendo no acertiva la solución en la métrica de condiciones no factibles. Si se usa una distancia grande se vuelve un problema que genera grupos con variables lejanas al problema siendo poco útil dando resultados no logicos para las variables de interés, como un

costo con tramos de pendiente negativa.

Capítulo 5

conclusiones

Los valores de disimilitud que permiten evaluar la función de costo no viene necesariamente explícita en los datos, y es necesario deducción de algún método estadístico como se ha usado aca con la disimilitud de frechet y la euclideana que son el primer paso para realizar el clustering.

Al ser un problema local es necesario filtrar por distancia las variables que influyen en la agrupación, dejando de lado los valores que estén muy distante de la línea de interés para diseñar el SPS. Para el corredor que se estudió se obtiene como mejor filtro de distancia para hacer la agrupación la distancia 8 que es un punto medio entre tener que usar toda la información del sistema y tomar un valor muy pequeño como la distancia 2 que se está perdiendo mucha información y entrega curvas menos suaves.

Se aplicó con éxito la herramienta de agrupación de condiciones de operación a un caso real reducido, lo que resultó en un notable ahorro de tiempo en el cálculo de un SPS. Al utilizar los clusters, el tiempo requerido para obtener una solución del problema completo se redujo de aproximadamente 2 horas a menos de 2 minutos. Aunque es necesario considerar el tiempo de procesamiento de los datos, este sigue siendo un resultado muy favorable, ya que dicho tiempo es considerablemente menor a las 2 horas requeridas para resolver el problema en su totalidad.

En el caso 1 reducido, al aumentar en un paso el número de grupos durante la búsqueda de soluciones, se obtienen mejores resultados con la dimensión de generación. Al utilizar la distancia de Fréchet y el algoritmo jerárquico, se genera una curva que muestra un descenso suave en el número de condiciones no factibles, convergiendo a un valor estable. En contraste, el método de k-medoide presenta una mayor oscilación. Aunque el k-medoide puede ofrecer mejores resultados en algunos casos, su rendimiento es dependiente de los datos y puede ser impredecible, ya que en algunas curvas funciona bien y en otras no. Por otro lado, la combinación del algoritmo jerárquico y la distancia de Fréchet mostró un comportamiento estable para todas las dimensiones de agrupación.

En este caso se llega a usar más del 50 % de los puntos de operación para el diseño cosa que sería infactible para un caso con más de 600 puntos de operación.

En el caso 2 que es un sistema más grande con sobre 500 puntos de operación como entrada para el diseño de SPS, con los pasos de aumento de número de cluster de 30, no influye

tanto el algoritmo ni la distancia en cuanto a lo suave de la curva. En este caso la solución completa no converge por lo que se buscó acercarse lo más posible a una solución óptima. Con menos de 30% de los datos se puede diseñar un sip con un bajo margen de error siendo el número de condiciones no factibles menor al 1% del total de condiciones de operación. Los mejores resultados se obtienen con el algoritmo Jerárquico y la disimilitud de Fréchet. En cuanto a la dimensión de interés que se realiza la agrupación nuevamente la generación es la de mejor desempeño, esta vez, junto con el costo marginal.

Como recomendación final, basándose en las pruebas realizadas, se sugiere agrupar en función de la dimensión de generación, ya que esta captura la información clave sobre la potencia de los generadores y determina de manera más efectiva el perfil de cada punto de operación. Se recomienda utilizar la disimilitud de Fréchet, que refleja de mejor manera las similitudes entre condiciones de operación, junto con el algoritmo jerárquico, que, al formar los clusters, proporciona curvas más estables a medida que aumenta el número de grupos. Además, se aconseja aplicar un filtro de distancia para aprovechar la naturaleza local del problema.

Aunque puede parecer engorroso y costoso explorar las numerosas alternativas de agrupación y criterios, esta herramienta resulta muy útil en comparación con el enfoque de abordar el problema completo, que es computacionalmente inabordable. Permite aproximarse a una solución factible que es manejable en términos de cálculo, proporcionando una estimación del costo del SPS. Con los resultados obtenidos, se establece como una herramienta efectiva para el diseño de SPS, añadiendo una lógica que facilita la búsqueda de soluciones viables para el solver del problema de optimización.

Bibliografía

- [1] Ñancupil Quirilao, I., “Marco para la identificación y diseño de esquemas especiales de protección para el manejo de congestiones del sistema de transmisión”, Tesis M.Sc., Departamento de Ingeniería Eléctrica, Universidad de Chile, Santiago, Chile, 2023.
- [2] Zhang, W. *et al.*, “Clustering and decision tree based analysis of typical operation modes of power systems”, *Energy Rep.*, vol. 9, pp. 60–69, Sep. 2023.
- [3] Rajabi, A., Eskandari, M., Ghadi, M. J., Li, L., Zhang, J., y Siano, P., “A comparative study of clustering techniques for electrical load pattern segmentation”, *Renewable Sustain. Energy Rev.*, vol. 120, p. 109628, Mar. 2020.
- [4] Kristiansen, M., Korpas, M., y Hartel, P., “Sensitivity analysis of sampling and clustering techniques in expansion planning models”, en *2017 IEEE Int. Conf. Environ. Elect. Eng. 2017 IEEE Ind. Commercial Power Syst. Europe (EEEIC / I&CPS Europe)*, (Milan, Italy), pp. 1–6, IEEE, Jun. 2017.
- [5] Fitiwi, D. Z., de Cuadra, F., Olmos, L., y Rivier, M., “A new approach of clustering operational states for power network expansion planning problems dealing with res (renewable energy source) generation operational variability and uncertainty”, *Energy*, vol. 90, pp. 1360–1376, October 2015.
- [6] González-Cabrera, N., Ortiz-Bejar, J., Zamora-Mendez, A., y Paternina, M. R. A., “On the improvement of representative demand curves via a hierarchical agglomerative clustering for power transmission network investment”, *Energy*, vol. 222, p. 119989, May 2021.
- [7] Gao, S., Hu, B., Xie, K., Niu, T., Li, C., y Yan, J., “Spectral clustering based demand-oriented representative days selection method for power system expansion planning”, *Int. J. Elect. Power & Energy Syst.*, vol. 125, p. 106560, Feb. 2021.
- [8] Moradi-Sepahvand, M. y Tindemans, S. H., “Capturing chronology and extreme values of representative days for planning of transmission lines and long-term energy storage systems”, en *2023 IEEE Belgrade PowerTech*, (Belgrade, Serbia), pp. 1–6, IEEE, Jun. 2023.
- [9] Poncelet, K., Hoschle, H., Delarue, E., Virag, A., y Drhaeseleer, W., “Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems”, *IEEE Trans. Power Syst.*, vol. 32, pp. 1936–1948, May 2017.
- [10] Li, C., Conejo, A. J., Siirola, J. D., y Grossmann, I. E., “On representative day selection for capacity expansion planning of power systems under extreme operating conditions”, *International Journal of Electrical Power & Energy Systems*, vol. 137, p. 107697, May 2022.

- [11] Liu, Y., Sioshansi, R., y Conejo, A. J., “Hierarchical clustering to find representative operating periods for capacity-expansion modeling”, *IEEE Transactions on Power Systems*, vol. 33, pp. 3029–3039, May 2018.
- [12] Mozafari, Y. y Rosehart, W. D., “Representative operating scenario selection with algebraic multi-grid clustering for integrated energy systems planning”, *International Journal of Electrical Power & Energy Systems*, vol. 146, p. 108767, March 2023.
- [13] Nazari-Heris, M., Asadi, S., Mohammadi-Ivatloo, B., Abdar, M., Jebelli, H., y Sadat-Mohammadi, M., *Application of Machine Learning and Deep Learning Methods to Power System Problems*. Cham: Springer International Publishing, 2021.
- [14] Bastías, J. P., “Cluster para la selección de puntos de operación críticos para estudios de estabilidad de tensión”, Tesis M.Sc., Universidad de Chile, Santiago, Chile, 2019.
- [15] Penna, R. H., “Análisis e implementación del modelo de line switching en el sistema eléctrico chileno”, 2017.
- [16] Pyclustering, “Pyclustering - python library for clustering algorithms”. <https://pyclustering.github.io/>.

Anexo

Anexo A. Interfaz

Se presenta a continuación capturas de la Interfaz.

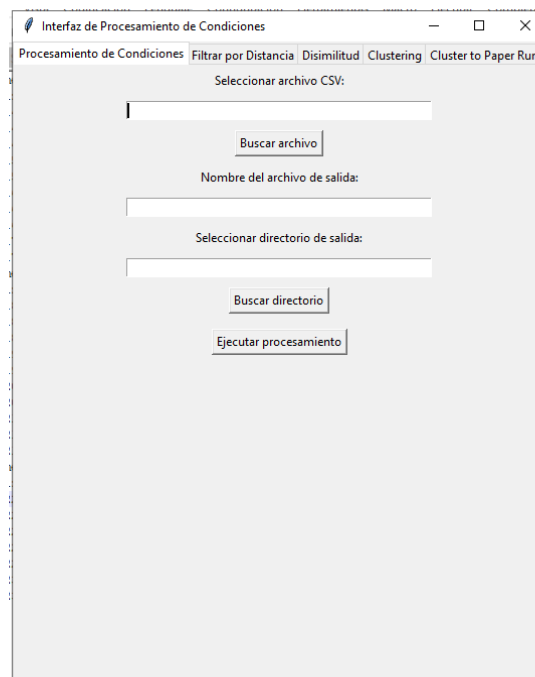


Figura A.1: Pestaña Procesamiento de Condiciones

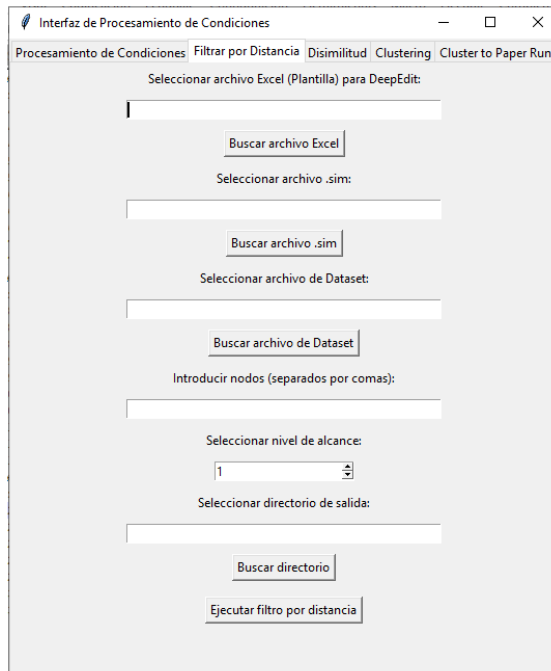


Figura A.2: Pestaña Filtrar por Distancia

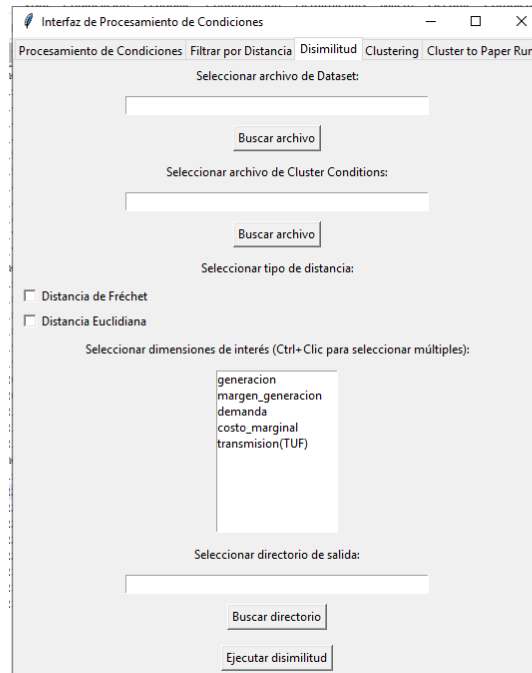


Figura A.3: Pestaña Disimilitud

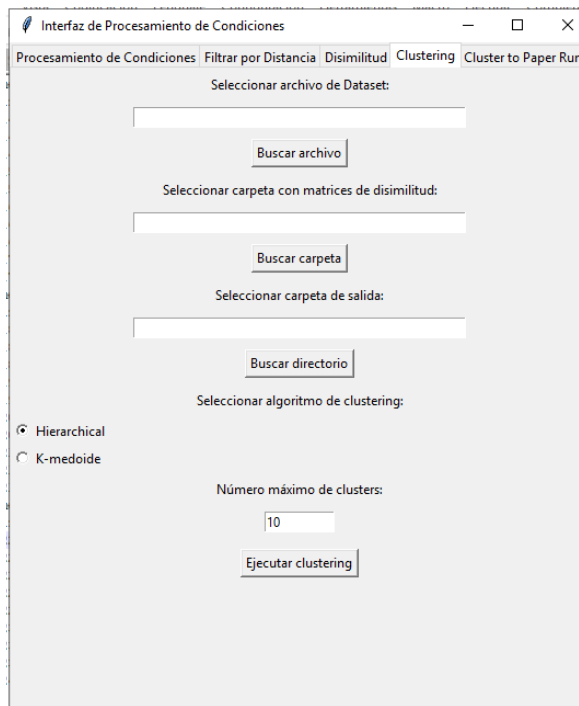


Figura A.4: Pestaña Clustering

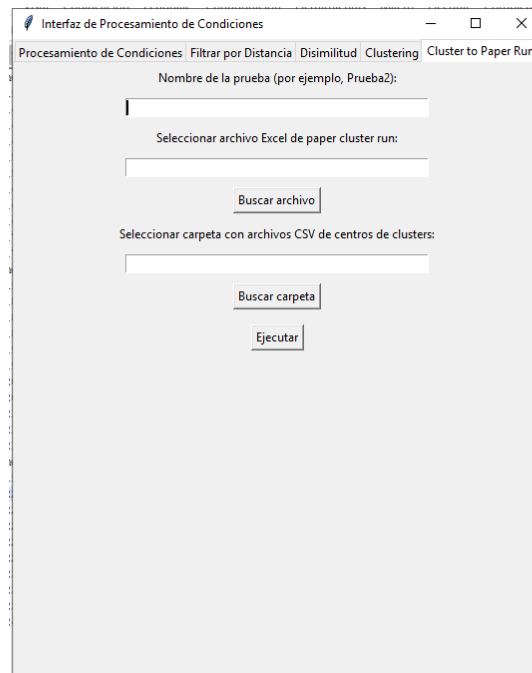


Figura A.5: Pestaña Cluster to paper